eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Oscillatory characterization of sensory wordform pre-activation in the visual and auditory domains.

Irene Fernández Monsalve

Euskal Hizkuntza eta Komunikazio Saila

Euskal Herriko Unibertsitatea

Supervised by

Nicola Molinaro and Mathieu Bourguignon

A thesis submitted for the degree of

Doctor of Philosophy

Donostia 2019

Acknowledgements

I would like to firstly thank my supervisors, Nicola and Mathieu, for their support and guidance, but especially to Nicola, for always respecting and encouraging a positive work- personal-life balance. Without such support I would not have been able to complete a thesis at this point in my life. I also want to thank Manolo and the predoc-selection team, for giving me an opportunity despite my un-grantable "viejuna" status.

A big thanks also to the support teams at the BCBL. The admin team, for always being encouraging and caring, and for their invaluable support in navigating bureaucracies. The lab teams, always ready to help us out and making the data collection process fun, but especially to the research assistants, Sara and Itziar, who were instrumental in the data collection process for the present thesis. And I really want to also thank the IT team, who lived in a small and crowded room in a lower floor, and answered a phonecall on my first day with a "Have you tried turning it off and on again?". Seriously, Iker, for always being there to resolve incidences big and small, and for going beyond, opening our eyes to python, git and other wonderful technologies. Borja, for getting my data back almost immediately every time I managed to accidentally erase months-worth of work. And Jose and Guti, for allowing and installing linux on my desktop computer, which at last made me feel totally at home.

I must also acknowledge the contribution of Ainhoa Bastarrika, that was not only part of the wonderful predoc community, but also shared with me her adaptation of the "ociamthesis" LaTeX template used in the current work, you saved me a lot of time in formatting! And finally my biggest, most heart-felt thanks must go to all my fellow pre-docs, for the super interesting discussions on the science, and the always warm and supportive environment. Sharing with you the hard times and the frustrations, and turning them into laughs over pintxo-pote made them not only bearable, but actually worth it.

Abstract

Prediction has been proposed to be a fundamental cognitive mechanism. However, empirical data regarding the neural correlates of the predictions themselves are still scarce. One outstanding question, especially debated within the language processing literature, is what degree of perceptual detail is carried by predictive representations. Despite the ubiquitous data showing the role of meaning-based expectations, their translation into word-form sensory predictions is contended. Most prior studies compared highly predictable words to non-predictable counterparts, imposing different attentional demands that could confound the results. Furthermore, studies often focused on post-target-word effects, thus examining the consequences, rather than the generation, of the prediction. In the present thesis, we investigated the presence of sensory wordform pre-activation by comparing activity before two highly predictable words differing only along a sensory dimension. Several versions of the same paradigm were conducted with written or spoken words to examine the role of sensory modality, and with either fixed or variable delays to explore the role of temporal predictability. We used time-frequency analysis of MEG signals, to relate our findings to prior work describing the implementation of predictive processing through neural oscillations.

Our results showed that pre-word activity was modulated as a function of expected word-form features both in the auditory and visual domains. Congruent with prior literature, effects arose over superior temporal areas in the theta band for auditorily presented words, and over ventro-occipital cortex in the beta band for written ones. Predictability of the temporal onset had a determining influence in the auditory, but not the visual domain. Differences across modalities may reflect preferential oscillatory signatures of different brain areas as well as specific characteristics of the speech and written texts, with time being a defining dimension in phonological categorization, but not in orthographic processing. These results contribute to the understanding of the role of predictive processing in general and in the linguistic domain specifically, showing that word-form pre-activations may be generated when context clearly biases towards a specific word candidate. By identifying the oscillatory frequencies that characterize such pre-activation, we hope these results will help interpret pre-target word brain activity in other more naturalistic set-ups.

Contents

Part I

Theoretical and empirical background

Chapter 1

Theoretical background

In the present chapter we provide the theoretical background to the main question addressed by our studies: can predictive sensory-level representations be generated before word onset. In the first section we introduce the concept of predictive processing, explaining some prominent models. We then frame the question of sensory pre-activation and the possible role of neural oscillations in this context. In the second section we shift our attention to neural oscillations, giving some general background on their cognitive interpretation, and in language specifically. Finally, in section 3 we hone in on prediction in language. We outline the main neuro-cognitive models of language processing and word recognition, examining the role for prediction in each of them. We review studies providing evidence for wordform pre-activation, and identify the limitations that we aim to address with the present work.

1.1 The proactive brain

Our senses provide us with vast amounts of data regarding our environment in a continuous manner. This data stream is complex and often ambiguous, so that interpreting it into meaningful precepts is no trivial task. Nevertheless, our environment is also highly structured, with important regularities. Dogs bark, objects fall downwards, in the English language, articles are followed by nouns. Predictive processing theories thus argue that perception must rely on internal world models representing such regularities in order to respond in real time to the complex and often incomplete information provided by the senses. In this way, perception is not just a passive mechanism for information gathering, but a constructive process

in which prior world knowledge is used to interpret incoming stimulation. The integration of current knowledge and incoming stimulation would be achieved through a process of probabilistic inference, where hypotheses about the likely causes of sensory stimulation are formulated and checked against actual sensory input. This comparison serves to prove or disprove the hypothesis, which can be revised in a recurrent fashion aimed at minimizing the prediction error.

The idea of perception as reverse inference was originally proposed by Helmholtz (1860), at least partly inspired by the 18th century philosopher Immanuel Kant (Swanson, 2016), but it is currently receiving increased attention. Converging evidence from the cognitive neuroscience of perception (e.g. Kveraga et al., 2007) and motor control (e.g. Flanagan et al., 2003) have shown the prevalence of predictive processing, with generative models being at the heart of both perception and action. This has fueled theoretical work that places prediction as a fundamental cognitive mechanism, able to provide the basic principle of brain function (Friston, 2005; Bar, 2007; Clark, 2013), with alterations in its function being able to explain a wide range of cognitive disorders, such as depression (Bar, 2009), schizophrenia (Brown et al., 2013), functional movement disorders (Edwards et al., 2012), hallucinations (Kumar et al., 2014), aphasia (Cope et al., 2017), etc.

In parallel, proposals for possible computational and neural (Rao and Ballard, 1999; Lochmann and Deneve, 2011) implementations of such active inference mechanisms have grounded predictive processing at the brain level. One of the leading models attempting to describe the neural and computational architecture of prediction is predictive coding. This account is based on two main ideas: that predictions develop over a hierarchically organized system, and that there is a segregation in the flow of predictions and prediction-errors across the hierarchy. Each hierarchical level would generate internal predictions that would be transmitted to the level below, trying to explain incoming information at that lower level. Any discrepancy between such predictions and the incoming information, the prediction error, would be transmitted back up to the higher level where the prediction could be revised with the new evidence. In this way, the only information that would be fed forward in the system would be the prediction error. At the neural level, superficial pyramidal cells, with their feed-forward connections would act as error units, propagating prediction error in a bottom-up fashion. On the other hand, pyramidal cells within deep cortical layers, with their feedback projections would carry predicted representations. The asymmetry in connectivity would be accompanied by asymmetric synchronization patterns, with gamma-range frequencies predominating in superficial cortical layers, and alpha or beta frequencies predominating in deep layers.

Pre-activation of sensory representations

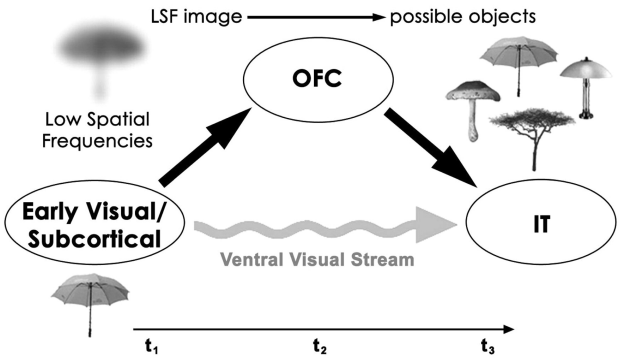

Figure 1.1: Prediction in the visual domain (Bar, 2007). Early low spatial frequency information (LSF) triggers the generation of predictions. OTC, orbito-frontal cortex; IT, inferior temporal cortex.

However, little is still known about the neural correlates of the predictions themselves. One outstanding question, that is the main focus of the present thesis, is to what degree of perceptual detail predictions may be generated before any sensory evidence is available. In stronger predictive processing accounts, prior expectations could lead to the pre-activation of expected precept representations, down to the earliest sensory processing areas (Kok et al., 2017). On the other hand, perceptual pre-activations may only be generated after initial sensory stimulation is available (Bar, 2007; Zylberberg et al., 2009), with the response to a stimulus taking place in two distinct steps: an initial bottom-up sweep would lead to the generation of top-down hypothesis that would guide subsequent processing (see Figure 1.1 for an example).

There is extensive evidence that valid prior expectations facilitate stimulus processing, and some studies have traced this facilitation back to primary sensory cortices. For example, Kok et al. (2012) showed that onset of an expected visual stimulus led to an attenuated but sharpened response in primary visual cortex. Participants were presented with grating stimuli of different orientations (see Figure 1.2) which elicit differential responses from orientation-selective neurons in primary visual cortex. Predictability of the orientation was manipulated using auditory tones as cues presented before the gratings. The overall neural response to expected gratings was reduced with respect to unexpected counterparts, but multivariate pattern analysis revealed better classification of orientation in the former, suggesting an enhanced orientation-selectivity in primary visual cortex for predicted stimuli. Stimulus omission studies go a step further, by examining response of primary sensory areas to omissions of highly expected stimuli. For example, in the auditory domain, SanMiguel et al. (2013) carried out an experiment where participants pressed a

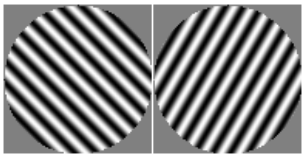

Figure 1.2: Example of grating stimuli with different orientations.

button to trigger a sound. Under the predictive condition the sound triggered by the button press was always the same, whilst in the random condition the tone of the sound varied randomly. Rare omissions were included in both conditions, and an auditory response to omissions was found in the predictive, but not the random conditions. This suggests that the observed response was a feature-specific surprise, only generated when prior knowledge could predict the identity of the upcoming sound, licensing a sensory detailed expectation.

However, few studies have attempted to identify these representations in the interval before stimulus delivery. One exception to this trend is a study by Kok et al. (2017), that trained classifiers on stimulus grating orientations with MEG data whilst participants were actually seeing the grating, and then tested the same classifiers before presentation of a grating where there was a high expectation of seeing a particular orientation. They showed above average decoding before stimulus onset, but this was only significant when subtracting the decoding performance when the expectation corresponded to the trained orientation from trials when the expectation contradicted the presented (and trained) orientation.

Another exception is a study by Stokes et al. (2009), who found selective activation of shape-specific neural subpopulations within visual cortex (lateral occipital complex) before onset of the actual targets. However, in this case, the detected activity was the result of an attentional-selection mechanism, rather than a predictive pre-activation: participant's task was to detect the presence of a certain target (letter 'X' vs letter 'O') in near threshold stimuli. However, conditional probability of the two possible stimuli was the same (25% each, with 50% of trials showing no letters). In this case, the detected activity could be attributed to feature-based attention (only one target was relevant) rather than feature-based expectation (both targets were equally probable). Summerfield and Egner (2016) highlight the importance of distinguishing between these two cognitive processes: feature-based attention, where one sensory feature is of special relevance for a given task, and feature-based expectation, where on feature is deemed highly probable given prior knowledge and context.

Summing up, although there is extensive evidence for the presence of predictive processes in perception, little is still known about the extent and neural correlates of these mechanisms. The present thesis aims to contribute to the evidence for predictive pre-activation by addressing the identified gaps in the literature: we focused on the study of neural activity before expected stimuli, and we kept conditional probability of targets at 100%.

1.2 Neural oscillations

The presence of on-going rhythmic electric activity in the brain has been known ever since electroencephalography was invented by Hans Berger in the 1930s, with his description of the alpha rhythm. It later became apparent that power at different frequencies correlated with cognitive and memory performance (Klimesch, 1999). Since then, evidence has been mounting for the close relationship between neural oscillations and the dynamic processes of cognition (Ward, 2003). In particular, they may be particularly relevant for predictive processes, with oscillations at different frequencies indexing top-down and bottom-up information flow (Bastos et al., 2012). These oscillations reflect rhythmic fluctuations in neuronal excitability, capable of modulating both output spike timing and sensitivity to synaptic input (Canolty et al., 2006). Neural population communication would be mediated through alignment of excitability states (Fries, 2005), which would be achieved through synchronization of their oscillating frequencies. Oscillations may thus play a key role in the functional linking and segregation of neural populations supporting a wide range of cognitive processes. Neural oscillations are characterized by their frequency and cortical origin, and are typically divided into discrete bands: delta (1-4 Hz), theta (4-8 Hz), beta (13-30 Hz) and gamma (>30 Hz). Local synchrony across neurons can be observed at the scalp with electrophysiological techniques as a power increase in certain frequency bands.

Synchronization in the gamma range was initially proposed to solve the perceptual binding problem. Given that different object properties would be encoded by neural populations in different brain areas, some binding mechanism would be necessary to generate a coherent, unified precept. The binding-by-synchrony hypothesis was thus proposed, with gamma band oscillations providing temporal synchrony for encoding of relations in the visual domain (Singer et al., 1997; Tallon-Baudry, 1999). However the importance of slower oscillations was soon also recognized (Başar et al., 2001). Whilst fast oscillations could synchronize cells within local neural populations, integration over larger spatial scales would require longer integration times, and thus slower frequencies such as theta and alpha (von Stein and Sarnthein, 2000). Both of these rhythms are readily observed, and have been associated with different cognitive processes.

Theta oscillations are a prominent rhythm in the hippocampus and the limbic system in general (Buzsáki, 2002), and have been linked to spatial navigation and memory encoding and retrieval. Alpha oscillations, originating in occipital cortex (but also motor and auditory regions) have been linked to working memory and attentional

processes (Klimesch, 1999). Initially attributed to an ”idling” state, increases in alpha power are currently associated with selective attention mechanisms (Jensen and Mazaheri, 2010). Power increases in this band would reduce excitability of task irrelevant areas, routing information flow through attended networks. In addition, alpha oscillations have been shown to modulate excitability changes in time, impacting the detection of near-visibility stimuli in a rhythmic fashion (Mathewson et al., 2012). Moreover, these oscillations could be modulated as a function of task requirements, achieving top-down control of excitability also along the temporal dimension (Wöstmann et al., 2016). Finally, alpha power also correlates positively with working memory load, which has been taken to reflect the inhibition of incoming stimulation in order to maintain in memory the relevant representations.

In a somewhat similar way, beta power modulations have been implicated in the balance between maintenance of an internal state and response to external stimuli (Engel and Fries, 2010). In the motor system, beta power decreases have been observed during planning and execution of movements, but beta increases are apparent during holding periods following movement and during sustained, repetitive actions. Under a different perspective, beta and alpha power decreases have also been proposed to be functionally responsible for memory encoding and retrieval (Hanslmayr et al., 2012): Power desynchronization in these bands following a stimulus would allow maximal information processing abilities, a requirement for successful encoding.

Neural oscillations may play a specially prominent role in the process of prediction (Arnal and Giraud, 2012). Given their periodic nature, they are ideally suited to implement temporal prediction, and the processing of rhythmic or quasi-rhythmic signals, by aligning excitability of sensory neural populations with the external stimulation (Schroeder and Lakatos, 2009). In addition, the bidirectional flow of information proposed by predictive coding theories could have distinct oscillatory correlates. The functional unit of neocortex, the canonical microcircuit, has been proposed to implement predictive coding computations (Bastos et al., 2012), with superficial cell populations calculating predictive errors, and deep cell populations encoding predictions. Importantly, these layers display specific oscillatory patterns, with gamma-band characterizing superficial layers and alpha/beta bands deep layers (Spaak et al., 2012; Wang, 2010). Indeed, a correspondence between top-down predictions with lower frequency oscillations and bottom-up predictions with gamma oscillations has been reported both for the visual and auditory domains (Kerkoerle et al., 2014; Bastos et al., 2015; Fontolan et al., 2014; Sedley et al., 2016).

Neural oscillations in language processing

Neural oscillations have also been found to have special relevance in language processing, in two distinct ways. Firstly, speech is a quasi-rhythmic signal, where linguistic information is embedded within the speech stream at different time-scales. Cortical oscillations have been shown to play a key role in speech comprehension through entrainment of ongoing neural oscillations to these stream regularities (Giraud and Poeppel, 2012). Synchronization has been observed to follow the pace of phonemes, syllables, and phrases (Meyer, 2018), through gamma, theta, and delta frequencies respectively. Importantly, this entrainment is not just a passive consequence of the input stream regularities, but is under top-down control and can be modulated according to task demands to enhance comprehension (Peelle et al., 2013; Park et al., 2015).

Secondly, the aforementioned domain-general roles for different frequency bands have also been reported in the language processing literature. Theta band power increases have been related to lexical retrieval (Bastiaansen et al., 2005), congruent with their general role in long term memory encoding. Alpha band power increases have been linked to verbal working memory, and to the maintenance of syntactic dependencies (Meyer, 2018), in line with its general role in inhibition and memory maintenance. Beta power decreases have been reported in response to unexpected words in word-processing odd-ball paradigms and after syntactic violations in a sentence (Weiss and Mueller, 2012). This effect could be related to the role of beta oscillations in the balance between external and internally-driven cognitive states, where the unexpected word signals the need for a change in the current cognitive state. From a different perspective, Schoffelen et al. (2017) report frequency-specific directed interactions for different brain areas during language processing: connections originating from temporal regions peaked at alpha frequencies, and those originating from fronto-parietal regions doing so at a beta range.

Finally, gamma oscillations present a somewhat puzzling picture. Whilst a domain-general role of gamma in the representation of prediction errors would predict a negative relationship between gamma and predictability, studies of sentence processing using MEG and EEG have found positive correlations between gamma and predictability of a word given its prior context (Wang et al., 2012; Monsalve et al., 2014). This positive relationship has been interpreted a coupling of near-by neuronal populations arising from successful predictive processing, where representations generated through top-down mechanisms are found to match those generated through bottom-up analysis of the stimulus (Lewis et al., 2015). On the other

hand, a link between gamma and prediction error has been reported for speech sensory processing, but only using depth-electrodes (Fontolan et al., 2014). To solve this incongruency, Lewis and colleagues (2015) propose that gamma band oscillations may be subdivided into two ranges with distinct functions. Low and high gamma bands would reflect synchronization within- and across-hierarchical levels respectively, the former indexing the matching of expected and encountered stimuli, and the latter indexing the transmission of prediction errors. The positive relationship between gamma and predictability seen with non-invasive techniques would This has been interpreted

1.3 Language comprehension and prediction

That context and prior knowledge play a fundamental part in language processing is uncontroversial, with decades of neuro- and psycho-linguistic research having shown contextual facilitation of sentence comprehension and word recognition. Language comprehension is dynamic and incremental: we interpret language as it unfolds, without waiting for the end of a sentence (or word) in order to understand it. Parsing such incomplete and potentially ambiguous fragments must involve predictions about possible completions at least at some representational level. The presence of such predictive processes has been acknowledged in different areas of psycholinguistic research, which we review in the following sections.

Theoretical work: Prediction in word recognition models

Most psycholinguistic models of word recognition acknowledge the presence of top-down influences along a hierarchically organized system, given the wealth of empirical evidence supporting the presence of predictive processes. From the first shadowing experiments, where participants instructed to repeat words did not wait to hear the complete word to initiate repetitions (e.g. Marslen-Wilson and Welsh, 1978), numerous priming studies, showing faster recognition of words after prior facilitating context, to low level perceptual effects, such as phoneme restoration, where participants actually report hearing an absent phoneme within a word (Warren, 1970).

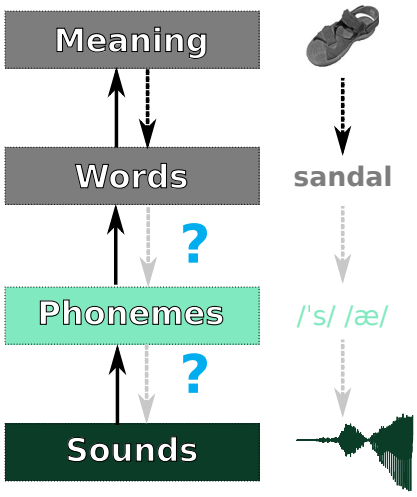

Figure 1.3: Feedforward and feedback connections within the language processing hierarchy.

Typically, these models describe several hierarchical levels of analysis, including phonological (linguistically-meaningful sound representations), lexical (abstract word representations), and semantic (meaning representation), but they differ on the extent of feedback connections within this hierarchy (see Figure 1.3). For example, the Cohort model (Marslen-Wilson and Welsh, 1978) limits the reach of top-down influences to the lexical level. The TRACE model (McClelland and Elman, 1986), a computational model using distributed representations, placed great importance to feedback connections from higher to lower levels of representation. However, under certain versions of this model, a first sound-feature analysis level is proposed that would not receive any top-down influences from the phonological level above. Currently, although these models have been refined and new models have been proposed, the question of whether predictive processes can percolate down to sub-lexical processing levels is still debated (Carreiras et al., 2014).

Brain networks for language processing

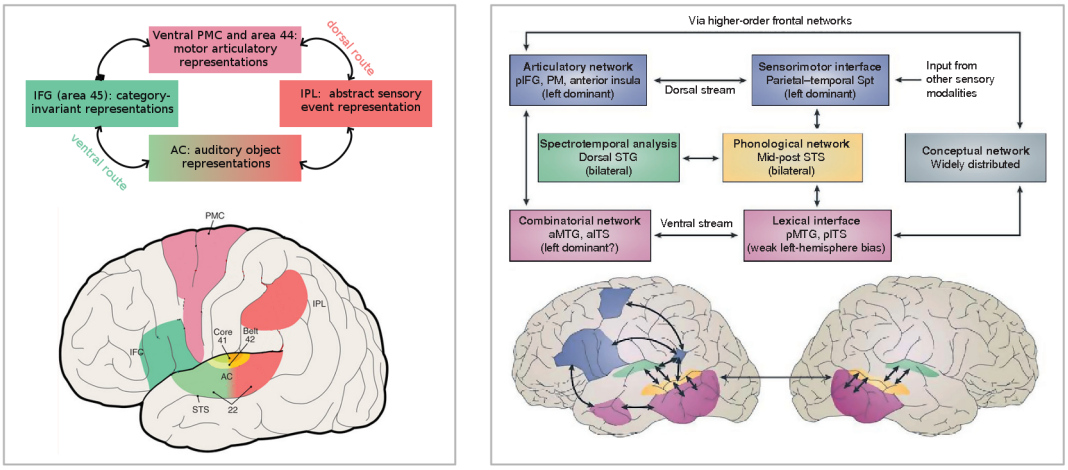

Figure 1.4: Examples of dual stream speech processing models. Left: model proposed by Rauschecker and Scott (2009); Right: model proposed by Hickok and Poeppel (2007). Images adapted from original papers. AC, auditory cortex; IFG, inferior frontal gyrus, PMC, pre-motor cortex; IPL, inferior parietal lobule; STG, superior temporal gyrus; STS, superior temporal sulcus; MTG, middle temporal gyrus; ITS, inferior temporal sulcus.

Psycholinguistic theory of hierarchical language processing has found grounding in neurobiological studies, that also describe language brain networks as a hierarchy progressing from sensory representations in superior temporal areas to more abstract representations and operations in frontal areas (Hickok, 2012; Friederici, 2011; Hagoort, 2016). Most current models propose a dual organization of these networks (Hickok and Poeppel, 2004; Scott and Johnsrude, 2003), whereby two streams

emanate from bilateral primary auditory cortex in the posterior temporal banks of the Sylvian fissure. A ventral stream would spread anteriorly along the superior temporal lobe, and a posterior stream would spread along superior temporal lobe and parietal areas, both streams converging in frontal structures. Gradual abstraction would take place along both streams, with a bilateral ventral stream connecting sound to meaning representations directly, and a mostly left-lateralized dorsal stream providing sensory-motor integration by the connection of sounds to their articulatory representations. In sum, these models assign semantic representations to brain-wide networks (sometimes proposing a temporal pole hub), lexico-semantics to middle-temporal-areas, and phonology to superior-temporal areas on the one hand, and to dorsal articulatory representations including parietal and frontal structures. Other frontal areas, namely the IFG, would be involved in abstract decision processes (see Figure 1.4 for examples of dual stream processing models).

In reading, these language processing networks are expanded to include visual processing areas, with a region in ventro-occipital cortex acting as a gateway between purely visual and linguistic processing. The connection between lexical and orthographic representation could be mediated by phonological representations, through a dorsal pathway linking occipital visual areas to temporal phonological processing sites, or directly through a ventral left-lateralized pathway linking visual cortex to frontal areas and to a brain-wide semantic network (Coltheart, 2006; Jobard et al., 2003).

Empirical work: Prediction in sentence processing

In the sentence processing literature it is well established that semantic and syntactic predictions can facilitate processing of upcoming words (Kuperberg and Jaeger, 2016). Words that are predictable under a given sentential context are read faster or skipped more often (Ehrlich and Rayner, 1981; Staub, 2015), and elicit a reduced neural response as measured by a the event related potential (ERP) technique about 400 ms after word presentation (Kutas and Hillyard, 1984). This effect, known as the N400 effect, is a negative deflection observed after averaging over trials scalp-recorded electrophysiological signals in response to a word. It has been extensively studied to the present date, and firmly linked to predictive processes (Kuperberg and Jaeger, 2016). However, given that lexical selection happens within 200 ms of word onset (e.g. Marslen-Wilson, 1987; Rayner, 1998) the timing of the N400 effect suggests it involves post-lexical semantic analysis, and cannot provide evidence regarding facilitation at lower representational levels (although see Pylkkänen and Marantz, 2003).

Although less numerous than the literature on the N400 component, effects of contextual constraint have indeed been reported on earlier ERP components, 100-200 ms after word onset (Sereno et al., 2003; Penolazzi et al., 2007; Dambacher et al., 2009; Kim and Lai, 2012; Molinaro et al., 2013). The modulation of these components, associated with sensory processing, shows that prior context can indeed lead to facilitated processing at sub-lexical stages. The fact that these early context effects are not nearly as ubiquitous as the N400 could signal that these modulations, although possible, are rare. On the other hand, their scarceness in the literature has also been attributed to methodological constraints (Penolazzi et al., 2007): variance on physical characteristics of the words, such as length, would heavily affect early ERP responses but not later ones, thus obscuring prediction effects only in the former.

Finally, a group of studies have cleverly used the N400 component to evidence the presence of phonological pre-activation. These studies have focused on the article preceding a highly expected noun, whose form (e.g. 'a' vs. 'an') depends on the phonological features of the upcoming noun (DeLong et al., 2005; Van Berkum et al., 2005). The N400 response to the article was found to be larger when it did not agree with the expected noun, showing that the phonological representation of the latter must had been pre-activated. Although the robustness of these effects is under some debate (Nieuwland et al., 2018), successful conceptual replications have been carried out using article-noun gender agreement properties (Wicha et al., 2004). These studies thus show that phonological pre-activation is possible, although the neural evidence they provide is the consequence of an unmet prediction, rather than the prediction generation itself.

The question of word-form pre-activation

In sum, neuroanatomical models of language processing have provided a biological basis for findings from the literature on sentence processing and word recognition, that acknowledge the presence of predictive effects in language. Furthermore empirical evidence has shown that prior knowledge may modulate sensory processing (e.g. phoneme restoration effects, contextual modulation of early ERP components). However, both interpretation of these effects, and their empirical relevance, are still questioned (e.g. Carreiras et al., 2014; Huettig and Mani, 2015). Firstly, sensory effects after word onset could be attributed to an attentional process whereby a prediction at the lexical level licenses a less exhaustive bottom-up processing. Alternatively, as was discussed in the first section of the current chapter, even if top-down connections may modulate sensory processing, an initial sweep

of bottom-up information may still be required. Lastly, even if possible, sensory pre-activation in language might be so rare that it is empirically irrelevant.

Answering these questions is hindered by the fact that most of the linguistic prediction literature focuses on post-target word effects, so that the neural mechanisms supporting the predictive process itself are largely unknown. A few studies in the language domain present exceptions to this general trend, having compared preparatory activity before a word rendered highly predictable by its context to a less predictable counterpart. In a picture-word matching paradigm, Dikker and Pylkkänen (2013) found increased magnetoencephalography (MEG) activity in the theta band before word onset in the left mid-temporal and visual cortex successively, when the previous picture represented a single object as compared to a picture of a group of objects. This finding was interpreted as evidence of lexical retrieval (given the role of the mid-temporal gyrus in lexical access: Friederici, 2011; Hickok and Poeppel, 2007) and visual feature pre-activation when the identity of the upcoming word could be predicted. Molinaro et al. (2013) compared EEG activity before target words embedded in multi-word fixed expressions, or in low-constraining sentences, finding increased functional connectivity in the theta band between frontal and occipital electrodes for the former, suggesting top-down modulation of sensory cortices prior to word onset when the target could unequivocally be anticipated. Fruchter et al. (2015) used adjective-noun pairs where the adjective provided the context for the following noun, in an MEG study. They focused their analysis in a middle temporal gyrus area, finding an increased response in evoked activity to highly predictive adjectives, and a subsequent decreased response to the highly predicted noun. They interpret this finding as a lexical pre-activation of the expected noun, but whether this led to an orthographic (visual) pre-activation was not examined.

Although the studies reviewed in the previous paragraph present suggestive evidence of sensory pre-activation, they also present certain limitations. Firstly, in some studies the paradigms included violations of expectations in 50% of the trials (e.g. Dikker and Pylkkänen, 2013), raising the attention vs prediction confound (Summerfield and Egner, 2016) that was discussed in the first section of the present chapter. Secondly, most of the studies used contrasts where high expectancy conditions were compared to low expectancy counterparts. The enhanced activity in sensory cortices found before a highly expected word as compared to a less expected one could be explained in terms of different allocation of processing resources when semantic information can guide subsequent visual analysis, without involving wordform preactivation per-se.

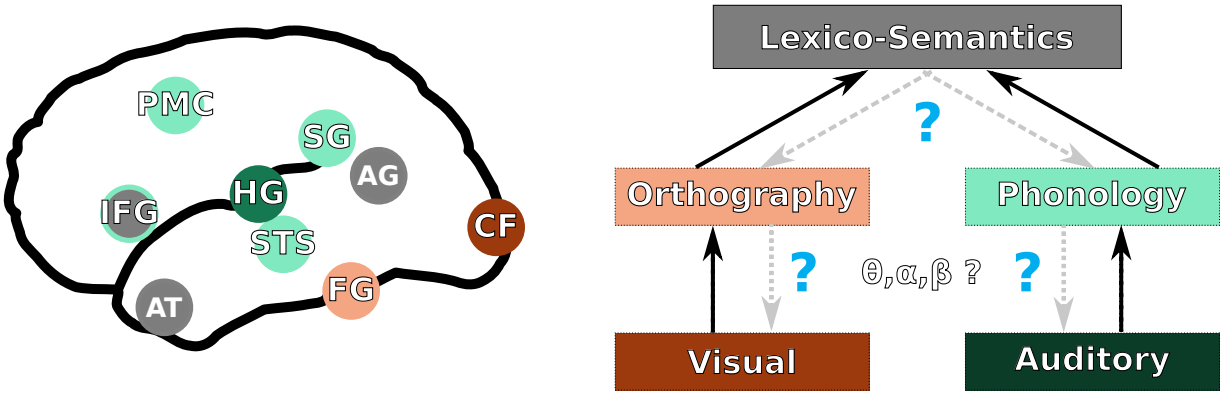

Figure 1.5: Questions addressed in the present thesis. Can prior knowledge lead to sensory pre-activation of wordforms? What are the oscillatory and anatomical correlates of these processes?

The main aim of the present thesis is to address the gaps identified both in the domain-general- and language-specific literatures, by focusing on the pre-target word interval by maintaining 100% validity of the cue, and by contrasting two high expectancy conditions that differ only on their sensory features (see Chapter 3 for more details on the experimental paradigm). Furthermore, by using MEG as the imaging technique, we aim to characterize the oscillatory and anatomical correlates of the prediction process. Specifically, we aim to answer the following questions (see Figure 1.5):

1. Can prior knowledge lead to sensory pre-activation of upcoming words?
2. If so, what are the oscillatory mechanisms and anatomical networks supporting the predictive process?

Chapter 2

Methods

2.1 Introduction to magnetoencephalography (MEG)

The MEG technique measures the magnetic fields generated by neural electrical activity, being part of the electro-physiological family of brain imaging techniques. Together with EEG and electrocorticography (ECoG), these methods have two main strengths: they provide a direct measurement of neural activity, and a temporal resolution in the order of milliseconds (Hämäläinen et al., 1993). In contrast, other techniques such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), have good or excellent spatial resolution, but provide indirect measures of neural activity (mediated by the hemodynamic response), with a temporal resolution that is well-below the brain's speed of processing.

Within the electrophysiological techniques, EEG and MEG use sensors placed outside the skull, whereas in ECoG, electrodes are placed directly on the surface of the brain. The former two methods have the great advantage of being non-invasive, but as a result of the distance to the generating sources, they can only detect activity of populations of tens of thousands of neurons acting in synchrony (Hämäläinen and Hari, 2002). Although the insights they can give into neural dynamics are at a coarser grain than those provided by ECoG, they are specially well-suited for the cognitive neuroscience field, providing an intermediate level of explanation crucial for bridging the gap between mind and brain.

EEG is a versatile and cheap technique, sensitive to currents flowing in any direction. MEG, on the other hand, can only measure sources that are tangential to the scalp due to a compete shadowing of primary currents by volume conduction

currents for radial sources. However, EEG has one important disadvantage: its limited spatial resolution.

The topographical distribution of electrical currents at the scalp can be used to infer their generating sources inside the head (in engineering terms, finding a solution to the inverse problem). However the skull is a poor conductor of electricity, smearing the electrical currents that can be measured at the scalp. This poses an important obstacle to solving the inverse problem with EEG, given the large number of possible source configurations that could give rise to a given pattern at the scalp. Although the multiplicity of solutions to the inverse problem is also present in MEG, it is of significantly smaller magnitude. The distortion of magnetic fields by the skull or other tissues in the scalp is minimal, making source reconstruction of MEG data possible with a spatial resolution of a few millimeters for focal sources (Hämäläinen et al., 1993).

2.2 Neural activity measured by MEG

The nervous system's main function is the transmission of signals for the coordination of bodily function and behaviour. Neurons, the cells making up the nervous system, are responsible for the generation and transmission of such signals, in the form of electrochemical waves (Ramón y Cajal, 1904). Neurons have long slender projections, named axons, that transmit these waves along the cell in the form of an action potential or spike. Action potentials are an all-or-none discrete electrical event generated by rapid falls and rises of the membrane potential. Once this wave reaches the cell body, chemical neurotransmitters are released into the synapse, binding to receptors in the so-called post-synaptic neuron. These transmitters generate slow, gradual fluctuations (depolarization) in the membrane of the receiving neuron, known as post-synaptic potentials. Hundreds of other neurons may provide such input to the post-synaptic neuron, contributing in a linear way to its depolarization. Once (if) the depolarization reaches a threshold, that neuron will in turn fire.

Electrical activity generated by a single neuron cannot be measured at the scalp. Given their transient nature ($\sim$1 ms), action potentials from different neurons will not coincide in time to generate a large-enough summed signal. In contrast, this may happen with the slower, gradual changes of post-synaptic potentials. However, temporal alignment alone is not enough: the spatial orientation of neurons must also be similar, in order for the currents not to cancel each other out. Luckily, a common neural type in the cortex, pyramidal neurons, have the required spatial configuration. This electrical activity in turn generates magnetic fields around it,

that when originating in pyramidal neuron populations tangential to the skull, can be measured at the scalp with MEG.

2.3 Instrumentation

The magnetic fields to be measured at the scalp are extremely small, orders of magnitude smaller than other environmental sources of electromagnetism, and a billion times smaller than the Earth's magnetic field. This poses two main challenges for MEG instrumentation: detecting the fields, and isolating them from environmental magnetic noise.

Most current MEG set-ups rely on sensors made of super-conductive material (SQUIDs: Super conducting interference devices). In order to maintain their super-conducting properties they must be kept at low temperatures, below 20 degrees Kelvin (-250˚C). To achieve this, sensors are contained in a vacuum flask (dewar) filled with liquid Helium. The magnetic field decays rapidly as a function of source-sensor distance. The dewar is shaped to partially surround the participant's head, and the sensors are placed close to the internal surface of the dewar.

Different strategies are used to minimize the impact of environmental noise on the recorded signal. In the first place, the MEG system must be contained in a magnetically-shielded room. Secondly, the configuration of the sensors themselves can filter out noise. Two main types of sensors are commonly used in current MEG systems: magnetometers and gradiometers. Magnetometers are made a of a simple loop. When a magnetic field crosses it a small current proportional to the magnetic field is generated and recorded. Gradiometers are made of two contiguous loops, both measuring magnetic fields, but only the difference between the two is recorded. This effectively reduces the impact of far away sources (which will be "seen" similarly by both loops), leaving only magnetic fields generated by sources closest to the coils. As a drawback, their ability to detect deep brain sources is also reduced.

The MEG used in the current thesis is a Vectorview (Elekta/Neuromag) system, equipped with 306 sensors arranged in triplets comprised of one magnetometer and two orthogonal planar gradiometers. Each gradiometer thus records one orthogonal component of sources located right beneath the sensor.

2.4 Analysis of MEG data

Pre-processing

After data collection, further steps for noise reduction are implemented. Firstly, The Elekta/Neuromag system provides its own software for this purpose, Maxfilter. This software implements a mathematical method for the separation of magnetic signals coming from a sphere inside the sensors from other sources (Signal Space Separation, SSS; Taulu et al., 2005). Furthermore, this method can be used to compensate for head movements. However, there may be sources of noise close to the sensors, for example, dental work in the participant. Using temporal information in addition to spatial patterns allows the identification and separation of such sources of noise. This method is referred to as tSSS, or temporal extension of SSS (Taulu and Simola, 2006).

One last source of noise must be rejected before data analysis. Participant blinks and heart-beats produce large magnetic perturbations that remain in the signal after cleaning with Maxfilter. Trials contaminated with artifacts may be rejected manually. Another alternative, used in the present thesis, is to use Independent Component Analysis (ICA), a computational method for separating a mixed signal into its additive sub-components. These components can be examined manually to identify those whose time-course and topography are consistent with ocular or heart activity. Alternatively, this can be done automatically, by correlating the ICA components with EOG and EKG activity that can be simultaneously recorded during the experimental session. The MEG signal can then be reconstructed after removing the components reflecting eye- and heart- related activity.

Sensor level

Evoked activity: time domain

A common analysis pipeline for electrophysiological data then proceeds by averaging over single trials for each participant and experimental condition, and expressing the data of interest relative to a pre-stimulus baseline. This averaging step works as “brain-noise” cancellation operation, that will filter out ongoing brain activity not directly related to the presented stimulus. This procedure results in a series of smooth positive and negative deflections in voltage or magnetic field. These deflections, or components, have been extensively explored in the field of EEG, and

their modulation by experimental conditions has led to important insights into the timing of perceptual and cognitive processes.

Originally, these components were interpreted as a sequence of discrete responses evoked by the stimulus after an otherwise flat baseline. However, this interpretation has proved problematic. No one-to-one correspondence between components and generating sources has been found, and the assumption that the components are present in single trials (but obscured by noise) has been challenged (Makeig et al., 2004). This makes linking the observed brain response to a specific cognitive process problematic.

Induced activity: time-frequency domain

An alternative view would explain the response to a stimulus as a modulation of ongoing brain dynamics. These dynamics would be implemented as neural oscillations at different frequencies, and their modulation could involve amplitude-, phase-, or frequency- changes. From these responses, only those at lower frequencies, and tightly time- and phase-locked to the stimulus would be apparent after ERP averaging (Tallon-Baudry, 1999). Other modulations of brain activity, often referred to as "induced", can be studied by transforming the pre-processed data into its frequency components. Such decomposition is possible using mathematical methods based on Fourier analysis. To obtain time-varying power spectra, Fourier analysis is performed over a sliding window, multiplying data by a Hanning taper in order to control spectral leakage and the amount of frequency smoothing. The length of the time window will determine the time and frequency resolution of the resulting representations. A longer time window will allow a higher frequency resolution at the expense of temporal resolution. In the present thesis, a 500 ms window window was employed, giving rise to a 2 Hz frequency resolution.

Analysis in the time- and time-frequency-domains can thus offer complementary insights into the underlying neural dynamics, highlighting the bottom-up response to a stimulus or the ongoing top-down processes respectively. In the present thesis, we will use averaging in the time domain to examine the early response to the stimulus, but in the time-frequency domain to study the predictive processes prior to it. From an empirical point of view, these processes might not be strictly time-locked to presentation of a stimulus, arising only after high-level processing of the cue. From a theoretical stance, the predictive mechanisms themselves could be implemented through neural oscillations.

However, at the sensor level both approaches have one important drawback: the mixing of the signal. Many different sources may be contributing to a given observation at the scalp, making interpretation of cognitive processes difficult. MEG in this respect offers an important advantage: source reconstruction will allow the un-mixing of the signal, and will provide a link to the underlying anatomy. Both aspects may be crucial in order to provide cognitive interpretations of the observed neural dynamics and to link results of individual MEG experiments to the literature from other imaging techniques.

Source reconstruction

Source reconstruction of MEG data requires finding a solution to the inverse problem: given a set of magnetic observations at the scalp, O, we want to obtain the brain sources I that generated them (see equations 2.1 and 2.2). We can model the brain sources themselves (populations of pyramidal neurons) as electric dipoles, i.e. currents with a position, orientation, and magnitude, but no spatial extent. These sources can be placed in a 3-D grid within the brain, obtaining one source per voxel. The next required step is to estimate the forward model G. We can model the observations at each sensor as a linear combination of contributions from all sources, since magnetic field varies linearly with current amplitude, and magnetic fields generated by several dipoles combine in an additive fashion. Therefore, given N_O number of sensors and N_I number of sources, the forward model, or leadfield L, will be an $N_O \times N_I$ matrix which can be right-multiplied by the $1 \times N_I$ vector of source dipoles to give the $N_O \times 1$ vector of observations (see equations 2.3). We can add an $N_O \times 1$ noise term, ϵ to account for any residual variance not accounted for by the forward model.

$$O = G(I) \tag{2.1}$$

$$I = G^{-1}(O) \tag{2.2}$$

$$O = LI + \epsilon \tag{2.3}$$

Forward model

The simplest way to estimate how a magnetic field at a given location would propagate across the head is to model it as a sphere. Within this sphere, different shells may be used to reflect the different conductivity or magnetostatic properties

of brain tissues. However, unlike electrical conductivity, the magnetostatic properties of the head are homogeneous across tissues, and only one shell is typically needed. Alternatively, head geometry can be modeled realistically, using a T1-weighted MRI image of the participant's head. There are different approaches to mathematically implement both sphere-based- realistic-geometry models, such as boundary element method (BEM) or finite difference method (FEM). BEM more suited for modeling magnetic propagation, and FEM for electric propagation. Finally, image co-registration in needed. MRI and MEG techniques localize the head of the participant using different coordinate systems. These must be aligned in a common reference frame. In the present thesis, forward models were constructed following the BEM approach, with realistic head models.

Inverse model

Even after developing an accurate forward model, inverting it is non-trivial. The number of sources far exceeds the number of observations (sensors), and there are infinite numbers of source configurations that would give rise to a given set of sensor observations. In engineering terms, the inverse problem is ill-posed. In order to select one solution, additional a-priori assumptions must be imposed. In the present thesis, two methods are used: minimum-norm methods (MNE) (Hämäläinen and Ilmoniemi, 1994) and linearly constrained minimum variance beamformer (Van Veen et al., 1997). Both of these methods linearly project sensor responses onto sources for which parametric source maps can be built. Their main advantage is that the subsequent source maps can be normalized to a common template, and averaged over multiple subjects. The MNE method assumes that sources are distributed, and that the brain functions in an energy-optimal way. The contribution from all sources are evaluated simultaneously, but the solution with the minimal energy configuration is selected. In practice, a regularization parameter adjusts the compromise between model fit, and energy minimization.

Beamforming methods take a different approach to finding a solution to the inverse problem. They circumvent it by estimating the contribution of each source to the scalp activity one at a time. The underlying assumption that licenses treating each source separately is that they are uncorrelated, ie. that no two brain areas display the same electrical activity. This is achieved by applying a spatial filter for each source, to pass brain activity from that location whilst attenuating activity from the rest of the brain. In the present thesis the MNE approach is employed for source localizing evoked activity and beamforming for oscillatory data. Although theoretically there are no reasons to prefer one model for each type of data, from

a practical perspective beamforming may work better for sustained brain responses (such as induced modulations of ongoing oscillations) and MNE for shorter-lived responses as evoked by a stimulus (Jensen and Hesse, 2010).

Statistical analysis

As we have seen in this chapter, the MEG technique provides extremely rich data to characterize brain function in time, space, and frequency. However, this poses important challenges for statistical appraisal. The number of dimensions across which experimental conditions can be compared leads to the multiple comparisons problem: the probability of incorrectly rejecting the null hypothesis increases with the number of tests. This can be addressed by implementing techniques to correct for multiple comparisons, or by limiting the statistical tests based on a-priori hypotheses. The former approach typically leads to loss of sensitivity (ability to detect true positives), whereas the latter may be difficult to implement given the individual and task-related variability in electro-physiological responses, and would miss any unexpected patterns in the data.

Furthermore, in MEG, the question arises of whether to perform statistics on sensor or source-localized data. Statistics on source-level data can be more sensitive, given the mixing present in the scalp signal. However, this will only be the case if the assumptions taken to solve the inverse problem are correct. In practice, the amount of data processing and potential sources of error involved in source localization make sensor-level statistical appraisal a more robust and preferable approach in most circumstances (Gross et al., 2013). In the present thesis, our planned analysis included (largely) unrestricted statistics on sensor level data, followed by source localization to explore the likely origin of the effects identified at the scalp.

Sensor-level inference

Statistical inference was carried out using cluster based permutation tests (Maris and Oostenveld, 2007). This technique controls for multiple comparisons whilst maintaining sensitivity by taking into account the temporal and spatial (also frequency if applicable) dependency of neighbouring samples. First, the data is clustered by performing pairwise comparisons (t-tests) between each sample (time-frequency-sensor or time-sensor point) in two conditions. Contiguous values exceeding a pre-defined threshold ($p = 0.05$ in our case) are grouped in clusters, and a cluster-based statistic is derived by adding the t-values within each cluster. Then, a

null distribution assuming full exchangeability (i.e. no difference between conditions) is approximated by drawing random permutations of the observed data (3000 in the present thesis) and calculating the cluster-level statistics for each randomization. Finally, the cluster-level statistics observed in the actual data are evaluated under this null distribution. Importantly, these tests only provide weak control of the false alarm rate (Maris and Oostenveld, 2007): they can provide evidence to reject the null hypothesis that two conditions are interchangeable, but not about where or when significant differences start or finish. However, the spatio-temporal (and frequency) characteristics of the clusters supporting the decision will be used to interpret the differences between conditions, both at the sensor and source levels.

Different restrictions on the windows for statistical analysis and source-localization windows were implemented in the present thesis:

Evoked by actual words. Statistical analysis was performed on a 500 ms second window centered around word onset, in order to include early responses to the word, and activity just prior to it. For source localization, the window was divided in windows of pre- and post-word activity of 250 ms each.

Induced by image-cue. Statistical analysis was performed over a time-frequency window ranging from image offset (250 ms) to the minimum trial length in Experiment 1 (1500 ms), and a low frequency range (1-30 Hz). Source localization windows were not set a-priori. They were based on the extent of the clusters identified in the sensor analysis.

Localization of sources of differences

In order to identify the anatomical regions displaying a different response to two conditions, it is common practice in the neuroimaging field to contrast parametric maps between conditions. However, this approach presents two problems. Firstly, such contrast maps mix differences due to amplitude and to location. Secondly, they may peak away from the sources that actually generated them, to an extent that scales with the smoothness of the maps (Bourguignon et al., 2018). This problem is especially worrying for the MEG technique, given the spatial leakage inherent to MEG source reconstruction.

Instead, we located regions of peak activity with respect to baseline, and restricted between-condition comparisons to those sites. Non-parametric permutation tests (Nichols and Holmes, 2002) were used to identify regions of significant change with respect to baseline, using both conditions pooled together. Within these regions, we

identified the coordinates of local maxima. We then calculated the mean difference between conditions at each maxima, in order to assess which sources were the main contributors to the effect observed at the scalp.

Part II

Experimental section

Chapter 3

Experimental design and common analysis methods

The main aim of the present work was to explore the presence of sensory predictions in language comprehension. To meet this aim, a collection of studies was performed using highly predictive contexts, where picture cues indicated the identity of the word that would be presented after a delay. The same basic paradigm was used in all experiments, varying modality of word presentation to explore effects in written and spoken language, and the delay between cues and targets to examine the influence of temporal predictability. This chapter presents an overview of said experimental paradigm, and of the main analysis approach. Chapters 4 and 5 then present each study with its specific variations, reporting the experimental results and conclusions: Chapter 4 presents data for the auditory domain, and Chapter 5 for the visual domain.

3.1 Experimental paradigm

Object pictures served as predictive cues to a word presented after a delay. The pictures were always congruent with the following word, generating a 100% conditional probability for the upcoming words. For example, a picture of a shoe would always precede the word "shoe". Crucially, the words were chosen to differ along one sensory dimension relevant to the modality employed: initial phoneme in the auditory experiments, and word length in the visual ones. This manipulation generated two experimental conditions, namely expect-fricative- and expect-plosive, where fricatives and plosives are two distinct consonant categories,

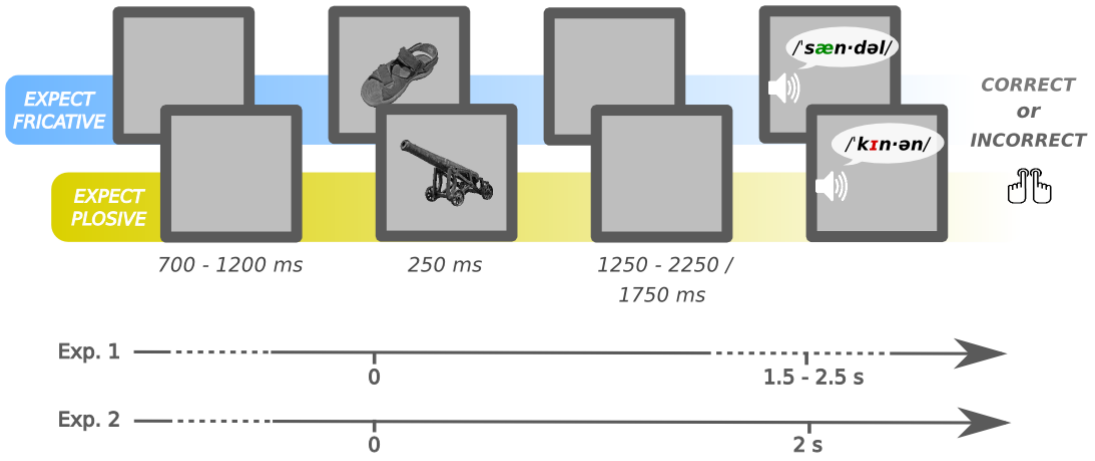

Figure 3.1: Experimental paradigm employed in auditory experiments. In the figure, words in English are included for clarity, but actual stimuli were in Spanish

or expect-long and expect-short words, where word length refers to the number of letters. The pictures that served as cues were carefully counterbalanced across conditions, so that any differences detected before word onset could only be attributed to sensory activation of the expected wordforms. In order to keep participants engaged in the task whilst maintaining predictive validity of the cue, the following words could contain a one letter replacement which participants had to detect (See Figure 3.1 for an example with spoken words). This paradigm was presented in two different experiments, one where the interval between the picture and the target word was fixed (Experiment 1), and another where it was variable (Experiment 2). Each experiment consisted of two experimental blocks, where the words were presented either visually (written words) or auditorily (spoken words, see Table 3.1).

Table 3.1: Summary of the Experiments reported in the present thesis

	No. of participants	Block
Experiment 1: variable timing	22	Auditory Visual
Experiment 2: fixed timing	26	Auditory Visual

Rationale

This experimental paradigm was designed to fill the gaps identified in the existing literature (see Chapter 1), and to maximize our ability to detect sensory prediction effects. To achieve these aims, the following design principles were followed:

Analysis of the pre-word interval. There is abundant data characterizing the *consequences* of prediction in language, using the response to target words. In the present studies we aim to characterize the predictive process before target onset, contributing to the understanding of *preparatory mechanisms* themselves.

Focus on wordform rather than semantic predictions, by comparing words that differ only along a sensory dimension. Although there is extensive literature regarding the presence of semantic, or meaning-based, predictions in language processing, there are still few studies focusing only on orthographic or phonological predictions.

Use of a "predictive vs. predictive" paradigm. Most studies so far have compared predictive to non-predictive conditions. Such differences in the conditional probability of the target word can generate concurrent differences in the attentional demands of the two conditions, leading to potential confounds (Molinaro and Monsalve, 2018).

Maximize predictive nature of the context. Pictures of objects served as 100% predictive cues to upcoming words. Our paradigm was designed to maximize the experimental ability to detect sensory predictions in the most favourable scenario, by using 100% valid cues. Furthermore, the presence of prediction violations could foster an additional top-down selective attention mechanism (Summerfield and Egner, 2009) which we wanted to avoid.

Explore the role of temporal predictability. Given the importance of temporal information for the oscillatory mechanisms supporting speech comprehension, temporal predictability might be required to observe phonological-feature pre-activation. On the other hand, it has been suggested that uncertainty regarding the bottom-up input could enhance reliance on top-down predictive mechanisms (Kuperberg and Jaeger, 2016) in language processing. In order to maximize the chances of detecting predictive effects, and disentangle between these two alternative hypothesis regarding the effects of timing on prediction, we performed two versions of each experiment: with fixed- or variable-delay between cues and targets.

Explore sensory predictions in the visual and auditory domains. Although we expect the same domain-general predictive processes to play a part in written and spoken word expectations, their possible instantiation as sensory pre-activations must necessarily differ, at least in the cortical regions involved. These differences might also include the corresponding oscillatory signatures, which might be characteristic of different cortical areas (Schoffelen et al., 2017), and/or reflect intrinsic differences in information coding in each modality (see for example, different coding preferences in spoken and signed languages: MacSweeney et al., 2008).

3.2 Experimental design

Data analysis was grouped according to presentation modality (see Table 3.2). This allowed us to assess the presence of sensory activation in each modality using the subjects of both Experiments pooled together, and to later evaluate whether temporal predictability modulated such pre-activation. The main contrasts performed are thus:

- Main effect of sensory feature: e.g. $Fricatives\ vs\ Plosives$
- Interaction between sensory feature and timing: e.g. $(Fricatives - Plosives)^{VariableTiming}\ vs\ (Fricatives - Plosives)^{FixedTiming}$

The results of the auditory data (auditory blocks of Experiment 1 and Experiment 2) will be presented in Chapter 4, and the results of the visual data (visual blocks of Experiment 1 and Experiment 2) will be presented in Chapter 5.

Table 3.2: Summary of the modality-specific analysis

Analysis	Manipulation	Conditions	Design
Auditory stimuli (Chapter 5)	Sensory feature	Fricative vs Plosive	Within subject
	Timing	Fixed vs Variable	Between subject
Visual stimuli (Chapter 6)	Sensory feature	Long vs Short	Within subject
	Timing	Fixed vs Variable	Between subject

As was discussed in Chapter 2, these contrasts were evaluated on the time-frequency (TFR) decomposed MEG data before word onset. In addition,

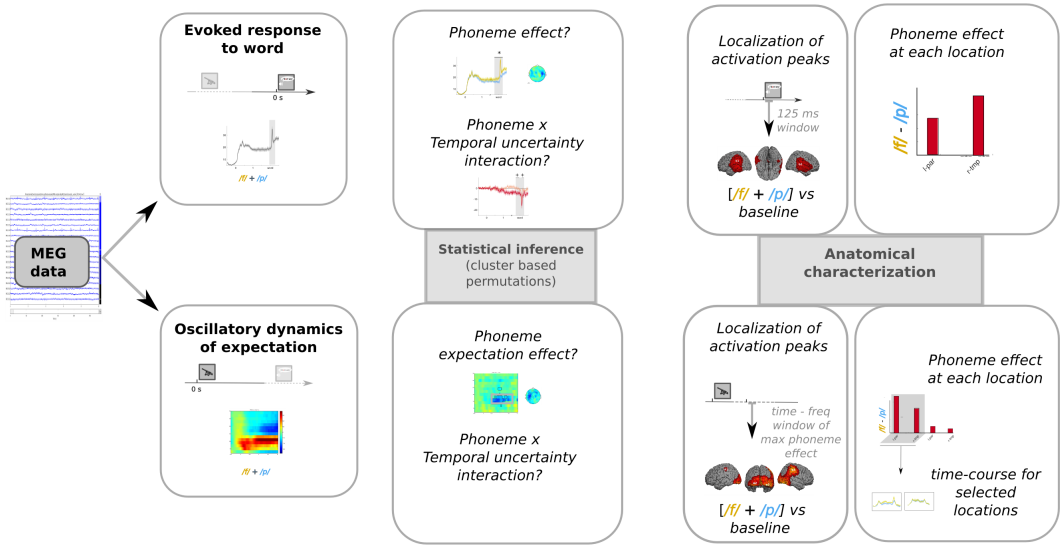

Figure 3.2: Analysis pipeline followed with the auditory data. Analysis of visual data proceeded in an analogous way, with word-length replacing initial-phoneme as the sensory contrast.

these contrasts were also applied to the event related fields (ERF) in response to the actual word, serving as a proof of concept that the sensory manipulation was indeed capable of eliciting a statistically different response over sensory areas. Statistical inference was carried out using cluster based permutations on sensor-level data, and any identified effects were subsequently anatomically characterized using source-localization techniques. Figure 3.2 shows a diagram of the main contrasts performed and the data analysis pipeline followed in these experiments, using the auditory studies as an example (for more information regarding the pipeline and choice of analysis techniques please refer back to Chapter 2). We also collected behavioural reaction time data, summarized in Appendix A.

3.3 Common Materials and Methods

Participants

Twenty-two participants took part in Experiment 1, and 26 in Experiment 2 (aged between 19 and 40, mean 24; and 20 to 39, mean 25, respectively). All participants provided informed consent in accord with the Declaration of Helsinki before starting the experiment, and received €10 in exchange for their collaboration. The present studies were approved by the BCBL Ethics Committee. Participants were

all right-handed native-Spanish speakers, with normal or corrected to normal vision and no history of neurological disease.

Stimuli and procedure

Each trial started with a blank screen presented for a variable interval from 700 to 1200 ms, followed by a picture depicting an object for 250 ms. After a variable (1250 to 2250 s: Experiment 1) or fixed (1750 ms: Experiment 2) interval, the word corresponding to the object was presented. The word was always congruent with the picture, but in 50% of the trials it contained a one letter or phoneme replacement. The resulting incorrect words were always phonotactically correct in Spanish. Various incorrect versions were generated for each actual word (see Chapters 4 and 5 for more details on the word stimuli). Participants were instructed to respond whether the word was correct or not, and were made aware that the non-correct words were identical to real ones but with one phoneme/letter replaced. They were encouraged to pay attention to the preceding images, explaining that these would always give valid cues as to the upcoming words. Participants had a maximum of 500 ms to give a response with a left/right index button press, and visual feedback was provided after incorrect trials (red cross in the center of the screen for 100 ms). Yes/no response side was counterbalanced across participants.

An additional non-predictive condition, not analyzed in the present thesis, was included in the experiments. Non-predictive trials were created using uninformative images as cues. These were created by trimming previous object images to the smallest possible bounding box and scrambling them using an in-house Matlab script employing the ImageProcessing library. We thus obtained a set of 20 images where no discernible object could be identified.

In total, 360 response trials were generated, including 120 in each of the two predictive conditions and a further 120 in the non-predictive. In addition to the response trials, 40 “no-go” catch trials were added to make sure that participants were attending to the image-cue, and not just waiting for the word. These were identical to the experimental trials, but the image presented was the original color-version instead of grayscale, and participants were instructed not to respond to the word in these cases. Trial order was pseudo-randomized, avoiding more than three repetitions in a row of the same image cue.

Ten pictures depicting words in Spanish were selected from the Bank of Standardized Stimuli (BOSS Brodeur et al., 2014) for each experimental condition. Different items were chosen for the visual and auditory blocks, making up a total

of 40 pictures. The color pictures obtained from the database were transformed to grayscale, trimmed, resized to a target diagonal of 500 pixels, and placed on a 550 by 550 pixel square with a medium-gray background using the imageMagick software package. Although this manipulation yielded object images of similar size, it made the smaller real-world objects harder to recognise, rendering them larger than their real-world size (for example, cherry). To ameliorate this problem we ranked real-world size of the objects in a 1-3 scale, and reduced diagonal size by 30% for rank 1 objects.

The pictures were balanced across conditions with respect to several ratings from the BOSS database (name- , object- , and viewpoint-agreement, familiarity, subjective complexity and manipulability), as well as several lower level image properties (contour complexity, number of pixels, and brightness). These properties were evaluated with Matlab 2012b and imageMagick software packages. The words were balanced across conditions within each block on frequency (obtained from the esPal database: Duchon et al., 2013), and semantic category (natural vs artifact).

The pictures created were presented within a black-background on a back-projection screen situated 150 cm away from the participant. Overall, the experimental session lasted approximately two hours, including participant preparation, MEG-recording, and debriefing. This session included two different experimental blocks, where the words following the pictures were presented either written on the screen or auditorily through ear-tubes. Each block lasted approximately 30 minutes, block order was counterbalanced across participants. Participant-controlled pauses were provided every 10 trials, in addition to two experimenter-controlled ones per block. An additional 10 minutes of resting state were recorded between blocks.

MEG data acquisition

Brain activity was recorded in a magnetically shielded room using a whole head MEG system (Vectorview, Elekta/Neuromag) with 306 sensors arranged in triplets comprised of one magnetometer and two orthogonal planar gradiometers. Participants were screened for magnetic interference prior to data collection, and instructed to limit head and face movements as much as possible, and to fixate in the center of the screen. Data was acquired with a 1000 Hz sampling rate and filtered during recording with a high-pass cutoff at 0.1 Hz and a low-pass cutoff at 330 Hz via the Elekta acquisition software. Head movements were monitored continuously using five head position indicator coils attached to the participant's head. Their location

relative to fiducials (nasion and left and right pre-auricular points) was recorded at the beginning of the session using an Isotrak 3-D digitizer (Fastrak Polhemus, USA). In addition, head shape was digitized to allow for alignment to each subject's structural MRI for subsequent source localization. Eye movements and heartbeats were monitored using vertical and horizontal bipolar electro-oculograms (EOG) and electrocardiogram (ECG).

MEG data preprocessing

MEG data were initially preprocessed using Elekta's MaxFilter 2.2 software, including head movement compensation, down-sampling to 250 Hz, and noise reduction using signal space separation method (Taulu et al., 2005) and its temporal extension (tSSS) for removing nearby artifacts (Taulu and Simola, 2006). Manually-tagged bad channels were substituted by interpolated values.

Subsequent data analysis was carried out in Matlab 2012b, using the FieldTrip toolbox (Oostenveld et al., 2010). The recordings were segmented from -1000 ms to 4000 ms (Experiment 1) or 3500 ms (Experiment 2) relative to the presentation of the picture, and low-pass filtered at 100 Hz. Since in Experiment 1 the delay between the cue and the target was variable, trials were then trimmed to exclude the response to the actual word. This allowed the generation of an image-locked average containing only preparatory activity. In addition, trials for this Experiment were re-segmented time-locked to word presentation, in order to allow examination of the event-related response to the actual word in subsequent analyses.

Eye movement, blink and electrocardiographic artifacts were reduced using independent component analysis (Jung et al., 2000), with subsequent visual inspection to remove any epochs with remaining artifacts. In the auditory analysis, fifteen percent of the trials were rejected in Experiment 1, and 9% in Experiment 2. There were significant differences in trial rejection between experiments ($F(1,39) = 11.0$, $p = 0.002$) but not between phonemes or their interaction with Experiment($F(1,39) = 1.2$, $p = 0.3$; $F(1,39) = 0.9$, $p = 0.3$). In the visual analysis, there were no significant differences in trial rejection between Experiments or its interaction with word length ($F(1,42) = 2.1$, $p = 0.2$; $F(1,42) = 0.1$, $p = 0.7$ respectively), but there were significant differences between the trial rejection in the short and long conditions (9% and 10% rejections for short and long conditions respectively; $F(1,42) = 4.4$, $p = 0.04$). This difference is probably the result of an increase in eye movement artifacts in the long word condition compared to the short word condition. However, given the reduced size of this effect (1%), we do not foresee any important

problems for subsequent analysis. Further sensor-data analysis was performed using gradiometer data only, but both magnetometer and gradiometer data were used for source-localization.

Chapter 4

Auditory experiments

4.1 Introduction

Existing evidence for pre-target word activation in language is not only scarce, but comes mostly from studies with visually presented words. Although top-down influences on sensory areas have been well documented in the auditory domain, previous studies have either used non-linguistic stimuli and omission paradigms (e.g. SanMiguel et al., 2013; Yokosawa et al., 2013) or have focused on response to the actual word (e.g. Sohoglu et al., 2012; Arnal et al., 2011). To the best of our knowledge, no study has yet explored the presence of sensory pre-activation of words in the auditory domain by looking at the pre-target word interval. In the following paragraphs we will review the publications with more relevance to our paradigm.

In a series of studies, Sohoglu and colleagues investigated the effects of prior knowledge and perceptual detail on auditory processing of single words. Participants were presented with acoustically degraded words after a written word that acted as a cue, and were asked to rate clarity of the spoken item. The acoustic degradation was parametrically manipulated to explore the effect of perceptual detail on perceived clarity, and the written cue, which could match or mismatch the auditory word, was used to study the effects of prior knowledge. Both valid prior information and increased acoustic detail correlated positively with perceived clarity (Sohoglu et al., 2014), but had opposite effects on the neural response at the superior temporal gyrus (STG) as measured with EEG and MEG (Sohoglu et al., 2012). Whilst perceptual detail enhanced activity at the STG, a phonological processing area, prior knowledge suppressed it. Furthermore, prior knowledge also modulated neural activity in the inferior frontal gyrus (IFG), taken to represent the most abstract node within the

speech processing network. Importantly, neural modulation at the IFG was observed before the differential activity detected at the STG. This pattern of results is coherent with predictive coding models, whereby prior expectations generated at an abstract level percolate down to concrete levels, thus reducing the neural activity needed to process the incoming stimulus once it arrives. Differential activity at the IFG as a function of cue validity was observed at 90-130 ms after word onset, suggesting very early engagement of top-down modulations. However, a lack of even earlier effects could indicate that prior knowledge only modulates auditory processing after an initial bottom-up sweep is processed. In contrast, in a further study using the same paradigm, Cope and colleagues (2017) found evidence for pre-target word preparatory activity in the beta band. They analyzed the oscillatory activity prior to target presentation by contrasting it with a pre-cue baseline. They report a beta power increase just before target word onset that correlated with precision of the expectations as modelled using Bayesian inference simulations with each subject's behavioural results. However the neural sources underpinning these effects were not analyzed, so that whether these effects are linked to sensory pre-activation, or to more abstract representations cannot be determined.

Crucially, all these paradigms employed matching and mismatching cues to the upcoming stimulus. As a result, the conditional probability of a given stimulus was 50% or lower, decreasing the sensibility of these studies to detect and describe sensory pre-activation. Furthermore, the presence of mismatching stimuli could trigger feature-based attentional mechanisms (Summerfield and Egner, 2016) which might be qualitatively different to purely predictive ones. An interesting exception comes from a study by Roll and colleagues (2017), who contrasted activity after word onset for word-beginnings with few possible completions to those with many possible candidates. They used event-related potentials (ERP) analysis, showing an increased negativity for the former as compared to the latter, which they interpret as a pre-activation effect. Furthermore, they used fMRI to characterize this effect anatomically, finding that the ERP pattern was associated with a blood-oxygen-level-dependent (BOLD) contrast increase in Broca's area (left IFG, pars opercularis) and left angular gyrus. This pattern was interpreted as a top-down influence aimed at inhibiting irrelevant left parietal activation during lexical selection. These results would thus reflect a pre-activation at the lexical level, but no evidence is reported for a pre-activation at the perceptual level.

The present study

With the present study we aimed to overcome limitations of the reviewed literature by focusing on the pre-target word interval and by maintaining 100% predictive validity of the cue. Instead of using a valid vs invalid cue comparison to isolate predictive mechanisms, we compared preparatory activity before highly expected words that differed only in their perceptual features. This perceptual contrast was achieved in the present experiment by manipulating the word's initial phoneme: words could either start with fricative or a plosive. These two types of phonemes differ in their manner of articulation: whilst in the former airflow is partially occluded, generating turbulent airflow and a gradual onset, the latter are produced by total interruption of airflow followed by the vowel, generating a sudden and short burst of energy. These differences in articulation generate specific sound profiles (see Figure 4.1) that would give rise to contrasting responses in auditory cortex and beyond (Toscano et al., 2010; Correia et al., 2015).

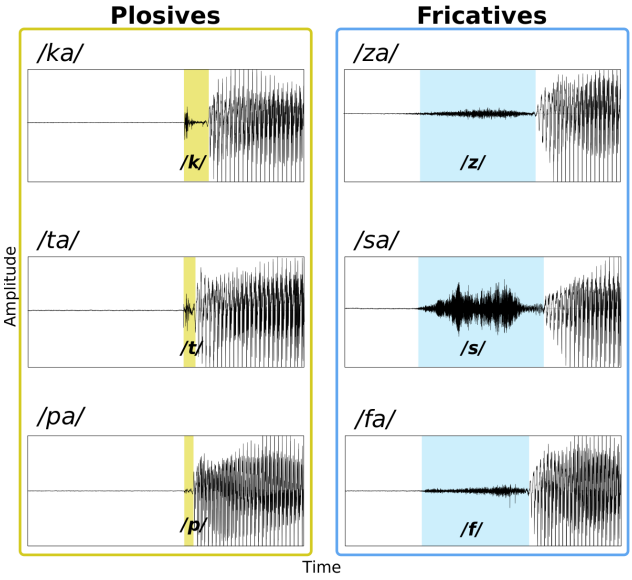

Figure 4.1: Sound waves generated by the different consonant categories. The examples shown were extracted from the actual experimental stimuli.

Anatomical and oscillatory hypotheses

If prior expectations involve pre-activation of word-form representations, we should be able to find differences between our two conditions before stimulus onset, in cortical areas involved in auditory processing. Whilst we expect the strongest effects around the STG, responsible for phonological processing (Hickok and Poeppel, 2007; Scott and Johnsrude, 2003), the dorsal articulatory pathway could also be involved in the passive representation of articulatory features (Correia et al., 2015).

The main neural response we examine in the present thesis is oscillatory activity. The beta band has been repeatedly implicated in predictive mechanisms in the

visual domain (Arnal and Giraud, 2012), but with auditory stimuli a more complex pattern has been observed. Whilst some studies report a prominent role of beta oscillations in auditory predictions (Cope et al., 2017; Sedley et al., 2016) others also show the role of lower frequencies (Arnal et al., 2011). Arnal and colleagues (2011) observed that whilst a slow oscillation regime (in delta and theta bands) developed over a high-order network when speech matched prior visual information, a faster local regime of beta and gamma arose over the STG when the speech violated prior expectations. The authors suggest that beta oscillations were involved in the reconciliation between prior expectations and prediction error and in the generation of new predictions, whereas delta and theta oscillations are involved in the stabilization of representations of the confirmed expectations. As a result, we expect differences between conditions to arise in beta and/or theta ranges.

4.2 Material and Methods

Participants

Twenty-one right-handed native Spanish speakers took part in the first experiment and 26 in the second (aged between 19 and 40, mean 24; and 20 to 39, mean 25 respectively). From the first experiment one participant was excluded due to excessive noise in the recordings, and from the second, one was excluded due to lack of compliance with experimental instructions, two due to excessive noise and artifacts, and two due to technical problems with the audio system detected after the experiment. This left 20 participants in the first experiment, and 21 in the second.

Stimuli

Table 4.1 shows the full stimuli used in the auditory blocks, classified according to the two initial-phoneme categories. The initial consonant was always followed by a vowel, and the words were 5-6 letters long. They were balanced across conditions on length in syllables, in addition to frequency and semantic category (natural vs artifact). Three different incorrect words were generated for each actual word when possible, but in some cases only one pseudo-word could be generated. The words were presented through plastic tubes and silicon earpieces to participants' ears. For more details regarding experimental stimuli, procedure and experimental design, please refer back to Chapter 3.

Table 4.1: Items in each phoneme condition. English translations within parenthesis.

Plosive-initial	Fricative-initial
toldo (awning)	sillón (armchair)
pelota (ball)	silla (chair)
camión (truck)	cereza (cherry)
cañón (cannon)	secador (hairdryer)
conejo (rabbit)	farola (streetlamp)
cometa (kite)	zapato (shoe)
camisa (shirt)	zorro (fox)
cohete (rocket)	salero (saltshaker)
tomate (tomato)	sierra (saw)
toalla (towel)	zueco (wooden shoe)

4.3 Results

Response to stimulus: word-locked event-related fields

Before examining our main contrast, involving the pre-target word oscillatory activity, we analyzed the ERF response to the actual words. We assessed the contrast between fricative- and plosive-conditions using cluster-based permutation tests over a 500 ms interval centered around word onset, and with the 41 subjects of both experiments pooled together. This contrast will be later referred to as the phoneme contrast, or the phoneme effect when deemed significant. We then evaluated whether phoneme effects were modulated by temporal predictability (i.e., whether there is an interaction between temporal predictability and sensory feature), by comparing the phoneme contrast between the two experiments. Finally, we performed source reconstruction with minimum norm estimate to identify the brain regions underpinning the effects found in the sensor-level analysis (see Chapter 2 for more details regarding the choice of analysis methodology).

Word presentation elicited a marked amplitude increase over a bilateral mid region and a left anterior area, that was more pronounced for plosive- than fricative-initial words. Cluster-based permutations comparing the brain response to fricatives and plosives (all 41 subjects) confirmed that these differences were significant (see Figure 4.2A), the largest cluster having a p-value = 0.003 (second largest: p = 0.2). Interestingly, this phoneme expectation effect seemed to start before word onset (the largest cluster spanned the whole analysis time window, and visual inspection of ERF plots suggest it was present before).

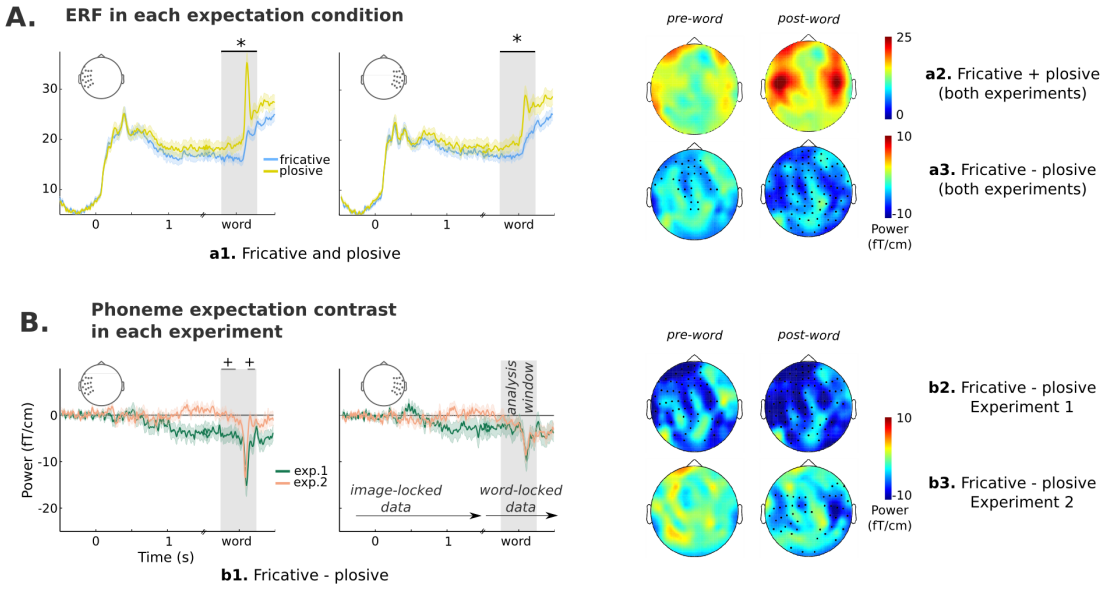

Figure 4.2: Peri-word ERF response. A — Phoneme effect with subjects of both experiments pooled together. On the left, ERFs for plosives and fricatives averaged over left and right temporal sensors. Shaded area indicates the analysis window, and black horizontal line indicates the temporal extent of the clusters reported in the results section ('+' indicates a p-value < 0.1, and '*' < 0.05). On the right plots a2 and a3 show the topography of the evoked response during the the pre- and post-word segments of the analyzed window. B — Interaction between phoneme and temporal uncertainty. On the left, plot b1 shows the ERFs of the difference between phoneme conditions in each experiment over left- and right-temporal sensors. On the right, plots show the topography of the phoneme expectation effect separately for each Experiment. Sensors that formed part of the clusters with $p < 0.1$ are marked in black. The shaded areas around the ERF plots indicate the standard error (for visualization purposes). The ERFs of Experiment 1 were constructed using image-locked data from the beginning of the interval up to 1600 ms, and word-locked data from 500 ms prior to word onset. There is 100 ms of overlap, marked as a discontinuity in the x-axis.

The test for an interaction between phoneme expectation and temporal predictability revealed a marginally significant effect (largest cluster -250 to -160 ms, p = 0.07; second largest cluster p = 0.09) that started before word onset and was largest over anterior and mid-left sensors (see Figure 4.2B). This marginal interaction seemed to be driven by a pre-word phoneme expectation effect that was apparent only when word onset was variable (i.e., in Experiment 1). Furthermore, this effect appears present even before our analysis time window (see Figure 4.2B), carrying over till after word presentation.

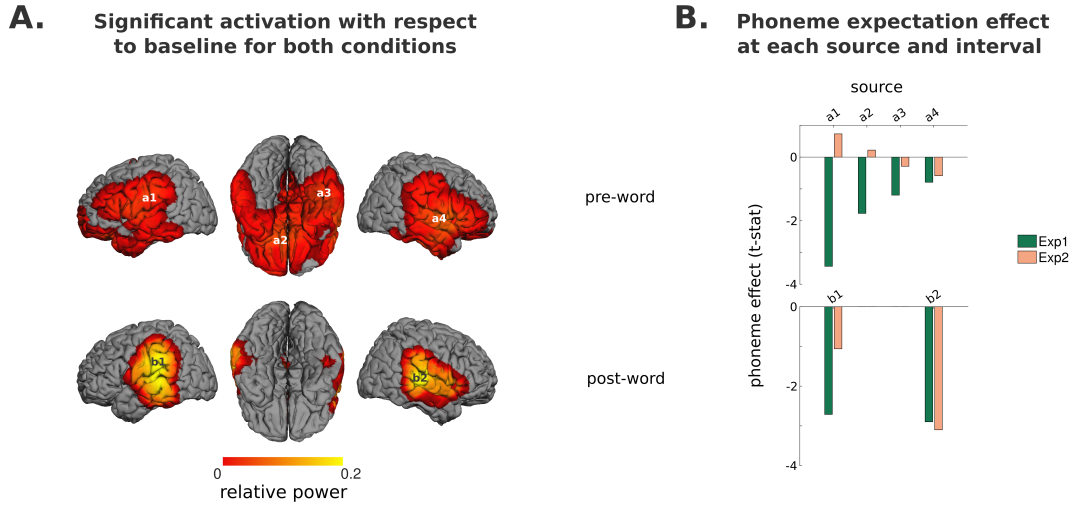

Figure 4.3: Source reconstruction of ERF effects. A — Regions of significant power change with respect to baseline (both conditions and experiments) at 125 ms windows just before and after word onset. The peak activity locations within these areas are marked ($a1$ to $b2$) within these images. B — Effect-size of the phoneme-expectation effect (estimated as the standardized difference between conditions) at each peak and time interval.

We performed source reconstruction in two separate 250 ms time-windows in order to separate brain activity before and after word onset. Peaks of significant activity with respect to baseline in these intervals were identified using permutation statistics and both phoneme conditions pooled together. In the pre-word interval we found extensive regions of significant amplitude change spanning bilateral perisylvian areas and frontal and temporal ventral regions (see Figure 4.3A top). Within these, a large number of maxima were found (15), so only the four largest were selected for further analysis: left parietal on the banks of the Sylvian fissure ([-50 -19 27]), left fronto-ventral ([-10 29 -19]), right temporo-ventral ([43 -31 -33]), and right superior temporal ([50 0 1]). We then examined the phoneme expectation effect at these sources separating subjects according to experiment, given that moderate evidence for an interaction with temporal predictability was found in the sensor level data (see Figure 4.3, panel B top). This revealed that the $fricative < plosive$ pattern over left-lateralized sources was present in Experiment 1 (being maximal over the left

parietal and fronto-ventral sources), but minimal or absent in Experiment 2. This is consistent with the topographies observed at the sensor level, where differences between ERFs to both phoneme conditions in Experiment 1 peaked over mid-left and left anterior areas (see Figure 4.2B).

In the post-word interval areas of significant amplitude change with respect to baseline spanned a more constrained bilateral peri-sylvian region surrounding auditory cortices. Two cortical maxima were found within these regions, in a left-parietal area in close proximity to the one identified previously, and a right superior temporal area, more posterior than the one identified in the pre-word interval (see Table 4.2 for MNI coordinates of these locations and their spatial relation to the ones identified in the ERF analysis). Both of the identified sources showed a $fricative < plosive$ pattern in both experiments. The effect appears to be right-lateralized in Experiment 2 but not Experiment 1 (see Figure 4.3B bottom), however this pattern could simply reflect the carry-over of the effect generated in the pre-target word interval in Experiment 1.

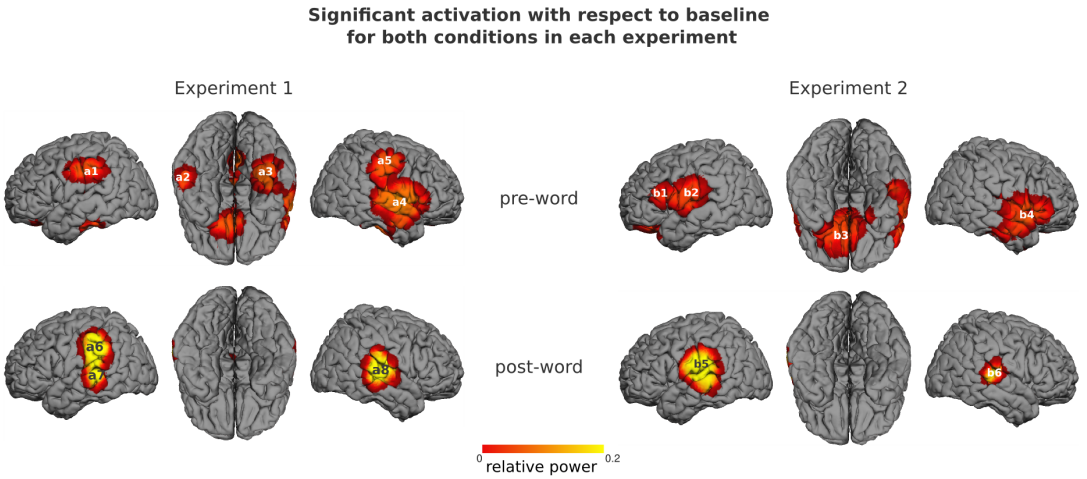

Figure 4.4: Source reconstruction of ERF effects for each experiment separately. Regions of significant power change with respect to baseline (both conditions) at 250 ms windows before and after word onset.

In sum, the ERF analysis showed that the phonological contrast was indeed capable of eliciting significant differences upon word presentation, and that these effects localized to auditory processing areas: superior temporal gyrus in the right hemisphere, and inferior parietal in the left. Note that the left source did not peak in or near primary auditory areas as was expected. This could be the result of the carry-over contribution from pre-existing prediction-related activity in Experiment 1. To confirm this interpretation, we repeated the source localization pipeline separately for both experiments. The qualitative pattern of activation with respect to

baseline was similar across experiments, except for a bifocal activation in Experiment 1 over the left hemisphere, with two clearly separable peaks over inferior parietal and superior temporal cortex, present both before and after word onset (see figure 4.4). This suggests that the contrast did indeed elicit a response over left auditory cortex, but the analysis using the data of both experiments pooled together conflated the pattern, giving rise to a unique peak between the activations observed in each Experiment separately.

Preparation for stimulus: image-locked time-frequency response

Time-frequency responses (TFR) were estimated in a pre-target word interval (from image offset to the minimum interval length: 250 ms to 1500 ms) and at frequencies in the range 3-30 Hz, including theta, alpha and beta bands. As in the ERF analysis, we assessed the phoneme expectation effect at the sensor level using all subjects pooled together, and then evaluated the presence of an interaction with temporal uncertainty over the same time-frequency window. Finally we performed source reconstruction using a minimum variance beamformer to identify the brain regions underpinning the effects detected in sensor data. Based on the literature and on results obtained with ERFs, we expected these to include STG and parietal areas.

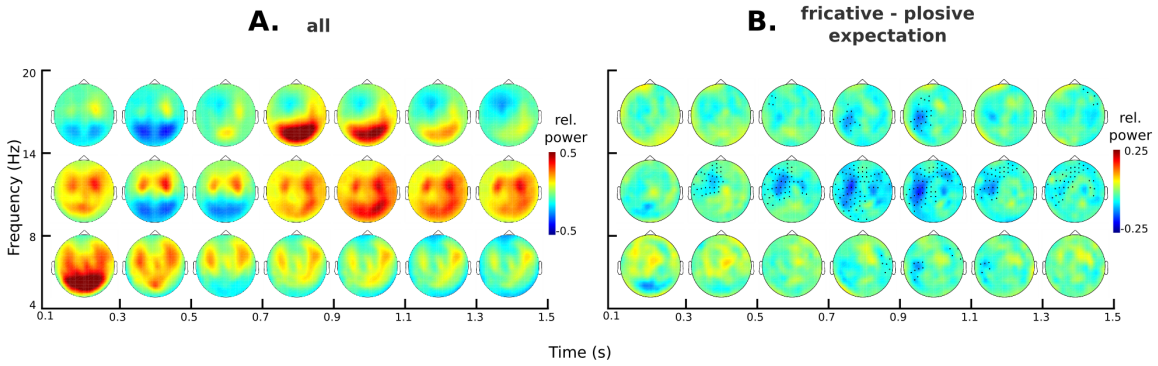

Figure 4.5: Image-locked time-frequency response. A — Power change with respect to baseline during the analyzed pre-word interval. B — Phoneme effect (fricatives minus plosives) during the same interval. Sensors that formed part of the cluster with $p < 0.05$ are marked in black.

Figure 4.5A shows the evolution of relative power with respect to baseline as function of time at theta, alpha and beta frequencies. Image offset induced a posterior power decrease at alpha and beta frequencies that turned to power increase at 700 ms. It also induced a bilateral anterior power increase at theta and alpha frequencies that was sustained throughout the analysis interval at alpha frequencies, but gradually diminished at theta frequencies. There was also a power decrease at beta frequencies over left anterior sensors that intensified as a function of time.

Figure 4.5B shows the evolution of power differences between fricative and plosive conditions. Cluster based permutations revealed that these conditions differed significantly (largest cluster, p = 0.02; second largest cluster, p = 1). Differences peaked 1 s after image onset and spanned mid-left and right anterior sensors, with fricatives displaying a smaller power increase than plosives. The largest identified cluster started 340 ms after image onset, was sustained throughout the analysis interval, and encompassed frequencies from 6 to 20 Hz with a mean of 11 Hz. Cluster based permutations comparing the phoneme expectation effect between experiments revealed no evidence for an interaction with temporal uncertainty (largest cluster, p = 1).

Based on sensor-level results, we evaluated source power at 900—1100 ms, at theta (4—8 Hz), alpha (8—14 Hz), and low-beta (14—20 Hz) frequencies using the data pooled across conditions and experiments. Significance was assessed in comparison with baseline power, with permutation statistics. Figure 4.6 A presents the three ensuing power maps, and Table 4.2 presents the coordinates of local maxima. The map at alpha frequencies had four clear peaks in inferior parietal and inferior occipital cortex, both bilaterally. In the beta band, there was a frontal peak of power suppression and bilateral occipital peaks of power increase. In the theta band we found bilateral parietal sources of power increase and a broadly distributed pattern of activity over a perisylvian area in the right hemisphere. Amongst these, we selected the two largest for further analysis (see Fig, 4.6 panel A), located in temporal and frontal areas (superior temporal sulcus and rolandic operculum respectively).

Although a statistical evaluation of the main effect of temporal uncertainty was outside the scope of this thesis, visual inspection of the sensor level data suggested there may be differences in the topographical distribution of power between experiments (see Figure 4.7). Therefore, as a final check we repeated the analysis pipeline described above for each Experiment separately. Power spatial distribution in both Experiments was very similar (see Figure 4.8) except for a theta source of power increase in left superior temporal cortex that was present in Experiment 1, but clearly absent in Experiment 2 (with a slight power decrease in this case). Given the theoretical relevance of such location, we decided to incorporate the coordinates of this peak (based on Experiment 1 subjects) to subsequent analysis (with subjects from both experiments together).

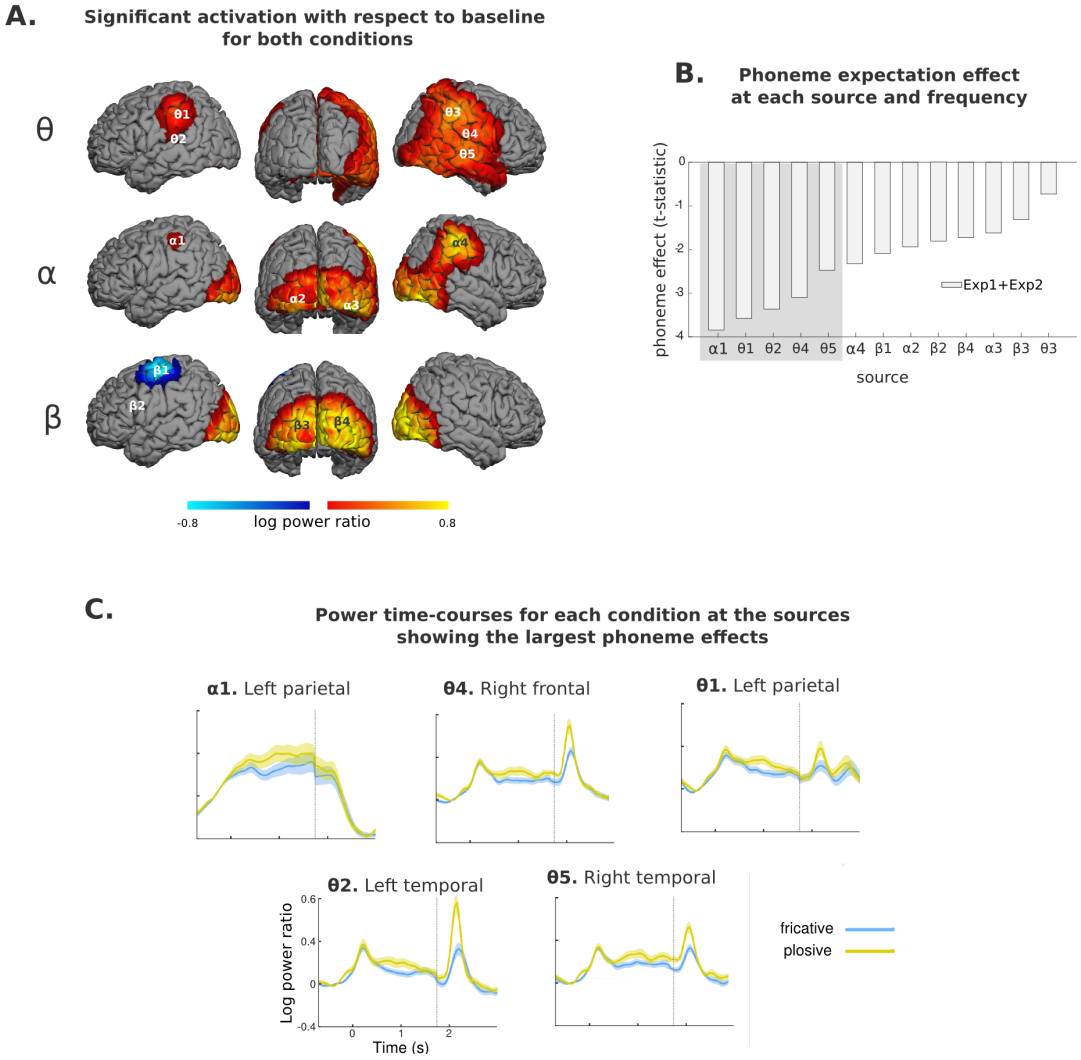

Figure 4.6: Source reconstruction of TFR effects. A — Regions of significant power change with respect to baseline for theta, alpha and beta bands, at 900—1100 ms. The peak activity locations are labelled within each map ($\theta 1$ to $\beta 3$). B — Effect-size of the phoneme-expectation effect estimated as the standardized difference between conditions in power averaged across 500—1500 ms. C — Time-courses along the whole trial for each phonological expectation condition, for the five sources showing the largest phoneme expectation effect. For visualization purposes, the standard error of the mean for each time-point is depicted as a shaded area around the main line. Vertical dashed lines within these plots indicate a discontinuity in the data shown: to the left the data reflects the mean calculated using both experiments, to the right only the data for Experiment 2 remains, given that word onset latencies in Experiment 1 were variable.

Table 4.2: MNI coordinates of significant maxima for post-word ERF and pre-word TFR sources maps. The last column provides the distance in mm of each TFR location to the peak ERF to the actual word in the same hemisphere. The TFR sources closest to the post-word ERF ones are highlighted in bold.

	Left hemisphere				Right hemisphere			
Analysis	*Location*	*Fig/label*	*MNI coords* [x y z]	*Distance to ERF*	*Location*	*Fig/label*	*MNI coords* [x y z]	*Distance to ERF*
Post word ERF	Parietal	3. b1	[-49 -27 26]	-	Temporal	3. b2	(62 -24 9]	-
Pre word								
Theta	Parietal	5. θ1	[-43 -33 40]	16.4	Parietal	5. θ3	[55 -30 45]	14.9
	Temporal	**5. θ2**	**[-55 -31 15]**	**13.2**	**Frontal**	**5. θ4**	**[59 -17 16]**	**10.3**
					Temporal	5. θ5	[58 -21 -5]	37.2
Alpha	Parietal	5. α1	[-47 -30 50]	24.3	Occipital	5. α3	[31 -77 -18]	67.1
	Occipital	5. α2	[-14 -82 -6]	72.6	Parietal	5. α4	[52 -30 50]	42.6
Beta	Frontal	5. β1	[-34 -13 66]	45.0	Occipital	5. β3	[27 -91 8]	75.6
	Occipital	5. β2	[-14 -91 1]	77.1				

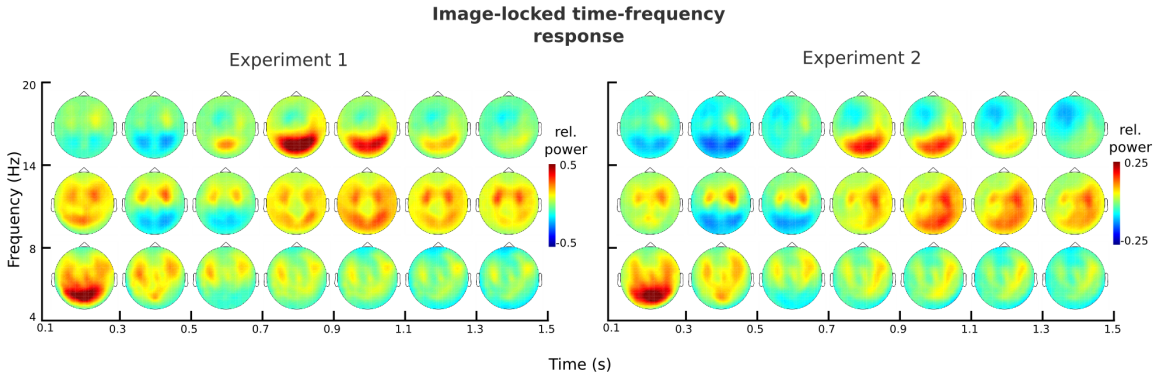

Figure 4.7: Image-locked TFR response for each experiment separately. Relative power change with respect to baseline during the pre-word interval for Experiment 1 (left) and Experiment 2 (right).

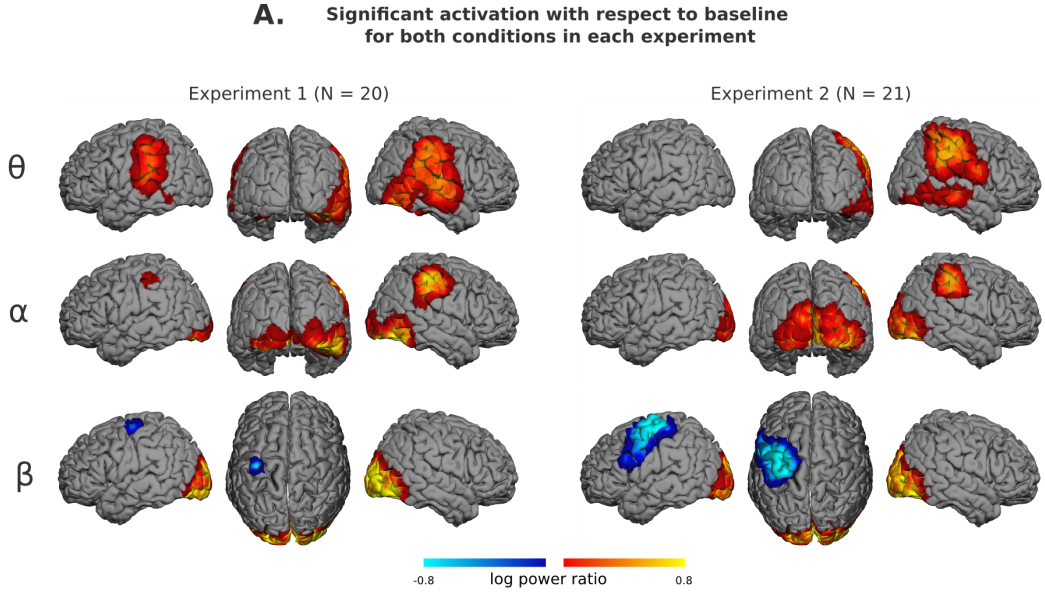

Figure 4.8: Source reconstruction of TFR effects for each experiment separately. Regions of significant power change with respect to baseline (both conditions) for theta, alpha and beta bands, at 900—1100 ms.

Considering all frequency bands together, twelve sources were identified. In order to compare the contribution of each source to the phoneme expectation effect, we extracted the time-courses at each location, and compared the power difference between conditions averaged over a time interval corresponding to the main cluster of phoneme expectation effect identified in the sensor analysis (0.34 to 1.5 s). Figure 4.6 B shows the standardized difference for each source, sorted according to z-score. Power was higher in the fricative than in the plosive condition in all examined sources (except for one). The largest contributors to the phoneme expectation effect were the left-parietal alpha and right-frontal theta sources. This is largely consistent with the data at the sensor level, where differences between phonemes concentrated over

mid-left areas in the alpha band, but also included right anterior sensors in the alpha and theta bands. Interestingly, the three subsequently-ranked peaks (all in the theta band) were located close to the sources of the ERF to the actual word: the left-parietal and left-temporal sources were 15 mm away from the left-parietal post-word ERF source, and the right temporal sources was less than 15 mm away from the superior temporal post-word ERF source (see Table 4.2). We thus selected the first five sources for further analysis.

Figure 4.6 C presents the power time-course for each phoneme expectation condition at these five location-frequency pairs for the whole trial duration. Congruently with the sensor-level data, differences between conditions were maximal around 1 s post-image onset, but appeared to be sustained until word onset in the right theta sources only. Upon word presentation, all sources except the left parietal alpha one displayed a marked power increase and an enhancement of the phoneme effect. This provides further evidence of their involvement in sensory processing of the word.

4.4 Discussion

In the present study we aimed to find evidence for predictive pre-activation of auditory words and to explore the oscillatory mechanisms generating them. We compared pre-stimulus activity between two high expectation conditions that differed only in the phonological features of the expected precept (fricative- vs plosive-initial words). Our results showed significant differences between expect-fricative and expect-plosive conditions before phoneme onset in a frequency range spanning alpha and theta bands, suggesting that sensory feature-specific preparatory activity was taking place. Sources of this effect were (amongst others) localized to superior temporal cortices, generally associated with phonological processing (Scott and Johnsrude, 2003). Furthermore, these locations were spatially very close to those showing the largest event-related response to the actual word, involving similar sources to those that show the largest modulation during processing of the actual stimulus.

Although the localization of the effects of interest included superior temporal cortices as we had hypothesized, multiple distant sources were found to contribute to the phoneme-expectation effect, painting a complex picture. The largest contributor was found to be a left inferior parietal source in the alpha band, followed by several perisylvian sources in the theta range over left and right hemispheres. However, we suggest that only the theta band sources reflect pre-activation of expected phonemes:

If the effects identified reflect a pre-activated representation we would expect an enhancement of the difference between phoneme conditions, and of overall activity with respect to baseline, once the actual phonological information was available at word onset. Whilst this was true of the theta sources, it was not the case in the alpha left parietal location (compare time-course *b1.* with *a1.* to *a4.* in Figure 4.6). The possible functional relevance of these sources will be discussed in the next paragraphs.

Anatomical findings

Although multiple possible contributors to the sensor-level phoneme effect were found, the largest of these were distributed following locations along the dorsal speech processing stream, as described by dual route speech processing models (e.g. Scott and Johnsrude, 2003). Along this stream, gradual abstraction from acoustic properties would take place posteriorly from primary auditory cortex along the superior temporal lobe to parietal areas. These sources are not limited to the temporal areas we had hypothesized, but are thus nevertheless consistent with the pre-activation of the expected phonological features at different levels of representation. Several features of this pattern of activation are worth discussing. Firstly, within the theta band sources, the right frontal motor source displayed the largest phoneme expectation effect. Its location, in the Rolandic operculum, is related to tongue motor control and speech articulation (Brown et al., 2008). Although controversial, the role of motor areas in passive speech perception is a recurrent topic in speech processing research (Liberman and Mattingly, 1985; Correia et al., 2015). Interestingly, the involvement of the motor system in perception could be mediated by the prediction process itself, that under some accounts would use motor simulations to generate full representations of their sensory effects (Schubotz, 2007; Molinaro et al., 2016), or at least their temporal dynamics (Morillon et al., 2015).

Secondly, it is interesting to note that only the sources distributed over the right hemisphere displayed a sustained effect that lasted until word onset. If differences between conditions indeed reflect a predictive mechanism aimed at facilitating perception, we would expect them to persist until word onset. Cope et al. (2017) for example, suggest that preparatory mechanisms may occur just before expected stimulation. An intriguing possibility would thus be that the right hemisphere would play a more prominent role in the representation of sensory predictions. This prominent role of the right hemisphere could be related to findings from studies looking at cortical entrainment to speech and non-speech auditory streams, showing

stronger theta coherence over right- than left- superior temporal cortices (Molinaro and Lizarazu, 2018).

Oscillatory findings

The involvement of the theta band in the pre-activation process is in agreement with the study by Arnal et al. (2011), that found an oscillatory regime in lower frequencies prevailed in the absence of expectation violations, and with studies looking at the pre-target word interval in the visual domain (Dikker and Pylkkänen, 2013; Molinaro et al., 2013). While in these previous studies theta band activity could also be linked to lexical retrieval, given the predictive vs non- (or less-) predictive contrasts, in our case it can be more directly linked to a pre-activation at the phonological level. We did not observe the involvement of beta-band frequencies in the phoneme expectation effect, which may be surprising given its association with predictive processing across modalities (e.g. Sedley et al., 2016). However, our pattern of results may be reconciled with such literature if beta oscillations are associated with the generation of predictions at an abstract level, but not with the activation of predicted representations at lower levels in the hierarchy. Indeed, our source-level analysis of activation peaks with respect to baseline identified beta band sources at the left IFG, that did not contribute largely to the phoneme expectation effect. This pattern is thus consistent with predictive activity at an abstract node in the language processing hierarchy that would be equivalent for both expectation conditions, since they only differed in sensory characteristics that would be realized at more concrete levels of representation.

One remaining question is the role of the alpha source, located at a left inferior parietal area. Alpha oscillations have been extensively associated to attentional selection (Jensen and Mazaheri, 2010; Wöstmann et al., 2017), both in the auditory and in the visual domains. Alpha increases have been linked to the suppression of task-irrelevant brain areas, acting as spatial or temporal filters. In addition, the left inferior parietal area has been linked to attentional selection in time (Coull and Nobre, 1998; Coull et al., 2016). Although timing of target onset should not differ between our phoneme conditions, fricatives and plosives do display distinct temporal profiles, gradually increasing in the former and sharp burst in the latter. These distinct temporal profiles could modulate attentional processes in left inferior parietal cortex. It is important to remember, on the other hand, that an inferior left parietal source was also identified in the theta band, in close proximity to the alpha band peak. The theta band temporal dynamics at this site, however, were very similiar to the other theta band sources, with an enhancement of overall activity and of differences

between phonemes once the target word was presented. In this way, activity in the left inferior parietal cortex could be functionally segregated in distinct frequency bands: a time-based attentional filter, implemented through alpha band activity, and a linguistic pre-activation of the expected target implemented in the theta band.

Influence of temporal uncertainty

In the present study we also assessed to what extent temporal uncertainty regarding word onset modulated phoneme pre-activation mechanisms. Although we found no evidence for the influence of word timing on the difference between phoneme expectation conditions, the evoked response during the peri-word interval suggests that such an interaction was indeed present. Therefore, even if pre-activation occurs in both cases, the nature of the representations, or the processes generating them may be modulated by the predictability of its temporal onset. Source localization revealed that the phoneme expectation effect in the ERFs found only in Experiment 1 (variable delay) originated in left inferior parietal cortex, near the alpha source that was found to be responsible for differences between phoneme expectation conditions in both experiments. Given our previous interpretation of an attentional effect at this site, we suggest that additional feature-based attention processes (Summerfield and Egner, 2016) would be present in Experiment 1, on top of predictive pre-activation that would be present in both experiments. Interestingly, our data suggest a hemispheric dominance for each of these top-down mechanisms. Whilst the theta activity associated with the prediction process localized mainly to the right hemisphere, differences in the evoked activity prior to word onset exclusive to Experiment 1 localized to left parietal areas. Hemispheric asymmetries in auditory processing have been well documented (Zatorre et al., 2002; Boemio et al., 2005), with a left-hemisphere dominance for timing information (or fast integration windows), and right-hemisphere dominance for frequency (or slower integration windows). This is coherent with our interpretation, where predictive pre-activation of the what (frequency profile) would be right-lateralized, but attentional filtering based on timing would lateralize to the left.

Other considerations

It is important to acknowledge two main limitations in the present study. Firstly, all our effects could be attributed to differences in predicted timing, rather than phoneme identity. Indeed, subjects' task was to identify spelling error in the vowel

following the first phoneme. The crucial second phoneme arrived systematically later, and with more variable onset, in the fricative than in the plosive condition. Hence, in addition to the obvious difference in phoneme identity, our conditions also differed in the temporal predictability of task-relevant vowel. Secondly, the source contributing the most to the significant phoneme expectation effect observed at the sensors was the alpha parietal location, under our interpretation responsible for implementing an attentional filter rather than a predictive pre-activation. The theta effects observed at the source level alone might have been insufficient to generate a statistically significant effect at the sensors. The subtlety of these effects could be in part responsible for the scarceness of pre-target word evidence in the literature, and should be taken into account in future studies.

In sum, our results show that the sensory features of expected phonemes modulate anticipatory activity over several different anatomical locations and frequency bands. This is in line with the view that predictive mechanisms involve brain-wide networks. Crucially, temporal auditory areas were an important part of this network, suggesting that phonological representations were activated before word onset. However, in order to confirm this interpretation, further studies are necessary to deal with the aforementioned limitations.

Chapter 5

Visual experiments

5.1 Introduction

The question of how far can predictions modulate early stimulus processing has a strong reflection in the visual word recognition literature. Models that propose purely feedforward processing have co-existed with fully interactive ones, where representations at every hierarchical level could exert modulatory influence on lower ones. Although there is currently extensive evidence confirming that higher order information can influence orthographic processing, whether these effects can be explained by a fully or partially limited interactive model is still an open question (Carreiras et al., 2014). This debate has also been tackled from the sentence processing literature, where the possibility of orthographic pre-activation is still a contended topic (as discussed in the general introduction to the present thesis), and anatomically, where the debate has intensified around the role of the so-called visual wordform area, a portion of the left fusiform gyrus (ventral occipito-temporal cortex, vOTC) that responds selectively to words and pseudowords. Whilst some accounts propose that the VWFA stores and computes abstract prelexical visual representations using primarily feedforward information (Dehaene and Cohen, 2011), other accounts propose that higher order information, including semantics, can modulate orthographic processing at this visual area (Price and Devlin, 2011).

Both in evolutionary and in developmental terms speech precedes reading. Reading networks must build on previously-existing speech processing networks, connecting these to visual processing areas. Models of visual word recognition propose that whilst during reading acquisition the connection between the lexical and the orthographic level of processing would be mediated by phonological

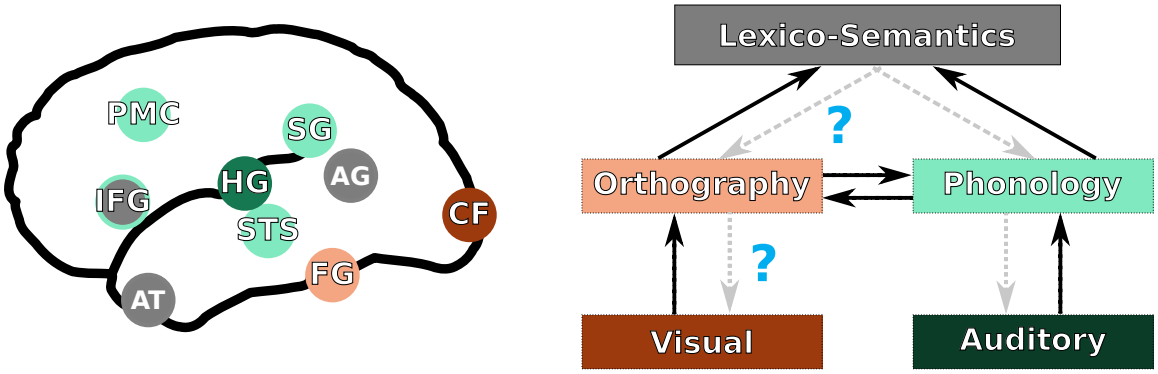

Figure 5.1: Brain networks for language and reading.

representations, expert readers would also be able to map directly lexical representations into visual codes (Coltheart, 2006; see Figure 5.1). These models have been confirmed by neuroimaging studies (Jobard et al., 2003), that describe a dorsal pathway, linking visual areas to superior temporal phonological areas and to frontal areas through parietal cortex, and a left lateralized ventral pathway that would link visual areas directly to frontal and a brain-wide semantic network through the vOTC. These dorsal and ventral pathways previously described for reading in fact piggyback upon well described routes of visual object recognition, with the ventral route being typically associated with the "WHAT" and the dorsal pathway with the "WHERE" (or "HOW"). As was the case for auditory cortex, a gradual abstraction of visual features takes place along these pathways, with the left fusiform gyrus acting as the gateway from vision to language.

The present study

In the present study we aimed to determine if top-down influences can modulate processing of words at a sensory level, by focusing on the pre-target word interval. As was discussed in the general introduction to the present thesis, studies in language have generally analyzed the interval after word onset, thus observing the consequences rather than the prediction process itself. A few studies present exceptions to this general trend, and have focused on the pre-target interval. Dikker and Pylkkänen (2013) used a picture-word matching paradigm, finding evidence of visual feature pre-activation. However, they used a paradigm including mismatching items, so that attention, rather than prediction could account for their findings. Molinaro et al. (2013) compared EEG activity before target words embedded in multi-word fixed expressions, or in low-constraining sentences, finding evidence for frontal to occipital directed interactions. However, given the predictive vs less-predictive paradigm, observed effects could again be interpreted as attentional

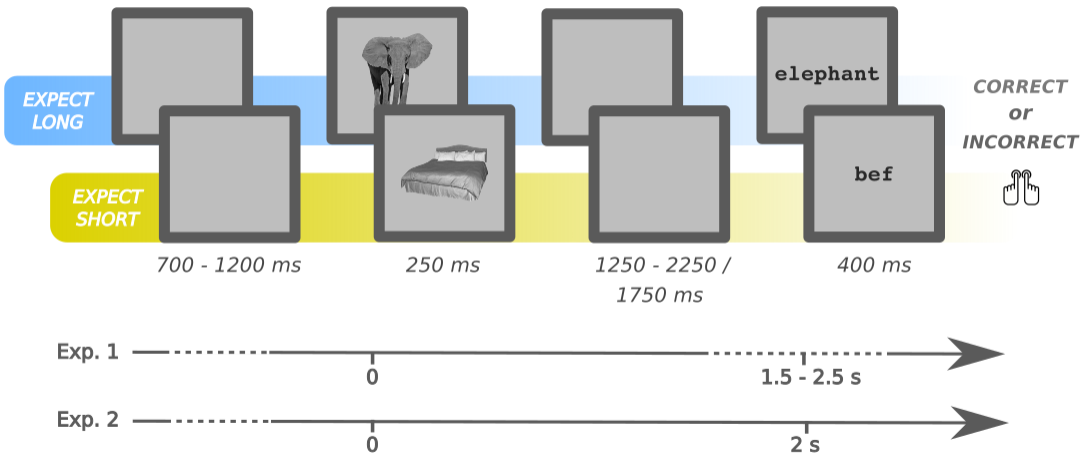

Figure 5.2: Experimental paradigm employed in visual experiments. In the figure, words in English are included for clarity, but actual stimuli were in Spanish

modulations. Finally, Fruchter et al. (2015) used adjective-noun pairs where the adjective provided the context for the following noun, in an MEG study. They provide evidence for lexical pre-activation, but did not analyze the presence of sensory-specific preparatory activity. In the present study we aim to contribute to the body of work examining pre-word activity, overcoming some of the aforementioned limitations. As was the case in the Auditory Experiments presented in the previous chapter, we compared preparatory activity before highly expected words that differed only in their perceptual features, maintaining 100% cue validity (see Figure 5.2). The perceptual contrast was achieved in the present experiment by manipulating the word's number of letters: words in the short expectation condition were 3-4 letters long, and in the long expectation conditions were 7-8 letters long. This yielded a difference of X°in the subtended visual angle, which given the retinotopic organization of primary visual cortex is expected to elicit an early differential neural response (Joukal, 2017). Any differences between the length expectation conditions over occipital areas prior to word-presentation could thus be attributed to stimulus-specific pre-activation.

Anatomical and oscillatory hypotheses

The visual-orthographic contrast employed modulates processing in early (≈100 ms) post-stimulus activity, and may thus have effects in primary visual cortex, also known as the striate area. This early visual processing area, located around the calcarine fissure in the occipital lobe, contains a retinotopic map of the visual field,

and approximately 50% of that area represents only the central 5°of the visual field (Joukal, 2017). Studies focusing on word length contrasts have localized the peak of these effects at secondary visual processing areas in the fusiform or lingual gyri. For example, Schurz et al. (2010) found that a length contrast for words led to a relatively small occipital cluster with activation maxima in the lingual gyri rather than the VWFA. Interestingly, Lerma-Usabiaga et al. (2018) conducted a study to try and re-concile differing accounts regarding the exact location of the VWFA within ventral occipito-temporal cortex. They employed a word vs pseudoword contrast to isolate areas responding to words as abstract linguistic units, and word vs checkerboard contrast to additionally include word visual features. They found a more anterior area was isolated with the former contrast, and a more posterior one with the latter. We therefore expect differences between our word length expectation contrast to peak at such a posterior portion of the left fusiform or lingual gyri.

As was discussed in the general introductory chapters, beta band oscillations have been typically linked to predictive processing in the visual domain. For example, Bastos et al. (2015) report top-down directed influences within primate's visual system are carried by low-beta-band synchronization. As a result, we expected to find evidence of prediction within this range. The few studies assessing pre-target word oscillatory mechanisms with linguistic stimuli report effects in the beta band (Molinaro and Monsalve, 2018), but also in theta ranges (Dikker and Pylkkänen, 2013; Molinaro et al., 2013), so as was the case for the previous auditory study, we will analyze oscillatory patterns in low frequencies including theta to beta ranges.

5.2 Material and Methods

Participants

Twenty-two right-handed native Spanish speakers took part in the first experiment and 26 in the second (aged between 19 and 40, mean 24; and 20 to 39, mean 25 respectively). From the first experiment two participants were excluded due to excessive noise in the recordings, and from the second, one was excluded due to lack of compliance with experimental instructions. This left 19 participants in the first experiment, and 25 in the second.

Stimuli

The experimental manipulation in the visual block consisted of length of the word in letters which could be either long (7-8 letters) or short (3-4 letters; see Table 5.1 for full stimuli list). The incorrect words were created by substituting consonants or vowels by a same-category member, and formed phonotactically legal words in Spanish. Up to four different incorrect versions were generated for each word when possible.

Table 5.1: Items in each length condition. English translations provided within parenthesis.

Long words	Short words
rodilla (knee)	oso (bear)
medalla (medal)	pez (fish)
hormiga (ant)	ojo (eye)
ardilla (squirrel)	ajo (garlic)
manzana (apple)	pan (bread)
telefono (telephone)	bota (boot)
calabaza (pumpkin)	foca (seal)
elefante (elephant)	pera (pear)
guitarra (guitar)	cama (bed)
aguacate (avocado)	pila (battery)

5.3 Results

As with the previously described auditory experiments, we first analyzed the evoked response to our two main experimental conditions, as a validation that the perceptual contrast employed did elicit a differential response in sensory cortices. Then we analyzed the oscillatory response during the pre-target interval, to assess the presence of length expectation effects.

Response to stimulus: word-locked ERFs

We performed statistical inference at the sensor-level using cluster-based permutations over a 500 ms interval centered around word onset (from now on referred to as the peri-word interval). We assessed the contrast between long- and short-word conditions, using the 44 subjects of both experiments pooled together. This contrast will be referred to from now on as the length contrast, or the length effect. We then evaluated whether length effects are modulated by the degree of temporal predictability, by comparing the difference between long and short

words between the two experiments. Finally, we performed source reconstruction with minimum norm estimate to identify the brain regions underpinning the effects identified in sensor data (see Chapter 2 for more details on the methods and choice of analyses).

Word presentation elicited a strong power increase over posterior sensors around 0.07 s after word onset (see Figure 5.3). This increase was larger for long- than short-words. Cluster-based permutations comparing both conditions revealed only moderate evidence against the null hypothesis (largest cluster: p=0.08; second largest cluster: p = 0.1). The largest cluster showed more power for long than short words over posterior and mid-central sensors, from 0.09 to 0.12 s after word onset. Interestingly, long words elicited more power than short words over posterior sensors well before word-onset, with differences between the two conditions being sustained until word presentation (see Figure 5.3 a1). However, our statistics provided no evidence that this may be a true effect, being captured by three small clusters over posterior sensors with p-values $>$ 0.3 (spanning the following intervals: -0.18 to -0.19; -0.02; -0.248 to -0.244). We found no evidence for an interaction between length and temporal uncertainty, with cluster-based permutations comparing the length effect between Experiments 1 and 2 in the peri-word interval identifying no clusters with p <0.05 (largest cluster: p = 0.23).

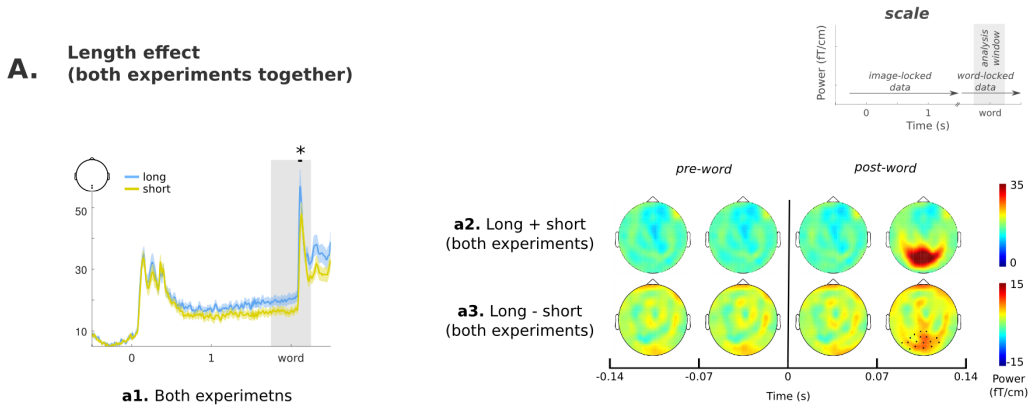

Figure 5.3: Peri-word ERF response. A — Length effect with subjects of both experiments pooled together. On the left, ERFs for long and short words averaged over sensors showing the largest effects. Shaded area indicates the analysis window, and black horizontal line indicates the temporal extent of the clusters reported in the results section ('*' $<$ 0.05). The shaded areas around the ERF lines indicate the standard error (for visualization purposes). On the right plots a2 and a3 show the topography of the evoked response during the the pre- and post-word segments of the analyzed window.

We performed source reconstruction over a 0.07 s window corresponding to the largest effect identified in the sensor analysis (0.07 to 0.14 s). In addition, we also

Table 5.2: MNI coordinates of the peri-word ERF activity peaks, during the three temporal windows examined. The label column refers to location markers used in Figure 5.4

Window 1: -0.07 – 0 s		Window 2: 0 – 1.07 s		Window 3: 0.07 – 1.14 s	
MNI coordinates [x y z]	Label	MNI coordinates [x y z]	Label	MNI coordinates [x y z]	Label
[-50 -18 26]	a1	[-53 -17 27]	b1	[1 -94 15]	c1
[-5 -62 3]	a2	[-6 -63 -6]	b2		
[56 -2 -24]	a3	[32 -17 -38]	b3		
[48 4 -5]	a4	[51 7 -8]	b4		
[28 32 31]	a5	[28 31 30]	b5		

source localized activity in two windows of equal length just before and just after word presentation (-0.07 to 0 s and 0 to 0.07 s), in order to perform exploratory analyses on these intervals. To achieve this, we first localized regions of significant change with respect to baseline using permutation statistics with all conditions pooled together. Then, we compared activity between long and short conditions at the peak sources of activity.

The first two windows (-0.07 to 0 and 0 to 0.07) revealed very similar distribution of activity, with extensive regions of significant activation with respect to baseline including bilateral perisylvian, bilateral occipital, and right frontal and ventral areas (see Figure 5.4A). Of five different maxima identified in each window (see Table 5.2), two were located within occipital cortex: left lingual gyrus and right fusiform gyrus. All the peaks examined except one (left parietal) displayed the $long > short$ pattern to some extent, which was maximal over left lingual and right temporal sources (see Figure 5.4B). The distribution of the pattern is consistent with the data observed at sensor level, and as in the sensor analysis, we found no strong evidence for the significance of the $long > short$ pattern at any of the examined peaks. During the third window (0.07 to 0.14) a focused pattern of activation over primary visual cortex was found (see Figure 5.4A). The peak activation with respect to baseline was found at the left calcarine sulcus (see Table 5.2), with long words eliciting a considerable larger increase than short ones (see Figure 5.4B).

Summing up, the statistics performed at sensor level showed that as expected, long words elicited a stronger evoked response over posterior sensors than short ones. Although this effect was not big enough to reach an $\alpha = 0.05$ level of significance, both its topographical distribution and its direction ($long > short$) match our a priori hypotheses. Source localization identified primary visual cortex as the main source of the differences between long and short words, providing further evidence that this was a true effect: the length contrast at this peak had a t-value of $\approx$4. Sensor level data showed that a posterior $long > short$ trend was present well before word onset,

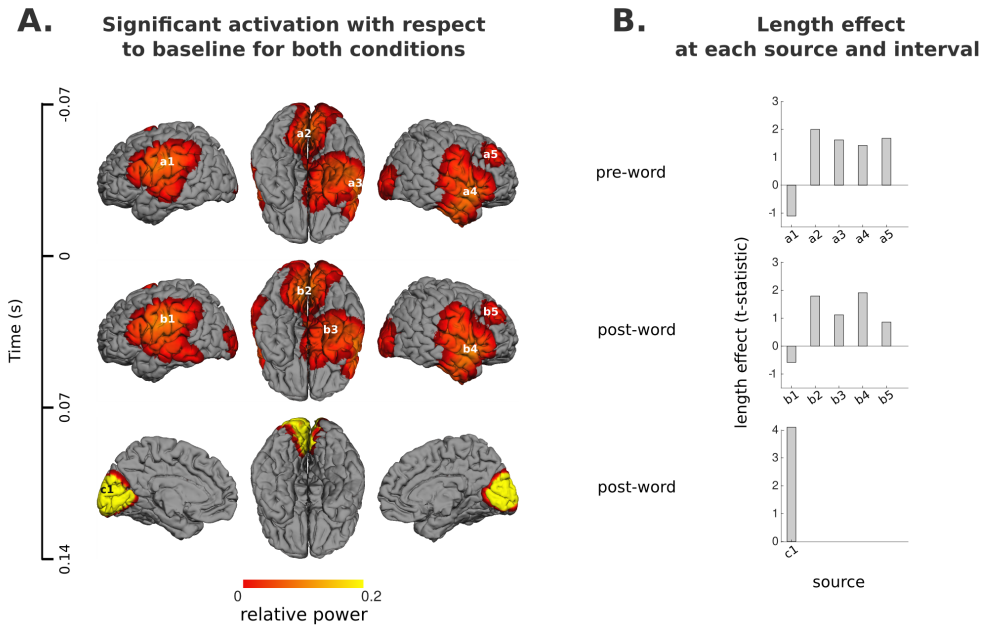

Figure 5.4: Source reconstruction of ERF effects. A — Regions of significant power change with respect to baseline (both conditions and experiments) at 0.07 s windows just before and after word onset. The peak activity locations within these areas are marked ($a1$ to $c1$) within these images. B — Effect-size of the length-expectation effect (estimated as the standardized difference between conditions) at each peak and time interval.

and these differences were traced back left secondary visual cortex (lingual gyrus). However, the difference between long and short words at this point was small, and not enough statistical evidence was found to interpret it as a true effect.

Preparation for stimulus: image-locked time-frequency response

As in the ERF analysis, we assessed differences between length conditions at the sensor level using all subjects pooled together. We refer to this contrast from now on as the length-expectation effect. In addition, we assessed the interaction between length-expectation and temporal predictability (long minus short in Experiment 1 vs long minus short in Experiment 2). Given that in the ERF analysis the length effect whilst viewing the word was not strong enough to reach significance when considering the whole scalp, analyses in the predictive period were topographically restricted to posterior sensors. Time-frequency responses (TFR) were estimated in a pre-target word interval (from image offset to the minimum interval length: 250 ms to 1500 ms) and at frequencies in the range 3-30 Hz. Finally we performed source reconstruction using a minimum variance beamformer to localize the brain regions underpinning the effects identified in sensor data. Based on the literature and on

results obtained with ERFs, we expected these to include visual cortices up to the visual word form area.

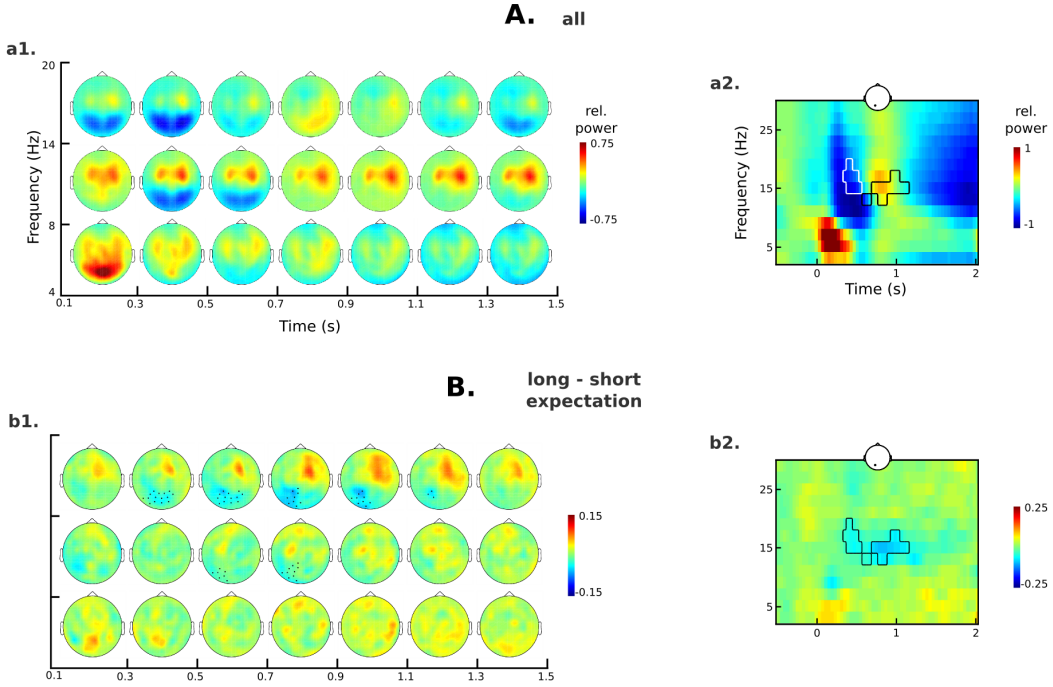

Figure 5.5: Image-locked time-frequency response. A — Power change in all conditions with respect to baseline during the analyzed pre-word interval. B — Length expectation effect during the same interval. Sensors that formed part of the clusters with p <0.1 in the long vs short analysis are marked in black. Plots on the right (a2, b2) display time-frequency profile for a representative posterior sensor, and include an outline of the previously mentioned clusters.

Figure 5.5A shows the evolution of relative power with respect to baseline as function of time at theta, alpha and beta frequencies. Image offset (at 0.25 s) induced a posterior power decrease at alpha and beta frequencies that turned to power increase in the beta band at 0.6 ms. At beta frequencies there was also a power increase over right anterior sensors over the whole interval, and a decrease that intensified as a function of time over left anterior sensors. Image offset also induced a bilateral anterior power increase at theta and alpha frequencies that was sustained throughout the analysis interval at alpha frequencies, but gradually diminished at theta frequencies. Figure 5.5B shows the evolution of power differences between long and short conditions. The largest differences were found in the beta band with two spatially distinct distributions displaying opposite patterns: less power for long than short words over left-posterior sensors, and more power for long than short words over right-anterior ones.

The topographically restricted cluster-based permutations provided moderate evidence in favour of rejecting the null hypothesis. The clustering algorithm identified two negative clusters capturing the previously described posterior differences between long and short words. These clusters (p = 0.05, p = 0.07) spanned contiguous time segments and similar frequency bins and sensors (0.34 - 0.58 and 0.58 - 1.14 s; 15 - 19 and 13 - 17 Hz; See Figures 5.5b1 and b2). The first cluster included the period of post-image beta decrease, which was stronger for the long- than the short-expectation condition (see Figure 5.5 a2). The second cluster was located over the subsequent rebound in the same frequency band, with long-expectation showing a weaker increase as compared to baseline than the short-expectation condition. Cluster based permutations comparing the word expectation effect (long minus short) between experiments revealed no evidence for an interaction (largest cluster, p = 0.2).

Summing up, expecting a long vs. a short word led to early differences over posterior sensors, that were not sustained over the whole analysis window. These differences concentrated in a low beta range, spanning a period of power decrease following image offset, and a subsequent rebound, with less beta power for the expectation of long- than short-words. We found moderate evidence that these differences were significant. This pattern appeared consistent in the two Experiments, with no evidence for an interaction between length expectation and temporal uncertainty. Differences between the length-expectation conditions were apparent over right anterior sensors and a beta frequency range too, but these had been excluded from the statistical analysis due to their topographical anterior distribution.

In order to identify the cortical sources responsible for the effects observed at source level we evaluated source power at at a low beta frequency range (8 - 14 Hz) and four 0.2 s windows spanning a 0.3 - 1.1 s interval. The chosen interval includes the period of maximal differences between expect-long and expect-short conditions, and was broken up into four windows in order to examine separately the different beta dynamics within it: initial power decrease (first window), period of rapid power increase (second window), rebound peak (third window), and subsequent decrease (final window).

Figure 5.7A presents areas of significant power change with respect to baseline (assessed with permutation statistics, both conditions pooled together) for the first three time windows, and Table 5.3 presents the coordinates of local maxima. The last window, from 0.9—1.1 s, was omitted from further analysis since no occipital regions presented significant change with respect to baseline. The first two windows show

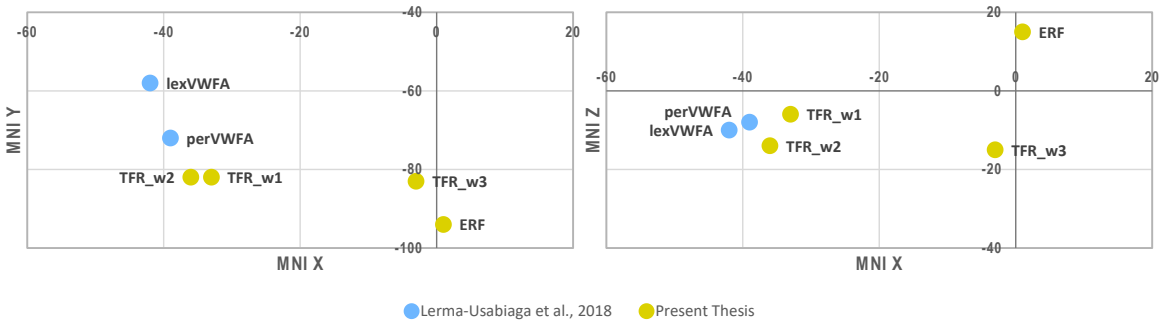

Figure 5.6: Coordinates of left hemisphere peak activations in our Experiments and their relationship to VWFA according to Lerma-Usabiaga et al. (2018). Data labels refer to: lexVWFA, perVWFA: lexical and perceptual VWFA respectively, as described by Lerma-Usabiaga et al. (2018). TFR_w1, TFR_w2, TFR_w3: peaks identified in our TFR source localization of the pre-word interval, ERF: peak of activation at calcarine sulcus in our ERF analysis of the post-word interval.

Table 5.3: MNI coordinates of the peri-word beta activity peaks, during the three temporal windows examined. The label column refers to location markers used in Figure 5.7

Window 1: 0.3 – 0.5 s		Window 2: 0.5 – 0.7 s		Window 3: 0.7 – 0.9 s	
MNI coordinates [x y z]	Label	MNI coordinates [x y z]	Label	MNI coordinates [x y z]	Label
[-33 -82 -6]	a1	[-36 -82 -14]	b1	[-3 -83 -15]	c1
[35 -88 2]	a2	[34 -68 18]	a2	[19 -86 16]	c2

a general power decrease with respect to baseline over occipital cortex, whilst the third shows power increase, concentrating around the calcarine fissure. The second and third windows show, in addition, frontal decrease with respect to baseline, and a focus of power increase over right supramarginal gyrus (note that in the figure, these frontal and parietal activations are not visible in the third window, since a medial lateral view is depicted to better represent the occipital activations and peak locations).

In each of the three windows two local maxima were found at occiptal areas, one per hemisphere. In the first window, the peaks were located in inferior occipital areas, close to the fusiform gyrus on the left. In the second window, the peaks at each hemisphere are shifted ventrally, both locating at the fusiform gyrus. In the third window, coinciding with the peak of the beta rebound, the peaks locate near the calcarine sulcus in both hemispheres. We subsequently compared the long and short conditions at each peak using t-tests, to check whether differences at these locations could account for the patterns observed at sensor level (See Figure 5.7B). The largest length-expectation effect was found in the second window (corresponding to the period of rapid beta increase), over bilateral fusiform giri, although it was also of a considerable magnitude ($t\text{-}statistic \sim -2$) in the first window (beta trough) at the left fusiform/inferior occipital site.

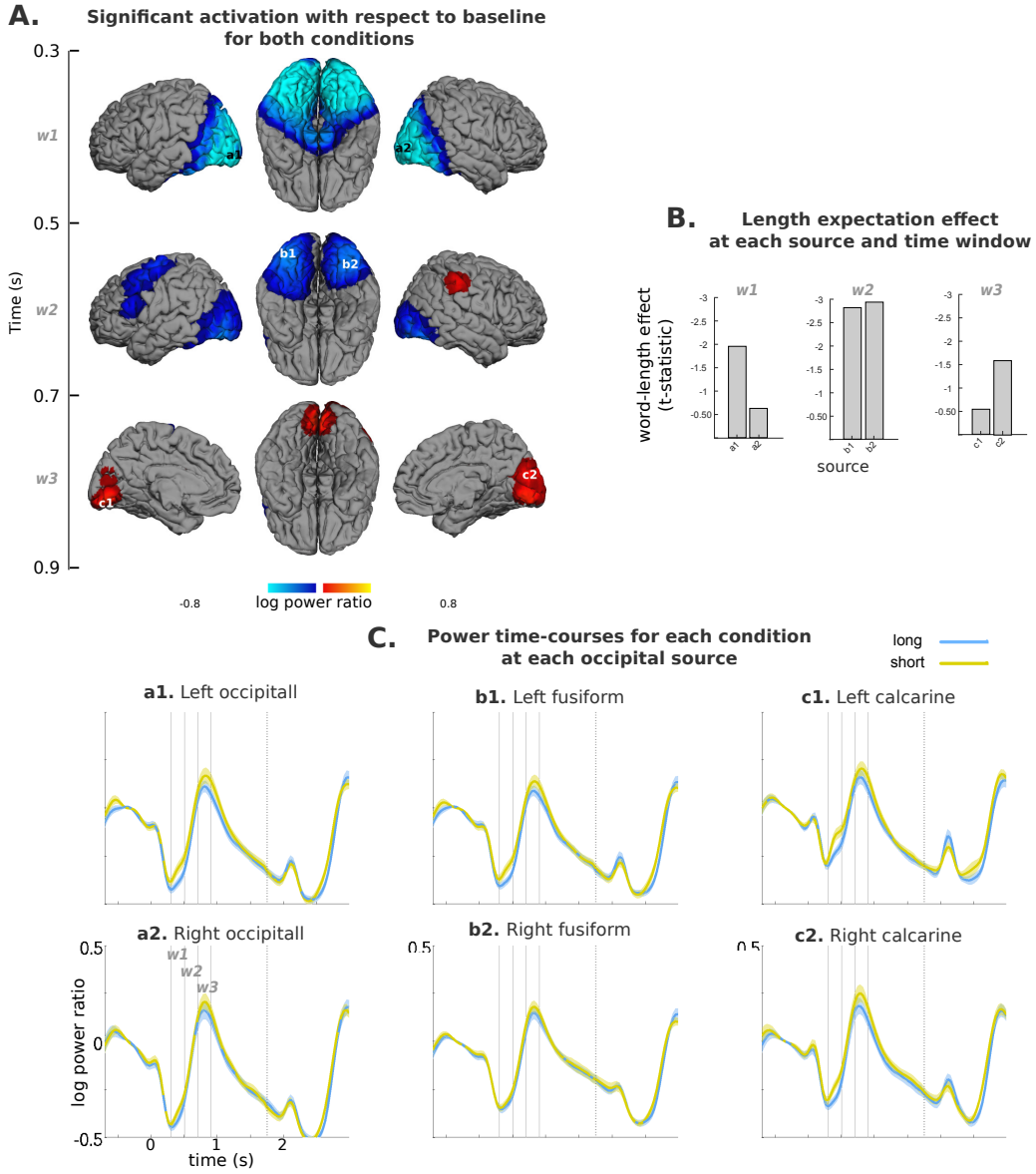

Figure 5.7: Source reconstruction of TFR effects. A — Regions of significant power change with respect to baseline for a low beta band at three different time intervals. The peak activity locations are labelled within each map ($w1$ to $w3$). B — Effect-size of the length-expectation effect estimated as the standardized difference between conditions in power averaged across window. C — Time-courses along the whole trial for each length expectation condition at each peak location. For visualization purposes, the standard error of the mean for each time-point is depicted as a shaded area around the main line. Vertical dashed lines within these plots indicate a discontinuity in the data shown: to the left the data reflects the mean calculated using both experiments, to the right only the data for Experiment 2 remains, given that word onset latencies in Experiment 1 were variable. Solid vertical lines indicate time windows $w1$ to $w3$.

Summing up, during the period of maximal long-short differences identified at sensor level, occipital beta-power undergoes important dynamic changes both in power magnitude and location of peak activity. Differences between long- and short-word expectation conditions were largest at bilateral fusiform giri in the second window, during a period of rapid power increase. However, they seem to be present to some extent over left occipital cortex at the beta power trough. Later, during the beta power rebound captured in the third window, t-values seem to drop as a result of higher individual variability (see width of shaded error area around the timecourses in Figures 5.7C). Interestingly, power peaks during this third window were located at primary visual cortex.

Post-hoc analyses

In addition to the planned analyses reported above, one pattern in the data merited further exploration. The plotted scalp topographies in Figure 5.5 show a marked positive difference between long- and short-expectation conditions in a low beta range, that being distributed over left anterior sensors was not assessed in our planned statistical analysis. Post-hoc cluster permutations statistics over the whole scalp identified a cluster with an un-corrected p-value of 0.03 capturing this pattern (0.5—1.4 s; 11—19 Hz; mean frequency = 16 Hz). To explore this trend further, or others that may be present in the data, we performed source localization for a low beta band in 0.2 s intervals over the whole window. A right parietal peak emerged in the windows examined from 0.6 s onwards. However, this synchronization peak did not seem responsible for the length expectation effect: the contrast between expect-long and expect short conditions at this site yielded t-values $\leqslant 1.7$ in all the windows (See Table 5.4a). In order to find the locus of this effect we re-examined the source maps with respect to baseline for both expectation conditions pooled together with a more liberal threshold with respect to baseline. This revealed a beta desynchronization source over a fronto-parietal area (should find a better anatomical term here), that was less intense for expect-long than expect-short conditions. The contrast at this site yielded t-values $\sim$2 in all windows examined, thus being a more likely candidate for the effects observed at the scalp (See Table 5.4b). Indeed, this is consistent with the scalp topographies, where it could be observed that the expectation contrast peaked over more anterior sensors than the synchronization with respect to baseline.

Table 5.4: Peak coordinates identified in the post-hoc analysis and t-values for the length expectation contrast at those locations

(a) Beta increase peaks

Time window	MNI coordinates [x	y	z]	t-value
0.5 – 0.7 s	[57	-31	45]	-1.0
0.7 – 0.9 s	[59	-31	39]	0.2
0.9 – 1.1 s	[57	-31	40]	1.4
1.1 – 1.3 s	[58	-30	41]	1.7

(b) Beta decrease troughs

Time window	MNI coordinates [x	y	z	t-value
0.5 – 0.7 s	[17	-11	61]	1.9
0.7 – 0.9 s	[28	-12	71]	2.5
0.9 – 1.1 s	[21	-11	66]	2.0
1.1 – 1.3 s	[16	-11	64]	2.1

5.4 Discussion

In the present study we aimed to find evidence for predictive pre-activation of visual words and to explore the oscillatory mechanisms generating them. We compared pre-stimulus activity between expect-long and expect-short conditions in a paradigm that maintained maximal predictive validity as to word identity. Indeed, we found significant but subtle differences at the sensor level between expect-long- and expect-short- conditions involving visual-processing areas in a low beta range. Both the anatomical localization of the effect at ventro-occipital cortex and the low beta frequency range it comprised are consistent with our pre-activation hypotheses. However the effect was early and not sustained until word onset, so that no clear relationship with actual word processing was apparent. Both the transient nature of the effect and the moderate evidence for its statistical significance hinder making strong interpretations regarding the presence of pre-activation as a result of perceptual prediction mechanisms, but the observed patterns reveal suggesting and sometimes intriguing possibilities that will be discussed in the following paragraphs.

In addition to the contrast between expect-long and short words, we examined the evoked response to the actual word and the influence of temporal predictability on item-pre-activation. The evoked response to the word showed that the length-contrast indeed elicited a differential response at occipital cortex, but this difference was small and only marginally significant. It is not surprising, therefore, that the effect in the pre-target word interval was also a subtle one. Interestingly, the $long > short$ pattern in the evoked response seemed to be present before word onset, but not enough statistical evidence was found to interpret it as a true effect. Finally, no interaction was found between temporal predictability and length expectation,

suggesting that predictive processes in the visual domain are not strongly modulated by temporal uncertainty.

Oscillatory findings

Differences between long and short expectation conditions were found in a low beta range, with power being lower for long- than short-word expectation condition. The involvement of low beta oscillations is consistent with prior work in visual prediction and with the predictive coding literature, that associate beta oscillations with feedback information flow. In addition, beta oscillations have been widely implicated in top-down modulation of perception, with desynchronization in this band being associated with the activation of representations (Hanslmayr et al., 2012). This would be consistent with our pattern of results: larger neural populations would be implicated in the representation of long- than short-words, given the retinotopic organization of visual cortex, and the larger portion of the visual field occupied by long words.

The observed beta effects spanned period of rapid overall beta power change after image onset, including three main trends that were source-localized separately. Initially, beta power decreased rapidly, reaching a trough at 0.3 s after image onset. At this point the length expectation effect started, with long-word expectation condition exhibiting a stronger decrease. This was followed by a period of rapid increase, which reached its peak at 0.8 s, to decrease again progressively until word onset. Interestingly, this tri-phasic pattern of beta modulations during a cue-target delay are strikingly similar to what has been previously found over sensory-motor cortex in tasks requiring a motor response. In a review of such motor response studies, Kilavik et al. (2013) examine paradigms using a cue-delay-target structure, describing a prominent post-cue decrease in beta power over sensory-motor cortex with relatively fixed timing at approximately 0.3 s after the cue, but insensitive of delay duration. This decrease is typically followed by a power increase, peaking around 0.8 s after the cue, and a new decrease until final GO target. Interestingly, the initial power trough is linked to movement preparation, being sensitive to the informativeness of the cue regarding the required response. This pattern is thus strikingly similar to our observations over visual cortex, whereby the initial post-cue power trough would index a visual preparation ahead of the upcoming stimulus in response to cue-information regarding the target.

The length expectation effect detected at the sensor level also included the subsequent beta increase. It is unclear, however whether this is a carry-over effect,

or it reflects further computations on the sensory word representations. In any case, differences between expectation conditions wear off after that, with the final pre-GO beta decrease being very similar for both conditions. This raises interesting questions regarding the prediction process. Although it has been previously argued that sensory predictions may be instantiated just in time ahead of expected stimulation, our data and the motor studies reviewed by Kilavik et al. (2013) seem to suggest that predictive sensory (or sensory-motor) representations may be activated as soon as prior information licenses a precise expectation. Beta band decreases would index this process, but maintenance of the expected representations, or their re-activation just in time ahead of stimulation, would be instantiated through different neural mechanisms. One possibility is that these sensory predictions are maintained as "activity-silent" representations, where functional connectivity, rather than population-level activity, would maintain precepts active in working memory (Stokes, 2015).

Anatomical findings

The previously described temporal beta dynamics were found to originate in also changing neural sources. The source analysis with respect to baseline showed that the initial post-cue beta desynchronization encompassed occipital cortex as a whole, but peaked over ventral occipital area in the first window analyzed. In the second window, during the beta transition, the occipital peaks localized to a similar albeit more ventral area. During this second window additional peaks were also found in left dorso-lateral cortex, including the IFG, commonly identified as the top-node within the language processing network. Finally, the third window examined, encompassing the subsequent beta increase, localized to primary visual cortex. Beta activity thus seemed to follow a trajectory from secondary to primary visual processing areas, suggesting that predictive top-down information flow is indeed capable of inducing sensory pre-activation at the earliest sensory processing areas, including and beyond the VWFA. Intriguingly, the expectation effect in secondary and primary visual cortices occurred during two distinct beta dynamics: a post-cue trough in the latter, and the subsequent increase in the former. If pre-activation is occurring in both cases, it seems that the mechanisms involved may be qualitatively different, even if they encompass a similar frequency band.

The time-course of peak activity within IFG does not seem entirely congruent with a top-down predictive information flow. If the IFG is the most abstract node within the language network, it should direct activity at lower hierarchical levels, and therefore activate before visual cortices. On the contrary, our results show that

IFG peaked in the second window, after the expectation effect was first observed within a visual area. This raises the possibility that semantic information extracted from object pictures could lead directly to visual wordform pre-activation, without intervention of the IFG. In fact, the connectivity within ventral occipito-temporal cortex (vOTC) lends plausibility to such an interpretation. Lerma-Usabiaga et al. (2018) propose that the vOTC displays a functional subdivision, with medial areas responding to the abstract properties of words as linguistic units, and more posterior areas responding the perceptual features of those words. Crucially, they found that the described functionally seggregated areas within vOTC were associated with different white-matter tracts: whilst the posterior portion was connected to the inferior parietal sulcus via the vertical occipital fasciculus, the middle area was associated with the posterior arcuate fasciculus, connecting it to the IFG. The source of our length expectation contrast, localized to a posterior location, could be structurally connected to the inferior parietal sulcus, which might be acting as the higher heriarchical node for this area rather than the IFG.

Post-hoc findings and additional considerations

In addition to the hypothesized effects over occipital areas, we observed a marked difference between expectation conditions over a right fronto-parietal source in the beta band, from 0.5 s onwards. During this time there was a bilateral beta decrease over fronto-parietal areas, that was more intense in the left hemisphere. However, differences between expectation conditions were only observed in the right, with long words exhibiting a less intense decrease than short ones. This beta decrease could be linked to motor preparation, maybe seccade planning that would be dependent upon word length. Indeed Ikkai and Curtis (2008) suggest that similar activity would ensue from gaze planning ahead of actual movement, or from covert attentional selection towards a spatial location. Right lateralized effects could arise as a differential spatial attention to the initial letters of the word, that would be further to the left for long than short words. However, the direction of the effect seems contrary to this interpretation: long words exhibited less beta decrease than short words.

Summing up, our data showed differences between expectation conditions that were apparent from 0.3 s onwards, in a low beta range over vOTC. Both the frequency range and the anatomical localization are consistent with the the presence of predictive wordform preactivation. Interestingly the timing of this effect suggests that item pre-activation occurs as soon an prior information licences an expectation, even if the expected item will only occur after a delay. A later effect was detected

over right motor areas, from 0.5 s on-wards, possibly as a result of differential spatial attention allocation for long than short words. Given its temporal profile, this attentional mechanism could rely on the previous activation of the expected word representations. The differences between long and short word expectation detected at occipital cortex were not sustained until word onset, so that we cannot draw conclusions regarding what predictive mechanism might facilitate perception once the expected word apears. Interestingly, in the ERF data, we did observe a $long > short$ trend that started 0.5 s after cue onset, that did not reach significance in the interval analyzed. An intriguing possibility might thus be that low beta modulations index the generation of the prediction, but maintenance of the representation may only be apparent in the evoked response.

Part III

Discussion and final remarks

Chapter 6

General discussion

The main goal of the present thesis was to find evidence for the pre-activation of sensory representations of expected words, and to characterize the oscillatory correlates of the prediction process supporting such pre-activation. The question of whether top-down influences can modulate early sensory representations in the absence of stimulation transverses different fields of neuroscientific research. It is a specially contended topic within the language processing literature, including word recognition models, sentence processing research, and functional characterization of anatomical language networks. However, direct evidence for (or against) sensory pre-activation in language is still scarce. Previous studies have often focused on the interval after word onset, thus examining the consequences of prediction rather than the preparatory process itself, and have used constraining vs non-constraining contrasts, which may be too coarse to isolate sensory pre-activation processes.

We aimed to overcome the limitations of prior work by focusing on the pre-target word interval, and by comparing two conditions where the upcoming words were always highly expected, but differed in their sensory features. Word expectations were induced by a picture cue, and these expectations were never violated. Conditional probability of target words was therefore kept at 100%, boosting our ability to detect probabilistic pre-activation. The use of MEG allowed us to characterize the oscillatory and anatomical correlates of the prediction process, in order to relate our findings to previous work regarding prediction in general, and anatomical organization of the language network in particular. Finally, we explored the generation of representations in the visual and auditory domains, and the role of timing, by using spoken and written words as targets, and by conducting experiments with fixed and variable delays between picture cues and word targets.

6.1 Summary of results

In Chapter 4 we explored auditory pre-activation by manipulating expected words' initial phonemes, which could either be fricatives or plosives. These two consonant types differ in their manner of articulation, and generate contrasting auditory waveforms. Our results showed significant differences between these two conditions in the pre-target word interval, spanning theta and alpha ranges. Source reconstruction showed a brain-wide network involving several frequency bands, including bilateral superior temporal areas commonly associated with phonological processing in a theta range (see Figure 6.1). The event-related fields analysis of the response to the word showed an interaction between initial phoneme and temporal predictability. Only when the delay between cue and target was variable did differences in the ERFs between the two phoneme conditions emerged before word onset.

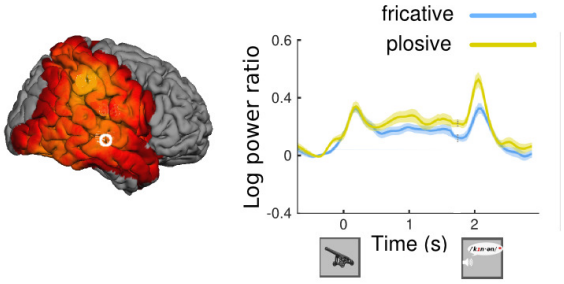

Figure 6.1: Theta power modulations with respect to baseline (left) and timecourse as a function of expected phoneme at marked peak location (right)

In Chapter 5 we explored written word pre-activation by manipulating expected words' length. Expecting a long vs. a short word led to early power differences over posterior sensors, that were not sustained over the whole analysis window. These differences concentrated on a low beta range, spanning a period of power decrease following image offset, and a subsequent rebound, with less beta power for the expectation of long- than short-words (see Figure 6.2). Source localization showed that the effect originated near the lingual/fusiform gyrus area, close to the so-called visual wordform area. However, this effect had a transient nature, and was not sustained till word onset. On the contrary, the event-related fields on the peri-word interval appeared to show that differences between long and short words emerged before word onset, but not enough statistical evidence was found to deem

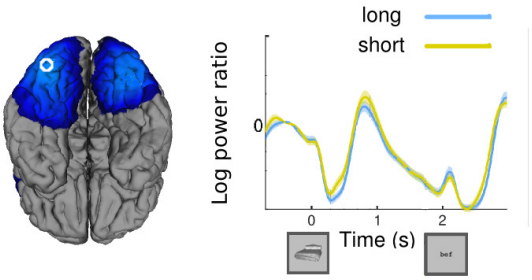

Figure 6.2: Beta power modulations with respect to baseline (left) and power time-course at marked peak location as a function of expected word length (right)

this a true effect. Finally, no evidence was found for an interaction between temporal uncertainty and word length.

6.2 Main questions addressed

Can prior knowledge lead to sensory pre-activation of upcoming words?

Our results revealed a differential neural response to cues that generated expectations for words differing only along a sensory dimension over the corresponding sensory cortices, both in the visual and auditory domains. These results thus confirm the presence of wordform-related neural activity before word onset, suggesting that predictive activation of expected precepts may be instantiated at a sensory level. This pattern is congruent with stronger predictive processing accounts, that argue for the presence of sensory templates even before initial stimulation (Kok et al., 2017). Within the language processing domain, our data support fully interactive word recognition models in the visual domain (Price and Devlin, 2011; McClelland and Rumelhart, 1981), with differences between conditions reaching primary visual areas (see Figure 6.3). In the auditory domain differences were traced back to the STS, a phonological processing area, but no evidence was found that pre-activation could encompass an earlier, non-phonological auditory feature level. However, this difference in the "hierarchical reach" of predictions between the visual and auditory domains can be attributed to the experimental contrasts employed, rather than to qualitative differences between the sensory systems themselves. Whilst the distinction between short and long words is purely visual, with no linguistic meaning, phonemes are linguistic sound categories. It might therefore be unsurprising that whilst the visual contrast elicited differences in early visual areas, the auditory contrast did so at a higher level of abstraction, where sounds are identified under linguistic categories.

However, even in our paradigm, with 100% cue validity, the detected effects were subtle. Firstly, in the auditory experiments, the source contributing most to the effect observed at the sensors was an alpha parietal location, in our interpretation responsible for implementing an attentional filter, rather than a predictive pre-activation. The theta effects observed at the superior temporal area alone might have been insufficient to generate a statistically significant effect at the sensors. Secondly, in the visual domain, the statistical evidence for the the initial effect was not very strong, and no evidence was found of its maintenance until word

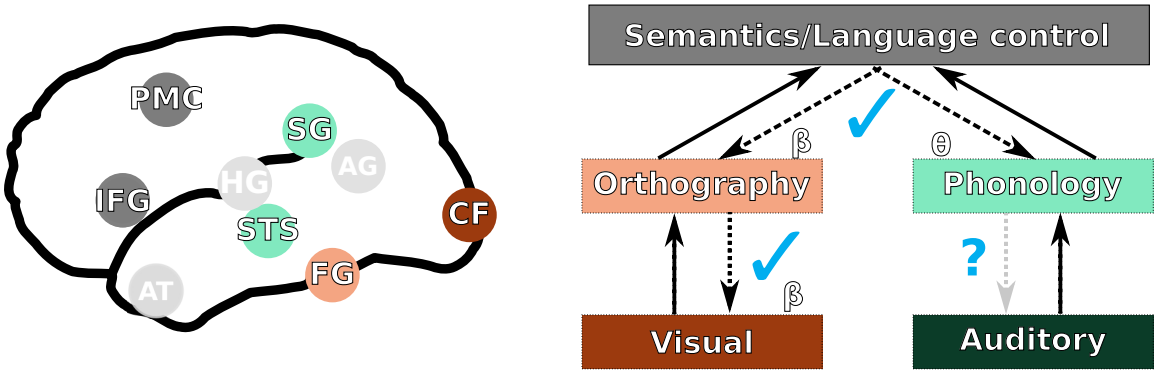

Figure 6.3: Anatomical nodes and hierarchical system observed in our studies. Nodes in dark grey indicate locations where a peak of activation was observed with respect to baseline, but did not exhibit a sensory feature effect. Locations in light grey represent nodes in the language network that did not come up in our analyses. Left hemisphere is shown as an example, but the figure should be interpreted as "hemisphere-agnostic".

onset. The subtlety of these effects could be in part responsible for the scarcity of pre-target word evidence in the literature and should be taken into account in future studies.

Finally, there could be alternative interpretations for the effects observed in our studies. On the one hand, the transient effects observed in the visual data could reflect an automatic process of orthographic activation upon selection of a lexical candidate not related to preparatory activity for future perception. On the other hand, a strategic inner naming process related to our task could be responsible for the effects observed in the auditory experiment, bearing no relation to automatic predictive perception processes. In our view, these interpretations are not very likely: Firstly, debriefing questionnaires conducted after the study did not reveal any evidence to support the presence of strategic inner naming in the auditory experiments; Secondly, as discussed in the previous chapter, the oscillatory patterns observed in the visual data are suggestively similar to beta dynamics attributed to predictive processes in non-linguistic studies (Kilavik et al., 2013). Nevertheless, further work is needed to rule out alternative explanations to the effects observed, and to directly link the early post-cue effects to subsequent perception of actual words.

Does temporal predictability influence pre-activation of expected words?

Our results revealed a different answer to this question depending on presentation modality of the expected words: predictability of the temporal onset had a

determining influence in the auditory, but not the visual domain. An interaction between time predictability and expected phoneme was detected in the ERF analysis of the auditory experiments, with differences between phoneme conditions before word onset observed when the delay between the cue and the target was variable, but not when it was fixed. The differential modulation of pre-activation effects by timing in our auditory and visual data may reflect the different characteristics of the sensory modalities themselves (MacSweeney et al., 2008). Whilst information coding in speech signals relies heavily on timing, temporal information is irrelevant for orthographic coding, that instead relies on static spatial light/dark patterns. It is therefore not surprising that timing-related effects could be found in the auditory data, where temporal dynamics are inherent to phoneme-identification, but not with written targets.

Source localization of the ERF phoneme effect detected only under temporal uncertainty revealed a left inferior parietal source, near the alpha power peak that was found to be responsible for differences between phoneme conditions in both experiments. We previously suggested that this left parietal source may be implicated in attentional processes as a result of the different temporal dynamics of our manipulated consonants (fricatives vs plosives). The detected interaction in the ERF at this source could thus reflect the deployment of an additional attentional mechanism in cases of higher uncertainty, when the temporal features of the upcoming stimulus could not be reliably predicted. This would support the dissociation between feature-based attention and feature-based expectation proposed by Summerfield and Egner (2016), whereby these would be two cognitively distinct processes implemented through distinct neural mechanisms.

Characterization of the predictive process

Evidence for sensory pre-activation was found both in the auditory and visual data, but the oscillatory correlates and temporal dynamics of the effects were different in each modality. The theta band was associated with pre-activation in the auditory domain, with differences between expect-fricative and expect-plosive conditions that started $\sim 0.3s$ after cue onset and that were sustained until word presentation over right hemisphere locations. These differences were source-localized to several peri-sylvian regions, including right STS and the rolandic operculum, areas implicated in phonological representation and articulation. In contrast, differential activity to long vs short words in the visual domain were transient -they started $\sim 0.3s$ as well, but weaned off $\sim 1s$ after cue onset- and encompassed a low beta range. Furthermore, both the beta modulation and its

sources changed dynamically, with an initial decrease localized to ventral visual areas associated with perceptual wordform processing, and a subsequent increase peaking at a striate area, corresponding to primary visual cortex.

These differences across modalities may reflect preferential oscillatory signatures of different brain areas (Schoffelen et al., 2017), and are in fact congruent with previous literature in each domain, with beta oscillations indexing predictions in the visual domain, and lower frequencies predominating in the auditory modality (Bastos et al., 2015; Fontolan et al., 2014). Furthermore, these differences may reflect contrasting characteristics of each sensory modality, such as their temporal-spatial resolution, and the subsequent differential information encoding in spoken and written language. Intriguingly, differences between modalities were not restricted to their frequency correlates, but also their temporal dynamics and "hierarchical reach". Whether this reflects inherent differences in the predictive processing of speech and written texts, or is a trivial consequence of the specific contrasts employed in our experiments is an open question in need of further research.

Finally, despite the differences discussed in the previous paragraphs, the patterns observed in both experiments also present several similarities. Firstly, feature-expectation effects started 0.3 s after cue onset in both modalities, suggesting that pre-activation processes may be triggered as soon as prior information licences a sensory-specified prediction. Secondly, frontal beta activity with respect to baseline was similar in both experiments: a significant decrease was detected in both cases, encompassing inferior frontal regions, that started 0.5 after image onset and was intensified until word onset. Intriguingly, this frontal region, often considered a top node within the language processing hierarchy appeared to be activated after wordform-related differences in sensory cortices became apparent. These data seem to suggest that although the IFG seems to be playing an important part in the preparation process, it does not modulate wordform access after the cue, given that its activation follows the sensory effects detected in our study. This raises the possibility that access to phonological or visual word representations could follow directly after object identification takes place in visual recognition areas.

6.3 Concluding remarks

Overall, our results suggest that sensory pre-activation may be observed when prior information licenses a sensory-defined, precise, expectation. These results contribute to the understanding of the role of predictive processing in the linguistic

domain and beyond, showing that sensory pre-activations may be generated when context clearly biases towards a specific precept. However, even in our 100% predictability context, the sensory pre-activation effects were subtle, which may account for the scarcity of data supporting their presence in the literature. By identifying the oscillatory frequencies that characterize such pre-activation, we hope these results may help interpret pre-target word brain activity in other experimental set-ups, and contribute to the understanding of neural mechanisms supporting predictive processing in general.

Part IV

Appendices

Appendix A

Behavioural results

Participants in both experiments showed high accuracy in detection of wrong words, with a mean of 95% hits the auditory blocks in both experiments (standard deviation of 5 and 4% in Experiment 1 and 2 respectively) and 91% in the visual blocks (standard deviation: 6% in both experiments). Mean response time across experiments and conditions was 750 ms (with a standard deviation of 68 ms) in the auditory blocks, and 548 ms (standard deviation of 52 ms) in the visual blocks.

We further analyzed the behavioral data using the free software statistical package R (R Core Team, 2013), and the lme4 library (Bates, 2010). Two different models were buildt one for the auditory blocks, and one for the visual blocks. Given the controversy behind calculating degrees of freedom and corresponding p-values in these types of models, we evaluated the significance of predictors using the normal approximation ($|t > 2|$). Trials with incorrect responses and reaction times (RT) under 0.2 s were removed before model fitting. The resulting RTs served as the dependent variable against mixed effects multiple regression models were built. Our independent variables of interest included the following bivariate categorical variables: Sensory feature (phoneme: fricative or plosive; or lenght: long or short), Image cue (predictive or nonpredictive), wordStatus (correct or incorrect) and Experiment (1 or 2). These variables were coded using treatment coding, making plosive/long, predictive, correct, Experiment 1 trials as the reference level.

Auditory blocks

We attempted to fit maximal mixed-effects models, using reaction times as the dependent variable and our main experimental manipulations and their bivariate interactions as fixed effects and as by-subject random slopes, in addition to random by-subject and by-item intercepts. This maximal model did not converge. We subsequently removed the by-random slopes for the interaction terms in an intermediate model and built a final model in which only interactions with significant fixed effects ($|t - value| > 2$) were added as random slopes (see Table 1 for final model specification).

The final model showed clear modulations of reaction times by the predictive value of the image cue and by initial phoneme condition, being slower for non-predictive than predictive image-cues and for fricative- than plosive-initial words (see Table A.1). There was a marginal effect of word status, with incorrect words receiving slower responses than their correct counterparts, and an interaction between predictiveness and word Status, with the wrong word effect being considerably larger for trials with non-predictive than predictive image cues. Temporal predictability of word onset did not influence reaction times, with no significant effect of experiment or its interactions with the other independent variables.

Table A.1: Fixed effects for RT model. Highlighted rows indicate significant ($|t| > 2$) and marginally significant predictors in final model. The predictors, all bivariate categorical variables were coded using treatment coding, making plosive, predictive, correct, Experiment 1 trials as the reference level. Final model specification: $RT\ Feature * (Experiment + Cue + wordStatus) + Experiment * (Cue + wordStatus) + Cue * wordStatus + (1 + Feature + Experiment + Cue + wordStatus + Cue * wordStatus | subject) + (1|Cue)$.

Fixed effect	Estimate (ms)	Standard error (ms)	t-value
Intercept	**722.0**	**18.6**	**38.9**
Feature_plosive	**−53.5**	**16.2**	**−3.3**
Experiment_2	−3.8	22.7	−0.2
Cue_non-predictive	**141.9**	**19.9**	**7.5**
wordStatus_wrong	**15.8**	**9.3**	**1.7**
Feature_plosive:Experiment_2	−3.7	5.2	−0.7
Feature_plosive:Cue_non-predictive	1.9	22.5	0.1
Feature_plosive:wordStatus_wrong	4.5	4.9	0.9
Experiment_2:Cue_nonpredictive	−3.1	13.5	−0.2
Experiment_2:wordStatus_wrong	5.0	11.3	0.4
Cue_non-predictive:wordStatus_wrong	**19.4**	**6.5**	**3.0**

Visual blocks

We attempted to fit maximal mixed-effects models, using reaction times as the dependent variable and our main experimental manipulations and their bivariate interactions as fixed effects and as by-subject random slopes, in addition to random by-subject and by-item intercepts. This maximal model did not converge, neither did a reduced one where by-subject random slopes for interaction terms were removed. As a result, we report the full model with only random intercepts as by-subject and by-item terms. The final model showed no clear influence on interval timing on response times, but clear modulations by the predictive value of the image cue, with reaction times being faster when the identity of the words could be predicted, by the length of the word, with long words eliciting longer response times than short ones, and by word status, with longer response times for incorrect than correct words. However, these effects were modulated by several significant interactions with predictiveness of the cue. Under predictive conditions, the effects of timing, word length, or wrong word status were all larger than under non-predictive conditions. Finally, an additional interaction was detected between word length (feature) and word status, with the wrong word effect being smaller for short words.

Table A.2: Fixed effects for RT model. Highlighted rows indicate significant ($|t| > 2$) and marginally significant predictors in final model. The predictors, all bivariate categorical variables were coded using treatment coding, making plosive, predictive, correct, Experiment 1 trials as the reference level. Final model specification: $RT\ Cue_n onpredictive * (Experiment + Feature + wordStatus) + Experiment * Feature + Experiment * wordStatus + Feature * wordStatus(1|subject) + (1|Cueimage)$.

Fixed effect	Estimate (ms)	Standard error (ms)	t-value
Intercept	**515.5**	**12.1**	**42.6**
Cue_nonpredictive	**104.2**	**5.2**	**19.9**
Experiment_2	-16.6	15.8	-1.1
Feature_short	**-40.4**	**4.6**	**-8.7**
wordStatus_wrong	**70.2**	**2.5**	**28.0**
Cue_nonpredictive:Experiment_2	**6.9**	**3.4**	**2.0**
Cue_nonpredictive:Feature_short	**22.5**	**6.5**	**3.5**
Cue_nonpredictive:wordStatus_wrong	**-35.7**	**3.4**	**-10.5**
Feature_short:wordStatus_wrong	**-16.9**	**3.1**	**-5.4**
Experiment_2:Feature_short	4.3	3.2	1.3
Experiment_2:wordStatus_wrong	-3.2	3.2	-1.0

Appendix B

List of publications derived from the thesis

Monsalve, I. F., Bourguignon, M., & Molinaro, N. (2018). Theta oscillations mediate pre-activation of highly expected word initial phonemes. Scientific Reports, 8(1), 9503. doi:10.1038/s41598-018-27898-w

Molinaro, N. & **Monsalve, I.F.** (In preparation). Beta decreases over ventral occipito-temporal cortex index early visual pre-activation of expected wordforms

Appendix C

Resumen en Castellano

Introducción

Nuestros sentidos nos proporcionan una cantidad ingente de información sobre el entorno de manera continua. Este flujo de datos es complejo y a menudo ambiguo, haciendo de su interpretación una tarea no trivial. Sin embargo, nuestro entorno es también altamente estructurado, con notables regularidades. Los perros ladran, los objetos caen hacia abajo, en la lengua española, los artículos preceden a los sustantivos. Las teorías de procesamiento predictivo describen la percepción no como un proceso pasivo de recogida de información, sino como un proceso constructivo mediante el cual el conocimiento previo sobre las regularidades del entorno se emplea para interpretar la estimulación entrante. Aunque la visión de la percepción como un proceso de inferencia no es nuevo (inicialmente propuesto por Helmholtz en 1860), está recibiendo especial atención en la actualidad, gracias a evidencias convergentes sobre su relevancia del campo de la neurociencia cognitiva (ej. Kveraga et al., 2007) y propuestas concretas de su implementación a nivel computacional y neural (Rao and Ballard, 1999; Lochmann and Deneve, 2011), destacando el papel de las oscilaciones neuronales en dichos mecanismos (Bastos et al., 2012). De hecho, la predicción se ha postulado como el mecanismo neural esencial, con la capacidad de proporcionar una explicación unificada de la percepción, la cognición, y el comportamiento (Clark, 2013).

Sin embargo, la evidencia empírica mostrando los correlatos neurales de la generación de predicciones es aún escasa. Una de las cuestiones pendientes es qué nivel de detalle perceptual pueden acarrear dichas representaciones predicitivas. Según propuestas más fuertes, las expectativas previas podrían desencadenar

pre-activación de los preceptos esperados incluso en regiones sensoriales primarias (Kok et al., 2017). Otras propuestas argumentan que dicha pre-activación sólo podría generarse en niveles sensoriales más abstractos, una vez se ha recibido información externa inicial sobre dicho precepto (Bar, 2007). En el campo del procesamiento del lenguaje, esta cuestión está especialmente candente (Huettig and Mani, 2015). A pesar de la abundante evidencia de la relevancia de predicciones a nivel del significado en la comprensión del lenguaje, la extensión de dichas predicciones a niveles sensoriales está en entredicho. La literatura existente presenta ciertas limitaciones que dificultan la resolución de este debate. En primer lugar, la mayoría de los estudios de predicción en el lenguaje analizan la respuesta a palabras predecibles, observando por lo tanto *las consecuencias* de la predicción, en lugar de la *generación* de dicha predicción. Por otro lado, dichos estudios suelen manipular la predictividad de las palabras, contrastando palabras altamente y escasamente predictivas. Es probable que dicha comparación incluya procesos mentales adicionales a la predicción, como por ejemplo diferencias en el nivel de atención (Molinaro and Monsalve, 2018).

La presente tesis pretende contribuir al debate sobre la presencia de predicciones sensoriales en el lenguaje, abordando las limitaciones detectadas en la literatura existente. Los experimentos realizados se centran en el análisis del intervalo previo a la palabra, para describir el proceso de predicción y no sus consecuencias, y las condiciones experimentales incluyen siempre escenarios altamente predictivos, difiriendo solamente en las propiedades sensoriales de las palabras esperada. La técnica de imagen empleada fue la magnetoencefalografía (MEG), y la medida independiente principal fue la actividad en tiempo-frecuencia, para poder caracterizar el proceso de predicción en base a la actividad neural oscilatoria. Además, para obtener una visión más completa del proceso de predicción se realizaron experimentos con palabras habladas (experimentos auditivos) y palabras escritas (experimentos visuales), y se exploró el rol de la predictibilidad temporal, a través de experimentos en los que la palabra siempre aparecía tras un intervalo de tiempo fijo, o tras un intervalo de tiempo variable.

Métodos y materiales

El paradigma experimental empleó fotos de objetos como indicación de la identidad de la palabra que se presentaría a continuación, tras un intervalo fijo o variable (ver figura C.1). La palabra siempre era congruente con la imagen previa, pero podía contener un error en una de sus letras. La tarea principal

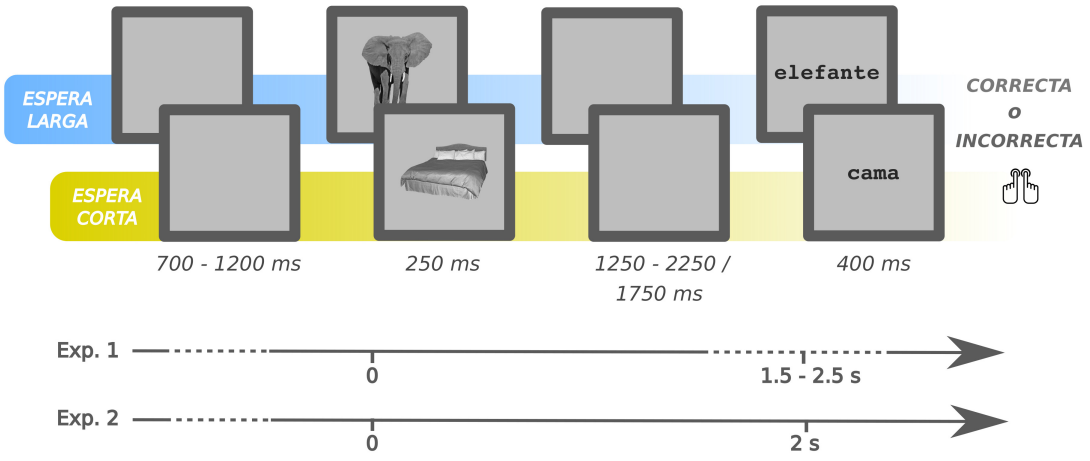

Figura C.1: Paradigma experimental.

para los participantes consistía en detectar palabras incorrectas. Esta tarea se incluyó para mantener la atención de los participantes durante el experimento. Sin embargo, la manipulación de interés en este paradigma consistía en diferencias en las propiedades sensoriales de las palabras esperadas: en los experimentos auditivos contrastamos palabras que comenzaban con fonemas fricativos (/f/, /s/, /c/) a palabras que comenzaban con fonemas plosivos (/p/, /k/, /t/), mientras que en los visuales comparamos palabras cortas (3-4 letras) con palabras largas (7-8) letras. Cualquier diferencia detectada entre las condiciones en el intervalo previo a la palabra, en áreas de procesamiento sensorial, puede ser atribuido a pre-activación de las palabras esperadas.

Los experimentos se realizaron con dos grupos distintos de participantes, siguiendo siempre el paradigma experimental descrito anteriormente. El primer grupo realizó las tareas con un intervalo de tiempo variable entre la imagen y la palabra, y el segundo un un intervalo de tiempo fijo. Ambos grupos realizaron un bloque visual y un bloque auditivo, estando el orden de dichos bloques contrabalanceado entre participantes. El análisis de datos se agrupó según modalidad de presentación (ver Tabla C.1), lo que permitió evaluar la pre-activación usando los participantes de las dos sesiones experimentales, y posteriormente su interacción con la predictabilidad temporal. La inferencia estadística se realizó con los datos a nivel de sensor, usando permutaciones basadas en clústeres (Maris and Oostenveld, 2007) sobre los registros transformados en tiempo-frecuencia. Estos tests sirvieron para evaluar la presencia de dichos efectos y para delimitar ventanas de tiempo-frecuencia para exploración posterior mediante localización de las fuentes neurales generadoras

de dichos efectos. Dicha localización se realizó mediante técnicas de beamforming (Van Veen et al., 1997), y se basó en la identificación de picos de actividad global respecto a una baseline, con la posterior exploración de las diferencias entre condiciones en los picos identificados (Bourguignon et al., 2018). Además, este procedimiento también se siguió para evaluar la respuesta evocada por la palabra en sí, como prueba de que la manipulación sensorial empleada realmente induce una respuesta diferencia en zonas sensoriales.

Análisis	Manipulación	Condiciones	Diseño
Estímulos auditivos	Propiedad sensorial	Fricativas vs Plosivas	Intra-sujetos
	Intervalo temporal	Fijo vs Variable	Entre-sujetos
Estímulos visuales	Propiedad sensorial	Largas vs Cortas	Intra-sujetos
	Intervalo temporal	Fijo vs Variable	Entre sujetos

Tabla C.1: Resumen de diseño experimental

Resultados

Experimentos auditivos

Las permutaciones basadas en clústeres mostraron diferencias significativas entre las condiciones de expectativa-plosiva y expectativa-fricativa en el intervalo previo a la presentación de la palabra. Las diferencias emergieron 300 ms tras la presentación de la imagen, y se mantuvieron durante todo el intervalo analizado, incluyendo un amplio rango de frecuencias (6-20 Hz). La reconstrucción de fuente reveló una amplia red anatómica y diversas bandas frecuenciales, destacando entre éstas áreas temporales superiores, comúnmente asociadas al procesamiento fonológico, en la banda de frecuencia theta. En estas fuentes se observó un aumento de la actividad en dicha banda con respecto al baseline, que fue más pronunciado para

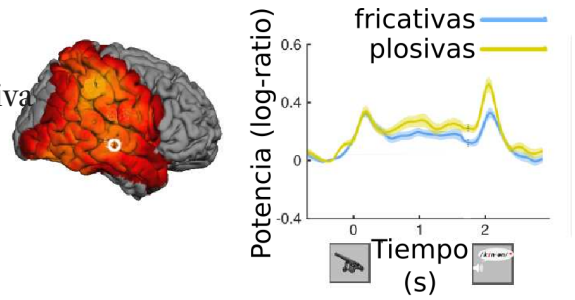

Figura C.2: Modulaciones de potencia en theta con respecto a la baseline (izquierda) y evolución temporal para cada condición en la localización cerebral marcada (derecha).

las plosivas que para las fricativas (ver figura C.2). Este patrón apareció con especial intensidad en el hemisferio derecho, incluyendo áreas temporales y frontales. El análisis de los campos evocados por la palabra reveló una interacción entre la condición de fonema inicial y la predictabilidad temporal. Sólo cuando el intervalo entre la imagen y la palabra era variable las diferencias entre fonemas emergían antes de la presentación de la palabra.

Experimentos visuales

Los experimentos con palabras escritas mostraron diferencias tempranas entre las condiciones de expectativa-larga y expectativa-corta, que no se mantuvieron en el tiempo. Éstas diferencias a nivel de sensor se reflejaron en una banda beta baja durante un periodo de rápidas modulaciones de potencia, incluyendo un periodo inicial de decremento tras la presentación de la imagen, y un periodo de incremento a continuación (el "rebote" de beta), dónde la condición de expectativa-larga presentó menos potencia que la de expectativa-corta. Las permutaciones de clústeres mostraron evidencia moderada de la presencia de un efecto en este intervalo, capturado en dos clústeres contiguos con p = 0.05 y 0.07, durante los intervalos 0.34 - 0.58 y 0.58 - 1.14 s. Este patrón apareció de forma consistente en los dos experimentos, sin evidencias de una interacción entre expectativa temporal y expectativa sensorial. La localización de fuente se realizó en ventanas temporales de 0.2 s, para capturar por separado las tendencias de beta (valle, incremento y pico). En las tres ventanas se encontraron picos occipitales, destacando localizaciones próximas a la "visual wordform area" (Dehaene and Cohen, 2011), entre el giro fusiforme y el giro lingual, y otras ubicadas en el córtex visual primario.

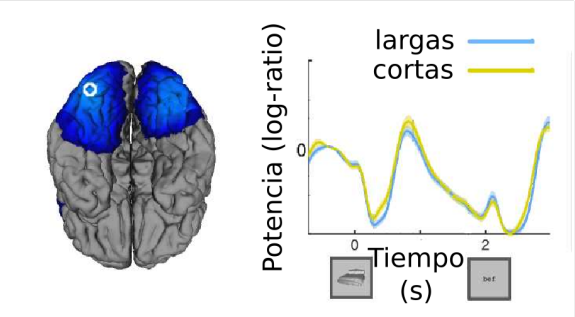

Figura C.3: Modulaciones de potencia en beta con respecto a la baseline (izquierda) y evolución temporal para cada condición en la localización cerebral marcada (derecha).

Discusión

Los resultados de los experimentos de la presenta tesis revelaron una respuesta neural a imágenes contextuales diferenciada según las propiedades sensoriales de

las palabras que iban a presentarse a continuación. Esta diferenciación se localizó en los córtices sensoriales responsables del procesamiento de dichas palabras, tanto para palabras escritas como para palabras habladas. Estos efectos sugieren que la pre-activación de palabras por el contexto puede instanciarse a nivel sensorial, en consonancia con teorías de procesamiento predictivo "fuertes", que proponen la presencia de esquemas sensoriales incluso anteriores a una estimulación inicial (Kok et al., 2017).

Aunque nuestros resultados mostraron evidencia de la pre-activación en las modalidades auditivas y visuales, los correlatos oscilatorios y la evolución temporal de dichos efectos fueron marcadamente diferentes. En los experimentos auditivos las diferencias emergieron 0.3 s tras la presentación de la imagen, abarcando un amplio rango de frecuencias y de fuentes neurales. Las diferencias en los experimentos visuales también emergieron 0.3 s segundos tras la presentación de la imagen, pero no se sostuvieron en el tiempo, y se concentraron en una banda frecuencial más estrecha (beta baja) y en fuentes próximas al córtex visual ventral. La caracterización oscilatoria diferencial en cada modalidad es congruente con estudios previos en cada campo, con oscilaciones en beta acompañando a predicciones en la modalidad visual, y frecuencias más bajas predominando en la modalidad auditiva (Bastos et al., 2015; Fontolan et al., 2014). Sin embargo, la caracterización temporal de dichos efectos resulta más sorprendente. En primer lugar, las diferencias emergieron en ambos casos poco tiempo después de la presentación de la imagen, sugiriendo que la pre-activación ocurre tan pronto como la información previa permite la formulación de una predicción detallada a nivel sensorial. Sin embargo, estas diferencias no se mantuvieron en el caso visual, lo que podría indicar la presencia de representaciones "silenciosas" (Stokes, 2015). En los bloques auditivos las diferencias sí se mantuvieron, pero la implicación de una amplia red frecuencial y cortical sugiere que nuestra manipulación incluyó procesos adicionales a los de predicción, probablemente atencionales.

Asimismo, la influencia de la predictibilidad temporal en la pre-activación sensorial resultó ser distinta para predicciones auditivas y predicciones visuales, siendo determinante en los primeros pero no en los segundos. La interacción entre la predictibilidad temporal y fonema esperado se encontró en el análisis de los campos evocados por la palabra: diferencias entre los fonemas esperados aparecían antes de la palabra solamente cuando el intervalo era variable. Éstas diferencias pueden reflejar sistemas de codificación de información específicos en cada modalidad: mientras que en el habla dicha codificación depende en gran medida de los aspectos temporales, el tiempo es irrelevante en la codificación ortográfica, que se basa únicamente en patrones lumínicos espaciales.

Sin embargo, incluso en condiciones de alta predictabilidad, con una validez del contexto previo del 100%, los efectos detectados fueron sutiles. En primer lugar, en los experimentos auditivos, la fuente que más contribuía al efecto detectado a nivel de sensor fue un pico en la zona parietal y en la banda frecuencial alfa, que interpretamos como un filtro atencional, más que como una fuente de pre-activación predictiva. El efecto en theta en la zona temporal superior podría haber sido insuficiente para generar un efecto en la superficie por sí solo. En segundo lugar, en los experimentos visuales, la evidencia estadística no fue muy grande, y no encontramos indicios de que las diferencias entre condiciones se mantuvieran hasta la presentación de la palabra. La sutileza de dichos efectos podrían explicar la ausencia de evidencias en la literatura previa, o podrían ser interpretados como falsos positivos. Además, podría haber interpretaciones alternativas. Por un lado, los efectos transitorios del bloque visual podrían reflejar un proceso de automático de selección léxica y activación ortográfica tras la presentación de la palabra, no relacionada con actividad preparatoria para la próxima percepción. Por otra parte, un procesamiento de nombramiento interno estratégico para la realización de la tarea podría explicar los resultados auditivos. Desde nuestro punto de vista, estas explicaciones no son probables. En primer lugar, cuestionarios post-experimento no mostraron ninguna evidencia de procesos de nombramiento estratégico. Por otro lado, los datos oscilatorios observados en los experimentos visuales tienen un marcado parecido con dinámicas atribuidas a procesos predictivos en estudios no-lingüísticos (Kilavik et al., 2013). En cualquier caso, más estudios son necesarios para descartar explicaciones alternativas, y para ligar directamente los procesos preparatorios con el procesamiento posterior.

En general, nuestros resultados muestran que la pre-activación sensorial puede observarse cuando la información previa permite la postulación de una hipótesis precisa, bien definida a nivel sensorial. Estos datos contribuyen al entendimiento del procesamiento predictivo en el campo del lenguaje y en la cognición en general, mostrando que las predicciones sensoriales son posibles cuando el contexto permite anticipar el precepto siguiente. Sin embargo, incluso en nuestro paradigma, con validez predictiva del contexto al 100%, los efectos de pre-activación fueron sutiles y difíciles de aislar, lo que puede explicar la ausencia de evidencias en la literatura. Al identificar las frecuencias oscilatorias que caracterizan dicha pre-activación, esperamos que nuestros resultados contribuyan a interpretar actividad neural previa a la palabra en otros experimentos, contribuyendo al entendimiento de los mecanismos neurales responsables del procesamiento predictivo en general.

Bibliography

Arnal L. H and Giraud A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16(7):390–398. doi: 10.1016/j.tics.2012.05.003.

Arnal L. H, Wyart V, and Giraud A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, 14 (6):797–801. doi: 10.1038/nn.2810.

Başar E, Başar-Eroglu C, Karakaş S, and Schürmann M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39(2):241–248. doi: 10.1016/S0167-8760(00)00145-8.

Bar M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7):280–289. doi: 10.1016/j.tics.2007.05.005.

Bar M. (2009). A cognitive neuroscience hypothesis of mood and depression. *Trends in Cognitive Sciences*, 13(11):456–463. doi: 10.1016/j.tics.2009.08.009.

Bastiaansen M. C, Linden M. v. d, Keurs M. t, Dijkstra T, and Hagoort P. (2005). Theta Responses Are Involved in Lexical—Semantic Retrieval during Language Processing. *Journal of Cognitive Neuroscience*, 17(3):530–541. doi: 10.1162/0898929053279469.

Bastos A, Usrey W. M, Adams R, Mangun G, Fries P, and Friston K. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, 76(4):695–711. doi: 10.1016/j.neuron.2012.10.038.

Bastos A, Vezoli J, Bosman C, Schoffelen J.-M, Oostenveld R, Dowdall J, De Weerd P, Kennedy H, and Fries P. (2015). Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron*, 85(2):390–401. doi: 10.1016/j.neuron.2014.12.018.

Bates D. (2010). lme4: Linear mixed-effects models using s4 classes. r package version 0.999375-33. *http://CRAN. R-project. org/package= lme4*.

Boemio A, Fromm S, Braun A, and Poeppel D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3):389. doi: 10.1038/nn1409.

Bourguignon M, Molinaro N, and Wens V. (2018). Contrasting functional imaging parametric maps: The mislocation problem and alternative solutions. *NeuroImage*, 169:200–211. doi: 10.1016/j.neuroimage.2017.12.033.

Brodeur M. B, Guérard K, and Bouras M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, 9(9):e106953. doi: 10.1371/journal.pone.0106953.

Brown H, Adams R. A, Parees I, Edwards M, and Friston K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4):411–427. doi: 10.1007/s10339-013-0571-3.

Brown S, Ngan E, and Liotti M. (2008). A Larynx Area in the Human Motor Cortex. *Cerebral Cortex*, 18(4):837–845. doi: 10.1093/cercor/bhm131.

Buzsáki G. (2002). Theta Oscillations in the Hippocampus. *Neuron*, 33(3):325–340. doi: 10.1016/S0896-6273(02)00586-X.

Canolty R. T, Edwards E, Dalal S. S, Soltani M, Nagarajan S. S, Kirsch H. E, Berger M. S, Barbaro N. M, and Knight R. T. (2006). High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science*, 313(5793):1626–1628. doi: 10.1126/science.1128115.

Carreiras M, Armstrong B. C, Perea M, and Frost R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, 18(2):90–98. doi: 10.1016/j.tics.2013.11.005.

Clark A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204. doi: 10.1017/S0140525X12000477.

Coltheart M. Dual route and connectionist models of reading: an overview, (2006). URL `http://www.ingentaconnect.com/content/ioep/clre/2006/00000004/00000001/art00002`.

Cope T. E, Sohoglu E, Sedley W, Patterson K, Jones P. S, Wiggins J, Dawson C, Grube M, Carlyon R. P, Griffiths T. D, Davis M. H, and Rowe J. B. (2017). Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nature Communications*, 8(1):2154. doi: 10.1038/s41467-017-01958-7.

Correia J. M, Jansma B. M. B, and Bonte M. (2015). Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions. *Journal of Neuroscience*, 35(45): 15015–15025. doi: 10.1523/JNEUROSCI.0977-15.2015.

Coull J. T and Nobre A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 18(18):7426–7435.

Coull J. T, Cotti J, and Vidal F. (2016). Differential roles for parietal and frontal cortices in fixed versus evolving temporal expectations: Dissociating prior from posterior temporal probabilities with fMRI. *NeuroImage*, 141(Supplement C): 40–51. doi: 10.1016/j.neuroimage.2016.07.036.

Dambacher M, Rolfs M, Göllner K, Kliegl R, and Jacobs A. M. (2009). Event-Related Potentials Reveal Rapid Verification of Predicted Visual Input. *PLOS ONE*, 4(3): e5047. doi: 10.1371/journal.pone.0005047.

Dehaene S and Cohen L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6):254–262. doi: 10.1016/j.tics.2011.04.003.

DeLong K. A, Urbach T. P, and Kutas M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8):1117–1121. doi: 10.1038/nn1504.

Dikker S and Pylkkänen L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*, 127(1):55–64. doi: 10.1016/j.bandl.2012.08.004.

Duchon A, Perea M, Sebastián-Gallés N, Martí A, and Carreiras M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45 (4):1246–1258. doi: 10.3758/s13428-013-0326-1.

Edwards M. J, Adams R. A, Brown H, Pareés I, and Friston K. J. (2012). A Bayesian account of 'hysteria'. *Brain*, 135(11):3495–3512. doi: 10.1093/brain/aws129.

Ehrlich S. F and Rayner K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20 (6):641–655. doi: 10.1016/S0022-5371(81)90220-6.

Engel A. K and Fries P. (2010). Beta-band oscillations — signalling the status quo? *Current Opinion in Neurobiology*, 20(2):156–165. doi: 10.1016/j.conb.2010.02.015.

Flanagan J. R, Vetter P, Johansson R. S, and Wolpert D. M. (2003). Prediction Precedes Control in Motor Learning. *Current Biology*, 13(2):146–150. doi: 10.1016/S0960-9822(03)00007-1.

Fontolan L, Morillon B, Liegeois-Chauvel C, and Giraud A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nature Communications*, 5:4694. doi: 10.1038/ncomms5694.

Friederici A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, 91(4):1357–1392. doi: 10.1152/physrev.00006.2011.

Fries P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480. doi: 10.1016/j.tics.2005.08.011.

Friston K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):815–836. doi: 10.1098/rstb.2005.1622.

Fruchter J, Linzen T, Westerlund M, and Marantz A. (2015). Lexical Preactivation in Basic Linguistic Phrases. *Journal of Cognitive Neuroscience*, 27(10):1912–1935. doi: 10.1162/jocn_a_00822.

Giraud A.-L and Poeppel D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517. doi: 10.1038/nn.3063.

Gross J, Baillet S, Barnes G. R, Henson R. N, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld R, Parkkonen L, Taylor J. R, van Wassenhove V, Wibral M, and Schoffelen J.-M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, 65(Supplement C):349–363. doi: 10.1016/j.neuroimage.2012.10.001.

Hagoort P. Chapter 28 - MUC (Memory, Unification, Control): A Model on the Neurobiology of Language Beyond Single Word Processing. In Hickok G and Small S. L, editors, *Neurobiology of Language*, pages 339–347. Academic Press, San Diego, (2016). ISBN 978-0-12-407794-2. doi: 10.1016/B978-0-12-407794-2.00028-6. URL `https://www.sciencedirect.com/science/article/pii/B9780124077942000286`.

Hanslmayr S, Staudigl T, and Fellner M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, 6. doi: 10.3389/fnhum.2012.00074.

Helmholtz H. *Handbuch der physiologischen optik.*, volume 3. (English trans., Southall JPC, Ed.), Dover, New York, (1860).

Hickok G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6):393–402. doi: 10.1016/j.jcomdis.2012.06.004.

Hickok G and Poeppel D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1): 67–99. doi: 10.1016/j.cognition.2003.10.011.

Hickok G and Poeppel D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402. doi: 10.1038/nrn2113.

Huettig F and Mani N. (2015). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, 0(0):1–13. doi: 10.1080/23273798.2015.1072223.

Hämäläinen M. S and Hari R. (2002). Magnetoencephalographic (MEG) characterization of dynamic brain activation. *Brain mapping: The methods*, pages 227–255.

Hämäläinen M. S and Ilmoniemi R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1): 35–42. doi: 10.1007/BF02512476.

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

Ikkai A and Curtis C. E. (2008). Cortical Activity Time Locked to the Shift and Maintenance of Spatial Attention. *Cerebral Cortex*, 18(6):1384–1394. doi: 10.1093/cercor/bhm171.

Jensen O and Hesse C. Estimating distributed representations of evoked responses and oscillatory brain activity. In Hansen P, Kringelbach M, and Salmelin R, editors, *MEG: An Introduction to methods*, pages 156–185. Oxford university press, Oxford, (2010). ISBN 978-0-19-530723-8.

Jensen O and Mazaheri A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience*, 4:186. doi: 10.3389/fnhum.2010.00186.

Jobard G, Crivello F, and Tzourio-Mazoyer N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage*, 20(2): 693–712. doi: 10.1016/S1053-8119(03)00343-4.

Joukal M. Anatomy of the Human Visual Pathway. In *Homonymous Visual Field Defects*, pages 1–16. Springer, Cham, (2017). doi: 10.1007/978-3-319-52284-5_1.

Jung T.-P, Makeig S, Humphries C, Lee T.-W, McKEOWN M. J, Iragui V, and Sejnowski T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(02):163–178. doi: null.

Kerkoerle T. v, Self M. W, Dagnino B, Gariel-Mathis M.-A, Poort J, Togt C. v. d, and 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. doi: 10.1073/pnas.1402773111.

Kilavik B. E, Zaepffel M, Brovelli A, MacKay W. A, and Riehle A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology*, 245: 15–26. doi: 10.1016/j.expneurol.2012.09.014.

Kim A and Lai V. (2012). Rapid Interactions Between Lexical Semantic and Word Form Analysis During Word Recognition in Context: Evidence from Erps. *J. Cognitive Neuroscience*, 24(5):1104–1112. doi: 10.1162/jocn_a_00148.

Klimesch W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2):169–195. doi: 10.1016/S0165-0173(98)00056-3.

Kok P, Jehee J. M, and de Lange F. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, 75(2):265–270. doi: 10.1016/j.neuron.2012.04.034.

Kok P, Mostert P, and de Lange F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1705652114.

Kumar S, Sedley W, Barnes G. R, Teki S, Friston K. J, and Griffiths T. D. (2014). A brain basis for musical hallucinations. *Cortex*, 52:86–97. doi: 10.1016/j.cortex.2013.12.002.

Kuperberg G. R and Jaeger T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1):32–59. doi: 10.1080/23273798.2015.1102299.

Kutas M and Hillyard S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947):161–163. doi: 10.1038/307161a0.

Kveraga K, Ghuman A. S, and Bar M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65(2):145–168. doi: 10.1016/j.bandc.2007.06.007.

Lerma-Usabiaga G, Carreiras M, and Paz-Alonso P. M. (2018). Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proceedings of the National Academy of Sciences*, 115(42): E9981–E9990. doi: 10.1073/pnas.1803003115.

Lewis A. G and Bastiaansen M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 68:155–168. doi: 10.1016/j.cortex.2015.02.014.

Lewis A. G, Wang L, and Bastiaansen M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*. doi: 10.1016/j.bandl.2015.01.003.

Liberman A. M and Mattingly I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36. doi: 10.1016/0010-0277(85)90021-6.

Lochmann T and Deneve S. (2011). Neural processing as causal inference. *Current Opinion in Neurobiology*, 21(5):774–781. doi: 10.1016/j.conb.2011.05.018.

MacSweeney M, Capek C. M, Campbell R, and Woll B. (2008). The signing brain: the neurobiology of sign language. *Trends in Cognitive Sciences*, 12(11):432–440. doi: 10.1016/j.tics.2008.07.010.

Makeig S, Debener S, Onton J, and Delorme A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8(5):204–210. doi: 10.1016/j.tics.2004.03.008.

Maris E and Oostenveld R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190. doi: 10.1016/j.jneumeth.2007.03.024.

Marslen-Wilson W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1):71–102. doi: 10.1016/0010-0277(87)90005-9.

Marslen-Wilson W. D and Welsh A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1):29–63. doi: 10.1016/0010-0285(78)90018-X.

Mathewson K. E, Prudhomme C, Fabiani M, Beck D. M, Lleras A, and Gratton G. (2012). Making Waves in the Stream of Consciousness: Entraining Oscillations in EEG Alpha and Fluctuations in Visual Awareness with Rhythmic Visual Stimulation. *Journal of Cognitive Neuroscience*, 24(12):2321–2333. doi: 10.1162/jocn_a_00288.

McClelland J. L and Elman J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1):1–86. doi: 10.1016/0010-0285(86)90015-0.

McClelland J. L and Rumelhart D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5):375–407. doi: 10.1037/0033-295X.88.5.375.

Meyer L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7):2609–2621. doi: 10.1111/ejn.13748.

Molinaro N and Lizarazu M. (2018). Delta(but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, pages n/a–n/a. doi: 10.1111/ejn.13811.

Molinaro N and Monsalve I. F. (2018). Perceptual facilitation of word recognition through motor activation during sentence comprehension. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 108:144–159. doi: 10.1016/j.cortex.2018.07.001.

Molinaro N, Barraza P, and Carreiras M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, 72:120–132. doi: 10.1016/j.neuroimage.2013.01.031.

Molinaro N, Monsalve I. F, and Lizarazu M. (2016). Is there a common oscillatory brain mechanism for producing and predicting language? *Language, Cognition and Neuroscience*, 31(1):145–158. doi: 10.1080/23273798.2015.1077978.

Monsalve I. F, Pérez A, and Molinaro N. (2014). Item parameters dissociate between expectation formats: a regression analysis of time-frequency decomposed EEG data. *Frontiers in Psychology*, 5. doi: 10.3389/fpsyg.2014.00847.

Morillon B, Hackett T. A, Kajikawa Y, and Schroeder C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology*, 31(Supplement C):230–238. doi: 10.1016/j.conb.2014.12.005.

Nichols T. E and Holmes A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1):1–25. doi: 10.1002/hbm.1058.

Nieuwland M. S, Politzer-Ahles S, Heyselaar E, Segaert K, Darley E, Kazanina N, Von Grebmer Zu Wolfsthurn S, Bartolozzi F, Kogan V, Ito A, Mézière D, Barr D. J, Rousselet G. A, Ferguson H. J, Busch-Moreno S, Fu X, Tuomainen J, Kulakova E, Husband E. M, Donaldson D. I, Kohút Z, Rueschemeyer S.-A, and Huettig F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7. doi: 10.7554/eLife.33468.

Oostenveld R, Fries P, Maris E, and Schoffelen J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011:e156869. doi: 10.1155/2011/156869.

Park H, Ince R. A, Schyns P, Thut G, and Gross J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, 25(12):1649–1653. doi: 10.1016/j.cub.2015.04.049.

Peelle J. E, Gross J, and Davis M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, 23(6):1378–1387. doi: 10.1093/cercor/bhs118.

Penolazzi B, Hauk O, and Pulvermüller F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74(3):374–388. doi: 10.1016/j.biopsycho.2006.09.008.

Price C. J and Devlin J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6):246–253. doi: 10.1016/j.tics.2011.04.001.

Pylkkänen L and Marantz A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, 7(5):187–189. doi: 10.1016/S1364-6613(03)00092-5.

R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, (2013). URL `http://www.R-project.org/`.

Ramón y Cajal S. *Textura del Sistema Nervioso del Hombre y de los Vertebrados*, volume 2. Madrid Nicolas Moya, (1904).

Rao R. P. N and 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. doi: 10.1038/4580.

Rauschecker J. P and Scott S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724. doi: 10.1038/nn.2331.

Rayner K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372–422.

Roll M, Söderström P, Frid J, Mannfolk P, and Horne M. (2017). Forehearing words: Pre-activation of word endings at word onset. *Neuroscience Letters*. doi: 10.1016/j.neulet.2017.08.030.

SanMiguel I, Saupe K, and 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. doi: 10.3389/fnhum.2013.00407.

Schoffelen J. M, Hulten A, Lam N, Marquand A, Udden J, and Hagoort P. (2017). Frequency-specific directed interactions in the human brain network for language. *bioRxiv*, page 108753. doi: 10.1101/108753.

Schroeder C. E and Lakatos P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18. doi: 10.1016/j.tins.2008.09.012.

Schubotz R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences*, 11(5):211–218. doi: 10.1016/j.tics.2007.02.006.

Schurz M, Sturm D, Richlan F, Kronbichler M, Ladurner G, and Wimmer H. (2010). A dual-route perspective on brain activation in response to visual words: Evidence for a length by lexicality interaction in the visual word form area (VWFA). *NeuroImage*, 49(3):2649–2661. doi: 10.1016/j.neuroimage.2009.10.082.

Scott S. K and Johnsrude I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2):100–107. doi: 10.1016/S0166-2236(02)00037-1.

Sedley W, Gander P. E, Kumar S, Kovach C. K, Oya H, Kawasaki H, Howard M. A, and Griffiths T. D. (2016). Neural signatures of perceptual inference. *eLife*, 5: e11476. doi: 10.7554/eLife.11476.

Sereno S. C, Brewer C. C, and O'Donnell P. J. (2003). Context Effects in Word Recognition: Evidence for Early Interactive Processing. *Psychological Science*, 14 (4):328–333. doi: 10.1111/1467-9280.14471.

Singer W, Engel A. K, Kreiter A. K, Munk M. H. J, Neuenschwander S, and Roelfsema P. R. (1997). Neuronal assemblies: necessity, signature and detectability. *Trends in Cognitive Sciences*, 1(7):252–261. doi: 10.1016/S1364-6613(97)01079-6.

Sohoglu E, Peelle J. E, Carlyon R. P, and Davis M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, 32(25):8443–8453. doi: 10.1523/JNEUROSCI.5069-11.2012.

Sohoglu E, Peelle J. E, Carlyon R. P, and Davis M. H. (2014). Top-Down Influences of Written Text on Perceived Clarity of Degraded Speech. *Journal of Experimental Psychology. Human Perception and Performance*, 40(1):186–199. doi: 10.1037/a0033206.

Spaak E, Bonnefond M, Maier A, Leopold D, and Jensen O. (2012). Layer-Specific Entrainment of Gamma-Band Neural Activity by the Alpha Rhythm in Monkey Visual Cortex. *Current Biology*, 22(24):2313–2318. doi: 10.1016/j.cub.2012.10.020.

Staub A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, 9(8):311–327. doi: 10.1111/lnc3.12151.

Stokes M, Thompson R, Nobre A. C, and Duncan J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences*, 106(46):19569–19574. doi: 10.1073/pnas.0905306106.

Stokes M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, 19(7):394–405. doi: 10.1016/j.tics.2015.05.004.

Summerfield C and Egner T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9):403–409. doi: 10.1016/j.tics.2009.06.003.

Summerfield C and Egner T. (2016). Feature-Based Attention and Feature-Based Expectation. *Trends in Cognitive Sciences*, 0(0). doi: 10.1016/j.tics.2016.03.008.

Swanson L. R. (2016). The Predictive Processing Paradigm Has Roots in Kant. *Frontiers in Systems Neuroscience*, 10. doi: 10.3389/fnsys.2016.00079.

Tallon-Baudry C. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3(4):151–162. doi: 10.1016/S1364-6613(99)01299-1.

Taulu S and Simola J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in medicine and biology*, 51(7):1759.

Taulu S, Simola J, and Kajola M. (2005). Applications of the Signal Space Separation Method. *IEEE Transactions on Signal Processing*, 53(9):3359–3372. doi: 10.1109/TSP.2005.853302.

Toscano J. C, McMurray B, Dennhardt J, and Luck S. J. (2010). Continuous Perception and Graded Categorization Electrophysiological Evidence for a Linear Relationship Between the Acoustic Signal and Perceptual Encoding of Speech. *Psychological Science*, 21(10):1532–1540. doi: 10.1177/0956797610384142.

Van Berkum J. J. A, Brown C. M, Zwitserlood P, Kooijman V, and Hagoort P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3):443–467.

Van Veen B. D, Van Drongelen W, Yuchtman M, and Suzuki A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on biomedical engineering*, 44(9):867–880.

von Stein A and Sarnthein J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology*, 38(3):301–313. doi: 10.1016/S0167-8760(00)00172-0.

Wang L, Zhu Z, and Bastiaansen M. (2012). Integration or Predictability? A Further Specification of the Functional Role of Gamma Oscillations in Language Comprehension. *Frontiers in Psychology*, 3. doi: 10.3389/fpsyg.2012.00187.

Wang X.-J. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiological Reviews*, 90(3):1195–1268. doi: 10.1152/physrev.00035.2008.

Ward L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12):553–559. doi: 10.1016/j.tics.2003.10.012.

Warren R. M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science*, 167(3917):392–393. doi: 10.1126/science.167.3917.392.

Weiss S and Mueller H. M. (2012). "Too Many betas do not Spoil the Broth": The Role of Beta Brain Oscillations in Language Processing. *Frontiers in Psychology*, 3. doi: 10.3389/fpsyg.2012.00201.

Wicha N. Y. Y, Moreno E. M, and Kutas M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, 16(7):1272–1288. doi: 10.1162/0898929041920487.

Wöstmann M, Herrmann B, Maess B, and Obleser J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, 113(14):3873–3878. doi: 10.1073/pnas.1523357113.

Wöstmann M, Lim S.-J, and Obleser J. (2017). The Human Neural Alpha Response to Speech is a Proxy of Attentional Control. *Cerebral Cortex*, 27(6):3307–3317. doi: 10.1093/cercor/bhx074.

Yokosawa K, Pamilo S, Hirvenkari L, Hari R, and Pihko E. (2013). Activation of Auditory Cortex by Anticipating and Hearing Emotional Sounds: An MEG Study. *PLoS ONE*, 8(11):e80284. doi: 10.1371/journal.pone.0080284.

Zatorre R. J, Belin P, and Penhune V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1):37–46. doi: 10.1016/S1364-6613(00)01816-7.

Zylberberg A, Dehaene S, Mindlin G. B, and Sigman M. (2009). Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model. *Frontiers in Computational Neuroscience*, 3. doi: 10.3389/neuro.10.004.2009.