

# **Is there a common oscillatory brain mechanism for producing and predicting language?**

Nicola Molinaro<sup>1,2</sup>, Irene F. Monsalve<sup>1</sup>, Mikel Lizarazu<sup>1</sup>

<sup>1</sup>*BCBL, Basque center on Cognition, Brain and Language, Donostia/San Sebastian, Spain*

<sup>2</sup>*Ikerbasque, Basque Foundation for Science, Bilbao, Spain*

Corresponding author:

Nicola Molinaro

BCBL, Basque center on Cognition, Brain and Language

Paseo Mikeletegi, 69, 2º

20009, San Sebastian/Donostia

Spain

email: [n.molinaro@bcbl.eu](mailto:n.molinaro@bcbl.eu)

phone: +34 943 309 300

## **Abstract**

Recent proposals have suggested that language prediction is supported by the neurophysiological mechanisms involved in language production. Both prediction and production in language imply information processing percolating down from abstract semantic representations to lower-level processing steps, either for articulation (action) or active sensation (perception). Language production studies have repeatedly reported desynchronization of oscillatory beta power (13-30 Hz) over the left frontal cortex during word generation. Crucially, predictive coding theories propose that the beta frequency channel mediates top-down propagation of information during prediction. The present study evaluates initial experimental evidence on pre-stimulus activity during speech production and discusses the similar oscillatory dynamics involved in preparation for perception of words. We try to better characterize what processing dynamics the pre-stimulus beta-band activity represents, illustrating with some results from our lab. This evidence motivates the need for more fine-grained psycholinguistic paradigms to better characterize whether prediction and production are supported by similar neurophysiological mechanisms.

**Running Head:** Producing and predicting by beta?

**Keywords:** Prediction, Language production, Language comprehension, Beta band activity, Electrophysiology

Meaning-to-form mapping is the processing skeleton for language production, in which the intention of communicating a message translates into the top-down information processing flow that converts a semantic representation (through syntactic, lexical and phonological intermediate representations) into an efferent motor command that drives articulation. On the other hand, language comprehension follows similar intermediate processing steps but in the opposite direction: the perceived form of a linguistic input is fed-forward to higher-level processing steps (orthographic/phonological, lexical, syntactic) in order to transmit the intended (semantic) message of the speaker.

— Please insert Glossary around here —

Interestingly, recent proposals have identified a common processing mechanism in the service of both production and comprehension, namely prediction. Pickering and Garrod (2007, 2013) suggest that the comprehension system might use the production system in order to make predictions. Pickering and Garrod (2013) illustrate this link, focusing on forward modeling of an action.

*We use the simple example of moving a hand to a target. The actor formulates the action (motor) command to move the hand. This command initiates two processes in parallel. First, it causes the action implementer to generate the act, which in turn leads the perceptual implementer to construct a percept of the experience of moving the hand. Second, it sends an efference copy of the action command to cause the forward action model to generate the predicted act of moving the hand. (page 333).*

Action monitoring is the constant comparison between the predicted sensation deriving from the forward model and what “the perceptual implementer” perceives on-line. Language production involves such internal monitoring processes that on-line compare the articulation act with the predicted model of the same motor act at the sensory level. In consequence, a prediction of the expected articulatory output is sent to the sensory system predicting what will be heard. During language processing, an addressee could possibly employ the same mechanisms (implementation of the articulatory command at the sensory level), as the basic processing algorithm through which the properties of the incoming linguistic input are anticipated. This would make speaker-addressee synchronization optimal and communication more efficient.

Dell & Chang (2014) propose that there is a direct link between the two sides of the “language coin” (i.e., production and comprehension). In their P-chain, they acknowledge that

language processing largely involves prediction processes. They differ from Pickering and Garrod (2007, 2013) in that they do not make any claims about the possibly motor nature of the prediction mechanisms. Importantly however, they claim that prediction *is* production, based on the fact that both involve top-down processes from intended meaning to linguistic forms. Chang, Dell and Bock (2006) computationally implement the idea that production abilities arise directly from learning to predict; in other words, without an efficient predictive system developed during language exposure, based on implicit learning, children would not be able to develop proper language production skills (Mani & Huettig, 2012).

Despite the relevance of these proposals, the neurocognitive mechanisms on which prediction relies during language processing and whether these possibly link perception and production are still poorly understood. In fact, classic behavioral studies have not been able to provide evidence about how our mind prepares for future events (either for articulation or for perception), since they focus on stimulus reactions. Recent eye-tracking studies (e.g., Altmann & Kamide, 1999) have provided initial evidence; however, the paradigms employed have important limitations (see Huettig, 2015, for a discussion). In addition, most neuroimaging language research on comprehension still employs the traditional paradigm focusing on post-stimulus brain activity. In recent theoretical approaches to prediction (see Predictive coding below) post-stimulus activity most likely represents integration and learning processes (Friston, 2005); the inferences that can be made about predictive processing are thus “contaminated” by additional adaptive strategies.

The aim of this manuscript is to discuss the possible common neurocognitive mechanisms supporting production and prediction (during comprehension). Based on recent proposals (Pickering & Garrod, 2007, 2013), the neurocognitive architecture supporting production (and more specifically production monitoring) should also support prediction. We will first introduce the predictive coding framework. Then, we will present the on-going debate about the role of prediction in language processing. Based on these theoretical premises we will discuss neurophysiological evidence from studies of the brain dynamics reflecting preparation for articulation in language production. We will mainly focus on high temporal resolution techniques. Based on recent experimental evidence from our laboratory, we will evaluate the neurophysiological mechanisms involved in predictive processing during language comprehension. Finally, we will advance theoretical inferences from the experimental evidence discussed.

## **Predictive coding**

Before advancing in further discussion, a clarification of how the term “prediction” is employed in this review must be provided, given the considerable degree of heterogeneity in the use of this term.

Predictive processing here refers to predictive *coding* that qualitatively differs from predictive *timing* (Arnal, 2012; Arnal & Giraud, 2012; Arnal, Doelling & Poeppel, 2014). Predictive timing refers to the phenomenon by which sensory regions typically show increased excitability in view of whatever event is going to happen in a specific modality (i.e., predicting “when”). In this paper, we mainly focus on the predicting “what”.

Predictive coding is typically framed as reflecting a process that originates from abstract internal representations driving the activation of sensory representations in the processing periphery (Figure 1). The *predictive coding* approach (Clark, 2013; Rao & Ballard, 1999) details the computations (grounded in Bayesian hierarchical inference theory, see Friston, 2005) involved in predictive processing. A core notion of this framework posits an asymmetry in the flow of information across the processing hierarchy: top-down propagation of information would activate representations of expected sensory input generated from internal generative models, while only the portion of sensory data not accounted for by the prediction at each stage - the *prediction error* - would be propagated forward.

In more detail (see Glossary), pre-activation (i.e., prediction) “spreads” across many different hierarchical substeps, converting internal abstract Forward Models into pre-stimulus Sensory Predictions (see Figure 1; Bubic, von Cramon & Schubotz, 2010; Friston, 2005; Poeppel, Idsardi & van Wassenhove, 2008). Based on the information represented in the Forward Model, the brain communicates top-down with more peripheral sensory regions dedicated to the interaction with the environment. In some proposals (Arnal, 2012; Pickering & Garrod, 2007; Schubotz, 2007), these sensory representations are typically employed to monitor on-line the outcome of an internally generated motor command. However, they are “available” for the perception/comprehension system even when there is no motor activity involved.

Prediction by definition is ‘prone to error’, since it can either match or mismatch with the physical properties of the incoming stimulus (processed in sensory regions). This determines the amount of Prediction Error (Figure 1). Prediction Error explains activation that propagates bottom-up to higher processing levels. Such ascending activation would be used to update internal models and subsequent predictions, until the prediction error is totally “explained away” and assimilated throughout the processing hierarchy (Friston, 2005). Prediction Error activity would propagate across hierarchically more general processing stages and cortical areas in the opposite direction to the prediction: from sensory to association areas of the cortex. Importantly, Prediction Error accounts for post-stimulus brain activity in a more biologically plausible way compared to bottom-up frameworks, following the principle of minimizing the brain’s free energy (for discussions see Friston, 2005; Rao & Ballard, 1999). As a consequence, matching information has enormous

processing advantages for all the multi-level perception and integration processing hierarchy. However, mismatching information has an even greater value for readjusting (Adaptation in Figure 1) the internal source representations of the environment based on external pieces of information (i.e., Learning, see Dell & Chang, 2014; Friston, 2005).

In this paper, we will focus on our claim that Prediction is reflected in *pre-stimulus* brain activity, while Learning (driven by the assimilation of the Prediction Error along the processing hierarchy) is reflected in *post-stimulus* activity. As a consequence, properties of the predicted stimulus should be decodable from the pre-stimulus modulations of brain activity.

— Please insert Figure 1 around here —

### **The role of prediction in language processing**

The mechanisms through which predictive processing affects language comprehension are still under dispute. Indeed, some authors have doubts about the existence of predictive processing in language.

For example, Jackendoff (2002) proposed that it would make much more sense for our brain to consume neurocognitive resources in the post-stimulus analysis of an input, since failed predictions are very common in natural language (in fact, one important property of humans is the ability to communicate creatively avoiding redundancies; Hockett, 1960). Predicting all the time during comprehension would result in constant processing costs when the input does not match the predictions. This would require a large amount of cognitive resources because predictions are more likely to be wrong than right. Consequently, the costs of predicting would outweigh the benefits (see also Huettig, 2015). In addition, given the arbitrary relation between form and meaning in the language system (words such *ant* and *mosquito* are semantically related but not form related), a semantic neighborhood would not directly map onto an orthographic one, and pre-activation of the visual features of one word would be of no benefit when processing conceptually similar items. In support of such criticism, Van Petten and Luka (2012) and Huettig (2015) indicate that there is no robust evidence in favor of the prediction hypothesis.

Even among the supporters of the prediction hypothesis, an unsolved question is what representations (only abstract representations, or sensory information) are pre-activated and in which conditions. In fact, some authors assume that prediction is an essential processing mechanism for comprehension (Levy, 2008; Dikker & Pylkkanen, 2013; see also the analysis-by-synthesis approach by Poeppel et al., 2008), while a more conservative view does not consider prediction a fundamental principle of language processing (Huettig, 2015).

However, if we accept the predicting coding view that considers the brain as a highly adaptive prediction machine, capable of constantly building up predictions based on available contextual information (Bar, 2007; Friston, 2005; Poeppel et al., 2008), why should this mechanism be “blocked” during language processing? In predictive coding, predictions at different levels of the neurocognitive hierarchy are “nested” (from being more abstract to more concrete) with different levels of uncertainty related to a Forward Model. Critically, prediction would *always* percolate down to early sensory areas, so that some kind of Sensory Prediction would always be generated. However, the amount of contextual information determines the uncertainty of the internal Forward Models and, in turn, the quality of the Sensory Prediction that can be either general (i.e., *I predict I will perceive a word*) or more specific (i.e., *I predict I will perceive the word “dog”*).

From experimental studies, initial (although not conclusive) indications of the existence of prediction processes have been reported.

1. *Predictions matter in language comprehension* (Altmann & Kamide, 1999, Kutas & Hillyard, 1984). A well-established effect reported in the language comprehension literature concerns the fact that contextual information (e.g., a highly constraining sentence fragment) can facilitate the recognition of a word, as evidenced by faster eye gaze in reading or a smaller N400 component (a negative-going event-related voltage deflection (ERP) that peaks around 400 milliseconds post-stimulus onset in the scalp-recorded electroencephalographic signal). Classically, this effect has been interpreted as reflecting modulation of lexical-semantic retrieval and integration (Kutas & Hillyard, 1984). Recently, the N400 modulation has been interpreted as reflecting facilitated recognition of a word as a result of contextual pre-activation of some of its lexical/semantic features (DeLong et al., 2005). Despite the fact that this hypothesis has many supporters, the N400 is a post-target word onset effect and can only be indirectly linked to predictive processes (through theoretical assumptions). Post-stimulus brain activity is likely to reflect retrospective use of information regarding previous events (see Prediction Error in Figure 1, for a similar proposal on post-stimulus ERPs, Friston, 2005). This observable, stable effect temporarily follows predictive processing and can represent adaptive strategies. Thus, the questions remain: are internal Forward Models translated into Sensory Predictions during language processing? And if so, in which language scenarios?

2. *Information originating from the internal forward model can lead to pre-activation of low-level linguistic information*. DeLong and colleagues (2005) presented words in high constraining sentence contexts that differed in their initial phonemes (a vowel or a consonant, e.g., *kite* vs. *airplane*) and in expectancy levels (higher for *kite* compared to *airplane*). The authors

reported larger N400 for the less expected word (*airplane*), but more importantly, they found the same ERP modulation in the article that preceded the noun and that varied depending on the initial phoneme of the expected word (*an* for *airplane* elicited a larger negative effect than *a* for *kite*): in other words, the N400 amplitude changed depending on whether or not the article was appropriate for the highly expected word. This result indicates that low-level phonological features of the expected word were pre-activated by the semantic constraints imposed by the sentence context, since the expectancy effect was measured on a word (*a/an*) that does not provide any semantic (but does provide phonological) information concerning the following expected word (*kite/airplane*). This effect indicates that low-level features of an expected word are pre-activated before target word presentation. However, does prediction involves word-form sensory representations? And which neurophysiological mechanisms are recruited for language prediction?

3. *Differential activation based on the strength of prediction can be observed in peripheral sensory regions.* A small number of studies (Dikker & Pylkkanen, 2013; Molinaro, Barraza, & Carreiras, 2013a; Piai et al., 2014) employing neuroimaging techniques (electroencephalography, EEG and magnetoencephalography, MEG) with high temporal resolution have unveiled the preparatory dynamics of the brain activity immediately before a target event. Furthermore, MEG provides information about the source of brain activity with good spatial resolution (~2 cm). In a picture-word matching paradigm, Dikker and Pylkkanen (2013) found increased MEG evoked activity in the visual cortex in the milliseconds before the presentation of a word when the previous (related) picture represented a single object compared to a group of objects. Importantly, in these predictable conditions the single object pre-activated a single word (and, possibly, its word form), while the group of objects pre-activated the overall semantic category (but not a specific word form). This evidence partially supports the predictive view, since it shows that in specific experimental conditions prediction leads to pre-stimulus activity in sensory regions. Nonetheless, this effect only indicates that visual regions are more active for the less complex condition (in which a single object preceded the word) and may represent more general (attentional) preparatory processes that do not carry information about the perceptual properties of the predicted word. So, does this sensory peripheral effect reflect prediction of linguistic input?

More importantly for the present discussion, practically no studies so far have directly focused on the most relevant moment of time at which predictions are generated, i.e. before stimulus appearance (*before time zero*). Recently, Boylan, Trueswell and Thompson-Schill (2014) followed this rationale in an fMRI study in which they studied predictive processing of word form, depending on lexical-syntactic category (noun vs. verb). By focusing on the brain volumes of the



so-called visual word form area (Cohen et al., 2000), sampled during the time lag before the presentation of an expected stimulus (either a noun or a verb), they were able to successfully classify (based on functional data from left inferior occipito-temporal regions) trials in which participants were expecting a noun, from trials in which they were expecting a verb. This study shows the possibility of decoding word form representations from pre-stimulus activity. However, these findings do not provide evidence about the time-course of neural activity through which predictive coding evolves up to the millisecond preceding presentation of an expected stimulus. To address this issue, high temporal resolution techniques with reasonable spatial brain resolution can provide critical experimental evidence. Techniques like MEG and EEG reflect the cortical activity of large populations of pyramidal cells. These measures have the great advantage of providing information about brain activity with the temporal resolution required to disentangle multiple successive and interactive cognitive events. This is extremely important in prediction studies, in which it is vital to track the temporal course of neurocognitive events until a stimulus appears. Specifically, the analysis of oscillatory brain activity has great potential to bridge the gap between different levels of explanation in neuroscience research. Recent proposals (Buszáki & Draguhn, 2004; Wang, 2010; Ward, 2003) identify in the rhythmic activity of the brain the basic unifying principle coordinating activity of single cells into larger populations of neurons eliciting both local field potentials and long-distance synchronization. Buszáki & Draguhn (2004) review how such network oscillations bias input selection, temporally linking neurons into assemblies, and facilitating synaptic plasticity. This, in turn, is associated to cognitive functions such as multisensory integration, working memory, and selective attention (Wang, 2010).

### **Oscillatory dynamics reflecting preparation for articulation**

Pre-motor activity for action execution is a fundamental aspect of speech production. Motor planning has been associated with modulation of oscillatory brain activity in the beta range (13-30 Hz; but also with the mu rhythm, 7.5-12.5 Hz, Gestaut, 1952). Beta activity inversely correlates with the likelihood of an action to be performed, with larger power decrease reflecting stronger motor readiness. In such conditions, less synchronized activity of such populations, across multiple different motor events, does not “sum up” and triggers less power compared to resting conditions in which such activity is more synchronous (see Glossary). Interestingly, beta band activity is not only observed when an action has to be performed, but also when it is expected or imagined (Hari et al., 1998). Jenkinson and Brown (2011) observe that beta band desynchronization reflects estimation of the probability of new processing demands, either in the motor or perceptual domain. Engel and Fries (2010) reviewed the role of beta band activity both within and outside the motor domain,

proposing that change in the processing mode is correlated with power decrease in the beta band, compared to either (i) conditions in which the current processing routines remain unaltered, and beta power does not change, or (ii) higher task demands for the same cognitive operations recruiting more cognitive resources, reflected in higher beta power. Beyond the motor domain, Siegel, Donner, Oostenveld, Fries & Engel (2008) observed beta band modulation across the visual dorsal attentional network (involving frontal, parietal and visual regions). Importantly, regional beta band desynchronization was related to better performance in a visual attention task, highlighting the role of beta band activity in top-down modulation of sensory processes. Finally, Hanslmayr, Satudigl and Fellner (2012) proposed the *information via desynchronization hypothesis*. Based on information theory, beta synchronization effects would be related to richness of information processed in the brain. In other words, “the more information that needs to be encoded, the more desynchronized the firing of local neural assemblies needs to be” (page 7). When processing complex heterogeneous motor commands, but also high-level abstract cognitive information, beta desynchronization could thus enable encoding and retrieval of such brain-stored representations.

Weiss and Mueller (2012) recently detailed four different roles of beta activity in language processing: beta activity would reflect the semantic and sensory-motor features of word stimuli; it would be involved in complex sentence-level operations; beta oscillations would play a crucial role in working memory operations; more importantly, beta activity would play a critical role in top-down mechanisms during language processing. In the present paper, we focus on the parallel between production and prediction in language, based on the dynamic principles of forward modeling a motor command. Given the relevance of beta band activity in the motor domain and the importance of motor planning for articulation, it is not surprising to observe that a number of studies have reported beta band effects related to speech production (Fisher et al., 2008; Singh, Barnes, Hillebrand, Forde & Williams, 2002). Recently, Findlay et al. (2012) employed a verb generation task in which participants were presented with a noun and had to produce a related verb. They observed reduced beta power in the left inferior frontal regions prior to the generation of the verb. Liljeström, Kujala, Stevenson & Salmelin (2015) analyzed the brain network configuration preceding picture naming. They did not focus on local power modulations, but analyzed the functional connectivity in a larger speech production network. After picture presentation (and before naming) they observed increased large-scale coherence in the beta-range for over production. They claimed that such an effect reflects selection processes during speech planning. These two studies highlight the important role of beta oscillations in language production, showing both power reduction (desynchronization) and increase of cortical connectivity in this frequency band. Jenson et al. (2014) required their participants to produce syllables and words both overtly and covertly. In

both conditions they observed reduced beta power before production, with weaker effects in covert production. Interestingly, pre-stimulus beta effects similar to covert production were observed also during a perceptual discrimination task. Thus, the authors concluded that pre-stimulus desynchronization in the beta range could represent early predictive coding based on internal forward modeling.

Recent language production studies have employed experimental paradigms typically exploited in comprehension experiments. In language comprehension, the processing of a target word is influenced by the contextual constraint of the previous sentence fragment (typically measured with cloze-probability, i.e., the proportion of participants who report a word as continuing a sentence fragment in a paper and pencil questionnaire; Kutas & Hillyard, 1984). Recent proposals suggest that it reflects lexical pre-activation during language comprehension in cases of highly predicted upcoming linguistic information (DeLong, Urbach & Kutas, 2005). Piai, Roelofs and Maris (2014) asked participants to name a picture after word-by-word visual presentation of a sentence that could present different degrees of semantic constraint (high or low) towards the expected picture. This study thus combines the classic sentence constraint paradigm and word production in a picture-naming paradigm. By measuring EEG, the authors observed relative power modulation in the beta (and also alpha, ~10 Hz) frequency range before picture presentation in left frontal electrodes, i.e., less beta power while participants read high-constraining contexts (preceding the picture) compared to low-constraining ones. Assuming the “prediction is production” account (Dell & Chang, 2014; see also Pickering & Garrod, 2007), it is possible that the beta modulation reported by Piai and colleagues (2014) does not only reflect stronger motor preparation of the (to be named) target word in the high constraining context, but also the development of a more detailed internal forward model (i.e., an acoustic representation) of that same word (as suggested by Jenson et al., 2014).

Klein et al. (2014; Wheat, Cornelissen, Frost, & Hansen, 2010) reported MEG beta desynchronization in left inferior frontal regions during the first 100 ms after presentation of a visual target word in a masked priming paradigm. The effect was stronger (i) when the target word was preceded by a masked letter string starting with the same initial phoneme and (ii) when participants were asked to name the word aloud as compared to a lexical decision task. The authors interpreted this effect as reflecting early access to articulatory representations during visual word recognition. These findings indicate how fast pre-articulatory activity (reflected in beta-band desynchronization) is during word production. More importantly, the beta power reduction was significant (but less sustained in time) in the lexical decision task in which no (overt or covert) production was required of the participants. In this study, the preceding masked prime string was presented ~90 ms before the target word. As a consequence, the prime influenced the processing of

the target word at ~200 ms. Masked priming paradigms assume that primes pre-activate features of the target stimulus. For example, in the P-chain (Dell & Chang, 2014) comprehension of a prime involves prediction (and leads to implicit learning, that is useful for diminishing prediction error during processing in similar future situations, see also Figure 1). The stronger beta effects observed in production by Klein et al. (2014) could thus reflect the additional pre-stimulus involvement of the articulatory brain circuits; however, the observation of similar (but reduced) beta band effects when no articulation is needed possibly shows the development of the internal forward model supporting both sensory prediction (during comprehension) and monitoring of articulation (during production). In addition, Klein et al.'s (2014) findings highlight the role of left lateral prefrontal brain regions in planning the internal forward models (see also Hickok & Poeppel, 2007, quoted in Pickering & Garrod, 2007).

### **Oscillatory dynamics reflecting pre-stimulus prediction**

Studies of language production have highlighted the relevant role of beta band in forward modeling future sensory information for articulatory monitoring purposes (Jenson et al., 2014). If such Forward Models are similarly recruited for predicting future sensory inputs (Pickering & Garrod, 2007, 2013), we should observe similar oscillatory effects for processing in language scenarios in which active predictive processing is at work that does not involve any motor or monitoring (of motor) activity.

In this vein, recent studies have provided preliminary neurophysiological evidence of beta-band activity reflecting top-down processing (for a review, Bastos et al., 2012). Bastos et al. (2015) analyzed the oscillatory dynamics of the primate visual system employing electrocorticography recordings from grids implanted along the whole visual system of monkeys. They observed that increased coherence in the beta band was associated with feedback influences from higher processing regions to primary visual regions. Beta band activity modulated the relative “functional hierarchy” between lower and higher-order regions during pre-stimulus and post-stimulus visual processing.

In an MEG study, Arnal, Wyart & Giraud (2011) found that violating intermodal speech expectations was accompanied by increased beta oscillatory activity in a multi-sensory brain region (temporal regions), and increased high gamma activity in early sensory regions (see also Fontolan, Morillon, Liegeois-Chauvel & Giraud, 2014). Frye, Wu, Liederman and Fisher (2010) reported that greater connectivity in the beta band from the left inferior frontal regions to other cortical areas was significantly related to better phonological performance in dyslexic readers. In addition, previous proposals (Bubic et al., 2010; den Ouden, Kok & de Lange, 2012; Rahnev, Lau & de Lange, 2011)

indicated that pre-frontal brain regions drive the recruitment of source information to develop predictions and could then show beta band modulations related to the development of forward models (as in language production studies) in the pre-stimulus time interval (Bastos et al., 2015). Overall, these findings highlight the role of the beta-band frequency channel in predictive processing: beta effects have been observed both in the domain of local power modulations (post-stimulus beta power increase, as in Arnal et al., 2011) and in the analysis of larger-scale brain connectivity (increase pre-stimulus beta connectivity, as in Bastos et al., 2015). However, if the internal forward models employed in language production are involved also in language prediction we should observe pre-stimulus beta power desynchronization during language comprehension (see also pre-stimulus beta synchronization associated to predictive *timing*, Arnal et al., 2014).

Desynchronization effects during prediction in natural language would be in line with the information via desynchronization hypothesis (Hanslmayr et al., 2012) in which retrieval of complex conceptual representations (as a consequence, for example, of a linguistic prediction) would correlate with beta desynchronization. These authors indicate that a larger amount of information would be better coded by a neural population in which single cells fire with a lower degree of synchrony. This is the case of natural language, in which complex semantic representations corresponding to each content word should be retrieved and combined with the overall discourse-level representation of the message. In these conditions, more desynchronization is expected for the retrieval of more semantically complex (information-rich) representations. In the prediction domain, the pre-activation of representationally more complex information would then determine less beta power (single neurons' firing rates are highly desynchronized) compared to both scenarios in which the predicted representation is "simpler" (without activating large-scale distributed representations in semantic memory) and scenarios in which prediction is "weaker" (i.e., there is more uncertainty concerning the predicted representation, possibly due to the competition of multiple functionally similar representations). In these latter conditions, either the predicted representation is less complex (less information to code across the neural population) or multiple (functionally similar) representations trigger overlapping firing rates (across cells), causing overall larger synchrony and increased beta power. Thus, both information-rich and "information-poor" stimuli (such as beats or basic visual objects, Arnal et al., 2014; and the references in Bastos et al., 2012) would correlate with beta-band modulations, but the amount of desynchronization would be larger for the prediction of more complex stimuli.

In our lab, we recently carried out a set of studies exploring language prediction during sentence comprehension. We employed specific linguistic constructions that permeate everyday language experience, i.e., Multi-Word Units (MWUs) which, depending on definitional criteria,

roughly constitute 60% of spoken and 55% of written language. MWUs are continuous or discontinuous sequences of words which are, or appear to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by language compositional processes. A variety of phrase chunks constitute this category, from collocations (*make a difference*) to discourse organizers (*on the other hand*) and social routine clichés (*nice to meet you*) among others.

In Molinaro et al. (2013a; Monsalve, Pérez & Molinaro, 2014) we recently evaluated, in an EEG sentence comprehension experiment, the correlates of reading a target word that followed a regular non-fixed semantic compositional context (Semantic condition, as in Kutas & Hillyard, 1984; DeLong et al., 2005) as compared to when the same target word “concluded” a MWU (MWU condition). These strings were selected based on the fact that the last word of the string was highly predictable after reading the string until the second-to-last word (cloze-probability higher than 85%). Thus, in both our experimental conditions the same specific target word was similarly highly expected (for details see Molinaro et al., 2013a).

Despite the levels of expectation (cloze-probability) being the same in Semantic and MWU, the former condition relies on rule-based compositional processes (the meaning of the previous sentence fragment has to be computed word by word to figure out what piece of information is coming next), while the latter condition relies on memory-based processes (recognition of the MWU that would not require rule-based computation). In a number of studies (Molinaro, Vespignani, Canal, Fonda & Cacciari, 2008; Molinaro, Canal, Vespignani, Pesciarelli & Cacciari, 2013b; Molinaro & Carreiras, 2010; Roehm, Bornkessel-Schlesewsky, Rösler & Schlesewsky, 2007; Vespignani, Molinaro, Canal, Fonda & Cacciari, 2010) this has been associated to the fact that readers, before the presentation of the target word in multi-word scenarios, can develop a “deterministic” expectation. On the other hand, in the semantic constraining conditions, the context would pre-activate a semantic field (i.e., a set of semantically-related candidate words) and this would determine a more “probabilistic” expectation for the incoming word stimuli. MWUs are thus ideal for studying lexical prediction during language comprehension, compared to the Semantic condition, in which multiple functionally similar (semantically-related) lexical representations compete for prediction (DeLong et al., 2005; Federmeier & Kutas, 1999). According to the predictive proposal discussed above, prediction during MWUs should be stronger and more specific compared to the Semantic condition (“weaker” prediction).

In Molinaro et al. (2013a) we asked our participants to read sentences for comprehension. Their task was to answer a yes/no comprehension question every five sentences on average. Comprehension questions appeared randomly throughout the whole experiment. In addition, the

YES/NO button positions appeared randomly either on the left (half of the times) or on the right (half of the times). This experimental paradigm does not imply any motor preparation, since participants have no time-pressure when answering the comprehension questions. Based on EEG data, we provided initial indications about predictive processing (and, importantly, pre-stimulus evidence) in a language paradigm involving sentence comprehension. However, the available EEG data cannot be directly compared to the production studies reporting beta band effects during language production. First, oscillatory activity in these studies was estimated focusing on phase synchronization (Phase Locking Value: Varela, Lachaux, Rodriguez & Martinerie, 2001), with the aim of studying the network dynamics involved in predictive processing. Language production studies, on the other hand, have mainly focused on power (and coherence) modulations of oscillatory brain activity. Second, previous studies on language production (see *Oscillatory dynamics reflecting preparation for articulation* above) provide inferences about the underlying brain regions involved in pre-articulatory activity mainly by focusing on MEG recordings. EEG recordings, however, do not offer comparable evidence concerning the underlying brain regions involved in specific computations. To better evaluate the reliability of the prediction/production parallel (Pickering & Garrod, 2007), it is important to identify the brain source of beta-band effects, that both production (Klein et al., 2014) and prediction studies (Rahnev et al., 2011) would identify in frontal brain regions.

In a MEG study, we employed the same experimental paradigm as that described in Molinaro et al. (2013a). Here, we report the analysis of data acquired from a group of 24 participants. Data acquired at the sensor level were analyzed focusing on amplitude modulations of oscillatory brain activity across a large spectrum of frequency bands (from 1 to 90 Hz, frequency resolution of 1 Hz) from 1 second preceding the target word to 600 ms post-stimulus onset. Compared to the EEG topographical distribution, MEG sensor-level distributions (and specifically the data collected with planar gradiometers) are more reliable concerning the possible brain source of an effect. Still, the effects we report here should be interpreted with caution, since the underlying brain sources can be inferred only at a macroscopic level (along the left-right hemispheric lateralization and anterior-posterior dimensions).

— Please insert Figure 2 around here —

In Figure 2 we present the two main effects (cluster-based corrected) that emerged at the sensor level (while considering a pre-sentence baseline in which the fixation point was displayed to the participant). In the time interval between  $-670$  to  $-360$  ms preceding the presentation of the predicted target word, we observed larger beta band desynchronization for the MWU condition

compared to the Semantic in the frequency range from 17 to 23 Hz. This beta desynchronization is anteriorly distributed, possibly reflecting a bilateral effect involving frontal brain regions, with a larger involvement of left-lateralized regions. This effect was followed by a larger alpha band (8-12 Hz) synchronization for the MWU compared to Semantic condition in the time interval corresponding to the detection of the target word (from -10 to 330 ms). Compared to the beta effect, the alpha synchronization shows a more posterior distribution (with a slight right-hemispheric lateralization).

The beta band desynchronization in the present study possibly reflects more detailed preparatory dynamics for the processing of the incoming stimulus in the MWU condition compared to the Semantic. Importantly, we observed effects in a similar frequency channel as that observed for production studies, providing initial indications that the oscillatory mechanisms supporting prediction employ a similar oscillatory mechanism to the one observed in production. Following Pickering and Garrod's (2007, 2013) proposal, the beta frequency channel would play a role in developing internal Forward Models of future sensory events. This explanation is supported by the following effect in the alpha band. Alpha band effects have been related to modulation of visual attention (Jensen, Gips, Bergmann & Bonnefond, 2014); typically, recruitment of visual-attention resources inversely correlates with alpha power (alpha synchronization). In our study, the higher visual attentional resources allocated to processing the incoming word in the Semantic condition imply a relative reduction of alpha power time-locked to the presentation of the target word in posterior brain regions. On the other hand, in the MWU condition, the prediction is so strong that fewer attentive resources are required to visually decode the expected stimulus.

In order to better estimate the relation between the two effects, we selected the time-frequency windows for each of the two effects described above and the sensor in which the effect showed its peak of activity. We then computed the Pearson correlation between the two arrays and we observed the expected inverse correlation between the beta and the alpha effect ( $R = -0.51$ ,  $p = 0.018$ ). This provides a clear indication that the preparatory dynamics observed in frontal brain regions for MWUs are related to the reduced visual attentional resource time-locked to the presentation of the target word. It should be underscored that these findings do not provide evidence for the underlying brain networks triggering these oscillatory effects. Nevertheless, based on the data discussed, we can sketch initial inferences about the time-frequency oscillatory dynamics involved in predictive processing.

### ***To be(ta) or not to be(ta): That is the question***

Preliminary findings from our lab thus provide initial evidence that the oscillatory dynamics supporting language prediction employ a similar frequency channel to the ones involved in



language production. Beta band activity has been reported in production studies and in studies manipulating the contextual constraints during sentence comprehension. These last findings are in line with predictive coding theories pointing to the crucial role of beta-band activity in top-down modulating of the perception of incoming stimuli. Beta band activity has been associated to top-down brain activity (Bastos et al., 2012, 2015) and the evidence for such effects in frontal regions could reflect high-level planning of subsequent peripheral interactions with the external environment (Clark, 2013; Pickering & Garrod, 2007).

This effect observed before the presentation of an expected stimulus fits with the Prediction framework illustrated in Figure 1. In this model, predictive processing develops before stimulus onset and is composed of multiple processing stages. Based on the development of an internal Forward Model, activation propagates downwards across the multiple levels of the processing hierarchy (Friston, 2005) and can trigger Sensory Predictions. Beta-band effects would mediate such top-down propagation of activation. In our sentence comprehension MEG experiment we interpret the pre-stimulus beta-band power decrease as the pre-activation of high-level abstract representations. The spatial distribution of the effect (possibly associated to frontal lobe activity) supports the proposed interpretation of the findings. Frontal regions would constitute a critical hub collecting abstract information preliminary to production/prediction. Previous studies (Bastos et al., 2015; Rahnev et al., 2011) in fact show that pre-frontal brain regions would drive the recruitment of source information to develop Forward Models and Sensory Predictions, i.e., pre-activation of low-level information (in peripheral sensory regions) related to the physical properties of the predicted stimulus. Enhanced expectations could thus modulate visual attention and the perception of an incoming stimulus (for a discussion of the relation between expectation and attention see Summerfield & Egnér, 2009).

Nevertheless, evidence of pre-stimulus activity related to the post-stimulus perceptual processing of a stimulus is only partial evidence in favor of prediction during language processing. What such an effect represents in the complex neurophysiological dynamics driving predictive processing is still an open question. Previous studies focusing on beta band oscillations have related beta activity to visual attention (Siegel et al., 2008). It is thus feasible that the beta effects we are targeting in this paper reflect the recruitment of different attentional processes in more or less predictive scenarios, but do not necessarily reflect predictive *coding* of a specific stimulus, i.e., modulation of the pre-stimulus oscillatory activity based on the distinctive properties of a predicted representation (at each level of the processing hierarchy). Although this mechanism is relevant for prediction, the beta band oscillations could reflect predictive *timing* (Arnal, 2012).

In addition, evidence for Sensory Prediction (as claimed in Dikker & Pylkkanen, 2013; Molinaro et al., 2013a) involving differential predictive activity in primary sensory regions is still

missing. Interestingly, Siegel and colleagues (2008) observed a correlation between beta band activity in primary visual regions and the subsequent behavioral performance in the task. Despite the involvement of primary visual regions (where Sensory Predictions effects should be evident before stimulus presentation), this peripheral effect could still reflect recruitment of additional neurophysiological resources, depending on the amount of task needs.

We are now working on a set of experiments in which we manipulate the (physical, orthographic, phonological, lexical, semantic) properties of a predicted item in experimental conditions that recruit the same amount of attentional resources (i.e., similar contextual constraints). Evidence for differential activation in high- and low-level (Sensory related) brain regions would provide clear-cut indications of the different amounts of resources recruited for qualitatively different stimuli. For example, Piai and colleagues (2014) observed oscillatory effects in the sentence constraint-picture naming paradigm in the theta band (~4-7 Hz) for items with variable lexical frequency. Such manipulation in a sentence comprehension paradigm (excluding the naming task) is similar to what is typically done in language production electrophysiological experiments and would provide indications of prediction of different items. Based on previous studies on prediction, successful stimulus decoding in (or driven by) frontal regions would likely reflect development of internal Forward Models (Pickering & Garrod, 2007), while effects in primary sensory regions could reflect Sensory Predictions. Assuming what predictive coding proposals suggest (Bastos et al., 2012), we should be able to decode the predicted stimulus properties from oscillatory activity in the beta-band channel. However, we do not exclude achieving our aim (i.e., stimulus decoding from pre-stimulus oscillatory activity) by analyzing a lower frequency band (< 10 Hz, Fontolan et al., 2014; Piai et al., 2014).

Additional evidence should emerge from studies focusing on connectivity measures (Bastos et al., 2015). Communication between brain regions could be an important neural mechanism supporting predictive processing (Molinaro et al., 2013a). Patterns of brain connectivity (not necessarily reflected in amplitude modulations, but for example in synchronous phase oscillations) represent the coupling between brain regions that could support top-down communication to primary sensory regions (Sensory Predictions). Such long-range connectivity has been reported for low-frequency phase oscillations (in the theta range ~4-7 Hz) in reading (Molinaro et al., 2013a) and speech perception (Peelle & Davis, 2012). In addition, cross-frequency coupling between low-frequency phase entrainment (delta, <4 Hz) and the amplitude of high-frequency neural activity (gamma, 30-80 Hz) has been associated with enhanced perceptual attention (Lakatos, Karmos, Mehta, Ulbert & Schroeder, 2008).

## **Conclusions**

In this paper, we have reviewed recent findings emerging from oscillatory analysis of high-temporal resolution measures (electrophysiological brain activity) collected before the presentation of an expected stimulus, to explore the parallels in processing between prediction and production. We have focused on the role that beta-band oscillatory activity has played in both the motor and the prediction domain and the possible evidence for top-down processing in both language production and language prediction. Importantly, Siegel and colleagues (2008; similar findings in Bastos et al., 2015) link such beta band activity observed in frontal-to-posterior brain regions to preparatory processing, probably reflecting attentional modulation of perception (observed in occipital regions and in different frequency bands). It is thus not clear if the beta frequency channel encodes properties of a predicted stimulus (predictive coding). We thus argue the need for further studies in which the feature of an expected stimulus should be modulated to provide more reliable indexes of predictive coding. Testing (within participants) production and prediction of linguistic items varying along different psycholinguistic dimensions could better detail the relation between production and prediction (Dell & Chang, 2014; Pickering & Garrod, 2007, 2013).

## Glossary

**Top-down processing:** Information processing in which higher hierarchical levels influence lower levels. Typically, information related to the physical properties of a stimulus is processed in modality-specific lower levels, while general abstract information is processed in higher processing levels. Bottom-up processing reflects information processing in the opposite direction. Top-down activity involves feedback/descending propagation of activation, while bottom-up activity is supported by feed-forward/ascending propagation of activation.

**Forward model:** In the motor domain, an internal Forward Model (based on a “efference” copy of a planned motor command) is involved in the estimation of the predicted body positions during a movement. The Forward Model derives from computationally abstract memory-based prospective knowledge of reality. The format of the memory information that shapes a Forward Model can be very heterogeneous and is informed by the external state of the environment or the internal state of the neurocognitive system (analogies and associations; Bar, 2007).

**Sensory Prediction:** Pre-activation of low-level information in sensory regions related to the physical properties of an incoming stimulus. Sensory Predictions are top-down activated by internal Forward Models.

**Prediction Error:** Amount of representational mismatch between the predicted information and the percept obtained from the external input. In the case of small amount Prediction Error (when the Sensory Predictions were correct), post-stimulus bottom-up processing is reduced and activation propagates across lower processing levels; when the Prediction Error is large (when the Sensory Predictions were either incorrect or not stimulus-specific enough), bottom-up processing is costlier and can reach higher processing levels.

**Adaptation:** Modification of internal representations stored in long-term memory to achieve a better fitness of the individual to the external environment and the possibility of developing more detailed predictions.

**Bayes' law:** The theorem is mathematically stated as follows:  $P(Pr|In) = [P(In|Pr) \times P(Pr)] / P(In)$ , where  $Pr$  and  $In$  are two sources of information.  $Pr$  can be the Prediction and  $In$  the sensory Input.  $Pr$  can be seen as the hypothesis that we are interested in testing and  $In$  the piece of evidence that confirms or disconfirms the hypothesis.  $P(Pr)$  is defined as the prior probability of  $Pr$  and  $P(Pr|In)$

is defined as the posterior probability of  $Pr$ , i.e., the probability assigned to  $Pr$  after taking into account the new piece of evidence,  $In$ . This theorem provides the estimation of a determined statistic in relation to prior observed probabilities of events constituting that statistic.

**Predictive coding (predicting “what”):** In neuroscience, the predictive coding hypothesis proposes a functional interpretation of how information is represented (hierarchically) and processed (iteratively). Predictive coding implies that the neural system actively predicts upcoming sensory information rather than passively registering it, while only the unpredicted portion of the sensory signal is recorded.

**Predictive timing (predicting “when”):** Refers to the ability of the brain to estimate the time intervals in which to deploy resources for the processing of an input stimulus. The rhythmic features of the oscillatory neural activity are assumed to reflect such predictive timing.

**Event Related Potentials:** Estimation of neural activity (voltages) time-locked to the presentation of an event, obtained by averaging samples of electrophysiological recordings across multiple repetitions of the same event.

**Frequency desynchronization:** Estimation of the change in power of the oscillatory neural activity in a specific frequency relative to a baseline period. Power increases are defined as synchronization effects, while power decreases as desynchronization. Power decrease has traditionally been considered as reflecting the desynchronized activity of large populations of neurons that are actively processing information independently (Pfurtscheller & Da Silva, 1999).

**Functional connectivity:** Connectivity between brain regions that share functional properties, defined as the temporal correlation between spatially remote neurophysiological events. It can be estimated in different ways such as spectral coherence and phase locking value, among others.

## **Acknowledgements**

This work was supported in part by grant PSI2012-32350 from the Spanish Ministry of Economy and Competitiveness and an Ikesbasque Research Fellowship to N.M.

We are grateful to Oihana Vadillo for help with data collection and Margaret Gillon-Dowens for comments on previous versions of the manuscript.

## References

- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264. doi: 10.1016/S0010-0277(99)00059-1
- Arnal, L. H. (2012). Predicting “When” using the motor system’s beta-band oscillations. *Frontiers in human neuroscience*, *6*. doi: 10.3389/fnhum.2012.00225
- Arnal, L. H., Doelling, K. B., & Poeppel, D. (2014). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*. doi: 10.1093/cercor/bhu103
- Arnal, L. H., & 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., & 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
- 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
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695-711. doi: 10.1016/j.neuron.2012.10.038.
- Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., ... 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
- Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2014). Multi-voxel pattern analysis of noun and verb differences in ventral temporal cortex. *Brain and Language*, *137*, 40-49. doi: 10.1016/j.bandl.2014.07.009.
- Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*:25. doi: 10.3389/fnhum.2010.00025
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926-1929. doi: 10.1126/science.1099745
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review*, *113*(2), 234-272. doi: 10.1037/0033-295X.113.2.234.
- 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.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*(2), 291-307. doi: 10.1093/brain/123.2.291
- Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. doi: 10.1098/rstb.2012.0394
- DeLong, K., Urbach, T. P., & 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
- den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*:548. doi: 10.3389/fpsyg.2012.00548

- Dikker, S., & Pylkkanen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*, 127(1), 55-64. doi: 10.1016/j.bandl.2012.08.004
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations – signalling the status quo? *Current Opinions in Neurobiology*, 20(2), 156-165. doi: 10.1016/j.conb.2010.02.015
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: long term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469-495.
- Findlay, A. M., Ambrose, J. B., Cahn-Weiner, D. A., Houde, J. F., Honma, S., Hinkley, L. B., Berger, M. S., Nagarajan, S. S. & Kirsch, H. E. (2012). Dynamics of hemispheric dominance for language assessed by magnetoencephalographic imaging. *Annals of neurology*, 71(5), 668-686. doi: 10.1002/ana.23530
- Fisher, A. E., Furlong, P. L., Seri, S., Adjamian, P., Witton, C., Baldeweg, T., ... Thai, N. J. (2008). Inter-hemispheric differences of spectral power in expressive language: a MEG study with clinical applications. *International Journal of Psychophysiology*, 68(2), 111–122. doi: 10.1016/j.ijpsycho.2007.12.005
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & 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.
- 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
- Frye, R. E., Wu, M. H., Liederman, J., & Fisher, J. M. (2010). Greater pre-stimulus effective connectivity from the left inferior frontal area to other areas is associated with better phonological decoding in dyslexic readers. *Frontiers in systems neuroscience*, 4. doi: 10.3389/fnsys.2010.00156.
- Gestaut, H. (1952). Etude electrocorticographique de la reactivite des rythmes rolandiques. *Revue Neurologique*, 87, 176-182.
- Hanslmayr, S., Staudigl, T., & 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
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Science U.S.A.*, 95(25), 15061-15065.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402. doi: 10.1038/nrn2113
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203, 89–97.
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain research*. doi: 10.1016/j.brainres.2015.02.014
- Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press, New York.
- Jenkinson, N., & Brown, P. (2011). New insights into the relationship between dopamine, beta oscillations and motor function. *Trends in Neuroscience*, 34(12), 611–618. doi: 10.1016/j.tins.2011.09.003
- Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neuroscience*, 37(7):357-369. doi: 10.1016/j.tins.2014.04.001.



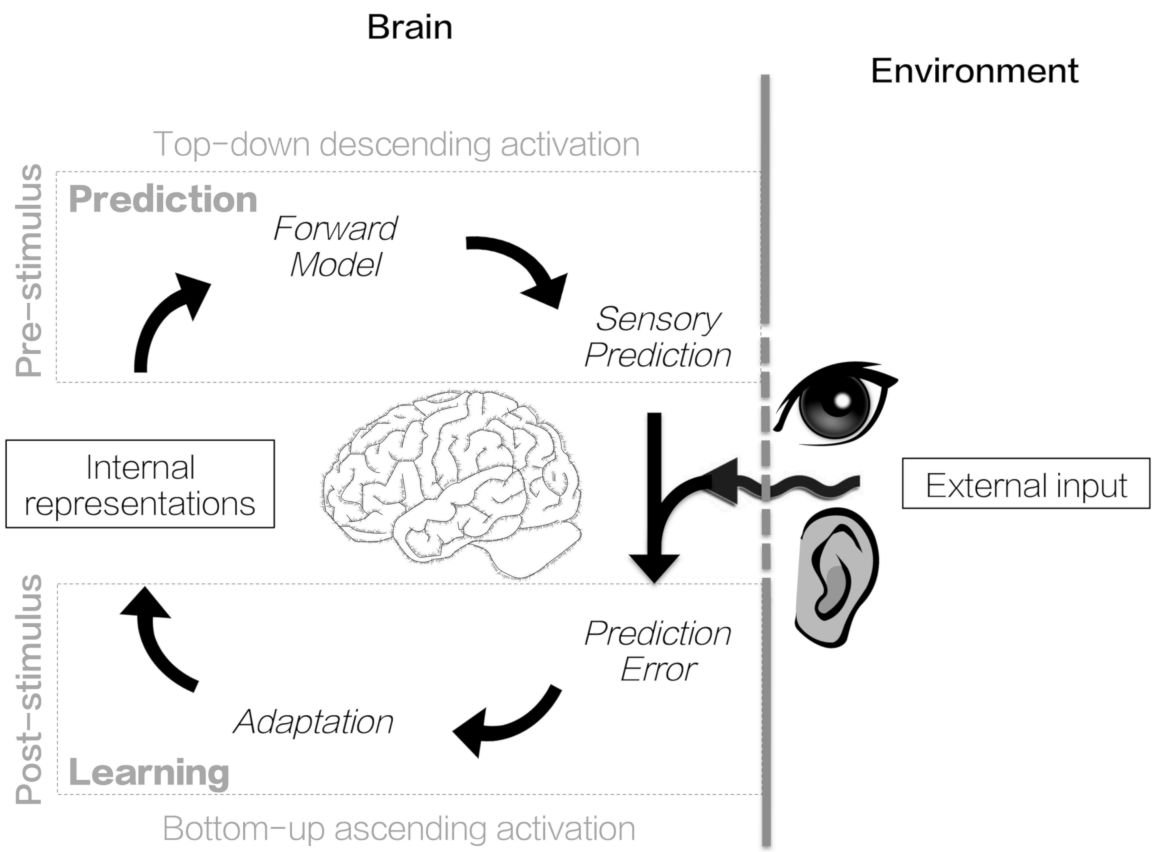
- Jenson, D., Bowers, A. L., Harkrider, A. W., Thornton, D., Cuellar, M., & Saltuklaroglu, T. (2014). Temporal dynamics of sensorimotor integration in speech perception and production: independent component analysis of EEG data. *Frontiers in psychology*, *5*. doi: 10.3389/fpsyg.2014.00656
- Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I. G., Hansen, P. C., & Cornelissen, P. L. (2014). Early Activity in Broca's Area During Reading Reflects Fast Access to Articulatory Codes From Print. *Cerebral Cortex*, Epub. doi:10.1093/cercor/bht350
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials reflect word expectancy and semantic association during reading. *Nature*, *307*(5947), 161-163.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science*, *320*(5872), 110-113. doi: 10.1126/science.1154735.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177. doi: 10.1016/j.cognition.2007.05.006
- Liljeström, M., Kujala, J., Stevenson, C., & Salmelin, R. (2014). Dynamic reconfiguration of the language network preceding onset of speech in picture naming. *Human Brain Mapping*. doi: 10.1002/hbm.22697
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843-847. doi: 10.1037/a0029284
- Molinaro, N., & Carreiras, M. (2010) Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*, 176-190. doi:10.1016/j.biopsycho.2009.12.006
- Molinaro, N., Vespignani, F., Canal, P., Fonda, S., & Cacciari, C. (2008). Cloze-probability does not only affect N400 amplitude: The case of complex prepositions. *Psychophysiology*, *45*(6), 1008-1012. doi: 10.1111/j.1469-8986.2008.00694.x
- Molinaro, N., Barraza, P., & Carreiras, M. (2013a). 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., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of Collocational Complex Prepositions. *Language and Cognitive Processes*, *28*(6), 762-788. doi: 10.1080/01690965.2012.665465
- Monsalve, I. F., Pérez, A., & Molinaro, N. (2014). Item parameters dissociate between expectation formats: A regression analysis of time-frequency decomposed EEG data. *Frontiers in Psychology*, *5*:847. doi:10.3389/fpsyg.2014.00847
- Peelle, J. E., & Davis, M. D. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*:320. doi: 10.3389/fpsyg.2012.00320
- Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, *53*, 146-56. doi: 10.1016/j.neuropsychologia.2013.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in cognitive sciences*, *11*(3), 105-110. doi: 10.1016/j.tics.2006.12.002
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-392. doi: 10.1017/S0140525X12001495

- Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, *110*(11), 1842-1857. doi: 10.1016/S1388-2457(99)00141-8
- Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1071-1086. doi: 10.1098/rstb.2007.2160
- Rahnev, D., Lau, H., & de Lange, F. P. (2011). Prior Expectation Modulates the Interaction between Sensory and Prefrontal Regions in the Human Brain. *Journal of Neuroscience*, *31*(29), 10741-10748. doi: 10.1523/JNEUROSCI.1478-11.2011.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87. doi: 10.1038/4580
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259–1274.
- 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
- Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A.K. (2008). Neuronal Synchronization along the Dorsal Visual Pathway Reflects the Focus of Spatial Attention. *Neuron*, *60*(4):709-719. doi: 10.1016/j.neuron.2008.09.010.
- Singh, K. D., Barnes, G. R., Hillebrand, A., Forde, E. M., & Williams, A. L. (2002). Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *Neuroimage*, *16*(1), 103–114.
- Summerfield, C., & 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.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-90. doi: 10.1016/j.ijpsycho.2011.09.015
- Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The Brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229–239. doi:10.1038/35067550
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682-1700. doi: 10.1162/jocn.2009.21293
- 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
- Weiss, S., & 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*:201. doi: 10.3389/fpsyg.2012.00201
- Wheat, K. L., Cornelissen, P. L., Frost, S. J., & Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *Journal of Neuroscience*, *30*(15), 5229-5233. doi: 10.1523/JNEUROSCI.4448-09.2010

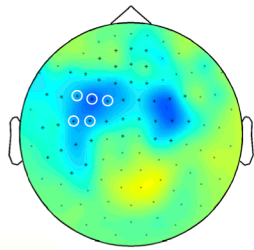
## Figure captions

*Figure 1:* Schematic model of the Prediction and Learning through which we represent the continuous interaction between the internal knowledge and the external environment. Prediction critically is reflected in the pre-stimulus brain activity; Learning is reflected in the post-stimulus brain activity

*Figure 2:* Oscillatory effects recorded in the MEG studies in which we studied deterministic prediction of lexical items (MWU) compared to probabilistic prediction (Semantic). In the top panel we report the sensor-level topography of the pre-stimulus oscillatory beta effects on the left, and post-stimulus oscillatory alpha effects on the right. In the underlying images, we report the time-frequency representation from the most relevant sensors (indicated by circles in the topoplots) for the two conditions and the differential effects.

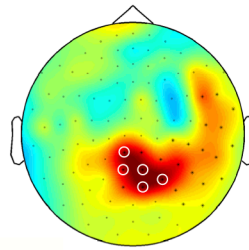


Pre-stimulus  
beta desynchronization  
for MWU vs. Semantic



time (sec) = [-0.67 -0.36]  
frequency (Hz) = [17 23]  
relative power change = [-0.15 0.15]

Post-stimulus  
alpha synchronization  
for MWU vs. Semantic



time (sec) = [-0.01 0.33]  
frequency (Hz) = [8 12]  
relative power change = [-0.15 0.15]

