The impact of human language on perceptual categorization: electrophysiological insights

Doctoral thesis by

Piermatteo Morucci

Supervised by

Nicola Molinaro and **Clara Martin**

Donostia 2021

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

NAZIOARTEKO BIKAINTASUN CAMPUSA

CAMPUS DE EXCELENCIA INTERNACIONAL

Piermatteo Morucci

Basque Center on Cognition, Brain and Language
Paseo Mikeletegi 69
Donostia-San Sebastián, Spain
June 2021

The impact of human language on perceptual categorization: electrophysiological insights

Doctoral thesis by

Piermatteo Morucci

Supervised by

Nicola Molinaro and **Clara Martin**

Donostia 2021

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

NAZIOARTEKO BIKAINTASUN CAMPUSA
CAMPUS DE EXCELENCIA INTERNACIONAL

bcbl
BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

This work received support from from “la Caixa” Foundation (ID 100010434) through the fellowship LCF/BQ/IN17/11620019, and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 713673.

Acknowledgements

I firstly want to thank my supervisors Nicola Molinaro and Clara Martin. I am deeply grateful to them for giving me the freedom to pursue my own ideas and work on the research questions that passionate me most. I thank Nicola for his guidance and support over these years. He accepted me in his group when I had basically no experience in electrophysiology and little knowledge about brain dynamics. He helped me to become a more mature and independent researcher. His view of science deeply shaped the way I look at brain and cognitive phenomena now. I thank Clara for her unconditional support, clear advice and sweet nature. Every exchange with her I've learnt something new.

I thank the members of the Brhyco Group for all the exchanges and discussions. It is a big luck to be surrounded by such smart and nice people in a working environment. I thank the postdocs for supporting my academic growth during this path. In particular Craig, for all the methodological and signal processing tricks he taught me. I had a lot of fun working together.

I thank Lucia Melloni for offering me the opportunity to join her research group. For anyone who is passionate about the neural bases of sequences and predictive processing, studying with Lucia is both a privilege and an honor. Her approach to data and research questions deeply influenced me.

A big thanks also to the support teams at the BCBL, especially the members of IT, Admin and the lab. Without them all my research projects would not exist.

To the PhD community and all my friends from the BCBL for all the experiences and moments together. To the La Caixa Team for their commitment in providing us with all the tools needed to achieve our career goals. To my friends from Kuraia gym – I spent quite a lot of my free time training with them during these years and I always felt at home. To my girlfriend Valeria and my friends from Viterbo and Rome, for being always by my side.

Finally, my never-ending gratitude to my family – my parents, my sister and Ivan. Especially my parents, who did everything to see me happy.

abstract

How does learning cultural systems like language affect perception and cognition? The last few years have seen increased interest into this topic, yet with little theoretical advance. One fundamental question concerns the nature of the neural mechanism through which language affects perceptual processes. Some accounts suggest that effects of language are “high-level”, meaning that language does not affect early perceptual processes, but rather interact at later conceptual or decision-making stages. More recent proposals posit that language can alter perceptual processes at early sensory levels. This latter account is in line with current predictive processing theories of perception, which suggest that sensory processes are largely influenced by prior knowledge and expectation. The present thesis aims at investigating whether and how language shapes perceptual processing. We focus on two specific types of language-perception interactions: (i) the effect of linguistic labels on the recognition of visual object categories; and (ii) the effect of linguistic knowledge on neural processing of rhythmic sounds. We address these questions by taking advantage of time-resolved electrophysiological measures like electroencephalography (EEG) and magnetoencephalography (MEG). We use these tools to investigate the interplay between language and perception by focusing on neural indices putatively associated to perceptual prediction (i.e., neural oscillations in the alpha/beta frequency bands) and prediction error signal (i.e., the Mismatch Negativity).

In the first study, we show that language boosts visual perception of congruent object categories to a larger extent than equally familiar natural sounds. Using EEG, we demonstrate that language impacts visual perception by preparing the brain for incoming input via the selective modulation of alpha and beta oscillations. These oscillatory indices carry content-specific representations, emerge in sensory regions before stimulus presentation, and are predictive of later recognition performance.

The second study investigates whether life-long exposure to certain linguistic patterns impacts neural processing of rhythmic sounds. By comparing MEG data from native speakers of typologically different languages (Basque vs. Spanish), we show that the auditory system relies on syntactic/prosodic patterns of native language to generate hierarchical predictions about incoming (non-linguistic) sounds. When an expected event disrupts a rhythmic sequence of sounds, the amplitude of the Mismatch Negativity varies orthogonally depending on the individual's linguistic background. This prediction error response occurs around 100 ms from deviant onset, and has its locus in auditory regions. This finding indicates that coding schemes employed to parse linguistic material are recycled by the auditory system to implement predictive models of the environment. This study also offers novel insights into the hierarchical organization of auditory predictions.

The study of the interaction between language and perception can provide novel insights into different domains of cognitive neuroscience, including the nature of conceptual representations activated during language processing, as well as the effect of high-level knowledge on lower-level processes. By identifying oscillatory and ERP components that characterize such interactions, we hope that these results will help to further define the implications of learning symbolic systems in sculping our knowledge of the world.

Contents

Chapter - 1 Theoretical background

1.2 Perception as inference

The history of science includes numerous challenging questions, including the question about the origin of our perception of the world. Before scientists could record brain activity and measure reaction times, philosophers have long reflected on this question. A common intuition is that perception provides us with a veridical representation of what is there in the external world. This position takes the name of direct realism. Philosophers like Aristotle and Thomas Aquinas were supporters of this position, with the former proposing that the forms of the objects in the world are the same of our concepts and percepts. An alternative position is called representationalism (also known as indirect realism or representative realism), and posits that our conscious perception does not reflect the real world itself, but a mere internal representation generated by the mind/brain in the attempts to find the causes of the external world. Among philosophers, Lock and Descartes (and maybe Kant) were the main supporters of this position.

Phenomena like dreams, hallucinations and visual illusions suggest the reality and our experience of it are not exactly the same thing. Take for instance the left-side of the image in Figure 1. The majority of people perceive the central paired tiles to have different tonalities of grey. However, as the image on the right-side of Figure 1 shows, this perception is illusory. This is the so called “Cornsweet illusion” and shows how our visual experience is not always veridical but can be biased by prior beliefs. In this example, the prior belief concerns the fact that the color of objects’ surface does not usually change its tonality but rather keeps a uniform tone. Thus, what we know about illuminance and reflectance may bias our perception of reality.

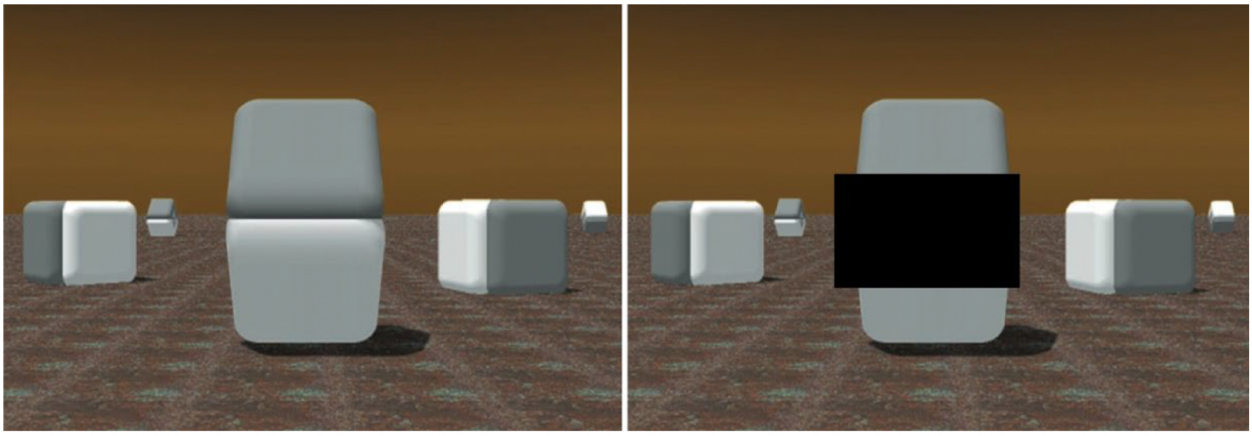


Figure 1. The image reflects a typical Cornsweet illusion (figure taken from Clark and Lupyan, 2015).

The idea of perception as a process of construction was further developed by German polymath Hermann von Helmholtz (1867), who proposed the "perception-as-inference" view. Helmholtz pondered on the problem of how we can generate accurate percepts given the ambiguity of sensory signals. He proposed that such problem could be solved via "unconscious inference", that is, the use *prior knowledge* to generate meaningful representations over noise data. This idea became later a building block of cognitive psychology. One prominent figure that contributed to the transition from representational realism as philosophical argument towards scientific theory was Richard Gregoory (1980). Gregoory compared the problem that our perceptual system has to face with the process of hypothesis in science: in the same way that scientists develop hypotheses to understand natural phenomena, the perceptual system tries to develop hypotheses about the ambiguous data it receives, in order to make meaningful models of reality. These theoretical approaches have been formalized more recently with the idea of the "Bayesian brain". This idea has its roots in the theorem of the British statistician Thomas Bayes, which provided a mathematical implementation of how to generate inferences by combining new data with prior knowledge. According to the Bayesian brain hypothesis, this type of probabilistic inferences emerges by using ("top-down") prior knowledge to interpret ("bottom-up") sensory information. The idea that the brain implements bayesian inference to support perceptual

processing is now a key assumption of predictive processing views of perception, a framework that is becoming extremely influential in cognitive neuroscience.

1.3 Predictive processing

For many years, the classical view in neuroscience has been provided by feedforward models. These models conceive the brain as a device that passively processes and registers external inputs. In contrast, the hypothesis of the brain as a predictive machine postulates that one of the fundamental functions of the brain consists in the anticipation of future events. This hypothesis is becoming increasingly influential in cognitive neuroscience. Different theoretical models have proposed different cortical architectures of how the predictive brain machinery can be implemented in cortical circuits, such as predictive coding, hierarchical temporal memory, and Bayesian inference (Friston, 2005; Hawkins and Blakeslee, 2004; Kording and Wolpert, 2004; Rao and Ballard, 1999; Spratling, 2010). Despite differing in their details, all predictive processing theories share the idea that the brain develops generative models of reality and uses such models to generate predictions about incoming events (Clark, 2016). Such generative models are conceived as a processing hierarchy: predictions are proposed to be conveyed through feedback signal stemming from higher to lower cortical areas, which are expected to have a suppressing effect onto incoming signals (see Figure 2). Predictions are then compared to bottom-up sensory signals at each level of the hierarchy. Only the difference, called prediction error (PE), is postulated to propagate through feedforward connections from lower to higher cortical areas in order to update internal models.

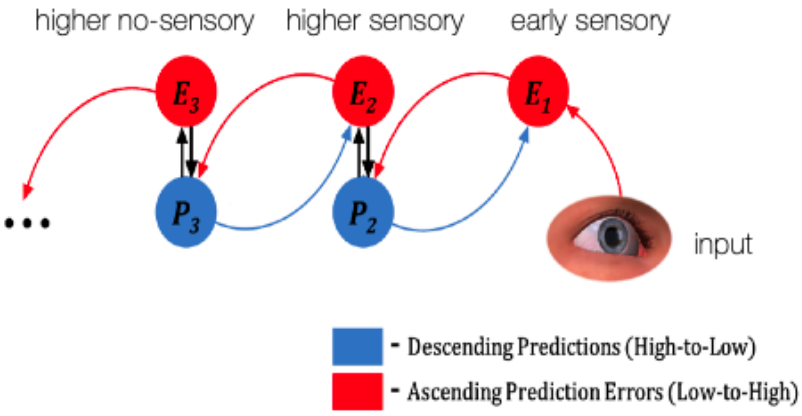


Figure 2. **How information flows through the cortical hierarchy in predictive processing.**

According to predictive processing models of perception, sensory information is trasmitted up through the hierarchy by prediction error signals (red) to adjust prediction (blue). Each processing stage of the hierarchy uses prediction error signals to adjust the internal model in order to minimise prediction errors (i.e., prediction activity at timepoint t reflects predictions and error signals from timepoint t-1). (adapted from Yon & De Lange, 2018).

Two neural units are needed to allow such implementation: prediction neurons trying to predict bottom-up activity in lower-level stages of the cortex via feedback signals; and prediction error units transmitting the difference between the predicted and incoming signals upward in the hierarchy. If prediction is accurate, prediction error signals are reduced. Both prediction and prediction-error signals are postulated to be, at least in sensory regions, feature-specific, meaning that they encode specific dimensions of a percept (e.g., the length of a tone in audition or the shape of an object in vision). Predictive processing can potentially bring different advantages to an organism. First, from a behavioral/cognitive perspective, being able to use prior knowledge to generate appropriate models of reality allows an organism to predict the future, thus facilitating the interaction with the environment. It does not only allow to predict the motor and sensory consequences of its action, but also the dynamics underlying the relations between objects and agents in certain contexts. For instance, after learning the association between two sensory stimuli like roaring with the

presence of a lion, just hearing a roaring would allow an animal to predict the likely presence of a predator. This would enable the animal to immediately select the most appropriate action scheme to interact with the environment (e.g., roaring → lion → run away). This and many other examples give intuitive support for the hypothesis that the brain is a predictive machine. However, predictive processing theories also fit nicely with the computational and anatomical properties of the mammalian brain. Studies on cortical computation have often highlighted the fact that computations using spikes are expensive in terms of metabolic costs (Lennie 2003). Thus, the brain must arguably select an appropriate strategy to reduce such costs. Using prior knowledge to predict activity related to incoming stimuli might offer a solution: a cortical mechanism that allows an expected stimulus to elicit a weaker response (that is, a smaller prediction error signal) than the same stimulus presented in an unpredicted context would allow to reduce redundant information, and as a consequence, a reduction of metabolic costs. That is, transferring only the unpredicted part of the signal is more efficient than transferring the whole bottom-up signal because fewer spikes are needed.

1.4 Language priors bias categorical perception

Considering perception as a predictive process implies considering perception as a "penetrable" process that can be influenced by prior knowledge and expectation, as far as such penetration is effective in reducing the prediction error. But what does count as prior knowledge? In humans, one form of prior knowledge is language. Several studies have shown that language can influence other non-linguistic systems such categorization, memory and perception. Rather than being naive or exotic as it is sometimes portrayed, the study of the interaction between language and perception offers a unique model to address several unanswered questions about the predictive nature of experience in the human brain. Language is indeed a high-level cognitive function unique to humans, arising from the

interaction of different brain networks. Language processing is highly hierarchical: information coming from early sensory regions is transformed into semantic representations across a series of processing steps. At the same time, syntactic/semantic information from higher stages of the hierarchy can affect low-level processes (e.g., Kuperberg & Jaeger, 2015). As such, language provides a perfect model to study the interaction between top-down prediction and bottom-up activity. Moreover, language is acquired at birth and develops over the whole life-spam of an individual. As such, it provides a good model to study the impact of life-long exposure to regularities on predictive processing.

At the most basic level, experience with language affects perception of language. When learning a linguistic system, we learn to map certain speech sounds into categorically distinct units. As a consequence, speakers of different languages categorize the same physical speech inputs (e.g., phonemes) in different ways depending on their linguistic background, thus changing their categorical perception of auditory events. For instance, speech sounds sharing the same voice onset time (in between 0 and 30ms) are perceived as voiced plosive (b, d, g) in English but as voiceless plosive (p, t, k) in Spanish.

Language-based prior knowledge can also affect visual processes such as object recognition, discrimination and detection (see Lupyan et al., 2020 for a review). Let's consider object recognition as example. The process of object recognition consists in relating the structure of an incoming visual stimulus to an internal, previously learned category or state. A typical paradigm that have been used to study the effect of prior knowledge on object concerns the use of ambiguous "Mooney" images (see Figure 3 below). The objects behind these images are difficult to recognize for 75% of people, as the sensory evidence does not provide unambiguous information to recognize them i.e., the perceptual system fails to assign meaning to these images. In a recent study, Samaha et al., (2018) demonstrated that priming "Mooney" images with linguistic material can disambiguate object recognition, resulting in an increase of 89% of recognition performance. Within a predictive

processing perspective, such effect can be explained by labels helping to form precise expectation against which the stimulus can be confronted, thus making an ambiguous perceptual stimulus completely interpretable.

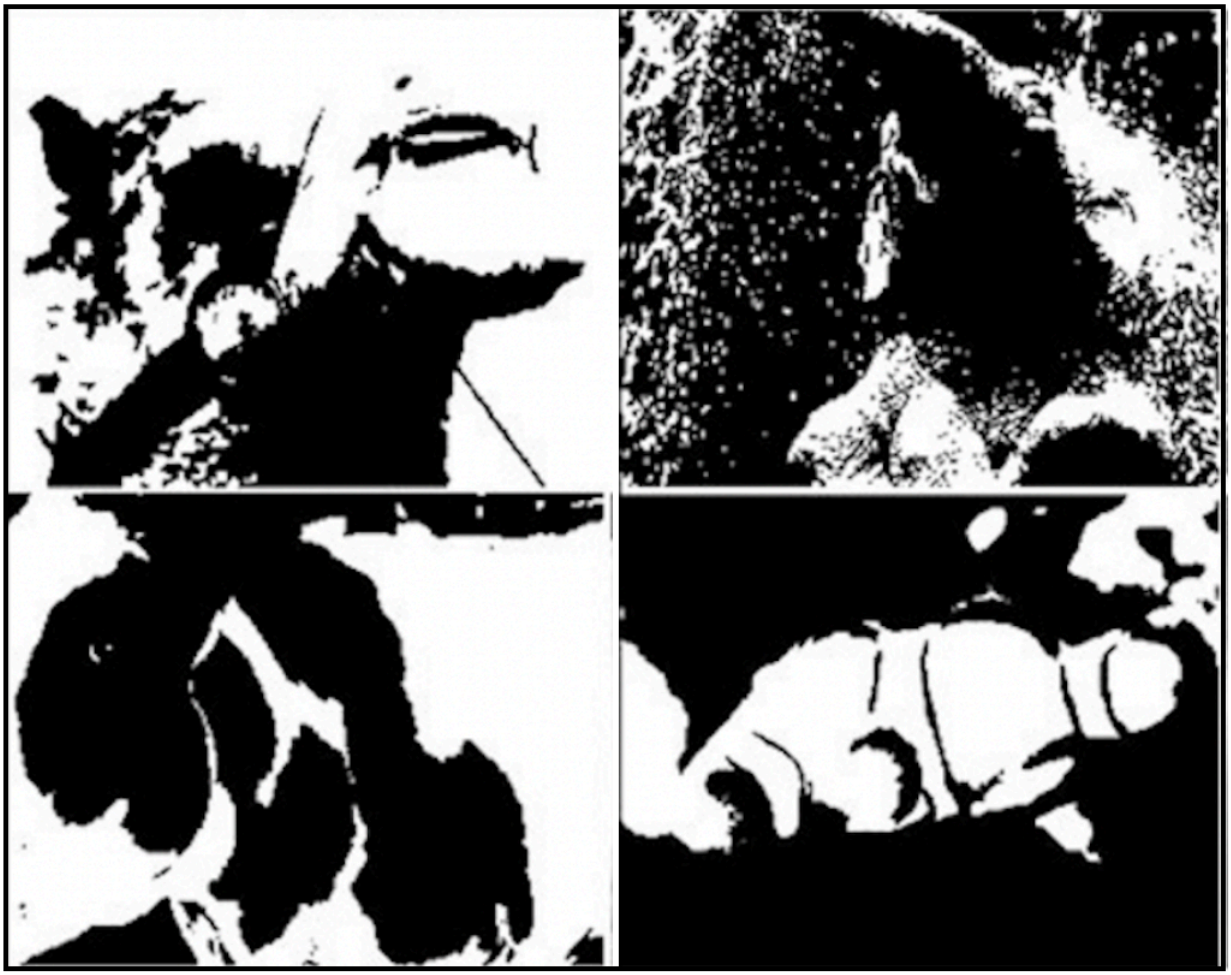


Figure 3. Examples of ambiguous “Mooney” images. The images on left refer to a camel (top) and a cow (bottom) respectively. The images on right refer to an elephant (top) and a fish (bottom) respectively. Images adapted from Hsieh et al. (2010).

Language can also affect low-level auditory processes, such as the categorization of simple sequences of non-linguistic sound. One example comes from auditory rhythm perception. When perceiving sequences of sounds, a basic operation of the auditory system consists in merging short sequences of tones into higher-level events – a phenomenon

known as perceptual grouping. In a seminal study, Iversen et al., (2008) showed that the way we group sequences of non-linguistic sounds is largely influenced by our native language. He tested how English and Japanese listeners perceive simple sequences of two tones alternating in duration (one short, one long, one short, etc) by asking them their preferential grouping pattern. They found out that the two groups provided asymmetrical responses: English speakers had a preference for a short-long pattern, while Japanese reported a long-short grouping bias. Crucially, these patterns mirrored the rhythmic/prosodic structures of the two languages. This study was subsequently replicated with Basque-dominant and Spanish-dominant speakers (Molnar et al., 2016). A possible mechanism explaining this effect is that regularities governing the rhythmic structure of a language become encoded in the tuning properties of the auditory system in the form of long-term priors, which are then used to generate expectation during auditory processing of rhythmic sounds. The studies on perceptual grouping show how life-long exposure to certain linguistic patterns can affect auditory perception at very early stages.

Despite the studies reported so far have been interpreted within a predictive processing perspective, there is not general consensus that this interpretation is correct. Indeed, many argue that language does not bias perception itself in a predictive way, but only interacts with later processes like working memory or categorical decision-making, which arise after visual processes (Pylyshyn, 1999; Klemfuss et al., 2012). On this account, the effect of language on perception are primarily post-perceptual. An example of bias on perception driven by decision responses is the Stroop effect, where the response about the stimulus color could overlap with the automatic response activated by the (irrelevant) lexical item, thus resulting in interference. Similarly, in the study by Samaha et al., (2018) mentioned above, the effect of linguistic hints on the recognition of ambiguous “Mooney” images could be coherent with an interaction at categorical decision-making stages. Linguistic hints could

have activated perceptual decision at higher-order stages rather that providing top-down guidance to the visual system. Predictive and the post-perceptual accounts of the effect of language on perception make often similar predictions at the behavioral level. In the next section, we discuss a possible solution on how to disentangle the two accounts.

1.5 How to measure predictive processing

As I mentioned above, it is often difficult to understand the mechanism underlying the effect of language on perception based only on behavioral data. However, time-resolved electrophysiological tools like electroencephalogram (EEG) and magnetoencephalogram (MEG) might provide some advantages over behavioral measures. Here we discuss two unique assumptions of predictive processing models that can be tested using neuroimaging tools which would help to understand the origin of the effect of language on perception.

1.5.1 The Mismatch Negativity: an index of cortical prediction error

The first assumption concerns the fact that predictive signals should modulate responses at each level of the cortical hierarchy, including early sensory areas. This can be measured, among others, by targeting event related potentials (ERPs). ERPs are positive- or negative-going waves that emerge in the electroencephalogram in response to certain events (for a more detailed discussion on the nature of ERPs, see the Methods chapter). Some ERPs components are putatively associated to low-level processes. For instance, in the visual domain, the P1 component (peaking around 100ms after the onset of the visual stimulus) is known to be generated in early visual cortices and being sensitive to low-level visual features like lightness and contrast (Luck, 2014). Similarly, in the auditory modality, a component that has been typically associated to low-level auditory processing is the mismatch negativity (MMN) and its neuromagnetic analog called MMNm. This ERP/MEG component is elicited by sudden changes in the acoustic environment. It is usually investigated using the Oddball

Paradigm or its variations, where a regular sequence of tones is disrupted by a novel unexpected event. It peaks at around 100-250ms from the event onset and shows a strong intensity in frontal and temporal regions, although it is generated in auditory areas. The MMN is usually calculated by subtracting the event-related response elicited by a standard event from the response of a deviant event. A large number of studies have shown that this component is strongly modulated by predictability of incoming input and transition probabilities (i.e., prior probabilistic knowledge; Garrido et al., 2009). Indeed, many studies have shown that the same auditory event presented in different contexts (e.g., predictable vs non-predictable) can generate a strikingly different MMN response. Because of its sensitivity to predictability and prior knowledge, the auditory MMN is considered a lower-level prediction error signal (Friston, 2005; Garrido et al., 2009), reflecting the difference between the brain response to an actual input and its prediction. It has been found to be generated even in non-attentive states, as well as during sleep, states of coma and anesthesia (Dehaene & Changeux, 2011; Bekinschtein, et al., 2009). Given its automatic nature, the MMN has been suggested by some to reflect a “primitive intelligence” of the auditory system (Näätänen et al., 2001).

1.5.2 Alpha and beta oscillations: an index of prediction

Another unique assumption of predictive processing models is that top-down prediction modulates brain activity before the presentation of a stimulus. For instance, if the barking of a dog could lead an individual to generate an expectation about the likely presence of a dog, then a neural signature of such content-specific expectation should be detected in brain activity before the individual identifies the actual dog in a scene. Based on several human and monkey studies, a candidate mechanism to carry this type of expectation are neural oscillations in the alpha and beta frequency bands.

Neural oscillations are electromagnetic signals that reflect the on-going rhythmic behavior of neural populations at different spatial and temporal scales. Despite their existence is known since the discovery of the alpha rhythm by Hans Berger in the 1930s, it is only recently that these rhythms have been suggested to play an important role in perception and cognition (Klimesch, 1999).These signals can be detected, among other, using non-invasive EEG and MEG, and are categorized based on their frequency: delta (2–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12– 30 Hz), and gamma bands (30–100 Hz). For a more detailed explanation on how this type of activity can be detected in the EEG, see the Methods chapter. Distinct oscillatory-frequency bands have been associated to different processes. For instance, gamma oscillations have been largely associated to feedforward signals, while alpha and beta traditionally reflect endogenous feedback projections (van Kerkoerle et al., 2014). Within the predictive processing framework, this frequency asymmetry reflects a functional asymmetry between top-down prediction and bottom-up prediction error.

As I mentioned above, one candidate mechanism to carry sensory predictions is the oscillatory activity in the alpha frequency band (8-12Hz). Alpha oscillations have been linked to different aspects of top-down processing, such as prediction, attention and working memory. Enhancement in alpha frequency have been reported for instance when attention is directed by a cue towards a specific feature or direction (Worden et al., 2000; Snyder and Foxe, 2010), or when a sequence of events is retained in memory (Jensen et al., 2002). There are currently two main non-inclusive accounts about the alpha rhythm. The first one considers alpha as reflecting states of inhibition and filtering of task irrelevant information (Jensen and Mazaheri, 2010; Klimesch et al., 2007). Such an inhibitory function of the alpha rhythm has been reported in different fields of attention, including spatial, feature- and object-based attentional selection (Thut et al., 2006; Snyder and Foxe, 2010; Knakker et al., 2015). For instance, when attention is directed towards a target in one side of space,

posterior alpha-band power increases at electrodes over the hemisphere ipsilateral to the target (Worden et al., 2000; Thut et al., 2006). Alpha has also been found to correlate with working memory load, which has been suggested to reflect inhibition of task-irrelevant information in order to “protect” task-relevant representations (Jensen et al., 2002). More recent proposals ascribe to neural alpha synchronization a variety of roles in top-down processing (Palva & Palva 2007; Klimesch 2012; van Kerkoerle et al. 2014). Enhancement of alpha waves in task-relevant regions may have excitatory effects reflecting selective amplification of neural representations of object categories (Mo et al., 2011). For instance, monkey studies demonstrated that sustained alpha power in the inferotemporal cortex increases before the onset of a cued target image, and that such increases are associated to a facilitation in the processing of a subsequent visual stimulus (Mo et al., 2011). These and many other findings suggest that alpha may support, among many cognitive functions, the endogenous deployment of perceptual knowledge in task relevant circuits.

Another brain rhythm which represents a candidate mechanism to carry perceptual prediction is the beta wave. Together with alpha waves, beta oscillations are ubiquitous in the brain. Initially implicated in sensorimotor planning and processing (Hari and Salmelin, 1997), beta oscillations have been recently associated to different top-down processing. For instance, beta waves have been proposed to mediate the balancing between internal states and response to external stimuli (Engel and Fries, 2010), the binding of neurocognitive network elements underling a given neural representation (Bressler and Richter, 2015), and the endogenously driven transitioning from latent to active cortical representations of categories (Spitzer and Haegens; 2017). Modulations of beta oscillations have been also associated to prediction of the timing and content of sensory events (Arnal and Giraud, 2012). For instance, some studies have shown that during processing of rhythmic sequences of sounds, beta power increases before the onset of each auditory event (Fujioka et al., 2012). Given its anticipatory nature, such beta bursts have been suggested to reflect

endogenous predictive signals. Interestingly, in contrast to alpha that has been traditionally associated to perceptual processes, beta oscillations have been suggested to encode also supramodal aspects of events. Several human and monkey studies reported that beta synchronization over parietal and frontal regions carry information about object categories (Antzoulatos and Miller, 2014, 2016). In particular, recent studies using categorization tasks have proposed that beta may encode abstract, supramodal properties of object categories (Wutz et al., 2018; Haegens et al., 2017). Similarly, working memory experiments on scalar magnitudes like stimulus duration, motion speed or approximate number showed that beta power modulations are sensitive not only to concrete sensory features of the stimuli, but also to high-level abstractions of the task-relevant magnitude (e.g., “being higher/lower than”; Spitzer et al., 2014).

Despite the precise role and functioning of the two neighboring bands is still unclear, these studies suggest a possible division of labor between alpha and beta in top-down predictive processing.

Chapter - 2 Methods

2.1 Introduction to magnetoencephalography (MEG) and electroencephalography (EEG)

EEG and MEG are neuroimaging tools used to map brain activity. EEG measures the electrical fields generated by brain cells (Berger, 1929), while the MEG allows the recording of brain magnetic fields (Cohen, 1972). The EEG and MEG are considered similar techniques as they both record the same type of signal, that is, ionic currents generated by biochemical processes at the cellular level. Together with electrocorticography (ECoG), these techniques belong to the group of electrophysiological tools that provide a direct measure of neural activity.

A main advantage of these techniques concerns the excellent temporal resolution. Indeed, these methods are able to track the temporal unfolding of brain activity in a milliseconds scale (Hämäläinen et al., 1993). These features complement those of other neuroimaging tools like functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), which have an excellent spatial resolution (in the order of millimeters) but limited temporal resolution.

EEG and MEG are non-invasive techniques. Indeed, they can record the neural electrical activity via electrodes placed on the scalp surface. On the contrary, ECoG, electrodes are placed directly on the brain surface of patients with epilepsy. However, given their distance from the sources generating the brain signal, EEG and MEG can only detect the synchronized activity of thousands of neurons acting in synchrony (Hämäläinen and Hari, 2002). Despite this limitation, MEG and EEG still represent a valid tool to investigate brain and cognitive phenomena arising at the network level.

The main difference between EEG and MEG lies in their spatial resolution. While EEG has a relatively poor spatial resolution, MEG methods allow to reconstruct brain source with a precision in the order of millimeters for cortical regions. The main reason behind this difference is that the scalp is a poor conductor of electrical signal. Thus, it is very challenging

to reconstruct the origin of the current from their topographical distribution. Moreover, the currents recorded by EEG electrodes come from different directions. This makes the isolation and source reconstruction of EEG signal recorded from the scalp even more challenging. On the other hand, magnetic fields are minimally distorted by the scalp, skull and other tissues, thus allowing to infer the origin of brain patterns with higher precision.

2.2 Brain activity recorded with MEG and EEG

Electrical activity generated by a single neuron cannot be recorded from the scalp. MEG and EEG techniques can indeed record only activity generated by large populations of neurons forming functional units. Concretely, this means that neurons within an assembly should fire in temporal synchrony. However, temporal synchrony is not enough for allowing neural recording from distant location: if neurons are not spatially oriented in a similar manner, their currents cancel out each other. Thus, in order to produce brain activity patterns that can be measured from the scalp, neural assemblies should also have a similar spatial orientation. A typology of neurons that provides all this set of features are pyramidal neurons. These neurons are organized in a sort of palisade structure, with the axes of their dendrites parallelly aligned, and their body perpendicular to the cortical surface. The synchronized activity generated by populations of pyramidal neurons produce laminar currents. Such currents give rise to electrical fields. In parallel, magnetic fields are generated around the electrical fields. Electrical and magnetic fields generated by neural populations of pyramidal neurons can be measured from the scalp, thus forming the basis of EEG and MEG signals.

The electrophysiological signal recorded from time-resolved techniques such as EEG and MEG is multidimensional, meaning that it contains different aspects of neural activity such as time, space, amplitude and frequency. Time-resolved analytical approaches such

as time-domain and time-frequency-domain analyses allows to extrapolate and analyze different dimensions of the signal.

2.2.1 Evoked activity: time domain

Time-domain analyses, such as ERP analyses (and their MEG equivalent, the ERF or Event-Related Fields), have been largely employed to study different aspects of perceptual and cognitive processes. The typical way to study ERP/F is by averaging single trials time-locked to different experimental conditions. The output of such averaging is then expressed in relation to a baseline period. By averaging brain activity time-locked to a specific event over multiple repetitions, brain patterns which are systematic in time and amplitude emerge in the signal over brain activity patterns not directly associated to the experimental condition. The output of such process results in a smooth positive or negative deflection in voltage or magnetic field, which reflects the so called ERP/F.

ERP/F are usually categorized based on their spatio-temporal properties. The main advantage of this approach is that it allows to study perceptual and cognitive processes with high temporal resolution. However, it also has certain limitations. By adopting a univariate approach based on an averaged-response procedure, this method reduces the complexity of the electrophysiological signal to a single variable (i.e., an ERP/F component). This means that this data analysis technique does not allow to track multiple processes occurring in parallel.

2.2.2 Oscillatory activity: time-frequency domain

A different approach to analyze the electrophysiological signal is by using time-frequency domain approaches. Compared to time-domain analyses, this approach takes into account the oscillatory nature of neural activity as an additional variable. Technically speaking, this approach captures neural activity that is time-locked (like ERPs) but not necessarily phase-locked to a specific event. Instead of reducing the brain signal in a single variable, this

method allows to decompose the brain signal into different components based on different frequencies i.e., neural oscillations. These oscillations are usually categorized according to their frequency bands: delta (< 4 Hz), theta (4-8 Hz), alpha (9-12 Hz), beta (13-30 Hz) and gamma (> 30 Hz). Indices of oscillatory activity are obtained by decomposing the brain signal via mathematical methods based on Fourier analysis. This method, applied to a sliding window, allows to obtain a time-varying power spectra. The resulting data can be then multiplied using Hanning taper in order to control spectral leakage and the amount of frequency smoothing. The length of the time window has an impact on the temporal and frequency resolution of the condition of interest. Longer time windows provide a higher frequency resolution at the expense of temporal resolution.

This frequency decomposition allows to differentiate distinct oscillatory patterns occurring at different timescales. Brain oscillations are often organized hierarchically, with oscillations in higher frequencies being nested within slower frequency oscillations. Being a multivariate approach, time-frequency decomposition allows to investigate parallel processes, that is, brain/cognitive dynamics carried at different frequency scales but occurring in the same temporal window. It must be noticed, however, that time-frequency methods provide a lower temporal resolution than time-based approaches such as ERP/F. Time-frequency and time-domain approaches thus provide complementary insights into the neural profile of cognitive dynamics, favoring the investigation of both top-down mechanisms and bottom-up responses.

2.3 Overview of content

In what follows, I will address some critical issues about the effect of language on perception from a predictive processing perspective. Chapter 3 concerns an EEG and behavioral study on highly proficient Basque-Spanish bilinguals. It focuses on a well-known behavioral effect,

the label-advantage in object recognition i.e., the fact that objects are recognized faster when cued by a word compared to an equally familiar natural sound. I will address whether such label-advantage arises at the perceptual versus semantic/decision-making level by exploring whether words facilitate object recognition by modulating prestimulus activity before the appearance of the actual object. Specifically, I will test whether neural oscillations play any role in deploying top-down priors during language-mediated visual object recognition. The bilingual component of the experiment will allow to extend the conclusions of this study to a second language, and to address some questions about semantic and top-down processing in bilinguals. In chapter 4, I will use language as a model to test whether the human brain generates predictions based on long-term priors during musical beat perception. Here I will take a different approach: I will compare MEG data from people with different linguistic backgrounds (Basque dominant bilinguals vs Spanish monolinguals) listening to rhythmic sounds intermitted by rare violations. The objective is to investigate whether the auditory system learns structural rules encoded in the rhythmic structure of language and uses them to generate predictions in other (non-linguistic) domain. Here I will focus on the MMN – an early ERP/F component putatively associated to prediction error signals.

Chapter - 3 Alpha and beta rhythms differentially support the effect of symbols on visual object recognition

3.1 Introduction

Hearing certain natural sounds (e.g., the croak of a frog) appears to automatically activate conceptual knowledge, enabling the perceptual system to quickly identify objects in the surroundings (e.g., the presence of a frog). Learning such cross-modal associations represents a crucial prerequisite for mediating interactions with the environment. In humans, conceptual representations can also be activated via language (e.g., “frog”). However, unlike natural sounds, linguistic symbols are categorical, making them uniquely suited to activate semantic information in a format that transcends within-category differences. Whether phylogenetically young systems like language exert similar effects on perception as natural sounds do, and which brain dynamics support such effects is still incompletely understood. In the present study, we test the hypothesis that language boosts visual processes by sharpening categorical priors via the modulation of alpha/beta oscillations.

Conceptual representations activated by auditory cues have been shown to interact with the visual system in different ways. For instance, hearing words and natural sounds can rapidly drive visual attention towards specific entities in a scene (Huettig and Altmann, 2007); facilitate the recognition and discrimination of congruent object categories (Edmiston and Lupyan, 2015; Boutonnet and Lupyan, 2015); lower the detection threshold for ambiguous objects (Lupyan and Ward, 2013); and even cause sensory illusions (Toskos Dils and Boroditsky, 2010). While this body of evidence suggests that both linguistic and non-linguistic cues activate content-specific representations, it is currently less clear whether these cues activate the *same* representations. Studies directly targeting this issue have often reported a “label-advantage” effect – that is, a facilitation in the recognition of objects when preceded by words compared to non-linguistic cues (Edmiston and Lupyan, 2015) – suggesting that language represents a more powerful tool to enhance visual processing.

To achieve these facilitatory effects on visual perception, linguistic categories could theoretically follow two possible pathways. On one account, language would not bias perceptual processes at early levels, but rather interact with later processing stages such as categorical decision-making. On an alternative account, words can affect visual processing by setting categorical priors with the effect of altering early perceptual processing (Simanova et al., 2016). Support for the latter account comes primarily from EEG studies showing that the better recognition of images preceded by congruent words was associated with modulations of early ERP such as the P1 (Boutonnet and Lupyan, 2015, Noorman et al., 2018) – putatively considered an electrophysiological index of low-level visual processes (Spehlmann, 1965). Yet, these experiments targeted the perceptual consequences that language has on visual behavior i.e., they focused on the post-stimulus time interval. The mechanisms that underlie prestimulus effects of language on visual perception are currently poorly understood.

Analysis of oscillatory activity provides an excellent opportunity to study prestimulus language-driven modulations in sensory areas. Based on previous human and animal studies, a candidate mechanism to carry perceptual priors are oscillations in the alpha/beta frequency range. Rhythmic brain activity in these bands has been repeatedly associated with top-down processes (Michalareas et al., 2016, Bressler and Richter, 2015; Arnal and Giraud, 2012).

In the present study, we used a cue-picture matching task to test the hypothesis that language enhances visual object recognition by setting categorical priors via the modulation of alpha/beta oscillations. In contrast to previous studies, we (i) focused on the time interval preceding the onset of the visual object, targeting top-down signaling directly; and (ii) included words in both first (L1) and second language (L2), in order to assess whether the previously reported label-advantage extends to language systems acquired later in development. We hypothesized that, if the label-advantage arises because words provide

refined categorical priors to the visual system, then differences in recognizing objects when cued by words vs. natural sounds should be associated with modulations of oscillatory alpha/beta dynamics before the onset of the target picture. Importantly, we should also expect such an oscillatory index to be linked to behavioral performance.

3.2 Materials and methods

Participants

We tested a total of twenty-five Basque-Spanish bilingual speakers. Notice that in earlier studies investigating the label-advantage in object recognition, a sample size of 15 participants was sufficient to detect the behavioral label-advantage effect (Boutonnet and Lupyan, 2015). Participants were native speakers of Basque who began acquisition of Spanish after three years of age (13 females; age range 18-33, mean: 25.66, SD: 5.45, age of acquisition of Spanish: 4.23 y.o., SD: 1.33). All participants were right-handed, with no history of neurological disorders. Their vision was normal or corrected to normal and received a payment of 10€ per hour for their participation. Before taking part in the experiment, all participants signed an informed consent form. The study was approved by the Basque Center on Cognition, Brain and Language (BCBL) ethics committee in compliance with the Declaration of Helsinki. Participants completed several language proficiency tests in both Spanish and Basque (see Table 1). First, participants were asked to self-rate their language comprehension (on a scale from 1 to 10, where 10 is a native-like level). All participants rated themselves as highly proficient in both Basque and Spanish. Participants also performed “LexTALE”, a lexical decision task (Izura et al., 2014; Lemhofer and Broersma, 2012) that tested their vocabulary knowledge. They displayed similarly high scores in both Spanish and Basque. In addition, participants had to name a series of pictures

of increasing difficulty in both languages (65 pictures in total). Here as well, participants achieved native-range scores in both languages. Finally, all participants were interviewed by balanced bilingual linguists who rated them on a scale from 0 to 5: no participants had a score below 4 in either language.

Measure	**Basque**	**Spanish**
Self-evaluation (0-10)	9.04 (0.16)	9.39 (0.24)
LexTALE Basque (0-50); Spanish (0-60)	46.04 (2.67)	54.09 (4.13)
Picture naming (0-65)	64.19 (1.47)	63.38 (1.62)
Interview (0-5)	5 (0)	4.95 (1.33)

Table 1. General proficiency assessment of the participants' linguistic profile.

Stimuli

The visual stimuli included 50 pictures from 10 object categories, referring to both animate (e.g., bird) or inanimate entities (e.g., camera). Each of the 10 categories was represented by 5 different highly recognizable images (.png extension, white background, 2000x2000 pixels): three color photographs obtained from online image collections, one normed color drawing (Rossion and Pourtois, 2004), and one "cartoon" image (Lupyan and Thompson-Schill, 2012). We selected different instances for each category in order to provide visual heterogeneity.

The audio stimuli included 10 words in Basque (L1), 10 in words Spanish (L2) and 10 natural sounds, each referring to one of the object categories. Words, both in Basque (L1) and in Spanish (L2), were recorded by a Spanish-Basque female speaker. Natural sound stimuli were downloaded from online libraries. Overall, the mean length of the audio stimuli was 0.8 ± 0.05 seconds (Word in L2, mean: 0.81 s, SD: 0.21; Word in L1, mean: 0.77 s, SD: 0.23; Natural Sounds, mean: 0.84 s, SD: 0.2).

In order to test that sounds and images were unequivocally identifiable, a side test was performed. A group of Basque-Spanish bilinguals (N=20) who did not take part in the experiment viewed several images and listened to different sounds. They were told to name the visual and audio stimuli they perceived with the first noun that came to their mind. For the present experiment, we only chose the images and sounds whose names were expressed by all 20 participants. In total, we selected 50 images from 10 categories, 10 words in Basque, 10 words in Spanish, and 10 natural sounds.

Procedure

The EEG study was run in a soundproof electrically shielded chamber with dim light. Participants sat on a chair, about sixty centimeters in front of the computer screen. Stimuli were delivered using PsychoPy software (Peirce, 2007). We followed the procedure illustrated by Boutonnet and Lupyan (2015). Participants completed a cued-picture recognition task composed of 300 trials (see Fig. 4). On each trial, a fixation point appeared on the center of the screen for one second, then participants heard an auditory cue: either a word in L1, (e.g., igela, “frog), a word in L2 (e.g., rana, “frog”) or a natural sound (e.g., a croak).

After 1s from the offset of the cue, a picture appeared on the screen, and participants had to respond “yes” or “no” by pressing one of two buttons on a keyboard to indicate whether the picture did or did not match the auditory cue at the category level. The picture

remained on the screen until the participant's response. In 50% of the trials the picture matched the auditory cue (congruent trials), while the other 50% was a mismatch (incongruent trials). In the case of incongruent trials, a picture belonging to a different category appeared on the screen. Stimuli presentation was randomized for each participant. The entire experiment lasted 40 minutes on average.

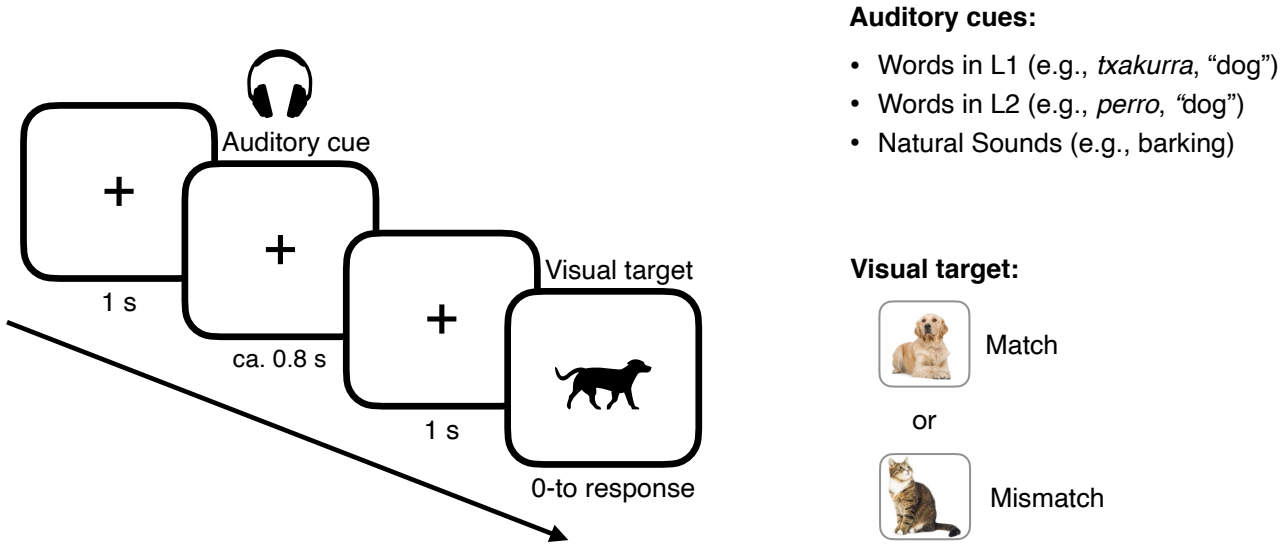


Figure 4. Illustration of the design and trial structure with an example for each possible auditory cue (words in L1, words in L2, natural sounds) and target object condition (Match, Mismatch).

EEG recording

Electrophysiological activity was recorded from 27 electrodes (Fp1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP1/2, CP5/6, P3/4, P7/8, O1/2, F/C/Pz) positioned in an elastic cap (Easycap) according to the extended 10–20 international system. All sites were referenced to the left mastoid (A1). Additional external electrodes were placed on the right mastoid (A2) and around the eyes (VEOL, VEOR, HEOL, HEOR) to detect blinks and eye movements.

Data were amplified (Brain Amp DC) with a filter bandwidth of 0.01-100 Hz, at a sampling rate of 250 Hz. The impedance of the scalp electrodes was kept below 5 kΩ, while eye electrode impedance was kept below 10 kΩ.

EEG preprocessing

All EEG data analysis was performed using Matlab 2014 with the Fieldtrip toolbox (Oostenveld et al., 2011; http: www.fieldtriptoolbox.org) and R (R Core Team, 2015; https: www.r-project.org). For data visualization, we used Matlab or FieldTrip plotting functions, R and the RainCloud plots tool (Allen et al., 2019). The recordings were re-referenced off-line to the average activity of the two mastoids. Epochs of interest were selected based on cue type (word in L1, word in L2, natural sounds) and congruency (match, mismatch), resulting in six different sets of epochs. They were computed from −3 s to 1.5 s with respect to image onset.

Trials in which subjects provided incorrect responses in the behavioral task were removed from the analysis. Spatial-temporal components of the data containing eye and heart artifacts were identified using independent component analysis and subsequently removed. Overall, we removed an average of 2.14 components per subject. We then identified epochs containing additional 'muscle' and 'eye blink' artifacts using an automatic artifact detection procedure (z-value threshold = 12). Trials selected as possibly contaminated by artifacts were visually inspected and removed (~8%). Finally, we removed a few additional trials containing artifacts using a visual inspection procedure (~0.11%). Three participants were excluded from the analysis because more than 25 % of the trials were rejected.

Statistical analysis

Behavior. We used the R environment (version 4.0.0; R Core Team 2020) and lme4 package (Bates et al., 2014) to perform mixed effect regression on reaction time data, following a procedure similar to that illustrated in Boutonnet and Lupyan (2015). Predicted reaction times (calculated from the onset of the target image up to the participant's response) were computed by fitting the model with cue-type (words in L1, words in L2, natural sounds), congruency (match, mismatch) and their interaction as fixed factors, and by adding by subject random slopes for the effect of cue type and congruency. Subsequent pairwise comparisons were performed using estimated marginal means (Bonferroni-corrected for multiple comparisons) with emmeans (Lenth, 2018). Because no reliable interaction was detected, post-hoc comparisons were based on a model with the same syntax as the one presented above but without including the interaction term, in order to facilitate the interpretability of post-hoc analysis. Accuracy was not analyzed statistically because it was near ceiling (98%). For the analysis of behavioral data, we excluded the same three participants that were excluded from the EEG analysis. Moreover, we excluded all incorrect trials (1.88%), as well as a few trials in which participants' responses exceeded 3 s (0.28%). These trials were also excluded from the EEG analysis. Before entering the statistical models, reaction times were log-transformed to improve normality.

Spectral power. A time-frequency analysis of artifact-free EEG trials was performed. Before applying spectral decomposition, the latency of each epoch was reduced to −1.5 s to 0.5 s with respect to image onset. The time-varying power spectrum of single trials was obtained using a Hann sliding window approach (0.5 s window, 0.05 s time steps) for the frequency range between 0 and 30 Hz, zero-padded to 1 s providing a frequency resolution of 1 Hz. Our focus on oscillatory activity up to 30 Hz was motivated by the fact that top-down processes are often associated with oscillations within this frequency band, while higher frequencies have been traditionally linked to bottom-up processing (e.g., Bosman et al., 2012). For the statistical analysis, we computed a single power spectral density estimate

for each participant, channel, frequency and epoch by averaging the spectral estimates centered on the -0.75 s to -0.25 s time interval. We selected this time-interval in order to yield more accurate spectral estimates, as activity here is largely uncontaminated by activity evoked by the preceding auditory event or the subsequent visual stimulus.

Grand-average power spectrum. In order to compute the power spectrum, spectral estimates corresponding to congruent and incongruent trials for each cue-type condition were combined, resulting in three different data sets for each cue-type (words in L1, words in L2, natural sounds). Note that subjects were not aware of incongruency in the prestimulus time window, thus the time-frequency representations at this stage should be indistinguishable for the congruent and incongruent conditions. Then, spectral estimates were averaged over trials, participants, channels and cue-type conditions, resulting in a single value for each of the 30 frequency bins (i.e., the grand-average power spectrum). A peak finding algorithm was used to identify spectral peaks as local maxima in the grand-averaged power spectrum. Two peaks, one at 10 Hz and one at 18 Hz emerged from this analysis (Fig. 5A). Based on these peaks, frequencies of interest (FOI) were obtained as the average of the frequency peaks ±1 Hz: that is, 9-11 Hz and 17-19 Hz respectively (Fig. 5B). We refer to these band estimates as the alpha and beta band power.

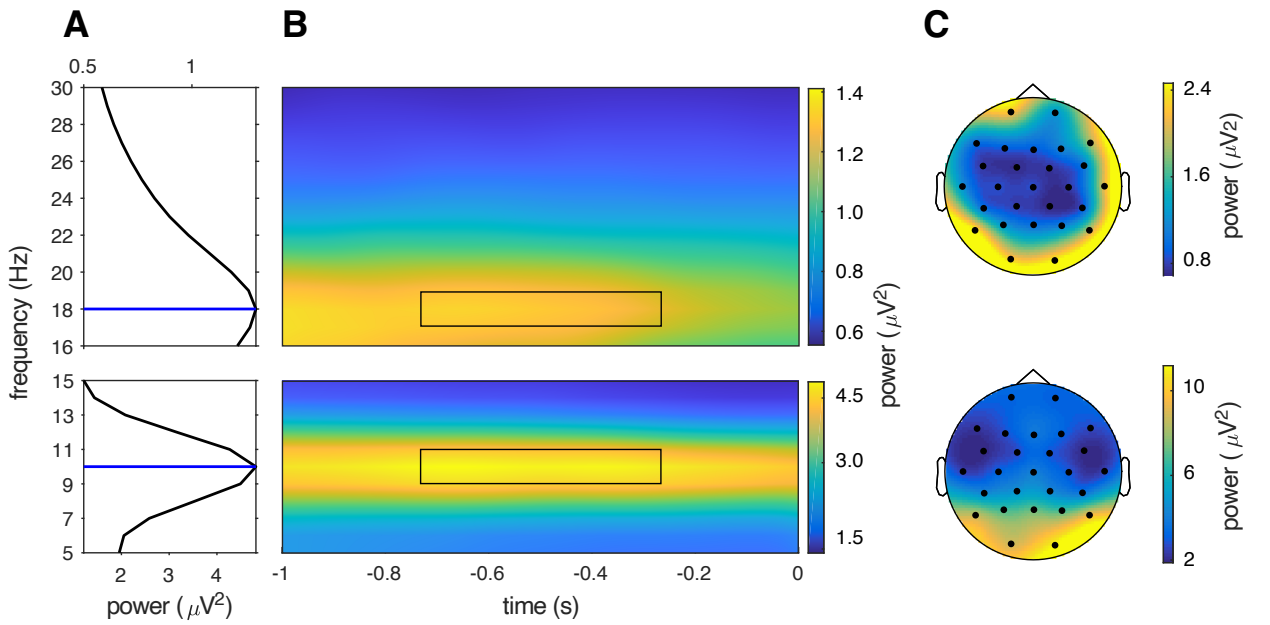


Figure 5. A) Alpha and beta peaks in the grand-average raw power spectrum of all epochs across conditions, during the -0.75 -0.25 pre-target time-interval. The blues lines indicate the power peak as local maxima. B) Time-frequency representation of grand-averaged data for the alpha and beta-band, in the 1 s time window between the offset of the auditory cue and the onset of the image. The black rectangle denotes the time-frequency interval selected for the statistical analysis. C) Topography of the time-frequency interval of interest.

Prestimulus spectral differences between cues. Spectral estimates for each cue-type (words in L1, words in L2, natural sounds) were averaged over trials. To reduce individual differences in overall EEG power, normalization was applied by converting the time-frequency power for each condition into percent signal change relative to the average power over all three conditions and channels, as performed by Bogaerts et al., (2020). This procedure removes individual differences in signal power, without distorting the relative magnitudes of the conditions, i.e. it functions as a baseline correction, when an appropriate baseline interval is not available. In order to test whether time-frequency representations in the prestimulus time-window differed across cue types, a non-parametric approach was selected (Maris and Oostenveld, 2007). For each FOI, we implemented a cluster-based permutation test based on a dependent sample F-test with the spectral data for each type

of cue (words in L1, words in L2, natural sounds) as the dependent variable. This approach is equivalent to a one-way ANOVA but allows to account for the spatial correlation between electrodes (i.e., no a priori region of interest needs to be defined). The minimum number of neighboring electrodes required for a sample to be included in the clustering algorithm was set at 2. The cluster threshold F-value (or t-value) was set at an alpha value at the 85th percentile of their respective distributions. Note that this parameter does not impact the false alarm rate of the test. Rather, it sets a cluster threshold for determining when a sample should be considered as a candidate member of a cluster. Small cluster thresholds usually favor the detection of highly localized clusters with large effect size, while larger cluster thresholds favor clusters with large spatio-temporal extent, and more diffusion of the effect (Maris and Oostenveld, 2007). Because alpha and beta rhythms usually emerge at the network level, we selected a relatively large cluster threshold, i.e. capturing what appears to be quite a globally distributed effect. The number of permutations for the randomization procedure was set at 100000. The critical alpha-level to control the false alarm rate was the standard $\alpha = 0.05$. All resulting p-values were Bonferroni corrected for the number of FOIs. For each FOI, one significant cluster was detected. In order to assess the directionality of the effect, post-hoc non-parametric pairwise comparisons were applied. Specifically, power values for each cue-type condition were averaged over all electrodes belonging to the significant cluster and compared pairwise using paired t-tests. The alpha-level for the three post-hoc t-tests was Bonferroni corrected for the number of comparisons. This procedure was applied to each FOI separately.

For both the alpha and beta band, post-hoc t-tests revealed that brain data elicited by symbolic cues (words in L1 and L2) come from a similar probability distribution, while both significantly differed from brain activity elicited by natural sounds. This motivated us to pool the data from the spoken word conditions together and contrast this average with that from the natural sound condition using a dependent sample cluster-based t-test (using the

same parameters as for the test based on the F-statistic). One significant cluster for each FOI emerged from this comparison: one including 23 (alpha) and 21 (beta) out of 27 electrodes respectively. The power over these clusters served for the analysis of brain-behavior correlation.

Brain-behavior correlation. In order to investigate the link between prestimulus brain rhythms and behavior, correlation analyses were performed over participants. For the correlation analyses reported below, only the trials belonging to the congruent condition were selected; i.e., those trials where the auditory cue matched the object picture. The same analysis pipeline was applied for each FOI.

Trials were averaged, providing an alpha, beta and RT for each participant and condition. To ensure the correlation was not driven by differences between conditions, participants' values were z-scored within conditions. The three conditions were then averaged, providing an alpha-RT and beta-RT pair for each participant. The Spearman correlation was then computed for these pairs for each FOI. To assess the statistical properties of the alpha and beta correlations, we bootstrapped the data over participants. We performed this 100000 times, generating a distribution of bootstrap values. Following Efron and Tibshirani (1986), we computed the percentile bootstrap 95% confidence intervals, and used this distribution to perform statistical tests to determine the difference between the observed correlation coefficients and zero. We finally conducted a two-sample bootstrap test to evaluate the difference between the alpha-RT and beta-RT correlation coefficients (Efron and Tibshirani, 1986).

3.3 Results

Effect of cues on visual object recognition. We first analyzed the accuracy. Overall, accuracy was high (98%) and similarly distributed across the three conditions (words in L1 = 98%; words in L2 = 99%, natural sounds = 97%). Participants were clearly at ceiling here; thus,

we focused on the analysis of the reaction times. Analysis of reaction times showed a main effect of Cue-Type ($\chi^2(2) = 31.9500$, $p < 0.001$) (Fig. 6). This was subsequently unpacked via post-hoc comparisons. Pairwise comparisons using estimated marginal means showed that object images preceded by symbolic cues in both L1 and L2 were identified faster compared to images preceded by natural sounds (words in L1 – natural sounds: $\Delta = -0.08$, $SE = 0.01$, $p < 0.001$; natural sounds – words in L2: $\Delta = 0.06$, $SE = 0.01$, $p < 0.001$). On the other hand, the pairwise effect between words in L1 and words in L2 did not reach the significance threshold (words in L1 – words in L2: $\Delta = -0.02$, $SE = 0.01$, $p = 0.06$). As in previous studies, we also observed a main effect Congruency ($\chi^2(1) = 7.0329$, $p < 0.01$), with matching cue-pictures pairs leading to faster responses compared to mismatching pairs. No reliable Cue-Type by Congruency interaction was detected ($\chi^2(2) = 1.5310$, $p = 0.46$).

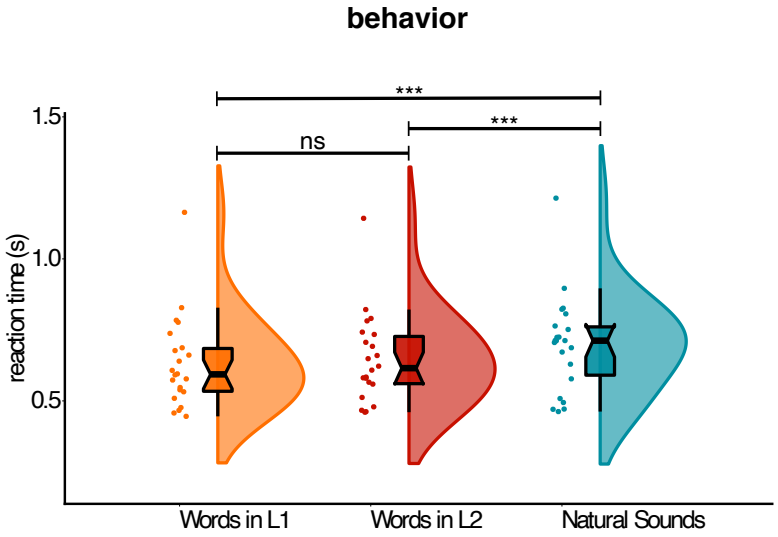


Figure 6. Mean reaction times (correct trials only) showing the main effect of cue-type on visual object recognition performance. Raincloud plots show probability density. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data). The notches indicate the 95% confidence interval around the median. Dots reflect individual subjects.

Effect of cues on prestimulus alpha rhythms. Differences between spectral power elicited by the three cue-type conditions were assessed using a cluster-based F-test for alpha and beta FOIs separately, focusing on the prestimulus interval. From the analysis of the alpha rhythm, one significant cluster was detected ($p < 0.01$, Bonferroni-corrected for the two FOIs) including several electrodes across the entire scalp (Fig. 7A). In order to assess the directionality of the effect, spectral power for each type of cue was averaged over all the electrodes belonging to the significant cluster and compared pairwise via t-tests. Pairwise comparisons showed that both words in L1 and L2 led to increased alpha power compared to natural sounds ($t(21) = 4.57$, $p < 0.001$ Bonferroni-corrected; $t(21) = 5.48$, $p < 0.001$ Bonferroni-corrected, respectively). No significant difference was detected between words in L1 and L2 ($t(21) = -1.70$, $p = 1$ Bonferroni-corrected).

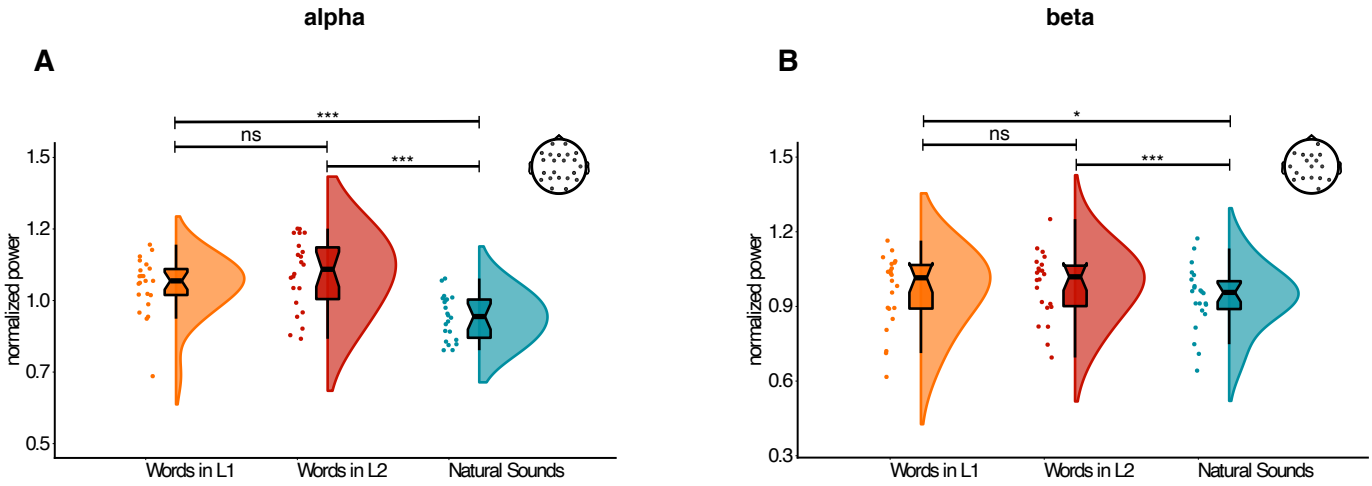


Figure 7. Effect of cues on pre-target alpha (A) and beta power (B) averaged over the electrodes belonging to the significant cluster. Conventions for the plot are the same of Figure 2. The electrodes belonging to the cluster are illustrated on the top-right of each figure.

Effect of cues on prestimulus beta rhythms. Analysis of the beta band revealed a pattern of results similar to the one that emerged from the analysis of the alpha rhythm. The

cluster-based F-test detected one cluster ($p < 0.01$, Bonferroni-corrected for number of FOIs) (Fig. 7B). Spectral power in the beta frequency range was averaged over the electrodes of the significant cluster for each type of cue separately and compared pairwise via t-tests. Beta power was larger when images were preceded by words in L1 and L2 compared to natural sounds ($t(21) = 2.68$, $p = 0.04$ Bonferroni-corrected; $t(21) = 4.68$, $p < 0.001$ Bonferroni-corrected, respectively), while no significant difference emerged when comparing words in L1 and L2 ($t(21) = -1.67$, $p = 0.33$ Bonferroni-corrected) (see Fig. 7B).

Relation between prestimulus alpha/beta rhythms and visual object recognition. The results reported so far point to a possible role of alpha and beta rhythms in supporting the label-advantage in object recognition. We further explored the relation between prestimulus alpha/beta oscillations and visual object recognition by correlating prestimulus spectral power and reaction times across participants.

Individual estimates for power and reaction times were correlated using the Spearman rank correlation. This method was selected because reaction time and beta power data significantly deviated from normality, as emerged from a Shapiro-Wilk normality test (reaction time: $W = 0.88$, $p = 0.01$; alpha power: $W = 0.92$, $p = 0.09$; beta power: $W = 0.89$, $p = 0.02$). We observed that both prestimulus alpha and beta power had a relation with reaction time performance (Fig. 8A). Yet, the directionality of the effect was opposite: alpha estimates were negatively correlated with reaction time performance in object recognition (Spearman's rho = -0.34, $p = 0.13$), while the relation between beta power and reaction time was positive (Spearman's rho = 0.33, $p = 0.13$). Though these correlations are not significant, the bootstrap percentile confidence intervals (Fig. 8B) provide marginal evidence that supports that the alpha correlation is less than zero, as shown by the 95% confidence interval only minimally exceeding zero (zero lies at the 93rd-percentile of the alpha bootstrap distribution). This can be expressed as a p-value using the bootstrap distribution to test the one-tailed hypothesis that the alpha-RT correlation is less than 0 ($p = 0.067$), which again

provides marginal evidence that the alpha-RT correlation is negative. Similarly, the bootstrap analysis shows evidence supporting the beta-RT relationship to be positive, with zero appearing at the 4.9th-percentile, which lies close to the border of the 95% confidence interval (Fig. 8B), and showing a significant difference from zero using the one-tailed hypothesis test ($p = 0.049$). Building on this evidence for opposing effects of alpha and beta power on RT, we explicitly tested that these correlations were in fact different. Figure 8B shows that the observed rho values for the alpha-RT and beta-RT correlations fall outside each others 95% confidence intervals. A two-tailed two sample bootstrap test was applied, which revealed that indeed the alpha-RT and beta-RT correlations are significantly different ($p = 0.012$), thus supporting the interpretation that these oscillations affect behavior differentially.

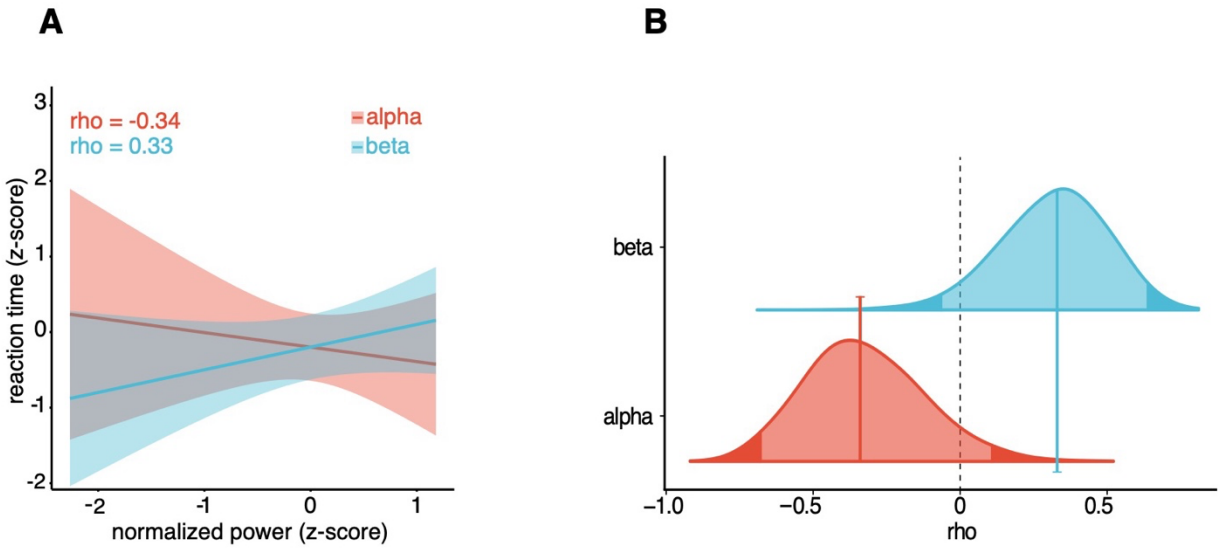


Figure 8. A) Correlation between alpha/beta power and reaction time (RT). Error bars represent 95% confidence interval. B) Bootstrap distributions of alpha/beta Spearman rho values.

3.4 Discussion

Spoken words are known to have facilitatory effects on visual object recognition, yet the mechanisms underlying such facilitation are incompletely understood. On one account, language does not influence perception itself but only later processes such as categorical decision making. On another account, language would bias perception at early sensory stages; specifically, via the amplification of category-specific priors in early sensory regions. A prediction of this latter model is that top-down priors evoke changes in neural activity before the presentation of visual stimuli. In the present study, we tested the hypothesis that neural oscillations can serve as a mechanism to carry language-driven priors about incoming object categories.

To test this hypothesis, we used EEG to measure prestimulus brain activity and characterize the oscillatory dynamics underlying the label-advantage in object recognition. We reasoned that, if objects are recognized faster because spoken words provide more refined categorical priors than natural sounds, then these cues should differentially modulate prestimulus oscillatory activity in the alpha/beta bands; and we should expect such an oscillatory index to be linked to object recognition performance.

Our results provide evidence that language affects visual perception by biasing prestimulus activity towards the identification of incoming input. We first replicated the previously reported label-advantage and showed that this behavioral effect persisted even when words were presented in a second language, indicating that the label-advantage relies on inherent properties of verbal symbols to deploy precise categorical information. Importantly, the reported behavioral advantage for spoken words was associated with an increase in the posterior alpha and beta rhythms in the time interval between the offset of the cue and the onset of the target object. Correlation analysis revealed that both rhythms contribute to visual object recognition performance. Yet, they appeared to operate in orthogonal directions: object recognition performance improved when alpha power increased, but decreased with the enhancement of beta rhythms. This finding suggests a

division of labor between alpha and beta rhythms in orchestrating language-mediated top-down guidance of visual behavior.

Enhancement of alpha oscillations have been largely reported in visual brain regions when top-down knowledge is directed by a cue towards a specific feature or direction (Worden et al., 2000; Snyder and Foxe, 2010). Two non-inclusive theoretical interpretations have been advanced to explain this effect. One prominent view is that enhanced alpha power reflects states of inhibition and filtering of task irrelevant information (Jensen and Mazaheri, 2010; Klimesch et al., 2007). More recent proposals however ascribe to neural alpha synchronization a large variety of roles in top-down processing (Palva and Palva 2007; Klimesch 2012; van Kerkoerle et al. 2014). Enhancement of alpha waves in task-relevant regions have been suggested to have excitatory effects reflecting selective amplification of neural representations of object categories (Mo et al., 2011). For instance, M/EEG studies have reported that alpha power increases in grapheme-processing regions as a function of predictability about the identity of letters (Mayer el al., 2016) or in the posterior cortex when meaningful hints precede the discrimination of ambiguous images (Samaha et al., 2018). Our results are in line with this interpretation, and suggests that alpha oscillations carry language-generated representations about the structure of visual objects. The functional role of prestimulus alpha waves in carrying object representations is also supported by the negative correlation between individual alpha power and reaction times, showing that participants with higher alpha power were overall faster in recognizing visual objects. Under this account, processing verbal symbols may help to form more precise object-representations than natural sounds against which the incoming visual input can be compared, thus resulting in a facilitation in subsequent object recognition.

A novel aspect in our results in contrast to previous similar studies concerns the cue-related differences in the beta rhythm between spoken words and natural sounds. Recent proposals suggest that beta oscillatory activity reflects endogenously driven transitions from

latent to active cortical representations of objects categories (Spitzer and Haegens; 2017), as well as the binding of neurocognitive network elements underling a given neural representation (Bressler and Richter, 2016). We speculate that the difference in beta modulations for spoken words vs. natural sounds may reflect a difference in the content of the activated states – and more importantly, in the amount of retrieved conceptual dimensions, e.g. the size of the neurocognitive network state, or load (Bressler and Tognoli, 2006). Behavioral and eye tracking experiments have indeed showed that spoken words activate a rich network of concepts during lexical processing (e.g., Huettig and Altmann, 2005). As a consequence, processing words might lead to the retrieval of knowledge dimensions far beyond purely sensory features of objects, such as conceptual, grammatical and lexical information. This is partially in line with human and monkey studies showing that beta synchronization over parietal and frontal regions carry supramodal information about object categories (Antzoulatos and Miller, 2014, 2016; Wutz et al., 2018). On this account, the positive correlation between beta power and reaction times may reflect the fact that activating a large space of possibilities is detrimental to successfully performing the current task, as it requires primarily the potentiation of visual features for faster recognition. Overall, these results suggest a division of labor between alpha and beta rhythms in top-down signaling during language-mediated visual object recognition, where alpha rhythms might function to amplify neural representations of object categories, while beta-frequency synchronization may maintain the neurocognitive network states elicited by the auditory cue.

The present findings inform the broad debate on whether language shapes perception at early or late stages of perceptual processing. At least to what concerns the effect of single words on visual object recognition, it seems unlikely that such biases arise at later semantic levels. Previous EEG studies showed that words affect visual processes by modulating ERP components such as the P1 (Boutonnet and Lupyan 2015; Noorman et al., 2018), which are traditionally associated with early visual processing. Similarly, fMRI

studies showed that language sharpens neural activity in visual regions associated with the processing of visual features, such as V4 for colors (Brouwer and Heeger, 2013); and object categories, such as the fusiform face area and parahippocampal place area (Puri et al., 2009). These results, together with our current study, suggest that the effect of words on visual perception arises at an early, sensory stage of processing – specifically, via the modulation of prestimulus activity in the alpha/beta frequency band.

Finally, a novelty in our study compared with previous categorization studies concerns the inclusion of words in L2 as an auditory cue. Our participants were indeed highly proficient Basque-Spanish bilinguals, with comparable levels of proficiency for the two languages, but with the L2 acquired later in development. The effect of top-down processing in bilinguals is currently debated, and largely dependent on factors like proficiency (Kaan, 2014; Hopp, 2013) and age of acquisition (Molinaro et al., 2017). Concerning semantic processing, despite that it is commonly believed that bilinguals access a common semantic system in both languages (e.g., Caramazza and Brones, 1980), recent studies have suggested that top-down processing may be reduced in a second language because of reduced access to perceptual memory resources (e.g., Hayakawa and Keysar, 2018), which are known to play an important role in the generation of visual expectation (Hindy et al., 2016). The reported comparable behavioral and neural responses on the effect of words in L1 and L2 on visual object recognition are in line with the idea that both languages access common conceptual representations, and deploy top-down guidance to the visual system in a similar manner.

On the contrary, the divergent effects of words and natural sounds challenges the hypothesis that these types of cues access a common nonverbal conceptual system. It is relevant to consider why cuing an object with a word results in enhanced visual recognition compared to a natural sound. Symbols have been proposed to be extremely effective in compressing semantic information in a format that transcends within-category differences,

thus leading to the amplification of those features which are relevant for distinguishing between exemplars of different categories. On the contrary, natural sounds are inevitably linked to their sources (e.g., the barking of a dog may trigger the representation of a specific exemplar of dog), thus being less effective at cuing a categorical state (Edmiston and Lupyan, 2015). Interestingly, ascribing labels to experience have been shown to also enhance other cognitive functions, such as the retention of items in visual working memory (Souza and Skóra, 2017), learning of novel categories (Lupyan et al., 2008), perceptual categorization across sensory modalities (Miller et al., 2018). These findings indicate that language acts as a powerful tool for compressing information, facilitating different operations important to a multitude of human cognitive processes (Clark 2012).

Chapter - 4 Language experience affects predictive processing during auditory rhythm perception

4.1 Introduction

Predictive coding (PC) (Friston, 2005, Rao and Ballard, 1999) is becoming a popular theory of perception. It assumes that each cortical areas extrapolates statistical regularities governing sensory input and uses them to build internal models of the environment. Such models are then used to generate top-down predictions about incoming sensory events. Predictions feedback from higher to lower cortical areas where they are compared with activity related to novel inputs. Only the difference, called the “prediction error” (PE), is transmitted via feedforward connections to higher cortical stages, where it can be used to adjust the internal model. The output is a bidirectional message-passing system that constantly updates its internal models at multiple hierarchical levels of processing.

In the auditory domain, great progress in the understanding of predictive capabilities of the auditory system has been made using variations of the Oddball design (see Heilbron and Chait, 2018 for a review). In these studies, subjects are usually presented with sequences of stimuli encoding some specific regularity (typically a repetition of tones sharing some physical features like pitch, duration or intensity), that is then violated by a ‘deviant’ event. Such deviant event elicits a novelty response in the EEG which has been defined “mismatch negativity” (or MMN), that arises around 100–250ms post-stimulus. Within the predictive coding framework, the MMN wave is suggested to reflect the violation of a prediction, that is, the cortical PE signal.

Interestingly, such error signals arise even when an expected input is omitted from a regular sequence of sounds. Such “omission responses” are difficult to reconcile with feedforward models, as they cannot be explained by the presence of incoming bottom-up sensory signals. As such, omission responses provide an elegant tool to investigate the implementation of top-down prediction decoupled from bottom-up input.

Experimental designs using variations of Oddball design – or similar designs such as optimum-1 (Näätänen et al., 2004), omission (Yabe et al., 1997), and roving-standard (Garrido et al., 2008) – have been important to unveil the sensitivity of the predictive system to local transition probabilities. However, these studies have primarily examined predictions generated over rules acquired in the context of an experimental task i.e., rules linked to short-term memories. These local, context-dependent predictions are known to be flexible, meaning that they can be easily updated or canceled out based on new sensory evidence. However, one core assumption of PC models is that the brain deploys also certain long-term predictions (Yon & De Lange, 2018), which are typically resistant to evidence-based updating. Such predictions may emerge via learning, through the extrapolation of regularities and physical patterns that are relatively constant throughout the lifespan of an individual. Because arising over long timescales, these priors become encoded into the tuning properties of early sensory cortices. Their being resistant to evidence-based updating reflects their overall computational goal: the optimization of the (long-term) PE (Friston, 2018; Lupyan and Clark, 2015).

The goal of the present MEG study was to investigate whether the auditory system generates predictions based on life-long exposure to auditory regularities, using patterns that extend beyond those acquired in the recent past. To that end, we compared MEG data from native speakers of Basque and Spanish performing a rhythmic version of the alternation paradigm with omission responses (see Fig. 9). In this paradigm, rhythmic sequences of tones alternating in duration are presented to participants, with rare tone omissions occurring randomly.

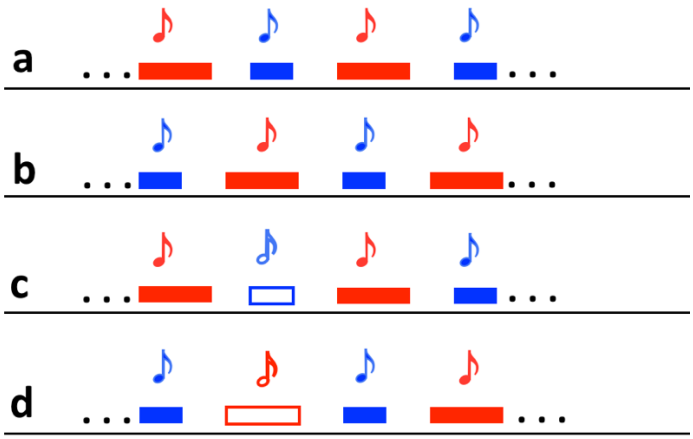


Figure 9. Experimental design

(a–b) Stimulus sequences of alternating long (437ms) and short tones (250ms) with intervals of 20ms. **(c–d)** Omission of a short and a long sound (500ms), respectively.

Importantly, Basque and Spanish are two languages that differ in their syntactic/prosodic structure, providing a unique testbed to study the effect of experience on auditory predictive processing. Spanish is a functor-initial language, in which short sounds (i.e., function words; e.g., “la”) usually precede long sounds (i.e., content words; e.g., “habitación”), whereas Basque is a functor-final language, in which short sounds (i.e., function words; e.g., “bat”) usually follow long sounds (i.e., content words; e.g., “logel”). Previous behavioral studies have shown that regular exposure to rhythmic regularities related to language syntax/prosody influences the automatic grouping biases of non-linguistic sounds (Iversen et al., 2008, Molnar et al., 2016), suggesting that such long-term, language-induced patterns may affect basic auditory processing in a top-down fashion.

We hypothesize that the auditory system partially builds long-term priors based on the regular patterns embedded in the prosodic structure of the language, and uses such knowledge to generate hierarchical predictions about incoming sounds. Based on this, we predict an interaction between language dominance of participants and the type of event

that is violated. This means that the same violation would elicit different MMN responses depending on language background. Because in Spanish's prosody long sounds usually follow short sounds, we predict that the omission of a long tone will elicit a larger MMN in Spanish dominant speakers as compared to Basque dominant speakers, as it represents the violation of two predictions in Spanish, but not in Basque: (i) a context-dependent (short-term or local) prediction based on the statistical regularities of previous stimuli (i.e., prediction of hearing a long tone after a short one in the specific context of the experiment), and (ii) an experience-dependent (long-term) prediction based on the statistical regularities of Spanish prosody. The same pattern is expected in Basque dominant speakers, when a short tone is omitted.

There is also another possible scenario that can be compatible with a hierarchical predictive processing account. Because during musical beat perception sounds separated by a short time interval tend to be merged into higher-level units (Litovsky et al., 1999), it may be that the auditory cortex is tuned to invest major predictive power on the onset of each higher-level event (thus on the onset of the first element of the chunk). For instance, when presented with a rhythmic sequence of two tones alternating in duration, Spanish speakers tend to chunk the sequence into "short-long" higher-level units. In this context, the auditory cortex may generate a long-term prediction about the onset of such higher-level unit (i.e., the "short" tone). This scenario would predict the opposite pattern of results as compared to the scenario presented above, with the omission of a short tone generating a larger MMN in Spanish compared to Basque dominant speakers (and the opposite pattern when a long tone is omitted). Despite making different predictions about the directionality of the effect, this second scenario still predicts that the amplitude of the MMN should be modulated orthogonally by the linguistic background of the participants, which is the core manipulation of the study.

We also provided participants with a control condition that has the same design than the test condition, but with the two tones differing in frequency and not in length. Here, no difference between the two populations is expected, as they should both rely on local (short-term) predictions, with no specific experience-dependent (long-term) prediction. Overall, the experiment takes the structure of a 2x2 design, with omission-type (long omission vs short omission) as within factor and language background (Basque vs Spanish group) as between factors.

4.2 Materials and Methods

Participants. In total, 20 Spanish dominant (mean age: 25.96 years, range: 20–33, 16 females) and 20 Basque dominants participants (mean age: 27.11 years, range: 21–40, 17 females) took part in the experiment. Participants were selected using a measure similar to that used in Molnar et al., (2016). The proficiency levels for each language were evaluated based on self-reported scores on a scale from 0 (=no knowledge) to 10 (=native proficiency). The exposure measures reflect the participants' self-report of the average exposure to the given language at the time of testing. In addition, most participants reported to have learned a second and third language in school settings (e.g., Basque, English, French, and Catalan for Spanish speakers; Spanish, English, French for Basque speakers).

Measure		Basque natives (N = 20)	Spanish natives (N = 20)
Proficiency (0-100)	Span	8.73 (0.78)	9.73 (0.54)
	Basq	9.6 (0.48)	//
Exposure (0-100)	Span	22.63 (10.45)	79.5 (9.98)
	Basq	70 (14)	//
Age of acquisition	Span	2.55 (2.87)	0.5 (0.85)
	Basq	0.15 (0.48)	//
Picture naming (0-65)	Span	57.8 (19.79)	64.82 (0.39)
	Basq	64.22 (1.11)	//

Table 2. General linguistic profile of participants in the two groups. Statistics on Basque knowledge in the Spanish group are not provided because only six participants reported to have learnt Basque as second or third language.

Stimuli and experimental design. Stimuli were created using Matlab Psychtoolbox and presented binaurally via MEG-compatible headphone. Experimental stimuli consisted of 60 sequences of two tones alternating in duration (short-tone: 0.250 s; long-tone: 0.437 s respectively) with fixed intervals (0.02 s). Tones had a frequency of 500Hz. The beginning and end of each tone were fade in and out of 0.015 s. Overall, each sequence consisted of 40 short-long tone pairs, for a total of 80 elements per sequence, and lasted around 30 s. The beginning and the end of each sequence was fade in and fade out of 2.5 s in order to mask possible grouping biases. Further, half of the sequences started with a long tone, and half with a short tone. In each sequence, 2 to 6 tones were omitted and substituted with a 0.5 s silence gap. The larger gap was introduced in order to avoid that activity related to the onset of the tone following the omission may overlap with the activity generated by the

omitted tone. Tone omissions occurred pseudorandomly, for a total of 240 omissions (120 short and 120 long). In the control condition, sequences consisted of tones alternating in frequency at fixed intervals (0.02 s). High frequency tones had a frequency of 700Hz, while low frequency tones had a frequency of 300Hz. Both high and low frequency tones had an overall duration of 0.343 s. This duration was selected in order to keep the overall length of the sequences equal to that of the test condition. As in the test condition, in each sequence 2 to 6 tones were omitted and substituted with a 0.5 s silence gap.

Overall, the experiment was divided into two main blocks: test and control. The order in which the blocks were presented was counterbalanced across participants. Each block consisted of 60 sequences and lasted around 35 minutes. Each sequence was separated by an 8 s silence gap. Every twenty sequences, a short pause was introduced. Further, the end of each block was followed by a longer pause.

Participants were requested to minimize movement throughout the experiment, except during pauses. Subjects were asked to keep their eyes open, to avoid eyes movements by fixating a cross. Similarly to previous studies, the only task that was asked to subjects was to count how many omissions were present in each sequence (e.g., Bekinschtein et al., 2009) - and report it at the end of the sequence during the 8 s silence gap. Participants only received instructions at the very beginning of the task (in their native language), and no verbal or written instructions was introduced during the task.

MEG Recordings. Measurements were carried out with the Elekta Neuromag NeuroSpin system (Elekta Neuromag), which comprises 204 planar gradiometers and 102 magnetometers in a helmet-shaped array. ECG and electrooculogram (EOG) (horizontal and vertical) were recorded simultaneously as auxiliary channels. MEG and auxiliary channels were low-pass filtered at 330 Hz, high-pass filtered at 0.1 Hz, and sampled at 1 KHz. The head position with respect to the sensor array was determined by four head-position indicator coils attached to the scalp. The locations of the coils were digitized with

respect to three anatomical landmarks (nasion and preauricular points) with a 3D digitizer (Polhemus Isotrak system). Then, the head position with respect to the device origin was acquired before each block.

Preprocessing. Signal space separation correction, head movement compensation, and bad channels correction were applied using the MaxFilter Software (Elekta Neuromag). After that, data were analyzed using the FieldTrip toolbox (Oostenveld et al., 2011) in Matlab (MathWorks). Trials were epoched from 1.2 s before to 1.2 s after the onset of each tone or omitted tone. Trials containing muscle artifacts and jumps in the MEG signal were detected and removed using a semiautomatic routine. Subsequently, independent component analysis (Bell and Sejnowski, 1995) was performed to partially remove artifacts attributable to eye blinks and heartbeat artifacts (Jung et al., 2000). To facilitate the detection of components reflecting eye blinks and heartbeat artifacts, the coherence between all components and the ECG/EOG electrodes was computed. Components were still checked visually before rejection. After artifact rejection, trials were low-pass filtered at 40 Hz and averaged per condition and per subject. ERFs were baseline corrected using the 0,05 s preceding trial onset. The latitudinal and longitudinal gradiometers were combined by computing the root mean square of the signals at each sensor position in order to facilitate the interpretation of the sensor-level data.

Statistical analysis. Statistical analyses and data visualization were performed using FieldTrip toolbox (Oostenveld et al., 2011) in Matlab (MathWorks) and R studio for post-hoc analysis and visualization. All comparisons were performed on combined gradiometer data. For statistical analyses we used a univariate approach in combination with cluster-based permutations (Maris & Oostenveld, 2007) for family-wise error correction. This type of test controls the type I error rate in the context of multiple comparisons by identifying clusters of significant differences over space and time, instead of performing a separate test on each sensor and sample pair. Two-sided paired- and independent-samples t-tests were used for

within- and between-subjects contrasts, respectively. The minimum number of neighboring channels required for a sample to be included in the clustering algorithm was set at 2. The cluster-forming alpha level was set at .05. The cluster-level statistic was the maximum sum of t-values (maxsum) and the number of permutations was set to 10000. To control for the false alarm rate, we selected the standard $\alpha = 0.05$. For the first analysis only in which we compared ERF generated by omission vs pure tones, we used a time-window between 0 and 0.250 s, and took into account the temporal dimension in the cluster-based permutation test. This explorative analysis was performed to assess the difference between activity elicited by a tone vs omission, as well as its temporal unfolding. In all the remaining analyses, MMN-responses were calculated by subtracting the ERFs of tones from the ERFs of omissions. Moreover, all analysis were conducted by averaging values in a time window between 0.100 and 0.250 s, which covers the typical latency of the MMN (Näätänen et al., 2007; Garrido et al., 2009). When multiple clusters emerged from a comparison, only the most significant cluster was reported.

Using this approach, we first assessed the elicitation of the omission responses by comparing the ERFs elicited by standard tones vs the ERFs elicited by omissions. Second, to assess for the presence of main effect of omission type we compared the MMNm responses elicited by omissions of short tones vs omissions of long tones. Third, we assess for a main effect of language background (Basque vs Spanish) by comparing the average of short omission and long omissions responses between groups. Finally, we assessed for the presence of an omission-type by language background interaction. As the cluster-based permutation test is designed to compare two conditions at a time, we tested for an interaction by subtracting the MMNm elicited by long omission from the MMNm elicited by short omission responses for each participant, and compared the resulting differences between groups. A significant cluster emerged from this contrast. In order to assess what drives this interaction, we ran post hoc t-tests on ERF data averaged over all the channels belonging

to the significant cluster and time points (0.100 – 0.250 s). All p-values resulting from these comparisons were FDR corrected. The same analysis pipeline was applied to the analysis of the control condition.

4.3 Results

We first look at responses evoked by the omissions of tones. Statistical analyses were ran by comparing the amplitude of the ERF elicited by tones vs omissions. Cluster analysis, as implemented in FieldTrip software, was used to identify clusters of neighboring sensors where a significant difference between the activity elicited by the two conditions. For this analysis only, we looked for spatiotemporal clusters in the 0 – 0.250 s time window. This analysis revealed the presence of a significant cluster ($p < .001$), indicating that, despite the absence of a physical stimulus, omissions generated a much larger ERF compared tones (see Fig. 10 A). Such difference emerged around 0.100 s after stimulus onset, and included several channels over the entire scalp (Fig. 10 B). ERF elicited by tones and omissions had similar topographes, with bilateral activations over the temporal regions. The ERF elicited by tones was more pronounced over the left hemisphere, while the one elicited by omissions was more pronounced on the right hemisphere.

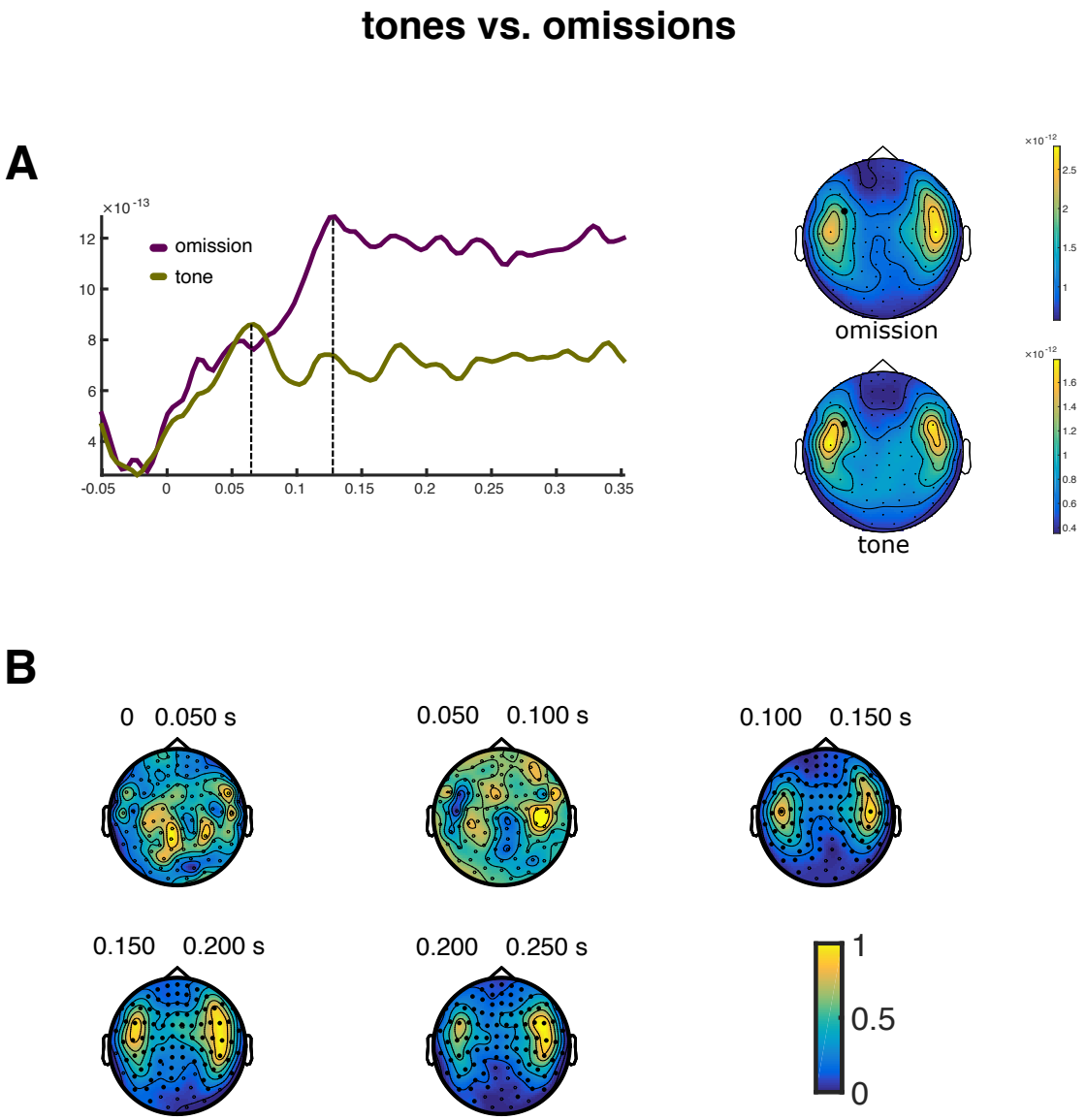
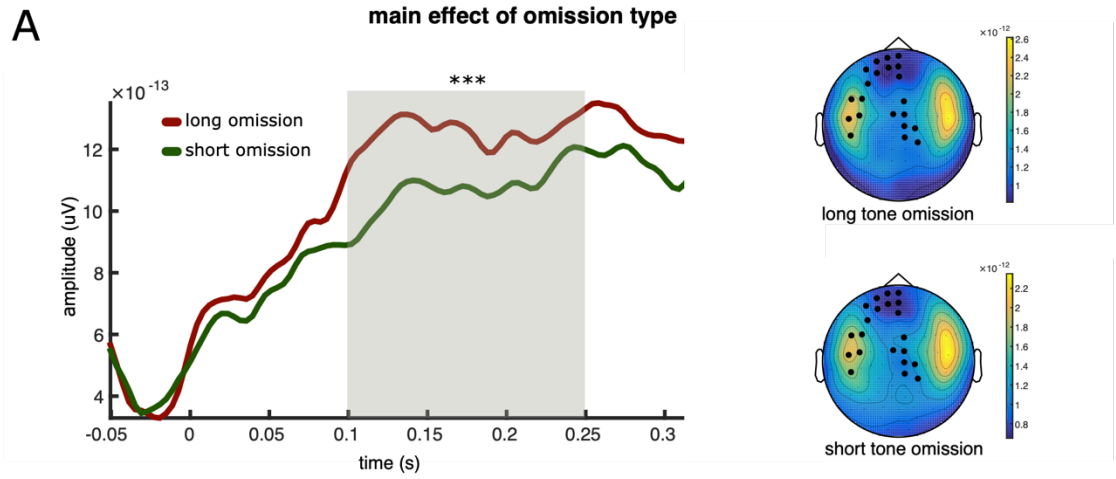


Figure 10. Panel A shows ERFs elicited by tone and omission responses. Dotted lines indicate the peak of each ERF. Topographies of the two conditions time-locked to the peaks are shown on right side. The marked channel indicates a representative sensor selected for visualizing the ERFs. Panel B shows the temporal unfolding of the topographical distribution of the cluster depicting the difference between ERFs elicited by tones vs. omissions. Channels contributing to the cluster are marked.

We then assess for a main effect of omission-type by comparing MMNm responses elicited by long tone omissions vs short tone omissions averaged across groups over the

0.100 – 0.250 s post (omitted) stimulus onset. Remind that MMNm responses were calculated by subtracting the ERF of tones from the ERF of omissions, as standard for the analysis of the MMN (Garrido et al., 2009). We found a main effect of long tone omission ($p < .003$) over several fronto-temporal and parietal channels, meaning that omissions of long tones generated a larger MMNm compared to omissions of short tones (See figure 11 A). The effect was consistent in the Basque ($p < .001$) but not in the Spanish group (no cluster detected).



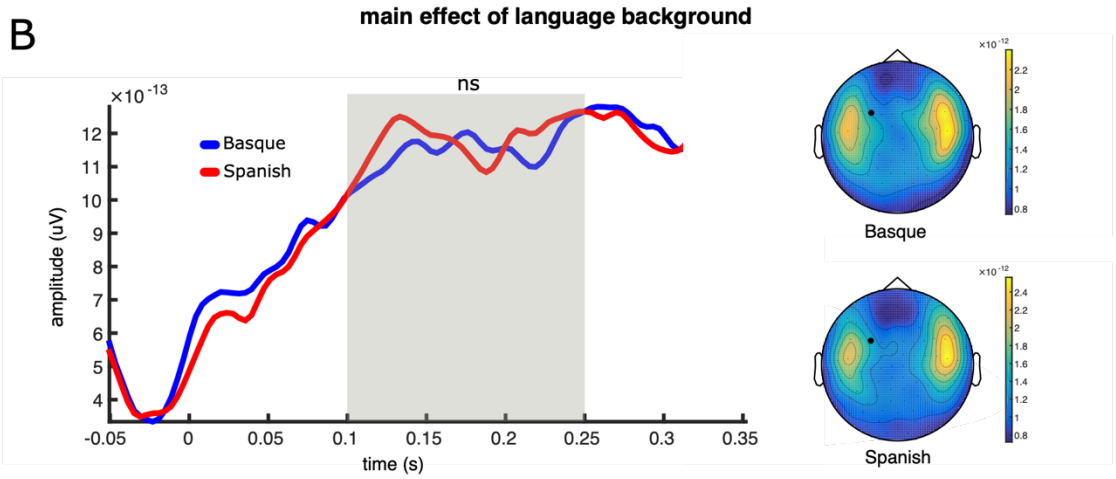


Figure 11. Panel A shows ERFs reflecting the main effect of long tone omission. The grey shadow part indicates the time-window of interest (0.100 – 0.250 s). The topographical distribution of the MMN is reported on the right, in which the channels contributing to the cluster are marked. Panel B shows ERFs reflecting the responses to omission in the Basque vs Spanish group. The grey shadow part indicates the time-window of interest (0.100 – 0.250 s). The topographical distribution of the MMN in one representative channel is reported on the right. The representative channel is marked.

Third, we assess for the presence of a main effect of language background (Basque vs Spanish group) by comparing the average of short tone omission and long tone omission responses between groups. No cluster emerged from this comparison (see figure 11 B). Finally, we investigate the interaction between omission-type by language background, which was the analysis of interest of our study. From this analysis, a significant cluster associated with a p-value of .03 was detected over a few channels covering left fronto-temporal regions (see figure 12 A, B). We then unpack such interaction effect by first averaging data for each participant over channels belonging to the significant cluster and time points of interest (0.100 – 0.250 s), and then comparing the conditions in the two groups using independent sample t-test. Specifically, we compare (i) omissions of long tones in the Basque group vs Spanish group, and (ii) omissions of short tones in the Basque group vs Spanish group. From these comparisons, it emerged that long tone omissions generated a larger MMNm response in the Basque compared Spanish group ($p < 0.05$ FDR-corrected; Fig. 12 A, C), while short tone omissions generated a larger MMNm in the Spanish compared to Basque group ($p < 0.05$ FDR-corrected; Fig. 12 B, D).

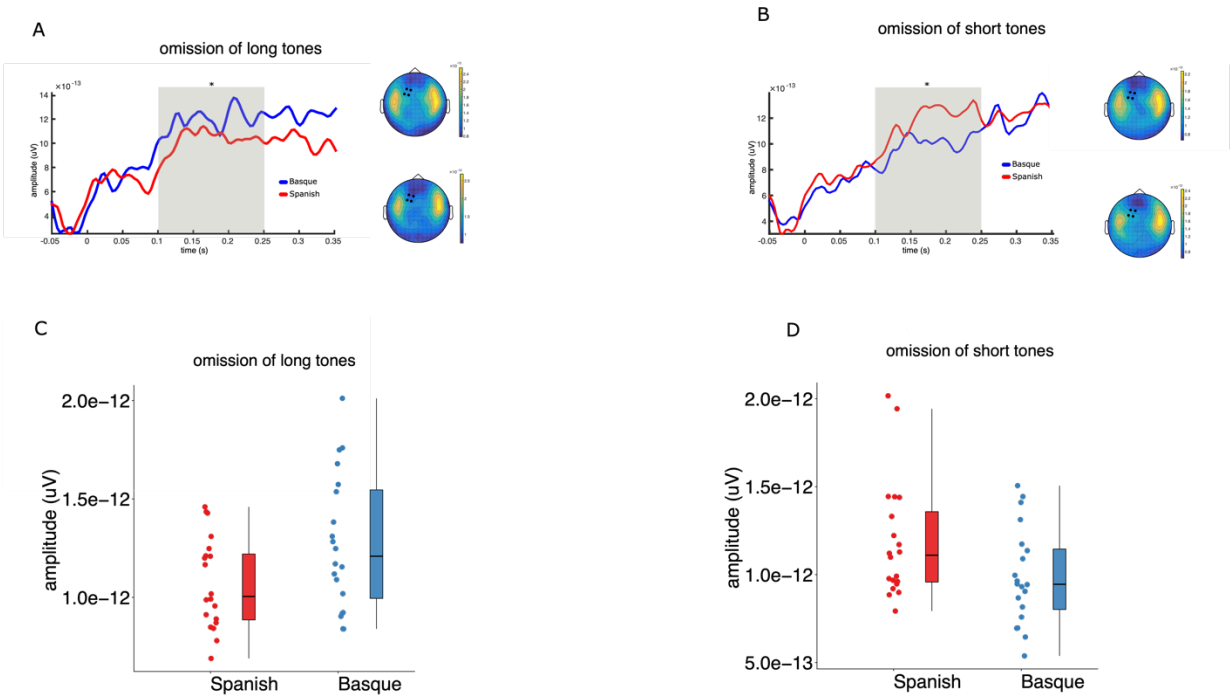


Figure 12. Panels A and B show ERF elicited by the omission of long (A) and short tones (B) in the Spanish and Basque group. Grey shadow part indicates the time-window of interest (0.100 – 0.250 s). The topographical distribution of the MMN is reported on the right, in which the electrodes contributing to the cluster are marked. Panels C and D show box plots on the effect of language background on long and short tone omissions over the channels belonging to the significant cluster and time-window of interest.

The same analysis pipeline was applied to control data. Here we found a main effect of omission type ($p < .001$), with omissions of high frequency tones generating a sharper MMNm response than the omission of low ones. Similarly to the test condition, no difference emerged when comparing the data from the two groups (no cluster). Crucially, no significant omission-type by language background interaction was detected (no cluster). To further check that no interaction was present in the control condition, we averaged the data over the channels and time points in which we detected a significant interaction in the test

condition, and ran independent sample t-test by comparing ERF responses to high and low frequency tone omissions in both groups. Even within this subset of channels, no significant difference between groups was detected when comparing omission of low ($p = 0.84$ FDR-corrected) and high frequency tones ($p = 0.19$ FDR-corrected).

A potential concern is that the observed electrophysiological differences between conditions may stem from differences in neural activity elicited by the tones preceding the omission, rather than from the hypothesized long-term predictions. This is unlikely because MMNm responses were calculated as the difference between omission and tone. However, to address this potential concern we analyzed the MEG signal before omitted tones for each condition and group. Epochs corresponding to 0.05 s of activity preceding omissions were averaged over condition (long omission, short omission) and group (Basque, Spanish). Differences between omission responses were assessed using a paired two-tailed permutation test at all time points of the 0.05 s pre-omission onset activity, first across all channels, then by focusing on the channels where the interaction was significant. Both analyses of the pre-omission activity – in both groups and in both test and control conditions – failed to detect any significant difference at any time points between the activity generated by tones preceding long and short tone omissions in the -0.05 to 0 s time window.

4.4 Discussion

In this study, we explored whether the auditory system generates predictions based on long-term priors (i.e., linguistic knowledge), using memories that extent beyond the recent past. To that end, we compared data from participants with life-long exposure to languages differing in their prosodic structure (i.e., Basque and Spanish), performing a rhythmic version of the alternation paradigm with omission responses as deviants. These two languages

provide an ideal model to study the effect of experience (linguistic in that case) on predictive processing. Both Spanish and Basque are part of the same cultural community in Northern Spain. These languages share almost the same sharing almost the exact same phonology and orthography. However, unlike Spanish, Basque is a non-Indo-European language (an isolated language) with no typological relationship with Spanish. It is thus very unlikely that cultural factor which are not language-specific (e.g., exposure to different musical traditions or educational and writing systems) may play any role in the present findings. We hypothesized that the auditory system relies on language-driven regularities to build predictive models of the environment and generate predictions about incoming sounds. A prediction of this hypothesis is that the same type of violation (i.e., short or long tone omission) should generate a different MMN response depending on subjects' linguistic background.

In line with this prediction, we found that language experience modulates the amplitude of the MMN elicited by omission responses – our measure of prediction error signal. When a long tone was omitted, the MMN was sharper in the Basque compared to the Spanish group. On the contrary, when a short tone was omitted, a sharper MMN response was elicited in the Spanish native group.

These results provide strong evidence that the auditory system relies on long-term priors to generate predictions during musical beat perception. Such long-term priors mirror abstract patterns present in the syntactic/prosodic structure of language. One possible interpretation of this effect is that experience with recurring prosodic patterns of native language shapes the tuning properties of early auditory regions. Such tuning is arguably important for reducing the prediction error during language processing, as it allows the auditory system to generate a functional coding scheme, or auditory template, against which the incoming linguistic input can be parsed. Such auditory template may be recycled by the auditory system to build top-down predictive models which can be applied to the processing

of non-linguistic auditory sequences. Considering the current results from a broader perspective, they support the idea that shared auditory top-down resources underlie speech, sound, and music processing (Asaridou & McQueen, 2013).

In the introduction, we hypothesized two possible scenarios in which the effect of language could bias low-level predictive processing. On a first scenario, the long-term prediction would be mainly deployed on the second element of the chunk. For instance, because Spanish speakers have a tendency to chunk a rhythmic sequence of tones alternating in duration using short-long grouping pattern, then the omission of a long tone would violate the long-term prediction that, according to Spanish's syntax/prosody, long elements usually follow short elements. Thus, the resulting MMN would reflect the accumulation of two violations: a context-dependent prediction based on the statistical regularities of previous stimuli, and a long-term prediction based on the statistical regularities of Spanish syntax/prosody; with the opposite pattern expected in Basque participants when a short sound is omitted. The second hypothesized scenario suggests that, since tones tend to be merged into higher-level chunks, then the auditory system is tuned to invest major predictive power on the onset of each higher-level event (i.e., the first element of the chunk). This is consistent with the fact that the onset of an acoustic stimulus always elicits a sharper peak in the auditory response (Jewett et al. 1970). For instance, because Spanish natives tend to chunk a rhythmic sequence of tones alternating in duration using short-long grouping pattern, then the long-term prediction would be deployed on the onset of the short sound. This would result in a larger MMN in the Spanish group when a short sound is omitted compared to a long sound, with the opposite pattern in the Basque group when long sounds are omitted. Our results indicate that the omission of a long tone generates a larger MMN in the Basque compared to the Spanish group, while the omission of a short tone generates sharper MMN in the Spanish than the Basque group. This pattern of results is thus in line

with the second hypothesized scenario, and suggests that long-term, experience-dependent priors build on recurring chunks to generate expectations.

Importantly, our study tested a prediction, unique to predictive coding models, that expectations are organized hierarchically. When two predictions, one experience-dependent and one context-dependent are violated, the amplitude of the MMN is larger compared to a scenario when only one stimulus- dependent prediction is violated. This result complements previous work showing that the same deviancy presented in different contexts may generate both local and global prediction error responses such as the MMN and the P3, generated in sensory and frontal/associative cortices, respectively (Wacogne et al., 2011). However, our results extend previous work by showing that hierarchical predictive processing emerges also within the same cortical system, depending on long-term priors and rules acquired over the life span.

Concerning, the more general debate on the interplay between language and perception, such results indicate that off-line effects of language on perception may arise via hierarchical predictive coding, with predictive signal coming from higher-level (linguistic) levels modulating signals at lower stages.

Our results also complement previous studies showing modulatory effects of (long-term) musical expertise on the MMNm (e.g., Vuust et al., 2005; 2009). These studies indicate that responses to violation during auditory rhythm perception are larger when the listener is an expert musician compared to a non-musician. In our study, we manipulated prediction orthogonally, with clear-cut predictions on the effect of language experience on predictive processing. Our results thus provide additional support for the idea that the auditory system generates top-down predictions based on (multiple) auditory experiences.

Omission paradigm have been used largely in the study of the behavior (Yabe et al., 1997; Raij et al., 1997) and predictive capabilities of the auditory cortex (Bendixen et al., 2009; San Miguel et al., 2013; Chennu et al., 2016; Wacongne et al., 2011; Todorovic and

de Lange, 2012). Within the predictive processing framework, omission paradigms offer an appealing advantage over classical oddball paradigm, as they allow to detect endogenous activity associated to certain cognitive functions without being confounded with activity generated by a deviant event. Yet, what reflects the activity elicited by omission responses is debated. On one account, such responses may reflect pure predictions: if the evoked response generated by a deviant sound within a regular sequence (as in classical oddball paradigm) reflects the difference between the top-down prediction and the bottom-up sensory signal (i.e., the PE signal), then when the bottom-up sensory stimulus is not present, the brain response should reflect a purely top-down predictive signal (Summerfield et al., 2008; Bendixen et al., 2009). However, recent studies have posited that responses to (unpredicted) omissions reflect prediction error signal. Even in the absence of a deviant sensory input, omission responses may reflect the detection of unfulfilled expectations (Hughes et al., 2001; Wacogne et al., 2011). While it is currently debated whether neural responses time-locked to an omitted sound reflect a pure prediction or PE responses, the latency and topography of the omission response in our data resemble those of a classical MMN (see Fig. 10). This provides some support for a PE interpretation. Indeed, if the response to an omission reflected a pure prediction, we should expect its ERF to have the same latency than the predicted tone (~0.05 s). However, the ERF response to omission in figure 10 peaks around 0.150 s. This suggests that its modulation reflects a novelty-detection mechanism, rather than a pure predictive signal. It should also be noted that, even if the response to an omitted sound would reflect a purely predictive signal, our main predictions about the effect of language experience on predictive processing would be the same: if subjects rely on both long-term plus contextual prediction when a certain sound is omitted, and only contextual prediction when another sound is missing, then the neural response should be larger in the first case, as it represents the overlap of two predictive signals (instead of two prediction error signal). It is also important to notice that other

mechanisms are likely to be at work at the same time when one tone is omitted. Given the rhythmic component of the paradigm, part of the evoked response might reflect some sort of rebound of cortical oscillators entrained to the previous stimuli (Wacogne et al., 2011). This would be in line with some current accounts of the MMN suggesting that its wave reflects a combination of active predictive and passive adaptation phenomena (May & Tiitinen, 2010).

One unexpected finding in our data concerns the main effect of tone omission, indicating that the MMN generated by the omission of a long tone was larger compared to that generated by the omission of a short one. Because such effect was consistent only in the Basque group, it is possible that it merely reflects a larger sensitivity of the auditory system of this group to long tone omission, as was predicted. Alternatively, we speculate that such effect could be driven by the fact that, during language processing, major predictive power is invested on long sounds compared to short sounds, as the former usually refer to content words i.e., semantically relevant events. As a consequence, the auditory system may apply a similar predictive scheme also during processing of non-linguistic sounds, independently of language background.

In summary, our results provide evidence that experience with syntactic/prosodic patterns of a certain language affect neural processing of rhythmic auditory sequences based on the simple repetitions of two tones alternating in duration. This language-mediated bias on perceptual process arises via the generation of predictive signals arguably employed to reduce prediction error during language processing. This is in line with hierarchical predictive coding view of perception, in which predictions at different levels in the cortical hierarchy influence activity at lower stages. Our results show that language biases can reach the lower sensory levels of auditory processing, as highlighted by the modulation of the MMN. The omission paradigm, by being sensitive to these responses, provides a flexible method to assess the hierarchical organization of cortical prediction.

Chapter - 5 General discussion

5.1 Summary of results

The objective of the current thesis was to investigate whether and how linguistic knowledge impacts perceptual processes. We investigated whether some previously reported effects of language on perception can be explained by current predictive processing models, which assume a bidirectional flow of information between areas at higher and lower cortical stages. In chapter 3, we assessed the electrophysiological mechanisms through which conceptual representations elicited by spoken words enhance visual processes like object recognition. We focused on a well-replicated behavioral effect, the label-advantage in object recognition (i.e., the fact that spoken words boost visual processes to a larger extent than natural sounds), and used EEG to investigate the neurophysiological dynamics underlying this effect. In contrast to previous studies that focused on the consequences that language-mediated prediction has on visual processing, we focused on the time-interval between the cue and the visual target. This allowed us to assess whether the facilitatory effect of words on object recognition arises from modulation of prestimulus activity in sensory regions (instead of later semantic or decision-making processes). We found that words and natural sounds differentially modulate prestimulus activity in posterior alpha and beta oscillations. Importantly, prestimulus alpha and beta rhythms correlated with behavioral responses, although showing an inverse relationship to behavioral performance. In chapter 4, we took advantage of two groups of Basque-dominant speakers and Spanish monolinguals to investigate whether and how life-long experience with language regularities can affect neural predictive processing of simple sequences of auditory sounds. We hypothesized that parsing strategies employed during sentence processing are recycled to process non-linguistic sounds. We recorded MEG activity time-locked to violations of regular sequences of tones mirroring the syntactic/prosodic structure of language (Basque and Spanish). As violation, we used omission responses instead of

novel deviant sounds. Brain responses to a violation by omission cannot be explained by the bottom-up features of the physical input, thus offering an appealing advantage to assess endogenous predictive mechanisms. When a sound was omitted from a regular sequence of tones, the amplitude of the MMN – an ERP/F component putatively associated to prediction error signal – differed orthogonally depending on subjects' linguistic background.

These studies provide some novel insights into the broad debate on the interplay between language and perception. In contrast to previous work, we focused on electrophysiological components traditionally associated to early sensory/predictive processing stages, such prestimulus alpha/beta oscillations and the MMN. This enabled us to assess whether the effects of language on perception have a perceptual vs conceptual locus, and whether such biases arise in a predictive-like manner. Moreover, we tested individuals with different linguistic profiles (i.e., bilinguals; speakers of typologically different languages), as only few studies have done so far in this research field. This allowed us to (i) make clear-cut predictions about the effects of linguistic experience on perception, as in the second study; and (ii) generalize our findings to the faculty of language more broadly, instead to the features associated to a specific linguistic system, as in the first study. Moreover, these are among the few studies that investigated the impact of language on perception at the neural level.

5.2 Discussion of results

Our two studies, despite differing noticeably on their details, provide some evidence that linguistic priors can bias perception at early stages. Importantly, these studies show how a predictive processing framework can provide an explanatory account of how high-level functions like language can influence putatively low-level processes like perception. The

current work contributes to the broad debate on the effect of language on perception by showing that: (i) language biases of perception may arise via (pre)activation of sensory priors in task-relevant regions; (ii) such priors are implemented via hierarchical predictive coding, with signal from higher (linguistic) regions affecting predictive processing in lower sensory stages.

The findings from the first study showed that the label-advantage in object recognition is associated to modulation of prestimulus activity in the alpha and beta bands. Importantly, these electrophysiological indices were correlated with later behavioral responses, suggesting that their modulation was not incidental for object recognition. The topography of alpha waves was strongly posterior, suggesting that language-mediated priors may be instantiated at the sensory level. This pattern is congruent with strong views of predictive processing accounts, which argue for the presence of sensory templates even before initial stimulation (Kok et al., 2017). Under this view, spoken words would provide more reliable predictions to the visual system about the structure of incmoming visual objects, thus being more effective in disambiguating whether a certain object belongs or not to a cued category. Previous EEG studied already suggested that language could bias perception at the sensory level. For instance, Hirschfeld et al. (2011) showed that ERP responses generated by incongruent sentence-picture pairs differed from those generated by congruent pairs around 170 ms and 400 ms after picture onset. Landau et al., (2010) showed that hearing sentences about faces influences face processing by modulating the N170 ERP component. Similarly, studies using the word-picture matching task showed linguistic modulations on picture processing as early as 100ms from the onset of the image (Boutonnet and Lupyan, 2015; Noorman et al., 2018). However, studies focusing on post-stimulus activity can also be coherent with a later semantic or decision-making account. Indeed, post-stimulus differences, even if very early, could still reflect rapid feed-forward integration of visual and linguistic information (Thierry et al., 2009). The prestimulus modulations of alpha waves thus

provide additional support for the idea that language biases on vision arise at an early perceptual level. More importantly, this finding provides a candidate mechanism underlying such biases: the activation of sensory priors in perceptual regions via the modulation of posterior alpha-band oscillations.

One open question concerns what features must be activated for successful object recognition. A recent ERP study have demonstrated that activation of object's shape is the main force underlying the facilitatory effect of symbols on visual object recognition (Noorman et al., 2018). Yet, it is not clear whether such shape template contains a holistic representation of the target object (e.g., a prototypical image of a bird's shape) or a set of visual features diagnostic of the target category (e.g., bird's wings, beak, legs). Future studies are needed to unveil what is the specific content of the top-down representations that mediate visual object recognition.

It is important to note that the reported results could be coherent also with other cognitive mechanisms that are not necessarily predictive. In our study, prediction is confounded with attention. Since in our task participants had to decide whether a certain image matches or not a cued category, it is likely that subjects were attending at those features which are critical to distinguish the target category. Under an attention account, our results can be cast in terms of category-based attention, with words being uniquely effective at deploying attentional guidance to visual categories (Lupyan, 2008; Zelinsky & Yang, 2009).

The increase of alpha oscillations can also reflect the implementation of visual information within a working memory template. Indeed, synchronization in the alpha band has been shown to be modulated by the amount of features retained in visual working memory (e.g., Jensen et al., 2002). Such an account, we believe, is not in contradiction with a predictive or attention-based account, but it may rather reflect a candidate mechanism supporting these processes.

More difficult in our opinion is the interpretation of prestimulus beta oscillations. Despite the fact that beta waves have been often been associated to visual prediction, they also have been associated to a large variety of top-down and cognitive processes. Moreover, the topographical distribution of beta synchronization is not very informative about the *type* of beta we observed. In chapter 3, we speculated that the different modulation of beta waves by sounds and words may reflect a difference in the features associated to these types of cues, that is, they size of the neurocognitive network, or load. Indeed, it is well established that words activate a larger space of features (e.g., gramamtical, lexical, etc.) during processing. An alternative but related interpretation can be that beta oscillations reflect the activation of sumpramodal category information associated to a current representation. Monkey studies on the neural bases of categorization have often associated beta rhythms in parietal and prefrontal cortices to high-level abstraction, i.e. categorization of objects' classes irrespective of their perceptual similarities (Antzoulatos and Miller, 2014, 2016; Wutz et al., 2018). The difference in beta modulations for spoken words vs natural sounds may indicate that words are more effective at activating supramodal-categorical states before stimulus presentation. Thus, while alpha rhythm may be responsible for carrying sensory features which are diagnostics of physically similar exemplars (e.g., birds), beta oscillations may be responsible for activating conceptual states (e.g., animals) about category members differing in their physical properties (e.g., birds and sharks). However, this interpretation remains merely speculative. Further, studies are needed to clarify the specific division of labor between prestimulus alpha and beta waves in language-driven object recognition and top-down signaling in general.

Our second study showed that life-long experience with language affects information processing of simple sequences of sounds alternation in duration. Similarly to the previous study, we found support for the idea that such biases arise at an early sensory level. Indeed,

linguistic schemes have been shown to modulate the auditory MMN – a component traditionally associated to low-level sensory processes.

Our experimental hypotheses were strongly influenced by previous work on perceptual grouping in Basque and Spanish (Molnar et al., 2016), showing that these groups of speakers have different grouping biases during auditory rhythm perception. Several behavioral studies have reported that non-linguistic perceptual grouping, despite showing some universal biases, is largely modulated by linguistic experience (Bhatara et al., 2013; Iversen et al., 2008; Yoshida et al., 2010). These language-specific influences also apply to other domains of auditory perception. For instance, native speakers of languages in which pitch carries phonemically meaningful information (i.e., tone languages; e.g., Mandarin Chinese) benefit from a behavioral advantage in non-linguistic pitch discrimination tasks as compared to speakers of non-tone languages like English (e.g., Bidelman et al., 2013). Similarly, experience with languages that use duration to differentiate between phonemes (e.g., Finnish, Japanese) can boost the ability to discriminate the duration of non-linguistic sounds (Tervaniemi et al., 2006). Despite the fact that our study lacks a behavioral part, it strongly suggests that these previously reported effects of language on perception may arise from a hierarchical predictive coding mechanism, with predictive signal coming from higher-levels (linguistic) levels interacting with predictions generated at lower stages. Specifically, the auditory system increases the weighting of low-level acoustic features which are relevant to parse language material, and re-applies such coding strategy to the processing on non-linguistic sounds. This interpretation is in line with the idea that shared auditory and computational resources underlie speech, sound, and music processing (Asaridou & McQueen, 2013). By differentiating between context-based and (linguistic) experience-based prediction, our experiment also demonstrates how the study of the interplay between language and perception can provide an excellent model to investigate the hierarchical organization of predictive processing in the neocortex.

One question for the future concerns whether these effects generalize also to other sensory modalities. Similar grouping patterns may emerge during natural reading. Speakers of functor-initial vs functor-final languages may chunk visual linguistic material differently depending on the phrasal properties of their native language. This may in turn affect how non-linguistic visual material is segmented into meaningful units. However, the effect of reading-based chunking on visual grouping is a largely unexplored topic, both at the behavioral and neural level. Future studies are thus needed to elucidate the impact of reading on visual grouping.

It is important to note that our two studies differ each other in many meaningful ways. For instance, in the first study the linguistic priors are implemented online, since the hypothesized predictive representations are triggered by a language cue (i.e., a word). In the second study, the effects of language on perception are off-line. The biases of linguistic experience on perception arise automatically, with no need to activate a language context. Despite these qualitative differences, both studies converge in showing that linguistic priors shape perceptual processes at early stages and in a predictive-like manner. However, it is important to note that some previous studies did not find evidence for such early modulations. For instance, Francken et al., (2015) used motion words to cue motion stimuli during a motion detection task. They reported a priming effect when the motion words and the motion stimuli were congruent. Such congruency effect was accompanied by a larger activation in the fMRI signal over the left middle temporal cortex, but not in motion-specific regions of the visual stream. Similarly, Tan et al. (2008) presented participants with a perceptual discrimination task on easy-to-name and hard-to-name coloured squares. Using fMRI, they found that color discrimination recruited regions selective for color knowledge and regions in the bilateral frontal gyrus. However, easy-to-name colors elicited stronger activation in the left posterior superior temporal gyrus and inferior parietal lobe compared to hard-to-name coloured squares, but no meaningful difference was found in occipital regions.

The left posterior superior temporal gyrus and inferior parietal lobe are regions typically associated to color naming, thus suggesting that the interaction between language and perception emerged in higher-level linguistic areas. These findings indicate that the interaction between language and perception may follow different routes depending on contexts and task demand. More studies are thus needed to unveil when and how linguistic biases on perception follow a certain route or another.

5.3 Concluding remarks

In the present thesis, we showed how the predictive processing framework can provide a valid account to investigate the impact of cultural systems like language on putatively low-level mechanisms like perception. The study of such interactions can provide novel insights into different domains of cognitive (neuro)science, including the broad debate on how culture shapes cognition and brain wiring. Beyond providing a research framework to investigate such empirical questions, the study of language-perception interaction offers a flexible model to study the influence of experience-based prediction on lower-level sensory processes. Understanding how brain circuits at different stages implement predictive algorithm might provide a solid grounding to understand disorders characterized by a disruption of the predictive machinery, such as dyslexia, schizophrenia or autism.

Chapter - 6 III Appendices

6.1 Appendix A: List of publications derived from the thesis

Morucci, P., Giannelli, F., Richter, C., Molinaro, N., (under review), Alpha and Beta Rhythms Differentially Support the Effect of Symbols on Visual Object Recognition.

Morucci, P., Martin, C., Molinaro, N., (in prep) Language Experience Affects Predictive Coding during Musical Beat Perception.

6.2 Appendix B: Resumen en Castellano

La historia de la ciencia abarca numerosas preguntas desafiantes, incluida la pregunta sobre el origen de nuestra percepción sobre el mundo: Cómo las poblaciones de neuronas dan lugar a percepciones. Antes de que los científicos pudieran registrar la actividad cerebral, los filósofos llevan planteando esta cuestión desde hace tiempo. Una idea común es que la percepción nos proporciona una representación verídica de lo que hay en el mundo externo —una postura conocida como realismo directo o ingenuo. Otra alternativa se denomina representacionalismo, y sugiere que nuestra percepción consciente no refleja el mundo real en sí, sino una mera representación interna generada por la mente/cerebro en un intento por encontrar las causas del mundo externo.

Fenómenos como los sueños, las alucinaciones y las ilusiones perceptuales sugieren que la realidad y nuestra experiencia con ella no son exactamente lo mismo. Más bien, estos fenómenos sugieren que nuestra percepción se asemeja más al proceso de construcción o inferencia, en el cual las creencias y expectativas previas moldean en gran medida la forma en la que experimentamos el mundo exterior.

La idea de la percepción como proceso de inferencia (von Helmholtz, 1867) es ahora un supuesto clave sobre las perspectivas de procesamiento predictivo sobre la percepción (Clark 2016; Friston, 2005; Rao and Ballard, 1999). En comparación con las teorías clásicas que conciben el cerebro como un dispositivo que procesa y registra de forma pasiva la información externa, la hipótesis del cerebro como máquina predictiva sugiere que una de las funciones fundamentales del cerebro consiste en la anticipación de futuros eventos. Esta hipótesis influye cada vez más en la Neurociencia Cognitiva. Las teorías sobre el procesamiento predictivo comparten la idea de que el cerebro desarrolla modelos generativos de la realidad y utiliza dichos modelos para inferir las causas que rigen el entorno externo (Clark, 2016). Tales modelos generativos se conciben como una jerarquía

de procesamiento: sugieren que las predicciones se transmiten a través de la señal de *feedback* o retroalimentación derivada de las áreas corticales superiores a las inferiores, las cuales se espera que tengan un efecto supresor en las señales entrantes. A continuación, las predicciones se comparan con señales sensoriales *bottom-up* en cada nivel de la jerarquía. Se postula que solo la diferencia llamada error de predicción podría propagarse a través de conexiones *feedforward* de las áreas corticales inferiores a las superiores para actualizar los modelos internos. Se cree que este tipo de computación podría ser canónica, lo que significa que cada parte de la corteza implementa este tipo de algoritmo predictivo.

Más allá de proporcionar un marco de investigación para estudiar una pregunta existencial milenaria, entender cómo surge el procesamiento predictivo de los circuitos corticales conlleva implicaciones médicas y éticas considerables. Entender cómo los circuitos cerebrales en distintas fases implementan algoritmos predictivos podría ofrecer una base sólida para comprender los trastornos caracterizados por una alteración de la maquinaria predictiva, como la dislexia, la esquizofrenia o el autismo.

Considerar la percepción como proceso predictivo implica considerar la percepción como un proceso «penetrable» que puede ser influido tanto por el conocimiento previo como por la expectación, en la medida en que dicha penetración sea efectiva a la hora de reducir el error de predicción. Sin embargo, ¿qué cuenta como conocimiento previo? En los seres humanos, una forma de conocimiento previo es el lenguaje. Diversos estudios han demostrado que el lenguaje puede influir en otros sistemas no lingüísticos, como la categorización, la memoria y la percepción. El estudio de la interacción entre el lenguaje y la percepción ofrece un modelo único para abordar varias cuestiones sin respuesta sobre la naturaleza predictiva de la experiencia en el cerebro humano, así como el efecto de la cultura en la cognición.

Una cuestión fundamental está relacionada con la naturaleza del mecanismo neural a través del cual el lenguaje afecta a los procesos perceptuales. Algunas hipótesis sugieren que los efectos del lenguaje son de «alto nivel», lo que significa que el lenguaje no afecta a los procesos perceptuales tempranos, sino que interactúan en las etapas conceptuales o de toma de decisiones más tardías. Propuestas más recientes sugieren que el lenguaje puede alterar los procesos perceptuales a niveles sensoriales tempranos. Esta última idea coincide con las teorías actuales sobre el procesamiento predictivo de la percepción, lo que sugiere que los procesos sensoriales están ampliamente influenciados por el conocimiento previo y la expectación.

La presente tesis tiene como objetivo investigar los mecanismos neurofisiológicos subyacentes a la interacción entre lenguaje y percepción. Nos centramos en dos tipos específicos de interacción lenguaje-percepción: i) el efecto de las etiquetas lingüísticas en el reconocimiento de las categorías de objetos visuales; y ii) el efecto del conocimiento lingüístico en el procesamiento neural de los sonidos rítmicos. Abordamos estas cuestiones mediante un enfoque interdisciplinar combinando medidas conductuales, de electrofisiología humana y enfoques estadísticos avanzados. Utilizamos medidas electrofisiológicas de resolución temporal como la electroencefalografía (EEG) y la magnetoencefalografía (MEG). Estas técnicas permiten registrar la actividad cerebral con excelente resolución temporal, por lo que nos ayuda a monitorizar los procesos cognitivos en desarrollo con una precisión temporal única. Con el fin de evaluar la naturaleza de los mecanismos computacionales subyacentes a las interacciones lenguaje-percepción, nos enfocamos en los índices neurales presuntamente asociados al procesamiento predictivo. Dentro del marco del procesamiento predictivo, la actividad oscilatoria en las bandas alfa y beta se han asociado tradicionalmente con los procesos *top-down*, por lo que representan un mecanismo que es candidato a contener predicción perceptual. De igual forma, algunos Potenciales/Campos Relacionados con Eventos tempranos, como el potencial de

disparidad auditivo, se han considerado en gran medida como indicadores de error de predicción cortical.

En el primer estudio investigamos cómo las etiquetas lingüísticas afectan al reconocimiento de las categorías de objetos visuales. Aprovechamos el hecho de que las palabras habladas incremental el reconocimiento de objetos visuales en mayor medida que los sonidos naturales —un efecto llamado *label-advantage*. Utilizamos la técnica EEG y el análisis tiempo-frecuencia para evaluar las dinámicas electrofisiológicas subyacentes a este efecto conductual. Al contrario que estudios anteriores, nos centramos en el intervalo de tiempo precedente a la aparición del objeto visual, estableciendo directamente como objetivo la predicción de arriba a abajo. Primero replicamos la ventaja de la etiqueta mencionada previamente y demostramos que este efecto conductual persiste incluso cuando las palabras se presentan en un segundo idioma, lo que indica que la ventaja de la etiqueta depende de propiedades inherentes de los símbolos verbales para implementar información precisa sobre la categoría. Cabe destacar que la ventaja conductual mencionada para las palabras habladas se asoció con un incremento en los ritmos alfa y beta posteriores en el intervalo de tiempo entre la desaparición de la pista y la aparición del objeto meta. El análisis de correlación desveló que ambos ritmos contribuyen al rendimiento en el reconocimiento del objeto visual. Sin embargo, parecía que operaban en direcciones ortogonales: el rendimiento en el reconocimiento de objetos mejoró cuando aumentaba la sincronización en alfa, pero disminuyó con la mejora de los ritmos beta. Este hallazgo sugiere una división de tarea entre los ritmos alfa y beta a la hora de orquestar la orientación lingüística *top-down* al sistema visual. Por tanto, las modulaciones pre-estímulo mencionadas de las ondas occipitales alfa y beta apoyan la idea de que las influencias lingüísticas sobre la percepción surgen de maniera *top-down*, y arrojan luz sobre un mecanismo candidato a subyacer dichas influencias: la amplificación de precedentes

sensoriales en regiones occipitales a través de la modulación de las oscilaciones de las bandas alfa y beta.

El segundo estudio investiga si la exposición vital a ciertos patrones lingüísticos impacta en el procesamiento neural de sonidos rítmicos. Comparamos los datos magnetoencefalográficos de hablantes nativos de euskera y castellano, que escucharon secuencias rítmicas de sonidos. Estas dos lenguas difieren en su estructura sintáctica/prosódica, por lo que era un modelo ideal para estudiar el efecto de la experiencia lingüística en el procesamiento predictivo auditivo. Nuestra hipótesis sugiere que el sistema auditivo extrapola diseños abstractos que subyacen a la estructura oracional del lenguaje, y utiliza este conocimiento para generar predicciones a largo plazo sobre los sonidos entrantes. Cabe destacar que, en nuestra manipulación, las secuencias de sonido codificaban patrones abstractos que reproducían la estructura sintáctica y prosódica del euskera y el castellano. Cuando un evento esperado interrumpe una secuencia rítmica de sonidos, la amplitud del potencial de disparidad varía ortogonalmente dependiendo de los antecedentes lingüísticos del individuo. Esta respuesta de error de predicción ocurre alrededor de los 100 ms del inicio del evento desviado, y su magnitud es mayor en regiones auditivas. Este hallazgo indica que los sistemas de codificación empleados para analizar el material lingüístico son reciclados por el sistema auditivo para implementar modelos predictivos del entorno. Asimismo, este estudio también ofrece perspectivas novedosas sobre la organización jerárquica de las predicciones auditivas.

El estudio de la interacción entre lenguaje y percepción puede aportar enfoques novedosos sobre diversos campos de la neurociencia cognitiva, incluyendo cómo la cultura modela la cognición, así como el efecto del conocimiento de alto nivel sobre los procesos de bajo nivel. Mediante la identificación de los componentes electrofisiológicos que caracterizan dichas interacciones, esperamos que estos resultados ayuden a definir con

mayor precisión las implicaciones de estudiar sistemas simbólicos a la hora de esculpir nuestro conocimiento del mundo.

6.3 Appendix C: Bibliography

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. Wellcome open research, 4.

Antzoulatos, E. G., & Miller, E. K. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. Neuron, 83(1), 216-225.

Antzoulatos, E. G., & Miller, E. K. (2016). Synchronous beta rhythms of frontoparietal networks support only behaviorally relevant representations. Elife, 5, e17822.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. Trends in cognitive sciences, 16(7), 390-398.

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. Frontiers in psychology, 4, 321.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural computation, 7(6), 1129-1159.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. Journal of Neuroscience, 29(26), 8447-8451.

Berger, H. (1929). Über das elektroenkephalogramm des menschen. Archiv für psychiatrie und nervenkrankheiten, 87(1), 527-570.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. Proceedings of the National Academy of Sciences, 106(5), 1672-1677.

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. The Journal of the Acoustical Society of America, 134(5), 3828-3843.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. PloS one, 8(4), e60676.

Bogaerts, L., Richter, C. G., Landau, A. N., & Frost, R. (2020). Beta-band activity is a signature of statistical learning. Journal of Neuroscience, 40(39), 7523-7530.

Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., ... & Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. Neuron, 75(5), 875-888.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: a label advantage in object recognition. Journal of Neuroscience, 35(25), 9329-9335.

Bressler, S. L., & Richter, C. G. (2015). Interareal oscillatory synchronization in top-down neocortical processing. Current opinion in neurobiology, 31, 62-66.

Bressler, S. L., & Tognoli, E. (2006). Operational principles of neurocognitive networks. International journal of psychophysiology, 60(2), 139-148.

Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. Journal of Neuroscience, 33(39), 15454-15465.

Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. Canadian Journal of Psychology/Revue canadienne de psychologie, 34(1), 77.

Chennu, S., Noreika, V., Gueorguiev, D., Shtyrov, Y., Bekinschtein, T. A., & Henson, R. (2016). Silent expectations: dynamic causal modeling of cortical prediction and attention to sounds that weren't. Journal of Neuroscience, 36(32), 8305-8316.

Clark, A., & Toribio, J. (2012). Magic words: how language augments hum an computation. In Language and Meaning in Cognitive Science (pp. 33-51). Routledge.

Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press.

Cohen, D. (1972). Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer. Science, 175(4022), 664-666.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. Neuron, 70(2), 200-227.

Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. Proceedings of the National Academy of Sciences, 107(37), 16396-16400.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. Cognition, 143, 93-100.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 54-75.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo?. Current opinion in neurobiology, 20(2), 156-165.

Francken, J. C., Kok, P., Hagoort, P., & De Lange, F. P. (2015). The behavioral and neural effects of language on motion perception. Journal of cognitive neuroscience, 27(1), 175-184.

Friston, K. (2005). A theory of cortical responses. Philosophical transactions of the Royal Society B: Biological sciences, 360(1456), 815-836.

Friston, K. (2018). Does predictive coding have a future?. Nature neuroscience, 21(8), 1019-1021.

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. Journal of Neuroscience, 32(5), 1791-1802.

Garrido, M. I., Friston, K. J., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Kilner, J. M. (2008). The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage, 42(2), 936-944.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. Clinical neurophysiology, 120(3), 453-463.

Hari, R., Salmelin, R., Mäkelä, J. P., Salenius, S., & Helle, M. (1997). Magnetoencephalographic cortical rhythms. International journal of psychophysiology, 26(1-3), 51-62.

Haegens, S., Vergara, J., Rossi-Pool, R., Lemus, L., & Romo, R. (2017). Beta oscillations reflect supramodal information during perceptual judgment. Proceedings of the National Academy of Sciences, 114(52), 13810-13815.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of modern Physics, 65(2), 413.

Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. Cognition, 173, 8-15.

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. Neuroscience, 389, 54-73.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nature neuroscience, 19(5), 665-667.

Hämäläinen, M., & Hari, R. (2002). Magnetoencephalographic (MEG) characterization of dynamic brain activation. Brain mapping: the methods, Ed, 2, 227-254.

Hawkins, J., & Blakeslee, S. (2004). On intelligence. Macmillan.

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing–MEG dissociates early perceptual and late N400 effects. Brain and Language, 116(2), 91-96.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. Second Language Research, 29(1), 33-56.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. Cognition, 96(1), B23-B32.

Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. Visual Cognition, 15(8), 985-1018.

Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., & Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. Neuroimage, 13(6), 1073-1089.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. The Journal of the Acoustical Society of America, 124(4), 2263-2271.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. Psicológica, 35(1), 49-66.

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. European journal of Neuroscience, 15(8), 1395-1399.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Frontiers in human neuroscience, 4, 186.

Jewett, D. L., Romano, M. N., & Williston, J. S. (1970). Human auditory evoked potentials: possible brain stem components detected on the scalp. Science, 167(3924), 1517-1518.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clinical Neurophysiology, 111(10), 1745-1758.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different?. Linguistic Approaches to Bilingualism, 4(2), 257-282.

Klemfuss, J. Z. (2015). Differential contributions of language skills to children's episodic recall. Journal of Cognition and Development, 16(4), 608-620.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain research reviews, 29(2-3), 169-195.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. Brain research reviews, 53(1), 63-88.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. Trends in cognitive sciences, 16(12), 606-617.

Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. Proceedings of the National Academy of Sciences, 114(39), 10473-10478.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. Nature, 427(6971), 244-247.

Knakker, B., Weiss, B., & Vidnyánszky, Z. (2015). Object-based attentional selection modulates anticipatory alpha oscillations. Frontiers in human neuroscience, 8, 1048.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. Language, cognition and neuroscience, 31(1), 32-59.

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The influence of language on perception: listening to sentences about faces affects the perception of faces. Journal of Neuroscience, 30(45), 15254-15261.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behavior research methods, 44(2), 325-343.

Lennie, P. (2003). The cost of cortical computation. Current biology, 13(6), 493-497.

Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. The American Statistician, 34(4), 216-221.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. The Journal of the Acoustical Society of America, 106(4), 1633-1654.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press.

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). Cognition, 108(2), 566-577.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. Journal of Experimental Psychology: General, 141(1), 170.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196-14201.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. Current Directions in Psychological Science, 24(4), 279-284.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. Trends in cognitive sciences.

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology, 47(1), 66-122.

Mayer, A., Schwiedrzik, C. M., Wibral, M., Singer, W., & Melloni, L. (2015). Expecting to see a letter: alpha oscillations as carriers of top-down sensory predictions. Cerebral Cortex, 26(7), 3146-3160.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods, 164(1), 177-190.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron, 89(2), 384-397.

Miller, T. M., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. Cognition, 171, 172-179.

Mo, J., Schroeder, C. E., & Ding, M. (2011). Attentional modulation of alpha oscillations in macaque inferotemporal cortex. Journal of Neuroscience, 31(3), 878-882.

Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. Cognition, 164, 61-73.

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. Cognition, 152, 150-159.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology, 38(1), 1-21.

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. Clinical neurophysiology, 115(1), 140-144.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical neurophysiology, 118(12), 2544-2590.

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. Scientific reports, 8(1), 1-10.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience, 2011.

Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscillations. Trends in neurosciences, 30(4), 150-158.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of neuroscience methods, 162(1-2), 8-13.

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. Brain research, 1301, 89-99.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. Behavioral and brain sciences, 22(3), 341-365.

Raij, T., McEvoy, L., Mäkelä, J. P., & Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. Brain research, 745(1-2), 134-143.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience, 2(1), 79-87.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. Perception, 33(2), 217-236.

Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. Scientific reports, 8(1), 1-14.

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted" what" but not" when". Frontiers in human neuroscience, 7, 407.

Simanova, I., Francken, J. C., de Lange, F. P., & Bekkering, H. (2016). Linguistic priors shape categorical perception. Language, Cognition and Neuroscience, 31(1), 159-165.

Snyder, A. C., & Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. Journal of Neuroscience, 30(11), 4024-4032.

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. Cognition, 166, 277-297.

Spehlmann, R. (1965). The averaged electrical responses to diffuse and to patterned light in the human. Electroencephalography and clinical neurophysiology, 19(6), 560-569.

Spitzer, B., Fleck, S., & Blankenburg, F. (2014). Parametric alpha-and beta-band signatures of supramodal numerosity information in human working memory. Journal of Neuroscience, 34(12), 4293-4302.

Spitzer, B., & Haegens, S. (2017). Beyond the status quo: a role for beta oscillations in endogenous content (re) activation. eneuro, 4(4).

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. Journal of neuroscience, 30(9), 3531-3543.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nature neuroscience, 11(9), 1004.

Tan, L. H., Chan, A. H., Kay, P., Khong, P. L., Yip, L. K., & Luke, K. K. (2008). Language affects patterns of brain activation associated with perceptual decision. Proceedings of the National Academy of Sciences, 105(10), 4004-4009.

Tervaniemi, M., Jacobsen, T., Röttger, S., Kujala, T., Widmann, A., Vainio, M., ... & Schröger, E. (2006). Selective tuning of cortical sound-feature processing by language experience. European Journal of Neuroscience, 23(9), 2538-2541.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. Proceedings of the National Academy of Sciences, 106(11), 4567-4570.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. Journal of Neuroscience, 26(37), 9494-9502.

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. Journal of Neuroscience, 32(39), 13389-13395.

Van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M. A., Poort, J., Van Der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proceedings of the National Academy of Sciences, 111(40), 14332-14341.

Vuust, P., Pallesen, K. J., Bailey, C., Van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter: pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. Neuroimage, 24(2), 560-564.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music–brain responses to rhythmic incongruity. cortex, 45(1), 80-92.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. Proceedings of the National Academy of Sciences, 108(51), 20754-20759.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex. Journal of Neuroscience, 20(6), RC63-RC63.

Wutz, A., Loonis, R., Roy, J. E., Donoghue, J. A., & Miller, E. K. (2018). Different levels of category abstraction by different dynamics in different prefrontal areas. Neuron, 97(3), 716-726.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. Neuroreport, 8(8), 1971-1974.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. Vision research, 49(16), 2095-2103.

Yon, D., de Lange, F. P., & Press, C. (2019). The predictive brain as a stubborn scientist. Trends in cognitive sciences, 23(1), 6-8.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. Cognition, 115(2), 356-361.