Phonological Prediction in Speech Processing

PhD Thesis

ma per seguir virtute e canoscenza

SISSA

Candidate: Yamil Vidal Dos Santos

Co-Supervisor: Dr. Tristan Bekinschtein

Supervisor: Dr. Jacques Mehler

October 26, 2016

Thesis Summary

Auditory speech perception can be described as the task of mapping an auditory signal into meaning. We routinely perform this task in an automatic and effortless manner, which might conceal the complexity behind this process. It should be noted that the speech signal is highly variable, ambiguous and usually perceived in noise. One possible strategy the brain might use to handle this task is to generate predictions about the incoming auditory stream.

Prediction occupies a prominent role in cognitive functions ranging from perception to motor control. In the specific case of speech perception, evidence shows that listeners are able to make predictions about incoming speech stimuli. Word processing, for example, is facilitated by the context of a sentence. Furthermore, electroencephalography studies have shown neural correlates that behave like error signals triggered when an unexpected word is encountered.

But these examples of prediction in speech processing occur between words, and rely on semantic and or syntactic knowledge. Given the salient role of prediction in other cognitive domains, we hypothesize that prediction might serve a role in speech processing, even at the phonological level (within words) and independently from higher level information such as syntax or semantics. In other words, the brain might use the first phonemes of a word to anticipate which should be the following ones.

To test this hypothesis, we performed three electroencephalography experiments with an oddball design. This approach allowed us to present individual words in a context that does not contain neither semantic nor syntactic information. Additionally, this type of experimental design is optimal for the elicitation of event related potentials that are well established marker of prediction violation, such as the Mismatch Negativity (MMN) and P3b responses.

In these experiments, participants heard repetitions of standard words, among which, deviant words were presented infrequently. Importantly, deviant words were composed by the same syllables as standard words, although in different combinations. For example if in an experiment XXX and YYY were two standard words, XXY could be a deviant word. We expected that if as we proposed, the first phonemes of a word are used to predict which should be the following ones, encountering a deviant of this kind would elicit a prediction error signal.

In **Chapter 3**, we establish that as we expected, the presentation of deviant words, composed of an unexpected sequence of phonemes, generates a chain of well established prediction error signals, which we take as evidence of the prediction of the forthcoming phonemes of a word. Furthermore, we show that the amplitude of these error signals can be modulated by the amount of congruent syllables presented before the point of deviance, which suggests that prediction strength can increase within a word as previous predictions prove to be successful.

In **Chapter 4**, we study the modulating role of attentional set on the chain of prediction error signals. In particular we show that while high level prediction (indexed by the P3b response) is strategically used depending on the task at hand, early prediction error signals such as the MMN response are generated automatically, even when participants are simply instructed to listen to all the words. These results imply that phonological predictions are automatically deployed while listening to words, regardless of the task at hand.

In **Chapter 5**, we extend our results to a more complex stimulus set that resemble natural speech more closely. Furthermore we show that the amplitude of the MMN and P3b prediction error signals is correlated with participant's reaction time in an on-line deviant detection task. This provides a strong argument in favor of a functional role of phonological predictions in speech processing.

Taken together, this work shows that phonological predictions can be generated even in the absence higher level information such as syntax and semantics. This might help the human brain to complete the challenging task of mapping such a variable and noisy signal as speech, into meaning, in real time.

Dedication

A mis viejos.

Porque hicieron lo mejor de su parte para que yo puede hacer lo mejor de la mia.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Frame- work Programme (FP7/2007-2013)/European research Council Grant Agreement 269502 (PASCAL) (to JM)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Prediction in speech processing

Auditory speech perception requires the fast mapping into meaning of a complex auditory signal that is variable, noisy and ambiguous [67, 12]. Despite this, under normal circumstances, speech perception is experiences as an effortless activity. This is in part explained by the fact that speech processing is supported by a truly vast network of brain regions involving temporal, parietal and frontal cortices [47, 48, 49, 62, 99].

One strategy that this brain machinery could use to process the incoming speech signal, is to generate predictions based on the information available at a given moment and past experience. In fact, prediction has shown to play a role in a wide variety of cognitive domains. In the case of visual perception, this idea has a long tradition and can be traced back to the concept of unconscious inference forged by Helmholtz in 1860 [50] and has gained new momentum in recent years [112, 1]. Currently it can be found in a variety of other domains such as motor control [95], placebo analgesia [16] and self recognition [4]. The concept of a predictive brain has become a guiding principle in neuroscience [15, 63, 30, 51, 52].

In the case of speech processing, the idea that prediction plays a role is not new, as several behavioural experiments starting in the late 70's and early 80's have shown. For example, in the now famous garden path effect, while reading a sentence, reading speed is reduced [39, 53] and comprehension accuracy drops [40, 41] when an ambiguity is encountered. Importantly, the magnitude of the garden path effect is dependent on the predictability of the forthcoming words given the preceding context.

The topic of prediction in speech as always been an arena of intense debate. Arguments against have centred in the fact that any given context can have several possible continuations, which implies that betting on one particular prediction would be costly, as this prediction would usually fail [43]. For some authors, predicting ahead of time would not only be an unnecessary waste of processing resources [111] but could even hinder performance [79].

Despite this, the general idea that prediction serves a role in speech processing has become widely accepted in recent years, and is now the focus of great research effort [58]. Authors agree that given the noisiness, ambiguity and speed of our linguistic input, prediction is the most efficient solution for fast, and accurate comprehension [67].

A revision of the current literature in speech processing shows that examples of prediction can be found across language processing levels. At the syntactic level, listeners' individual experience, knowledge, and beliefs influence sentence parsing and interpretation [108, 115, 37, 6]. In a similar way, lexico-semantic processing is affected by contextual predictability, as reaction times are faster for predictable words compared to unpredictable ones [110, 93].

Furthermore, electroencephalography studies dating back as far as 1980 have shown the existence of an event related potentials (ERPs) known as N400, which amplitude is inversely correlated with the semantic predictability of words given previous context [69, 110, 14, 70, 71, 45, 28]

1.2 Prediction at phonological level

In this thesis, we will focus on the study of prediction at the phonological level. This type of prediction can, for example, explain the phenomenon of phonemic restoration. If a phoneme in a word that is presented in a sentence is replaced by a non-speech sound (e.g., a tone), listeners tend to hallucinate its presence. Furthermore, the strength of this illusion can be influenced by preceding context which provides information about the likelihood of the missing phoneme. The more the context constrains in favour of a particular phoneme, the more listeners tend to hallucinate its presence [66, 57]. This phenomenon can be related to our ability to understand speech in noisy environments, for which it has been shown that a flow of top-down information constrains the interpretation of the incoming speech signal [24].

Similarly, Bendixen, Scharinger, Strauß and Obleser [10] have shown that when a phoneme is omitted from a word, this can elicit an event related potential known as the mismatch negativity (MMN) [76], which can be interpreted as a marker of violation of expectations [50, 118]. But importantly, this only happens if context in which the phoneme omission occurs, contains either semantic or probabilistic information that makes the omitted phoneme predictable.

In the examples shown so far, an expected phoneme is either omitted or replaced by a non linguistic sound. But the literature also contains examples of evidence of prediction in cases in which one phoneme is replaced by another phoneme. For example, in electroencephalography experiments, when words are presented sequentially in pairs and participants are asked to report if the words are identical, if the second word of the pair contains a phoneme change, this will elicit a mismatch response, but only if the phoneme replacement goes against a phonological rule of the native language of the listener [23, 104, 119]. This shows how the phonotactic rules of a language can constrain early stages of speech sound categorization favouring the prediction of particular phonemes.

1.3 Goal of this thesis

In brief, given the available body of evidence, is clear that the context influences the state of the speech processing system, across different levels, before the bottom-up input is observed [68]. This is also truth at the phonological level. Listeners' brains can use high level syntactic, semantic and phonotactic information to anticipate which are the likely incoming phonemes of a word while is still unfolding. We propose that as prediction seems to plays a fundamental role in speech processing [60, 80, 12], phonological prediction might be possible even in the absence of this high level information. Specifically, once a word form has been learned, even if it is not attached to any particular meaning (neither semantic nor syntactic), while listening to that word, preceding phonemes could be use to anticipate which should be the forthcoming ones.

To test this hypothesis, we performed three electroencephalography experiments with an OddBall design. This type of experimental design allowed us to present individual pseudowords in a context that does not contain neither semantic nor syntactic information. Additionally, this type of experimental design is optimal for the elicitation of event related potentials that can be interpreted as marker of prediction violation, such as the mismatch negativity (MMN) [76, 118, 50] and P3b responses [50].

The stimuli used in this experiments consisted in pseudowords that respected the phonotactic rules of the listeners native language (Italian). Importantly, deviant words were constructed by cross-splicing standard words. In this way, they were composed by the same syllables as standard words, although in different combinations. For example if XXX and YYY were two standard words, XXY could be a deviant word. This implies an advantage over previous experiments on the topic, because the phoneme over which prediction is tested, is not omitted or replaced by a non-linguistic sound, but instead, is replaced by a phoneme that belongs to a different word. In this way, if an effect is found, it cannot be ascribed to a difference in low level acoustic features.

To summarize, the use of an OddBall design in which deviant pseudowords were constructed by cross-splicing standard words, allowed us to constrain the information available for prediction to the phonological and lexical (word form) levels. This allowed us to test if the presentation of an unexpected sequence of phonemes elicited prediction error signals such as the MMN and P3b responses, which would imply that listeners' brains are able to predict incoming phonemes within a word, based solely in phonological and word form information.

The evidence of the deployment of phonological prediction in the absence of higher level information would help us better understand how the speech processing system can handle the variability and noisiness of the speech stream.

Chapter 2

Methods

2.1 Data acquisition setup

EEG data was collected using a 128-electrode-net system (Geodesic EEG System 300, Electrical Geodesics, Inc.) referenced to the vertex. EEG signal was bandpass filtered by hardware between 0.1 and 100 Hz, and digitalized at 250 Hz. Electrode impedance was kept below 100 kΩ. Participants were tested in a soundproof Faraday cage. They sat on a chair in front of a LCD 19 inches monitor. Sound was delivered by a loudspeaker located behind the monitor, at a comfortable sound intensity of approximately 60 dB. The experiment was programmed in MATLAB (MathWorks, Inc., Natick,MA, USA) using the Psychophysics Toolbox extensions [13, 86]. Word onset was marked on the EEG data by sending both a digital input signal (DIN) and a TCP/IP mark.

Participants were requested to minimize movement throughout the experiment, except during breaks between blocks. No particular instructions were given to the subjects with respect to when to blink, as eye blink artefacts can be easily removed using Independent Component Analysis [29, 18]. At the beginning of the recording session, we registered resting state brain activity in 2 blocks of 4 minutes, during which participants were instructed to fixate at a white cross presented centrally against a grey background. Resting state activity was registered for a connectivity analysis that is not included in this thesis.

2.2 EEG data pre-processing

EEG data preprocessing was performed in MATLAB using custom code and the EEGLAB toolbox [29]. After being imported, data was band-pass filtered (0.1-30Hz) and segmented. Bad channels were rejected using EEGLAB `pop_rejchan` function[1] [29]. Following this automatic cleaning, additional channels were rejected by visual inspection. Independent Component Analysis (ICA) was use to remove eye blinks [29, 18]. Following, data was re-referenced to the average of all electrodes and baseline corrected using the 300ms before word onset. Next, we performed trial rejection by eliminating trials containing extreme values ($\pm$ 200 μV) and improbable trials (4 σ).

Only after this cleaning procedure the data was divided into conditions. Given that different words from different conditions were presented with different frequencies, the datasets of each condition were pruned by randomly discarding trials to ensure the same number of trials per condition. Finally, missing channels were interpolated and the trials of all participants were concatenated for each condition. The final result of these procedures were 1 dataset per condition, all of which contained the same amount of trials.

2.3 Statistical analysis

Statistical testing was performed utilizing a nonparametric clustering methods, introduced first by Bullmore et al [17] and implemented in the FieldTrip toolbox for EEG/MEG analysis [82]. This method offers a straightforward and intuitive solution to the Multiple Comparisons problem. It relies on the fact that EEG data has a spatio-temporal structure. In other words, a real effect should not be isolated but should instead spread over different electrodes and over time. Instead of assessing for significant differences between conditions in a point by point fashion, which would lead to a very big number of comparisons, this method groups together adjacent spatio-temporal points.

The procedure is as follows. For every point in time and space, the EEG signal

[1]We used the 3 available methods. *Kurtosis* threshold was set to 4 σ, *Joint probability* threshold was set to 4 σ, and *Abnormal spectra* was checked between 1 and 30 Hz, with a threshold of 3 σ

of 2 conditions is statistically compared. In our case, we used a nonparametric permutation t test for this step. The t values of adjacent spatio-temporal points with p values < 0.05 are clustered together and a cluster-level statistics is calculated by summing the t values within a cluster.

Once these candidate clusters have been defined, their significance is assessed using a nonparametric permutation test. In this test, conditions are shuffled and cluster-level t values are calculated as before. This step is repeated 5000 times and on each iteration the most extreme cluster-level t value is retained. This allows to build an histogram of expected cluster-level t values under the null hypothesis of no difference between the conditions. The significance of the observed candidate clusters is calculated as the proportion of expected t values under the null hypothesis that are more extreme than the observed ones. This method offers 2 important advantages over parametric methods. It profits from the spatio-temporal structure of the EEG signal to effectively reduce the number of comparisons performed, and it does not require to make strong assumptions about the distribution of the data. For further details see [73].

Chapter 3

Experiment 1

3.1 Introduction

In agreement with Kuperberg & Jaeger have proposed [68], we can consider that a prediction has been made in speech processing whenever the context (preceding input) influences the state of the language processing system before the (forthcoming) bottom-up input is observed.

This system can generate predictions at different levels of speech processing and these different levels can influence each other. But a survey of the literature shows that there are few examples of prediction at the phonological level. Furthermore, such examples use contexts that include syntactic or semantic information, which means that even though what is being predicted is a particular phoneme, the prediction might depend on higher level information.

We propose that if the human brain possess a hierarchically organizes predictive system for speech processing, this system should be able to generate predictions about the incoming stimuli even at low levels of processing. To be more specific, we propose that when listening to words, the identity of preceding phonemes could be used to predict the forthcoming phonemes within that word, without the need of additional syntactic or semantic context. Additionally we propose that the strength such predictions could be modulated by the number of preceding congruent phonemes presented within the word.

To test these hypotheses, we performed an electroencephalography experiment with an OddBall design. We choose this type of design because is optimal for the elicitation of a MMN response [77, 76]. This event related potential is a well established marker of violation of expectations which can be interpreted as prediction error that the system has to minimize [20, 113, 54, 50]. Another advantage of using an OddBall design is that it allowed us to study prediction at the single word level, without introducing syntactic or semantic information.

Participants heard repetitions of 2 three-syllabic pseudowords (**STD** pseudowords). Among these pseudowords, we presented deviants of 2 types. While **XYY** deviants were composed of the first syllable of one word, and second and third of the other word, **XXY** deviants were composed of the first and second syllable of one word, and third of the other word. If as we propose the first phonemes of a word are used to predict the forthcoming ones, when a deviant is presented, the established prediction would fail, and this should in turn trigger a MMN response.

It is important to notice that, as the syllables in the deviant pseudowords were acoustically identical to the syllables in the **STD** pseudowords, if the presentation of a deviant would trigger an effect such as the MMN, this could not be driven by instantaneous low level features of the stimuli such as pitch or intensity. Instead, such an error signal could only be triggered by the violation of an abstract rule [84]. The rule in question would be the one that states that given a syllable X_1, the next syllable of the word should be X_2.

The use of deviants that are composed by the same phonemes of standard pseudowords implies an advantage over the stimuli used in phonemic restoration experiments. In such experiments, the expected phonemes are replaced by a non-speech stimulus, such as a tone, which on its own can trigger a MMN response, regardless of the predictability of the replaced phonemes. Instead, in the design we propose, the expected phonemes are replaced different phonemes, for which the only feature that defines them as deviant is that they are unexpected.

We expected that the presentation of deviants like the ones described in the preceding paragraphs would trigger a MMN response. We would interpret such response as an error signal generated by the violation of a prediction with respect to which should be the forthcoming phonemes given the previous ones presented in the word.

Besides the MMN response, we considered that the presentation of deviants could elicit a P3b response. This event related potential can de considered as an index of violation of expectations, although it represents processing of information at a higher level, as this ERP is highly dependant on top-down attention and is only elicited by unexpected events that are relevant for the task at hand (e.g. deviants to which a response is required in an OddBall design) [8, 90, 64].

In conjunction, probing for the presence of MMN and P3b responses in an OddBall paradigm like the one we propose, would provide valuable information to determine if the brain's speech processing system is predicting which should be the forthcoming phonemes within a word. Furthermore, evaluating the presence of these ERPs would be informative with respect to the degree of automation of the predictions and role of attention.

We included 2 deviant types in our design in order to test if the strength of a prediction could be modulated. We were interested in testing whether MMN amplitude could be modulated by the number of congruent syllables presented before the point of deviance. In the case of **XYY** deviants, hearing the syllable X_1 could lead to the prediction that the next syllable should be X_2, but when the syllable Y_2 is instead presented, the prediction proves to be wrong. In contrast, in the case of the **XXY** deviants, when the syllable X_1 is presented, if is predicted that the next syllable should be X_2, this prediction is fulfilled. As 2 congruent syllables are presented in this case, this could lead to a stronger prediction stating that the last syllable should be X_3. Thus, when the syllable Y_3 is instead presented, this could trigger a stronger error response than the one elicited in the case of a **XYY** deviant.

In brief, in this experiment we expected to find error signals such as the MMN and P3b, that would be taken as evidence that while listening to a word past phonemes are used to predict which should be the forthcoming ones. Additionally, we expected that the amplitude of these error signal would be modulated by the number of syllables congruent with a standard word, presented before the point of deviance, which would be taken as evidence in favour of a modulation of the strength of the predictions.

3.2 Participants

A total of 30 participants took part in this experiment (10 male and 20 female, mean aged 22.86 $\pm$ 3.42 years). All of them were Italian native speakers, right handed and reported no auditory or language-related problems. Participants were recruited from the city of Trieste and received a monetary compensation of 15 €.

After data preprocessing, 11 participants were excluded from analysis due to contributing with less than 30 clean trials per condition. Additionally, 1 participant was excluded due to poor behavioural performance. Therefore, 18 participants were included in the final analysis (6 male and 12 female, mean age 24.15 n $\pm$ 3.12 years).

3.3 Stimuli

A total of 6 pseudowords divided in 3 sets of 2 pseudowords each were used as stimuli in this experiment. The experiment followed an OddBall design in which while the pseudowords of one of the sets were presented 84% of the times (**STD** pseudowords), the pseudowords of the other two sets were used as deviants that were presented only 8% of the time each.

As all the subjects in our experiments were native Italian speakers, we applied a series of constrains in the construction of our stimuli to ensure that the resulting pseudowords would resemble real Italian words. First we consulted the phonItalia lexical database [56] to identify syllable candidates. In order to exclude monosyllabic words and onomatopoeias, we removed syllables with a token frequency above the 70th percentile. Next, in order to keep syllables that could take any position within a word, we removed syllables with initial, medial or final position token frequencies either bellow the 20th percentile or above the 90th percentile. Finally, syllables containing voiceless stop consonants were removed, as they can be perceived as a pause. This selection procedure allowed us to identify 24 syllable candidates that are not monosyllabic words and have an even frequency distribution at different word positions.

Using these syllable candidates, we constructed 2 trisyllabic pseudowords that contained no vowel or consonant repetitions. No syllables were repeated between pseudowords. Hereafter, these pseudowords will be referred to as **STD** (i.e. Standard) pseudowords. Taking these **STD** pseudowords as a base, we constructed 2 different types of deviant pseudowords. The first one, that we will refer to as **XYY**, consisted of the first syllable of a **STD** word and the second and third of the other **STD** word. The second type of deviant, that we will call **XXY**, consisted of the first and second syllable of a **STD** word, and the third of the other **STD** word. None of these deviant words contained either consonant or vowel repetitions.

In natural speech, phonemes are co-articulated. Hence, using cross-splicing to generate the deviant words could result in sharp transitions that would sound unnatural. Because of this, we took measures to obtain a natural render for our stimuli [101]. For the first and last syllable position, the vowels of both **STD** words had similar first and second formant. As one **STD** word had the vowel 'o' in the first syllable, the other **STD** word had the vowel 'u' at the same position. In the case of the third syllable, while one **STD** word used the vowel 'i', the other one used the vowel 'e'. In the case of the second syllable, both **STD** words had 'a' as the vowel (Figure: 3.1). Finally, for each syllable position, the consonants of both **STD** words had the same mode of articulation. This measures had the effect of reducing the difference between both **STD** words at the points of syllable transitions so that when we cross-spliced them to construct the deviant words, these didn't contain any sharp transitions, which would sound unnatural.

The audio of the 2 **STD** words was generated using the MBROLA speech synthesizer [33] and the Italian female diphone database it4[1]. Consonant and vowel durations were set to 150ms and 175ms respectively. In this way, word duration was 975ms. Once the 2 **STD** words were produced, deviants were constructed by cross-splicing the audio of the **STD**, with the precaution of cutting at a point close to zero amplitude.

The final set consisted of 2 **STD** words, 2 **XYY** deviants and 2 **XXY** deviants (3.1), and was checked by a native Italian speaker who specializes in linguistics to ensure that they sounded like plausible but not real Italian words.

[1] The it4 diphone database was created by the Istituto Trentino di Cultura, ITC-irst

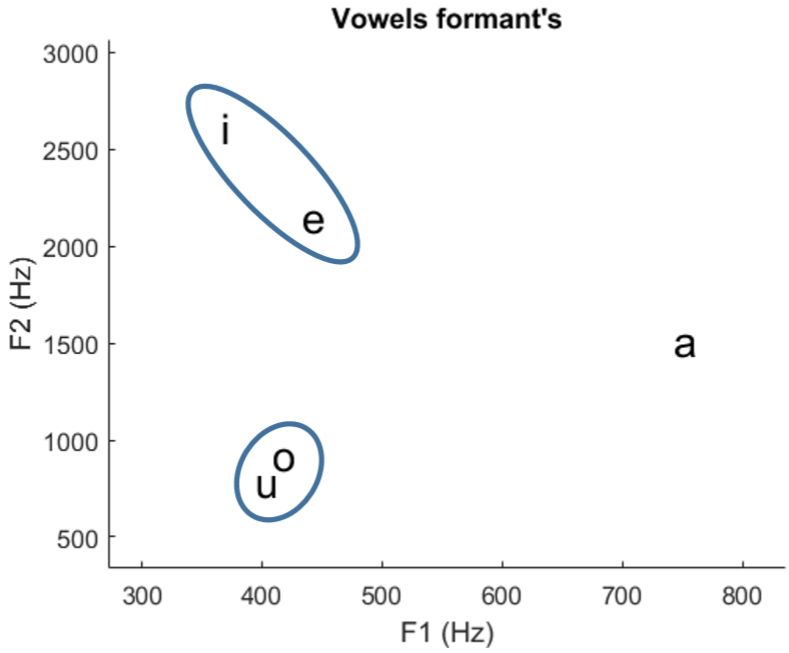

Figure 3.1: First and second formants of vowels used in the stimuli. While vowels 'o' and 'u' were used for the first syllable position, vowels 'i' and 'e' were use for the third syllable position. The vowel 'a' was used in the second syllable position of both Standard words. As deviant words were constructed by cross-splicing the Standard words, this prevented the occurrence of sharp formant transitions.

	STD	XYY	XXY
1.	foʃaʤe	foʣatsi	foʃatsi
2.	buʣatsi	buʃaʤe	buʣaʤe

Table 3.1: Stimulus set use for experiments 1 and 2 in IPA notation. A total of 6 words were used. Deviant words were produced by cross-splicing the 2 Standard words either at the end of the first syllable (**XYY**) or at the end of the second syllable (**XXY**).

3.4 Data acquisition setup

For a description of the data acquisition setup used in this experiment, see 2.1.

3.5 Procedure

The experiment followed an OddBall design. Stimuli were presented in 13 blocks with an average duration of 3.3 minutes each. During each block, a total of 98 words were presented, with an inter stimulus interval that varied between 900 and 1300 ms.

During the first of such blocks, only **STD** words were presented and participants were instructed to try to learn them. Subsequently, participants completed 12 blocks composed of 84% Standard words 8% **XYY** deviant words and 8% **XXY** deviant words. Within each block, word order was pseudo-random, with the restrictions of no word repeating more than 2 times consecutively and having between 2 to 4 **STD** words between deviant words. Participants were instructed to count the occurrence of "mistaken" words (i.e. deviant words) during each block and write down the result during the pause between blocks.

3.6 EEG analysis

3.6.1 Data pre-processing

Details about the procedure used for EEG data preprocessing and cleaning can be found in 2.2. Data was segmented into 1850ms long epochs starting 300ms before word onset. After cleaning, data was divided into **STD**, **XYY** and **XXY** conditions. The final result of these procedures were 1 dataset per condition, each containing 799 clean trials.

3.6.2 Regions of interest

We predefined 2 spato-temporal Regions of Interest (ROIs). The first one consisted on a **Fronto-Central** ROI comprised by 13 electrodes and spanned over a 325ms starting at the point of deviance of each deviant condition. With respect to word onset, this window spanned from 325ms to 650ms for the **XYY** condition, and from 650ms to 975ms for the **XXY** condition. This ROI coincided with the region were a MMN response could be expected [32, 77, 76].

The second region of interest consisted on a **Parietal** ROI composed by 21 electrodes and temporally extended from 200ms after the point of deviance of each deviant condition, to the end of the epoch. With respect to word onset, this window started at 525ms for the **XYY** condition, and at 850ms for the **XXY** condition. In this way, this ROI corresponded to the region were a P3b response would be expected [21, 90, 32] (Fig: 3.2).

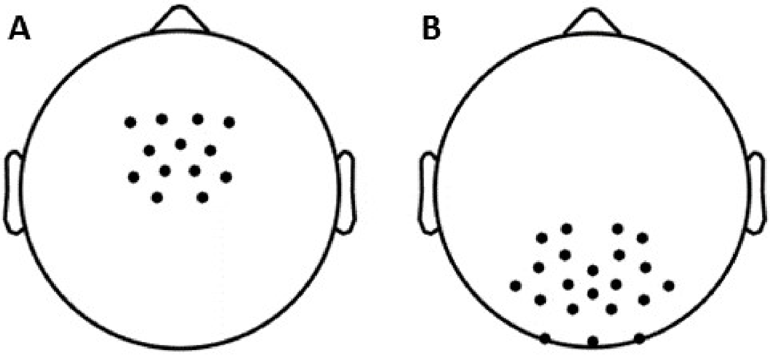

Figure 3.2: Regions of Interest for ERP analysis. **A)** Fronto-Central ROI. **B)** Parietal ROI

3.6.3 Statistical analysis

Unless otherwise noted, statistical testing was performed utilizing the procedure described in detail in 2.3. This nonparametric clustering methods offers a straightforward and intuitive solution to the Multiple Comparisons problem.

3.7 Results

3.7.1 Behavioural Results

During each block, a total of 16 deviant words were presented, and participants were requested to count them. On average, participants reported 15.22 deviant words per block (σ 2.56). For each participant, we checked the number of blocks with a deviant count further than 2 σ from the mean. While most of the participants reported a deviant count within these limits for all the blocks, 3 participants had 1 block with an abnormal count, and 1 participant had all 12 blocks outside the limits. This participant reported a mean of only 3.58 deviants per block, hence, we decided to exclude him from the analysis. After removing this participant, the mean number of deviants reported per block increases to 15.62 (σ 1.39).

Note that the method of asking participants to mentally count the occurrence of deviants does not allow us to determine with certainty neither the detection rate for each deviant condition, nor the occurrence of false alarms. In spite of this, given that the mean count of deviant was close to ceiling, we can safely conclude that the participants were able to perform the task with high accuracy for both deviant conditions.

3.7.2 EEG Event Related Potentials

In this subsection we will describe the results of our ERP analysis. Given that the **XYY** and **XXY** conditions differed with respect to the time point at which a word could be identified as a deviant, for clarity, we will refer to this point as zero, instead of taking the onset of the word.

Fronto-Central region of interest

Compared to **STD** words, **XYY** deviants elicited a negative deflection starting at 143ms (t = 41.77, $p < 0.05$). The topography of the difference wave between these 2 conditions, at the time point of the peak of this negative deflection corresponds to the one that characterizes a MMN response.

This was followed by a positive deflection starting at 219ms (t = -72.09, $p < 0.01$) whose topography is suggestive of a P3a response (Fig: 3.3), which is a positive-going potential with a maximum amplitude over fronto-central electrodes, with a peak latency between 250 and 280 ms [21, 85, 91, 90].

In parallel with these results, the comparison between the **STD** and **XXY** conditions also revealed a negative deflection, this time starting at 130ms (t = 102.90, $p < 0.01$) likewise displaying a topography compatible with a MMN. As previously, the negative deflection was succeeded by a positive one, starting 226ms after the point of deviance (t = -39.17, $p < 0.05$) with a topography resemblant to a P3a response (Fig: 3.4).

As we noted in the introduction (3.1) we expected that when a deviant word is presented, this would trigger an early error signal such as the MMN. These results fulfilled our expectations. Given that we were able to identify a MMN in both contrasts, we proceeded to compare both deviant types to test whether the amplitude of the MMN was modulated by the number of syllables congruent with a **STD** word, presented before the point of deviation.

Deviant conditions differed with respect to the time point at which the word presented revealed to be a deviant (350ms and 650ms from word onset for **XYY** and **XXY** conditions respectively). As the processing of a word has an intrinsic temporal

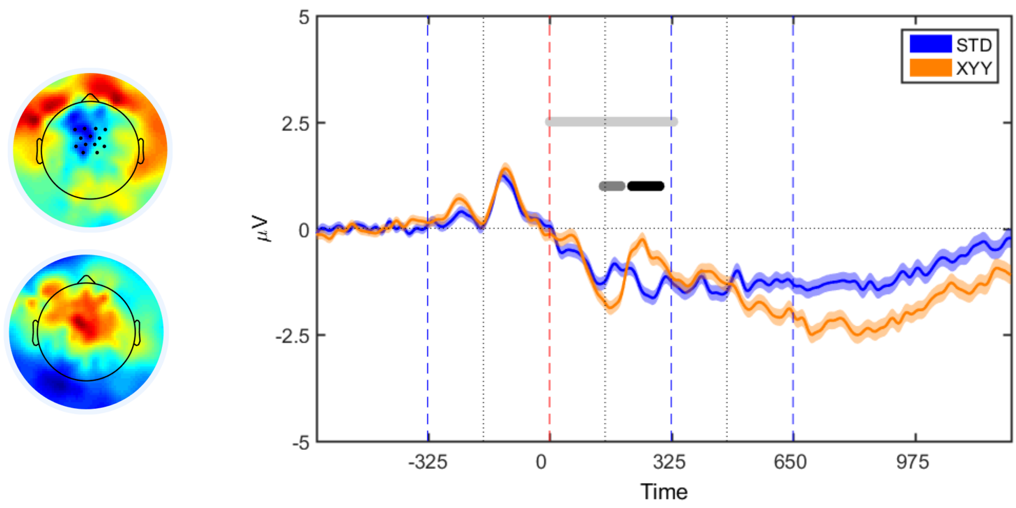

Figure 3.3: Experiment 1. Grand average over **Fronto-Central** ROI. **STD** vs **XYY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$. Middle grey horizontal line demarks $p < 0.05$. Left Top: Topography of the difference wave at the time point of the peak of the significant negative deflection. Left Bottom: Topography of the difference wave at the time point of the peak of the significant positive deflection.

dynamic, we eliminated this confounding factor by subtracting the activation elicited in the **STD** condition to each deviant condition. Moreover, we re-segmented the trials of both deviant conditions so that the point of deviation would be aligned. The resulting epochs had a length of 1148ms, starting 325ms before the point of deviance.

This comparison revealed that as we expected, the MMN elicited by the **XXY** condition had a bigger amplitude than the one elicited by the **XYY** condition (t = -46.91, $p < 0.01$) (Fig: 3.5). We expected that this would be the case, because as the **XXY** deviants start with two congruent syllables, this could lead to a stronger prediction that would be reflected by a stronger error response in case of failure of the prediction. On the contrary, the positive deflection that followed the MMN response, didn't differ across deviant types.

But an alternative explanation exists. We noticed that the topography of the MMN response elicited by the **XYY** deviant was slightly shifted to the left. Because of this, our pre defined **Fronto-Central** ROI might favour the **XXY** condition,

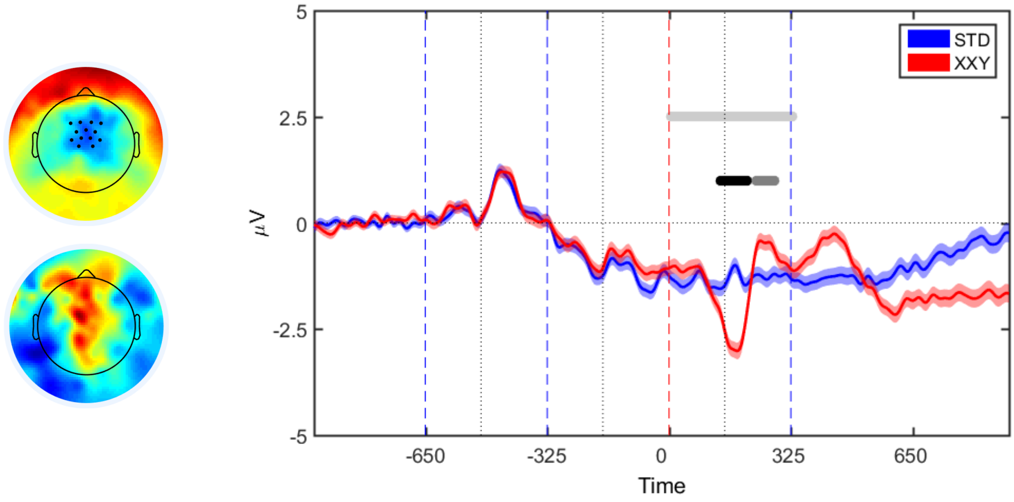

Figure 3.4: Experiment 1. Grand average over **Fronto-Central** ROI. **STD** vs **XXY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$. Middle grey horizontal line demarks $p < 0.05$. Left Top: Topography of the difference wave at the time point of the peak of the significant negative deflection. Left Bottom: Topography of the difference wave at the time point of the peak of the significant positive deflection.

which is better centred. To test this possibility, we repeated our analysis keeping the original **Fronto-Central** ROI for the **XXY** condition, but hand picking a ROI with the same number of electrodes but better aligned for the **XYY** conditions (Fig: 3.6).

Even after this intentional selection of the most negative electrodes in the **XYY** condition, we found that the amplitude of the MMN response elicited by the XXY condition was higher (t = -28.25, $p < 0.05$) (Fig: 3.7).

To summarise, our analysis showed that both deviant types elicited a MMN and a P3a response. The MMN is a well established marker of violation of expectations [84, 76] that can be interpreted as a prediction error signal [117, 50], reflecting a mismatch between a predicted stimulus and the one presented in reality. Therefore, we consider that the presence of a MMN in response to the presentation of a deviant evidences that a prediction about the forthcoming phonemes had been made. Moreover, we could determine that the amplitude of this MMN response, which would reflect the strength of the prediction, was modulated by the number of syllables congruent with

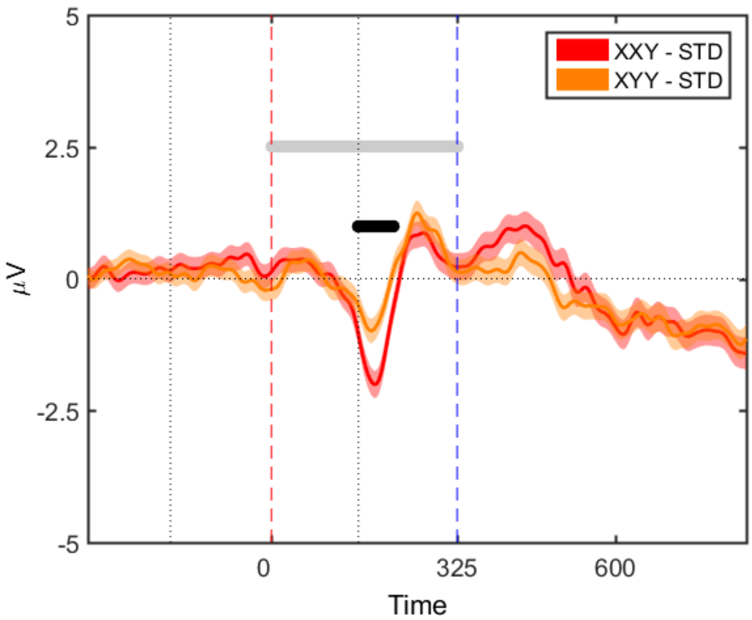

Figure 3.5: Experiment 1. Grand average over **Fronto-Central** ROI. **XYY** vs **XXY**. Trials were re-segmented and locked to the point of deviance, indicated by time zero. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$.

a **STD** word, presented before the point of deviance. Besides the MMN, we were also able to identify a P3a response, which can indicate novelty detection [72, 92, 21, 90]. Notice that in the case of the stimuli used in our experience, novelty cannot be defined in terms of low level features such as pitch. Instead, when a deviant is presented, what is novel is the occurrence of a syllable that is unexpected given the preceding one.

Parietal region of interest

Next we focused our analysis on the spatio-temporal region of interest tailored to detect a P3b component (see 3.6.2). We were able to determine that the **XYY** deviant evoked a long lasting positive deflection starting at 271ms and extending till the end of the epoch (t = -1297.6, $p < 0.001$). The topography of this deflection corresponds to a large parietal positivity, which allowed us to identify it as a P3b response (Fig: 3.8).

An equivalent positive deflection was present for the **XXY** condition (t = -1252.2, $p < 0.001$). This deflection started at 286ms and displayed its topography

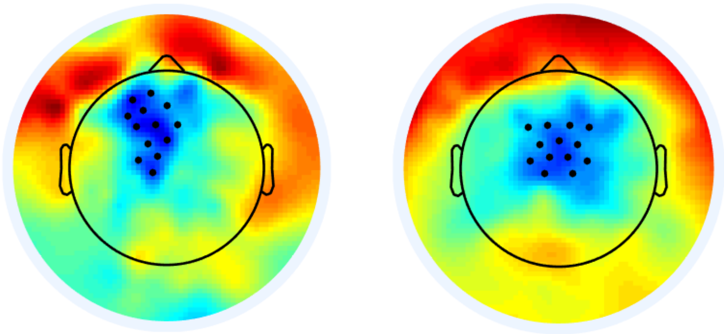

Figure 3.6: Alternative ROIs for comparison of **XYY** (left) and **XXY** (right) conditions. Instead of using the predefined ROI for the **XYY** condition, we hand picked electrodes that overlap with the MMN response in this condition.

consisted on a large parietal positivity. Hence, once more we could identify this effect with a P3b response (Fig: 3.9).

Our predictions for the occurrence of a P3b response were satisfied by the data. We proceeded to compare the P3b components detected in both deviant conditions using the same procedure that we used for the MMN. This allowed us to conclude that the amplitude of the P3b elicited by the **XXY** condition was higher than the one elicited by the **XYY** condition ($t = 177.93$, $p < 0.001$) (Fig: 3.10).

In brief, we could determine that both deviant types elicited a clear P3b response whose amplitude is modulated by the number of syllables congruent with a **STD** word, presented before the point of deviance.

3.8 Discussion

3.8.1 Prediction at the phonological level

The main goal of this experiment was to determine if while listening to words, the first phonemes presented are used to make a prediction regarding the identity of the forthcoming ones. We argued that if this would be the case, the presentation of deviants which start with the syllables of one word, but continues with the ones belonging to another word, should elicit an error signal such as the MMN, P3a or P3b. Additionally, we wanted to test whether this predictions could become stronger with successive successfully predicted phonemes within a word.

Overall, we found that the presentation of a deviant word consisting on an un-

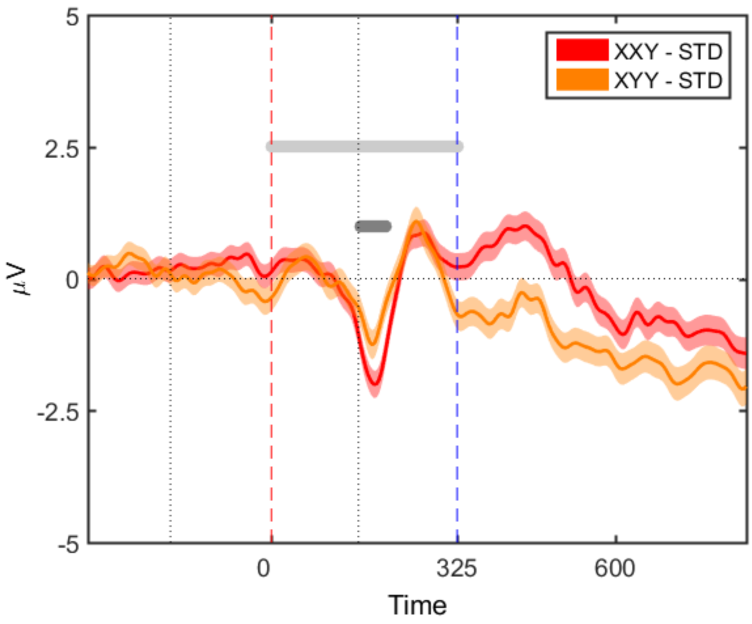

Figure 3.7: Experiment 1. Grand average over alternative ROIs. **XYY** vs **XXY**. (Using custom ROI for the **XYY** condition) Trials were re-segmented and locked to the point of deviance, indicated by time zero. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Middle grey horizontal line demarks $p < 0.05$.

expected sequence of phonemes, elicited a chain of error responses, including MMN, P3a and P3b ERPs. Furthermore, the amplitudes of the MMN and P3b responses were greater for the deviant type that contained 2 successive congruent syllables. These results suggest that as we expected, the first phonemes of a word are used to generate a prediction about which should be the forthcoming ones.

Given that both standard and deviant words were constructed using the same set of syllables, but in different combinations, the occurrence of a MMN in response to the deviants is of greater interest. It is know that the MMN can be elicited by manipulating a variety of stimuli features such as pitch, duration, intensity, location and the presence of a gap [78, 42, 2, 87, 88]. But in our stimuli, each syllable in a deviant was acoustically identical to a syllable in a standard word (see 3.1). Therefore, the MMN detected in our experiment cannot be explained by a mismatch of low level acoustic features, but is instead a response to an abstract rule violation [84]. The abstract rule being violated would be the one connecting sequences of phonemes to form words.

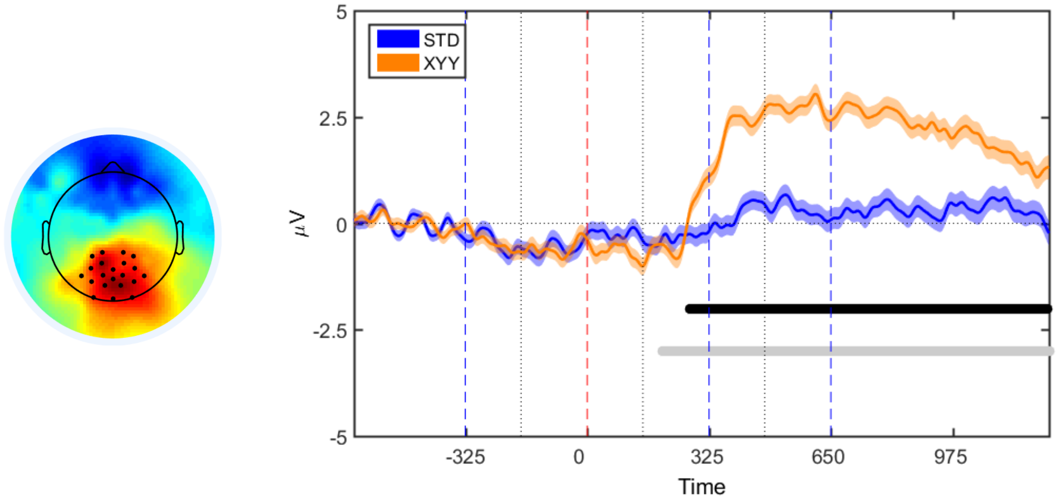

Figure 3.8: Experiment 1. Grand average over **Parietal** ROI. **STD** vs **XYY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$. Left: Topography of the difference wave at the time point of the peak of the **XYY** condition.

The presence of a P3a response in our experiment represents an unexpected finding. This ERP is usually elicited in 3 stimuli OddBall paradigms in response to distractors (deviants that are not targets) [91], and is associated with involuntary attentional shifts and the processing of novelty [21, 85, 90]. As in our experiment participants were explicitly instructed to deploy top-down attention to detect all deviant, we didn't expect to detect this ERP. Even if unexpected, the presence of this ERP is compatible with our hypothesis.

Finally a P3b was registered as we expected. As we argued before, this ERP can be interpreted as a marker of violation of expectations, but in contrast with the MMN response, which is largely automatic and can even be observed in patients with disorders of consciousness [9, 77, 19, 32, 65], this ERP is dependant on top-down attention and task instructions [90].

To summarise, we interpret the presence of MMN, P3a and P3b components in response to the deviants, as evidence in favour of our hypothesis which posits that, when processing speech, the human brain generates predictions about the incoming phonemes within individual words, even in the absence of higher level information such as syntax or semantics.

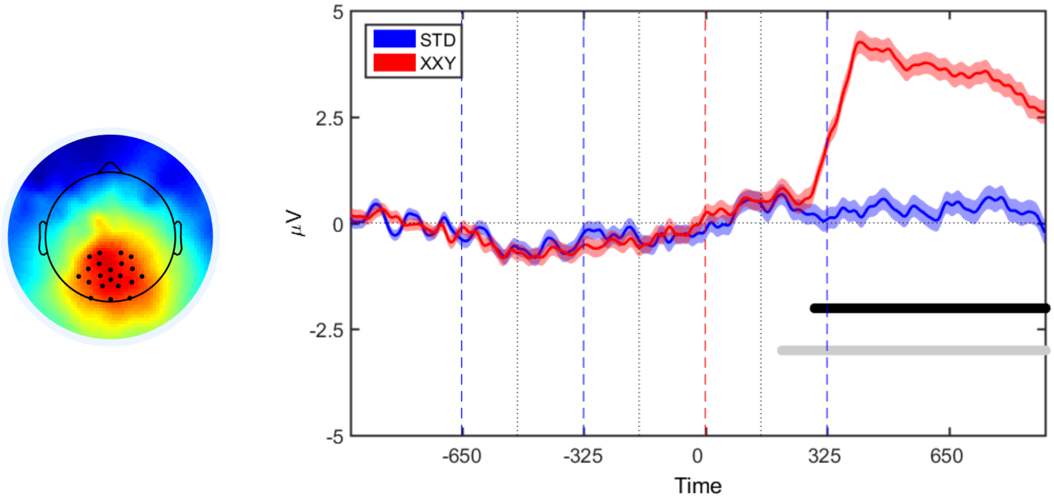

Figure 3.9: Experiment 1. Grand average over **Parietal** ROI. **STD** vs **XXY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$. Left: Topography of the difference wave at the time point of the peak of the **XXY** condition.

It is worth remarking that in our design, as we constructed the deviant stimuli by cross-splicing the standard words, the only feature that defined a word as deviant is that within that word, given a syllable X_n, instead of the usual syllable X_{n+1}, the syllable Y_{n+1} was presented. This represents an important difference with respect to previous studies in which the predicted phonemes are either omitted [10] or replaced by non-linguistic sounds [57].

To our knowledge, this is the first time in which it has been shown that the human speech processing system can generate predictions at the phonological level based solely in the identity of the preceding phonemes within a word.

3.8.2 Modulation of the strength of the predictions

As we described previously, the amplitudes of the MMN response (3.7.2) and the P3b response (3.7.2) were greater when 2 congruent syllables were presented before the point of deviance. We propose that this increase in amplitude could indicate that when a prediction based on a first syllable is corroborated by the presentation of a second congruent syllable, the prediction for the next incoming syllable inte-

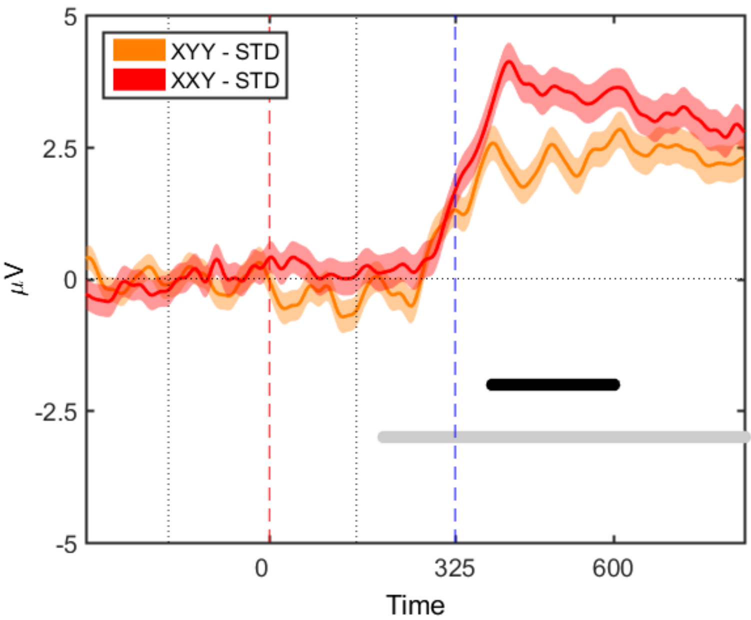

Figure 3.10: Experiment 1. Grand average over **Parietal** ROI. **XYY** vs **XXY**. Trials were re-segmented and locked to the point of deviance, indicated by time zero. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$.

grates information from the 2 precedent ones and becomes stronger. This result is particularly interesting in the case of the MMN response, as it shows that this early and automatic component can integrate information presented in (at least) the past 650ms. This implies that in order to predict the forthcoming phonemes within a word, the proposed predictive system can use information taken not only from the immediate past. The modulation of the P3b response is less surprising, as this ERP is contingent to top-down attention and even displays a tight correlation with conscious access [19, 96, 9, 26, 103, 38], a level of processing at which information can be integrated over very long time windows.

Alternatively, instead of the result of a strengthening of a prediction, the observed modulation of amplitude of the MMN and P3b responses could be due to a difference in detection rate between deviant types. If the **XYY** deviants would be more difficult to detect, a higher proportion of "miss" trials in this condition would lead to a smaller overall amplitude. However we find it implausible. With a mean count of 15.62 deviants per block (which corresponds to 97% of the deviants presented per block), the overall performance of the participants in the counting task was close to ceiling. This implies that unless the counting reported contains a high proportion of false alarms, even if there would be an asymmetry in the detection rate of the deviants, this alone would be unlikely to account for the observed difference in amplitude between the conditions. Furthermore, during a short debriefing after the experiments, participants tend to report that they notice the existence of both deviant types, and they didn't consider that they were presented with different frequencies. It should also be considered than even in our analysis using a custom ROI for the **XYY** condition, the amplitude of the MMN respond in this condition was 1.25 µV against 2 µV for the **XXY** condition, which represents a 60% of increase in amplitude for the **XXY** condition. Given all these reasons, we can rule out a difference in detection rate between deviants as an alternative explanation for the observe difference in amplitude of the MMN and P3b responses.

3.8.3 Experiment 1 summary and follow up

In this chapter we have exposed evidence suggesting that when listeners hear a word, the first phonemes perceived are used to build a prediction about the forthcoming ones, even in the absence of higher level information such as syntax or semantics. As our experiment shows, the presentation of a deviant word that is composed of an unexpected sequence of phonemes trigger a chain of event related potentials that are well known markers of violation of expectations.

In this experiment, we informed the participants about the presence of "mistaken" words, and we instructed them to count them. But while listening to natural speech, we don't (usually) pay particular attention to the individual words uttered by our interlocutors in an attempt to catch them making a mistake. As top-down attention can influence perceptual processes to a great extend, it remains possible that the responses we recorded during this experiment might behave differently if the participants were not explicitly instructed to count the occurrence of the deviants.

There are reasons to suppose that a MMN response would be present even under different task instructions. As we argued before, the MMN response has demonstrated to be independent from top-down attention [76], to the point that its presence can be corroborated across all stages of sleep [103] and even in patients with disorders of consciousness [65, 8]. Furthermore, source estimation of MMN using equivalent current dipoles suggests that its main generators are the auditory cortices [83, 54, 76]. These characteristics of the MMN response imply that the predictive system we propose can function at early stages of the processing hierarchy.

Nevertheless, the degree of automaticity of the MMN response registered in our experiment should be further evaluated. Some authors have proposed that the amplitude of the MMN response can be attenuated if stimuli are presented outside the focus of attention [5, 75]. Additionally, it should also be considered that the MMN response recorded in our experiment was not driven by changes of low level features of the stimuli, but by the violation of an abstract rule. It has been shown that in abstract rule paradigms, it is not always possible to trigger a MMN response in non-attentive blocks, unless participants complete at least 1 attentive block first [106, 105].

In the case of the P3a and P3b responses, it is well know that this components can be strongly modulated by task instructions and top-down attention [21, 91, 90, 19]. In particular the P3b response can disappear altogether if participants' attention is allocated somewhere else [32]. Because of these reasons, how task instructions can affect the chain of predictive error signals recorded in our experiment remains as an open question.

To answer this question, in the next chapter we will introduce an ecologically more valid version of our task which doesn't provide to the participants any information or particular instructions with respect to the occurrence of deviants.

Chapter 4

Experiment 2

4.1 Introduction

In the previous chapter (3) we exposed evidence in favour of our hypothesis which states that while listening to a word, the human brain builds predictions at the phonological level to anticipate the forthcoming phonemes, even in the absence of higher order information such as syntax or semantics. Using EEG in an experiment with an OddBall design, we showed that the presentation of deviants that were composed by an unexpected sequence of syllables, triggered a chain of well established mismatch responses such as the MMN, P3a and Pb3 (3.7.2).

As we argued in the discussion of the previous chapter (3.8.3), the behaviour, and even the presence of this ERPs, can be affected by how participants allocate attention, and this in turn, depend on task instructions. As in the experiment presented in the previous chapter we instructed participants to explicitly count "mistaken" (i.e. deviant) words, whether the same chain of mismatch responses would unfold under different instructions remained as an open question. Language processing is largely automatic [89], and even complex operations like the build up of local phrase structure [46], learning of nonadjacent dependencies [25] and speech segmentation [22, 107] are performed even when no explicit task instructions are provided. Because of this we considered that, even under different task instructions, incoming phonemes within a word would be predicted based on the preceding ones.

To test if this is the case, in this chapter we present an experiment (**Experiment 2**) with a similar design to **Experiment 1** (3), but this time asking participants to pay attention to all words, without informing them of the presence of deviants. The main goal of this experiment was to characterize the prediction error responses that would emerge while participants keep an attentional set closer the the one held while processing natural language.

We expected that under this conditions a MMN response would still be present, and its amplitude could even increase. It has been shown that when attention is focused on deviant detection, as in the case of **Experiment 1**, as the occurrence of deviants is expected, top-down attention can reduce the amplitude of the MMN response [20]. As in **Experiment 2** the participants were not informed about the presence of deviant words, these were fully unexpected.

We were also interested in testing whether the MMN modulation by number of congruent syllables found in **Experiment 1** (3.8.2) would persist. Even though source estimation studies show that the main generators of the MMN are the auditory cortices, a variety of additional regions can contribute, including frontal cortices, depending on the task at hand [83, 54, 76]. The construction of phonological predictions based on past information exceeding the adjacent phonemes might required the involvement of frontal regions that are only recruited when participants are explicitly trying to detect the deviants. Therefore, under instructions that don't distinguish between standard and deviant words, phonological predictions might be restricted to the use of the immediate preceding phonemes. In our experimental design, this would be reflected by the presence of a MMN that is not modulated by the number of congruent syllables presented before the point of deviance.

In the discussion of chapter 1 (3.8.1) we argued that as the P3a response is usually elicited by distractors in 3 stimuli OddBall paradigms [91], its presence in **Experiment 1** was unexpected. Instead, in the experiment we are about to present, it is possible that the deviants would trigger an involuntary attentional shift which usually is indexed by such ERP [21, 85, 90]. Therefore, if the response registered in **Experiment 1** is indeed a P3a, it should be also present in **Experiment 2**, and its amplitude might even be higher. Finally, with respect to the P3b response recorded in **Experiment 1**, without particular instructions regarding the deviants, it is possible that this ERP wouldn't be present [32].

In brief, the experiment that we will present in this chapter allowed us to characterize the prediction error signals that are present while participants listen to speech while in an attentional set more similar to the one used to process speech in everyday life.

4.2 Participants

A total of 29 participants took part in this experiment (9 male and 20 female, mean aged 23.24 ± 3.52 years). All of them were Italian native speakers, right handed and reported no auditory or language-related problems. Participants were recruited from the city of Trieste and received a monetary compensation of 15 €.

After data preprocessing, 4 participants were excluded from analysis due to contributing less than 30 clean trials per condition. Additionally, 1 participant was excluded from analysis due to having more than 3 adjacent bad electrodes in the predefined regions of interest. Therefore, 23 participants were included in the final analysis (7 male and 16 female, mean age 23.45 ± 3.26 years).

4.3 Stimuli

The stimulus set used for **Experiment 2** consisted of the same 6 words used for **Experiment 1** (2 **STD** words, 2 **XYY** deviants and 2 **XXY** deviants) (3.1), plus the addition of 2 New words that were only used during a test at the end of the experiment. These New words were composed by the first and last syllable of 1 **STD** word, and the second syllable of the other **STD** word (**XYX** structure). They were constructed following the same cross-splicing procedure that was use to construct the deviants (see 3.3).

4.4 Data acquisition setup

For a description of the data acquisition setup used in this experiment, see 2.1.

4.5 Procedure

The experimental setup and design were identical to those used for **Experiment 1** (see 3.5), with the exception of the instructions given to the participants and the addition of a behavioural test at the end of the experiment. Whereas in **Experiment 1** participants were instructed to count the occurrence of "mistaken words" (i.e. deviant words), in **Experiment 2** the participants were not informed about the presence of deviants and were simply instructed to pay attention to all the words presented and try to learn them.

At the end of the experiment, participants completed a forced choice test. Instructions were given verbally. On each trial, participants heard 2 words and were requested to choose the one that most likely was presented during the experiment. Participants completed 4 trials for each of 6 contrasts between conditions, for a total of 24 trials. The contrasts between conditions were "**STD** vs **XYY**", "**STD** vs **XXY**", "**XYY** vs **XXY**", "**STD** vs **New**", "**XYY** vs **New**" and "**XXY** vs **New**". Participants reported their answers verbally and the experimented entered them through keypress.

4.6 EEG analysis

4.6.1 Data pre-processing

Details about the procedure used for EEG data preprocessing and cleaning can be found in 2.2. Data was segmented into 1850ms long epochs starting 300ms before word onset. After cleaning, data was divided into **STD** (Standard), **XYY** and **XXY** conditions. The final result of these procedures were 1 dataset per condition, each containing 946 clean trials.

4.6.2 Regions of interest

The spatio-temporal regions of interest used for this experiment were identical to the ones used in **Experiment 1**. See 3.6.2 for details.

4.6.3 Statistical analysis

Unless otherwise noted, statistical testing was performed utilizing the procedure described in detail in 2.3. This nonparametric clustering methods offers a straightforward and intuitive solution to the Multiple Comparisons problem.

4.7 Results

4.7.1 Behavioural Results

In this experiment, participants were not informed of the existence of the deviants and were not instructed to perform any task with respect to them during the experimental blocks. Regardless of this, if our predictions were correct, the participants would prefer the **STD** words over both deviant categories. To evaluate this, at the end of the experiment we requested the participants to perform a force choice test in which each stimuli condition was contrasted against the others and against new words that were not presented during the blocks. The mean preference in each contrast was calculated for each participant and a one sample t test was performed at the group level to test against the null hypothesis of no difference from chance. Results were corrected for multiple comparisons using the Bonferroni-Holm method (See table 4.1).

Participants preferred the **STD** words over both deviant types. They choose **STD** words over **XYY** deviants on 67% of the trials ($t(28) = 3.57$, $p < 0.01$) and over **XXY** deviants on 69% of the trials ($t(28) = 4.07$, $p < 0.01$). When both deviant types were contrasted, they preferred **XYY** over **XXY** deviants on 62% of the trials, but this preference didn't reach a significant level ($t(28) =$ -2.32, $p > 0.05$).

Next, we contrasted the words used in the experiment against new words that were not previously presented. Participants selected **STD** words over new ones on 85% of the trials ($t(28) = 10.39$, $p < 0.0001$) and **XXY** deviants over new words on 64% of the trials ($t(28) = 2.99$, $p < 0.05$). **XYY** deviants on the contrary, could not be distinguished from new words and were preferred on only 55% of the trials ($t(28) = 1.03$, $p > 0.05$).

In brief, these results indicate that despite the fact that the instructions provided didn't distinguish between deviant and standard words, participants displayed a strong preference for **STD** words over both deviant types. Even though both deviant types had the same probability of appearance, while **XXY** deviants could be distinguished from new words, **XYY** could not.

Contrast	Pref.	Statistics
STD vs **XYY**	67%	$t(28)$ = 3.57 $p < 0.01$
STD vs **XXY**	69%	$t(28)$ = 4.07 $p < 0.01$
XYY vs **XXY**	62%	$t(28)$ = -2.32 $p > 0.05$
STD vs **NEW**	85%	$t(28)$ = 10.39 $p < 0.0001$
XYY vs **NEW**	55%	$t(28)$ = 1.03 $p < 0.05$
XXY vs **NEW**	64%	$t(28)$ = 9.99 $p > 0.05$

Table 4.1: Experiment 2 Behavioural Results. A forced choice test was performed contrasting each condition against the other and against new words that were not presented during the experimental blocks. The mean preference in each contrast was calculated for each participant and a one sample t test was performed at the group level to test against the null hypothesis of no difference from chance. Results were corrected for multiple comparisons using the Bonferroni-Holm method.

4.7.2 EEG Event Related Potentials

In this subsection we describe the results of our ERP analysis. Given that the **XYY** and **XXY** conditions differed with respect to the time point at which a word could be identified as a deviant, for clarity, we will refer to this point as zero, instead of taking the onset of the word.

Fronto-Central region of interest

We first directed our analysis to our **Fronto-Central** region of interest (see 4.6.2). As we argued in the introduction of this chapter, we expected that even without instructions to detect deviants, the proposed predictive system would generate a MMN response when a deviant is presented. Indeed, we could determine that both deviant types elicited a MMN response comparable to the ones found in **Experiment 1** (see

3.7.2).

A comparison between **STD** and **XYY** conditions revealed that the **XYY** deviants triggered a negative deflection starting at 135ms ($t = 44.35$, $p < 0.01$). As in the case of **Experiment 1**, the topography of the difference between the conditions matched with a MMN (Fig: 4.1).

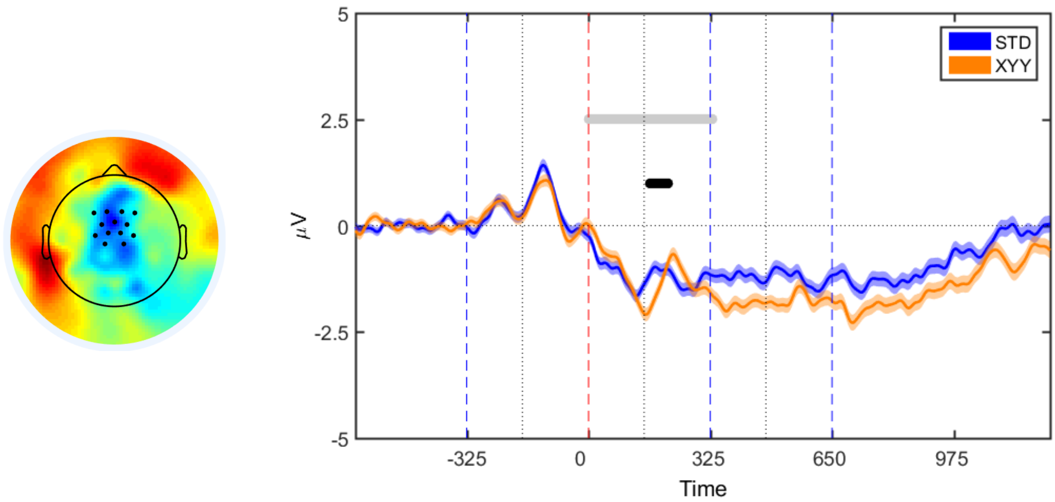

Figure 4.1: Experiment 2. Grand average over **Fronto-Central** ROI. **STD** vs **XYY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$. Left: Topography of the difference wave at the time point of the peak of the significant negative deflection.

Similarly, a negative deflection was also found for the **XXY** condition, in this case starting at 86ms ($t = 140.84$, $p < 0.001$). Once more, an examination of the topography of this deflection revealed a MMN response (Fig: 4.2).

Since we were able to establish the presence of a MMN response for both deviant types, we compared them to test whether a modulation as the one observed in **Experiment 1** was present (Fig: 3.5). As the point of deviance differed across conditions and the processing of a word has an intrinsic temporal dynamic, we eliminated this confounding factor by subtracting the activation elicited in the **STD** condition to each deviant condition and we align them by locking the signal to the point of deviance.

When compared, the amplitude of the MMN response triggered by the **XXY** condition proved to be bigger than the one elicited by the **XYY** condition ($t =$

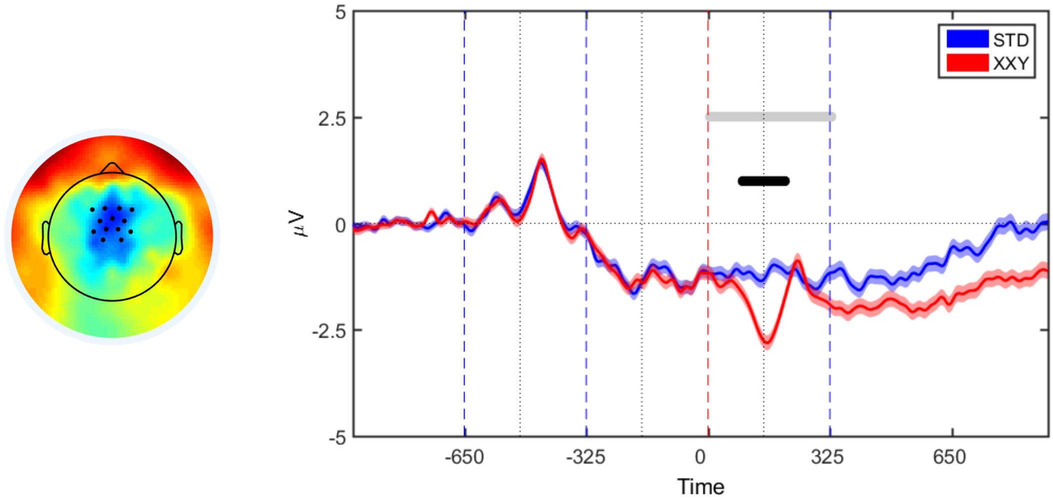

Figure 4.2: Experiment 2. Grand average over **Fronto-Central** ROI. **STD** vs **XXY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$. Left: Topography of the difference wave at the time point of the peak of the significant negative deflection.

-93.49, $p < 0.001$). This mirrors the effects found in **Experiment 1**. Even when instructions didn't distinguish between deviant and standard words, the occurrence of a word containing an unexpected sequence of phonemes triggered a MMN response, and this response was modulated by the number of syllables matching the **STD** word before the point of deviance (Fig: 4.3).

To summarise, both deviant types elicited a MMN response, whose amplitude was modulated by the number of syllables congruent with a **STD** word presented before the point of deviance. In contrast with the results of **Experiment 1** (3.7.2) no P3a was registered.

Parietal region of interest

In contrast with **Experiment 1**, we did not observe a P3b response. In the case of the **XYY** condition, we were only able to detect a small positivity, starting at 387ms, that approached significance (t = -30.74, p = 0.057) (Fig: 4.1).

Similarly, the **XXY** condition triggered a small but this time significant positivity starting at 410ms (t = -36.75, $p < 0.05$) (Fig: 4.5).

In the same way we did for the **Fronto-Central** ROI, we compared the re-

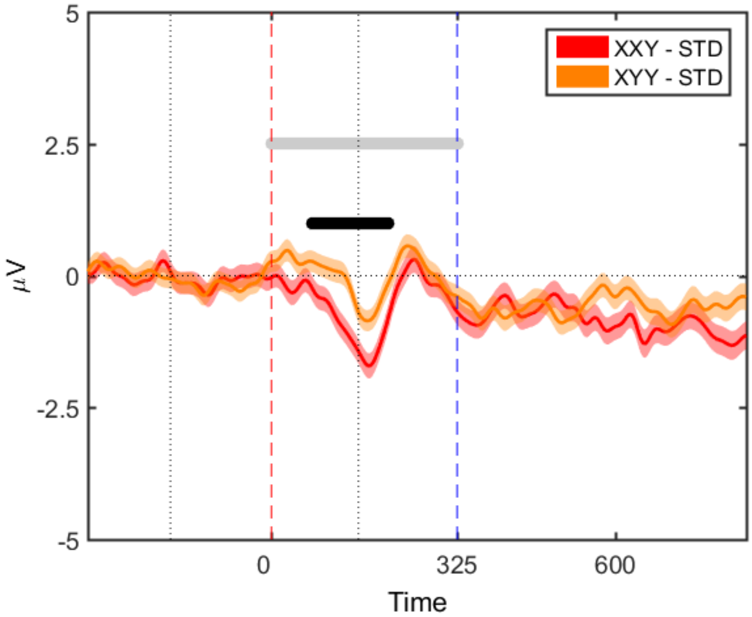

Figure 4.3: Experiment 2. Grand average over **Fronto-Central** ROI. **XYY** vs **XXY**. Trials were re-segmented and locked to the point of deviance, indicated by time zero. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line demarks $p < 0.01$.

sponses elicited by both deviant conditions in **Parietal** electrodes. We subtracted the activation elicited by the **STD** condition from each deviant type, and we align them by locking the signal to the point of deviance. This analysis didn't result in a significant difference.

4.8 Prediction error signals under different attentional sets

4.8.1 MMN response

In this chapter we have shown that even when listeners are not actively trying to detect mistakes in the incoming words, the presentation of a deviant containing an unexpected sequence of phonemes triggers a MMN response. This suggests that the human speech perception system generates predictions about the incoming phonemes of a word even when this is not required by the task at hand. But even if phonological predictions are deployed automatically, different attentional sets might

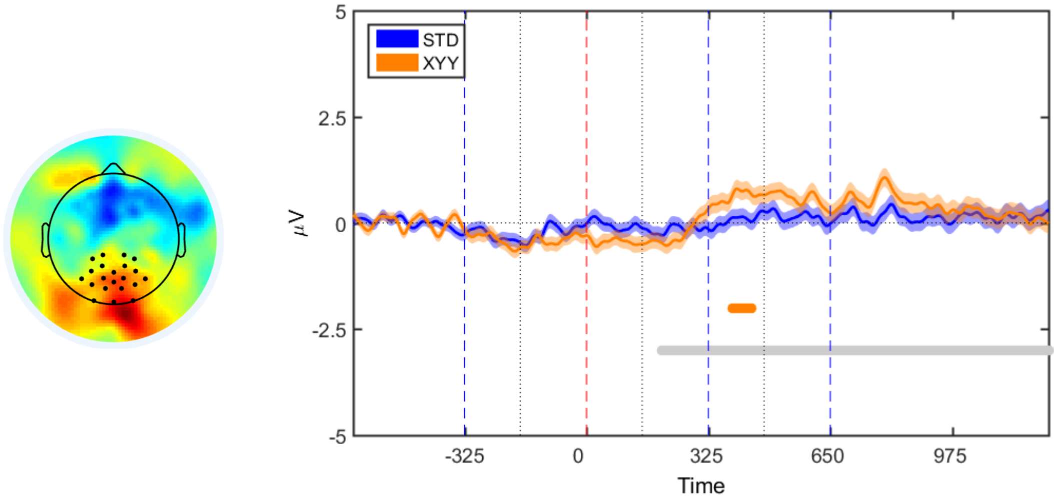

Figure 4.4: Experiment 2. Grand average over **Parietal** ROI. **STD** vs **XYY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Orange horizontal line demarks $p = 0.0576$. Left: Topography of the difference wave at the time point of the peak of the **XYY** condition.

have a modulating effect. For example, it has been shown when attention is focused in deviant detection, the amplitude of the MMN response can be reduced [20].

To test whether the attentional set of the listener could have a modulating effect in the strength of the predictions made about incoming phonemes in a word, we compared the 4 different MMN responses recorded in experiments 1 and 2. We performed a two-way ANOVA with a 2 x 2 design, with deviant type as a within participants factor with 2 levels (**XYY** and **XXY**), and attentional set as a between participants factor with 2 levels (”Count deviants” and ”Learn all words”). For each of the 4 conditions defined by this 2 x 2 design, taking the **Fronto-Central** ROI, we calculated for each trial, the mean amplitude in a 52ms window centred at the peak of the mean MMN response of that condition.

This analysis resulted in a main effect for deviant type (F(1,3486) = 20.41, $p < 0.0001$), no main effect of attentional set (F(1,3486) = 0.27, $p > 0.05$) and no interaction (F(1,3486) = 0.09, $p > 0.05$). In this way we can conclude that the number of syllables congruent with a **STD** word, presented before the point of deviance, modulates the amplitude of the MMN response, independently from task instructions. This suggests that when a word is presented, listeners can build a

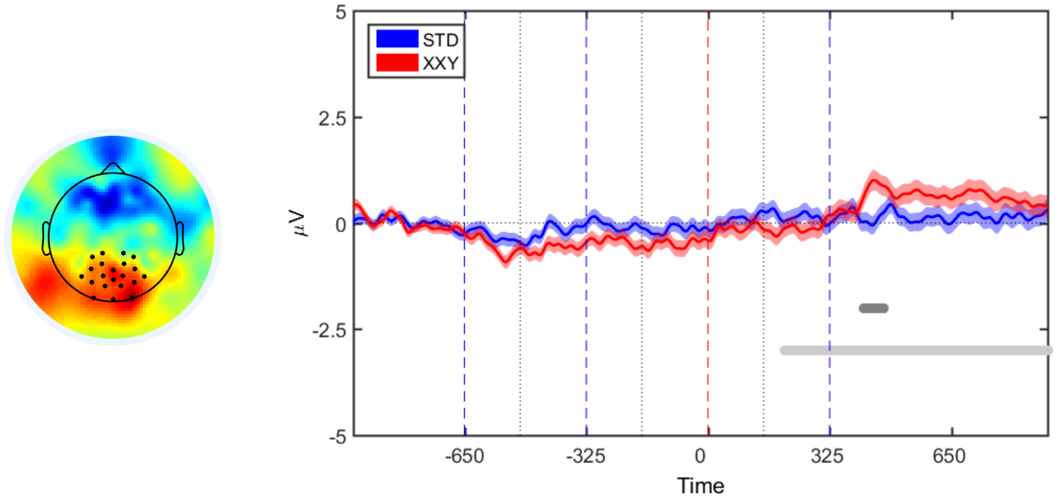

Figure 4.5: Experiment 2. Grand average over **Parietal** ROI. **STD** vs **XXY**. First, second and third vertical dashed lines indicate syllable onset times. Forth vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Middle grey horizontal line demarks $p < 0.05$. Left: Topography of the difference wave at the time point of the peak of the **XYY** condition.

predictive model of the incoming phonemes using at least the phonemes from the last 2 syllables, all this in an automatic way and independently from the task at hand. Figure 4.6 represents the mean MMN response amplitude for all the conditions.

4.8.2 P3a response

Contrary to our predictions, we didn't registered a significant P3a response in **Experiment 2**. To understand better the modulating role of attention on the this component, we performed a two-way ANOVA, as the one we performed for our analysis of the MMN response (4.8.1).

As before, we used a 2 x 2 design, with deviant type as a within participants factor with 2 levels (**XYY** and **XXY**), and attentional set as a between participants factor with 2 levels ("Count deviants" and "Learn all words"). For each of the 4 conditions defined by this 2 x 2 design, taking the **Parietal** ROI, we calculated for each trial, the mean amplitude in a 52ms window centred at the peak of the mean P3a response from **Experiment 1** and using the same time points for the data from **Experiment 2**.

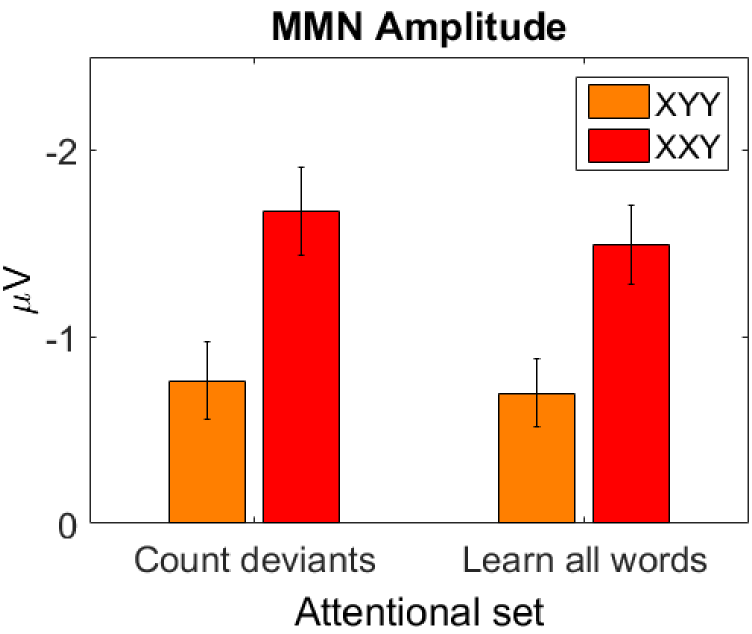

Figure 4.6: Comparison of MMN responses Error bars denote 1 SEM. As the ANOVA result indicates, the driving factor to determine the amplitude of the MMN response is the number of syllables that are congruent with a **STD** word, presented before the point of deviance. This modulation is independent from the instructions given to the participants with respect to the deviants

This analysis resulted in no main effect for deviant type ($F(1{,}3486) = 2.08$, $p > 0.05$), a main effect of attentional set ($F(1{,}3486) = 6.75$, $p < 0.01$) and no significant interaction ($F(1{,}3486) = 0.11$, $p > 0.05$). This implies that what we considered a P3a response in **Experiment 1**, is only present when participants are actively counting the occurrence of deviants (Fig: 4.7).

4.8.3 P3b response

Even though we did not registered a significant P3b response in **Experiment 2**, we did observe a small parietal positivity with a topography reminiscent of a P3b for both deviants (although it only approached significance for **XYY**). Because of this, in order to fully explore the modulating role of attention on the P3b response, we performed a two-way ANOVA for the **Parietal** ROI, as the one we performed for our analysis of the MMN response (4.8.1).

Once more, we used a 2 x 2 design, with deviant type as a within participants factor with 2 levels (**XYY** and **XXY**), and attentional set as a between participants

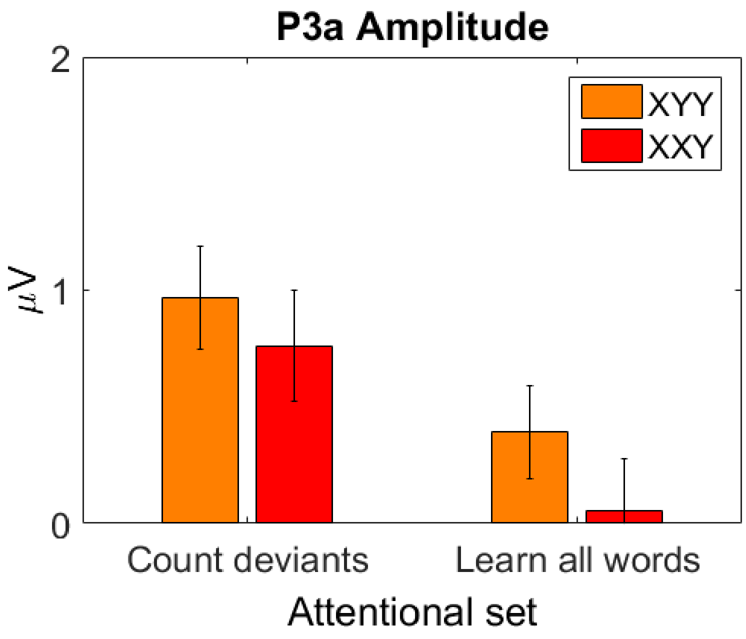

Figure 4.7: Comparison of P3a responses Error bars denote 1 SEM. What we considered a P3a response in **Experiment 1**, is only present when participants are actively counting the occurrence of deviants.

factor with 2 levels ("Count deviants" and "Learn all words"). For each of the 4 conditions defined by this 2 x 2 design, taking the **Parietal** ROI, we calculated for each trial, the mean amplitude in a 100ms window centred at the peak of the mean P3b response from **Experiment 1** and the peak of the clusters detected in **Experiment 2**.

This analysis resulted in a main effect for deviant type ($F(1{,}3486) = 4.52$, $p < 0.05$), a main effect of attentional set ($F(1{,}3486) = 54.33$, $p < 0.0001$) and a significant interaction ($F(1{,}3486) = 5.89$, $p < 0.05$). Looking at figure 4.8 we can conclude that while listeners are actively trying to detect deviants, is is possible to register a P3b that is modulated by the number of syllables congruent with a **STD** word before the point of deviance. When instead no particular instructions are given regarding the deviants, the P3b response disappears.

4.9 Discussion

As we argued in the introduction of this chapter (4.1), the main goal of **Experiment 2** was to study how a change in attentional set would affect the chain of predictive

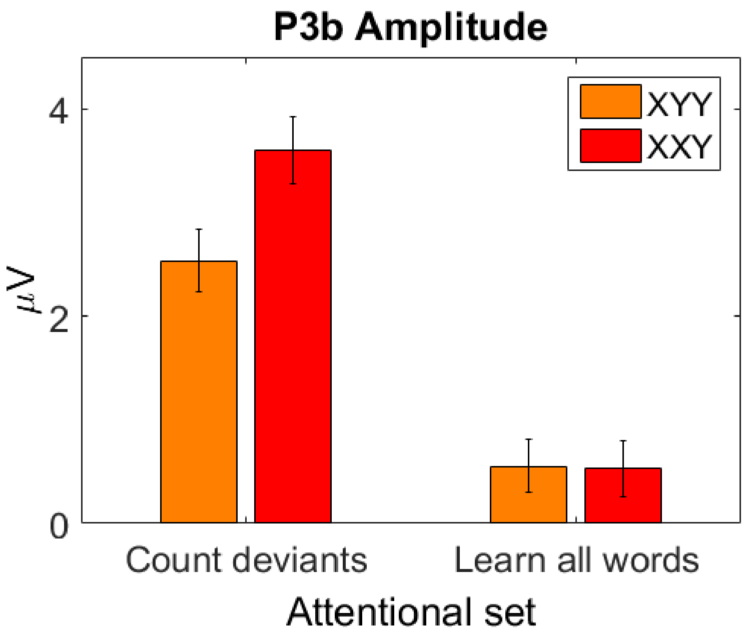

Figure 4.8: Comparison of P3b responses Error bars denote 1 SEM. As the ANOVA result show, in the case of the P3b response, the number of syllables that are congruent with a **STD** word, presented before the point of deviance, only have a modulating effect when listeners are actively counting the occurrence of deviants. When no instructions with respect to the deviants are given, there is no modulating effect of deviant type and the overall amplitude recorded at the **Parietal** ROI is much smaller.

error signals that are triggered by the presentation of a deviant word that contains an unexpected sequence of phonemes. We were particularly interested in testing if predictions about the forthcoming phonemes of a word would still be produced when task instructions don't distinguish between standard and deviant words. We were able to corroborate that even if participants are simply instructed to listen all words and try to learn them, the presentation of a deviant triggers a prediction error signal.

4.9.1 Phonological predictions within a word are performed automatically

In the experiment presented in this chapter, participants were not informed about the occurrence of deviant words, and the instructions given didn't distinguish between standard and deviant words. Despite this, the presentation of a word that contains an unexpected sequence of phonemes, elicited a MMN response. Furthermore, a comparison with the MMN responses elicited in **Experiment 1** showed that the responses recorded in **Experiment 2** had an equivalent amplitude (4.6).

In this way the results of **Experiment 2** not only confirm our previous findings but moreover they shows that listeners build predictions about the incoming phonemes of a word in an automatic way (i.e. even if is not required by the task at hand). As the attentional set held by the participants during this experiment resembles closely the one use in natural speech processing, these results implies that the language comprehension system proactively anticipates the incoming phonemes within individual words.

4.9.2 Modulation of the strength of the prediction

As in the case of **Experiment 1** (3.8.2) the amplitude of the MMN response elicited by a deviant was modulated by the number of syllables congruent with a **STD** word presented before the point of deviance, even if the listener is not actively trying to detect the deviants. This is of particular interest because it suggests that the prediction for the upcoming syllables integrates information from at least the two past syllables (i.e. 650ms in the past), and that this long range integration is performed automatically.

The behavioural results of **Experiment 2** are in line with the interpretation of a stronger prediction violation elicited by the **XXY** deviant. While participants were able to discriminate **XXY** deviants from new words in the forced choice test, they were not able to do it for the **XYY** deviants (4.7.1). We consider that **XXY** were more easily distinguishable from new words because as they generate more prediction error (i.e. MMN amplitude), there is a stronger effort to minimize that error, therefore they are encoded better and are easily remembered.

In brief, **XXY** deviants elicits a stronger MMN response and are more easily discriminable from new words. This suggests that predictions about the incoming phonemes within a word can integrate information from at least 2 syllables in the past. Is worth underlining that this predictions were produced in an automatic way, with new pseudowords that were learned in a period of minutes during the experiment and without the aid of higher level information such as syntax or semantics. This implies a formidable capacity and flexibility of the language system to learn the sequences of phonemes composing new words and generate predictions about upcoming phonemes based on the preceding ones.

4.9.3 Prediction error beyond the MMN response depends on listener's attentional set

In this chapter we have shown that the MMN response acts as an automatic prediction error signal elicited in response to a deviant that contains an unexpected sequence of phonemes (4.9.1) and that the predictive system that generates such response can integrate information from at least the 2 preceding syllables (4.9.2), all these even if listeners are not actively monitoring the words they hear looking for mistakes. On the contrary, when participants are not instructed to detect deviants, ERPs beyond the MMN response seem to vanish.

The absence of a P3a response went against our prediction. This ERP is usually elicited by distractors in 3 stimuli OddBall paradigms [91], and is thought to reflect automatic reorientation of attention. Given that we registered a P3a response in **Experiment 1** (3.7.2) even though the deviants in that experiment were targets to be counted, we considered that in **Experiment 2**, where from a top-down point of view deviants should receive the same attention as standard words, they were more likely to trigger an automatic bottom-up attentional shift. But instead, no P3a was present in **Experiment 2**. The fact that this positivity is only present when listeners are actively counting the occurrence of deviants calls for a revision of our labelling of it as a P3a response. In brief, this is a positive component peaking around 260ms after the point of deviance, which displays a fronto-central topography and is modulated by the attentional set of the participants, being present only when they are actively trying to count the occurrence of deviants. Unfortunately, with our current data, is not possible to further characterize this component or venture an interpretation of its behaviour.

Our ANOVA analysis of the P3b response (4.8.3) revealed that this component is present when participants actively try to count the occurrence of deviant words, but fades away when deviants receive the same treatment as standard words. A premature interpretation of this would be that this component is not an index of prediction, but instead only reflects the decision of categorizing a word as a deviant, or the result of the counting task itself. But it should be noted that when this component is present, is modulated by the amount of syllables congruent with a **STD** word presented before the point of deviance. Therefore, the P3b response can be considered as a marker of prediction in speech processing, but generated at a higher level of the processing hierarchy than the MMN response. As a high level event related potential [90], the P3b response can be modulated by different aspects of cognition, including predictive processes [50] and generation of responses [109].

4.9.4 Extending these results to a more complex scenario

The results of experiments 1 and 2 support our proposal that the human brain generates automatic on-line predictions about the identity of the forthcoming phonemes of words, even in the absence of higher level information such as syntax or semantics. Additionally our results suggest that the strength of such predictions can grow with the presentation of subsequent phonemes that confirm the prediction made. We were able to identify 2 different markers of prediction error that reflect different levels in the processing hierarchy. While the MMN response reflects predictive processes that are deployed automatically, the P3b response indexes higher level prediction that is contingent to the task at hand.

Having shown that the MMN and P3b responses elicited in an OddBall paradigm can be use to study prediction in speech processing, and considering that the amplitude of these components increased with subsequent successful predictions within a word, we decided to study another possible speech dimension that might modulate prediction strength. We refer to the frequency with which phoneme combinations are found in the lexicon. In the stimulus set used in experiments 1 and 2 (3.1) each syllable could be followed by 2 others, with 92% and 8% probability respectively. But in natural speech, each phoneme in a word can usually be followed by several competing candidates with different probabilities. In the next chapter we will present an experiment in which we used a richer stimulus set which included 2 possible deviants with different probabilities for the same syllable position. This allowed us, on one hand, to test whether prediction error signals can be modulated by how unlikely is a given syllable combination. On the other hand this allowed us to extend our results to a more complex stimulus set that better represents natural speech.

Chapter 5

Experiment 3

5.1 Introduction

One topic of intense debate in the general literature on prediction in speech processing is whether the predictive system generates a single prediction at a time, or if multiple parallel candidates can be established with different strengths. In general, older views of prediction consider that the predictive system picks only one candidate. The original explanations of the garden path phenomenon[1], for example, held that the reader predicted just one possible structure of the sentence, usually the most frequent and therefore the most likely structure [44]. Similar positions were held in early views of lexico-semantic prediction [43]. But this "single candidate" position was usually attached and used as an argument against prediction playing a major role in language processing. It was argued that given that any context have several possible continuations, betting on one prediction would be costly as it would usually fail [111].

In contrast with the "single candidate" view, prediction can be considered as a probabilistic phenomenon, in which different candidates are generated in parallel with different prediction strengths. In accordance with this view, evidence shows that predictive effects can be graded. The garden path effect, for example, can be modulated by how much a particular discourse context restrict possible candidates

[1]The *Garden Path Effect* refers to an increase in reading time in grammatically correct sentence that starts in such a way that a reader's most likely interpretation will be incorrect. Example: "The old man the boat".

[100]. In the same way, it is known that the amplitude of the N400 response elicited by an incoming word is inversely proportional to that word's probability considering the preceding context [70, 71]. In the example sentence *"He wanted to prepare coffee but he didn't have a clean..."*, while the word *cup* elicits a low amplitude N400 response, the word *knife* elicits a higher amplitude response, and importantly, the word *bowl* elicits a response with an amplitude that falls in between. One interpretation of this modulation is that while the word *cup* is predicted as the most likely candidate, the word *bowl* is also predicted, although with less prediction strength. These effects show that at least at the syntactic and semantic level, prediction strength can be modulated by contextual information in a graded fashion.

Although these examples of modulation of prediction refer to higher levels of speech processing, a similar modulation might be present at the phonological level. In natural speech, each phoneme of a word can be usually followed by several candidates, and such candidates occur with different frequencies. Therefore, considering the existing evidence in favour of graded prediction, it would be possible that the brain could build predictions that ponder different phoneme candidates, based on their frequency of appearance.

To test this, we performed an OddBall experiment in which we used 2 deviant types that differ in their probability of presentation. We expected that, if the phonological predictive system can keep track of different phoneme candidates, the magnitude of prediction error generated by an unexpected phoneme combination would be modulated by how unexpected (i.e. how infrequent) such combination is. In this way, a more unlikely deviant should generate more prediction error (i.e. higher amplitude MMN response), than a less unlikely deviant.

For this experiment, we used a stimulus set that is more complex than the one used in the experiments of chapters 3 and 4. Not only it contains more words, but also, as the point of deviance is the same for both deviant types, each syllable can be followed by 3 others, instead of 2. Therefore, finding evidence of prediction at the phonological level with this stimulus set would help us to extrapolate our results to natural speech.

In chapters 3 and 4, we have shown evidence that while hearing a word, listeners can make predictions about the forthcoming phonemes, based on the preceding ones, even in the absence of higher level information such as syntax or semantics. In particular, while the MMN response seems to act as an automatic prediction error signal, the P3b response can index higher level prediction that is strategic in nature (i.e. depends on the task at hand). One of the goals of the experiment presented in the current chapter was to test whether this predictions can actually serve a functional role by modulation behaviour. To test this, we introduced an on-line deviant detection task. Instead asking participants to count the occurrence of deviant words, we requested them to respond to the occurrence of deviants with a mouse click. This granted us a measure of reaction time that we could correlate with features of the MMN and P3b responses. Specifically, we expected that higher prediction error (i.e. higher ERP amplitude) would lead to shorter reaction times.

In brief, the current chapter presents an experiment that had 3 objectives. First, to test whether the phonological predictive system can keep track of how likely are different phoneme candidates. We accomplished this by including in our experimental design, deviants with different frequency of appearance. Second, the use of a more complex stimulus set (when compared to the one used for experiments 1 and 2) allowed us to test for the presence of prediction at the phonological level under conditions closer to natural speech. Finally, in order to test if the phonological predictive system can play a functional role, we performed a correlation analysis linking the amplitude of prediction error signals with the reaction time in an on-line deviant detection task.

5.2 Participants

A total of 34 participants took part in this experiment (13 male and 21 female, mean aged 24.88 ± 2.72 years). All of them were Italian native speakers, right handed and reported no auditory or language-related problems. Participants were recruited from the city of Trieste and received a monetary compensation of 15 €.

After data preprocessing, 3 participants were excluded from analysis due to contributing with less than 20 clean trials per condition, 4 participants were excluded from analysis due to missing more than 3 electrodes of interest and 2 participants were excluded due to poor behavioural performance. Therefore, 25 participants were included in the final analysis (9 male and 16 female, age 24.8 ± 2.44 years).

5.3 Stimuli

The stimuli of this experiment consisted of 3 sets of 4 bisyllabic words each, making a total of 12 words. Following the same procedure used for the construction of the stimuli for experiments 1 and 2 (see 3.3), we first constructed a set of 4 words based on which other 2 sets of 4 words were constructed using cross-splicing. In this 2 sets, each word was made of the first syllable from a word of the original set and the second syllable of another word, hence, they had a XY structure. The final product of this procedure were 3 sets of 4 bisyllabic words. Even though the words in all sets were composed of the same syllables, their combinations were different in each set. The 12 resulting words were checked by a native Italian linguist to ensure that they were plausible but not real Italian words.

	Set 1	**Set 2**	**Set 3**
1.	ʃaʤe	ʃatsi	ʃabo
2.	lutsi	lubo	luʣa
3.	moʣa	moʤe	motsi
4.	nibo	niʣa	niʤe

Table 5.1: Stimulus set use for**Experiment 3** in IPA notation. A total of 12 words were used. These words were divided in two sets. Sets 2 and 3 were produced by cross-splicing the 4 words belonging to Set 1.

The resulting stimulus set was more complex than the one used in experiments 1 and 2 (3.3). It was composed by 8 different syllables instead of 6, combined in 12 words instead of 6, and each syllable could be followed by 3 alternatives instead of 2. The higher variability in this stimulus set allowed us to test for the presence of phonological predictions under conditions that are more similar to natural speech.

5.4 Data acquisition setup

The experimental setup for EEG data acquisition was identical to the one use in experiments 1 and 2 (2.1). In addition, a Tobii T-120 Eye-tracker system was used to co-register pupil diameter, as we had plans for performing a pupillometry analysis that is not included in this thesis. In order to reduce artefacts participants were requested to minimize movement and use a chin rest positioned at 24cm from the screen throughout the experiment, except during breaks between blocks.

5.5 Procedure

After receiving instructions, the experimental session started with a 5 point calibration for the eye tracking system. Next, we registered resting state brain activity and pupil diameter in 2 blocks of 4 minutes, during which participants were instructed to fixate at a white cross presented centrally against a grey background. The analysis of this data will not be included in this thesis.

After resting state recording, the experiment followed an OddBall design. The 3 sets of words that we described in 5.3 were for each subject assigned randomly to one of 3 conditions (**STD**, **Dv1** and **Dv2**). Stimuli were presented in blocks with an average duration of 5.3 minutes. On each block, a total of 134 words were presented, with an inter stimulus interval that varied between 1800 and 2200ms. During the first of such blocks, only **STD** words were presented and participants were instructed to try to learn them. Subsequently, participants completed 12 blocks composed of 83% **STD** words 13% **Dv1** words and 4% **Dv2** words. Within each block, word order was pseudo-random, with the restrictions of no word repeating more than 2 times consecutively and having between 2 to 4 **STD** words between deviant words.

Participants were instructed to perform a mouse click each time that they heard a "mistaken" words (i.e. deviant words). We decided to include this on-line deviant detection task to have a measure of reaction time that we could correlate with features of the MMN and P3b responses. This allowed us to test if there is a functional link between prediction error signals and behavioural measures.

As in **Experiment 2** (4.5), at the end of the experiment, participants completed a forced choice test. Instructions were given verbally. On each trial, participants heard 2 words and were requested to choose the one that most likely was presented during the experiment. Participants completed 8 trials for each of 3 contrasts between conditions, for a total of 24 trials. The contrasts between conditions were "**STD** vs **Dv1**", "**STD** vs **Dv2**" and "**Dv1** vs **Dv2**". Participants reported their answers verbally and the experimented entered them through keypress.

5.6 EEG analysis

5.6.1 Data pre-processing

Details about the procedure used for EEG data preprocessing and cleaning can be found in 2.2. Data was segmented into 2700ms long epochs starting 300ms before word onset. After cleaning, data was divided into **STD**, **Dv1** and **Dv2** conditions. Given that this experiment included an on-line deviant detection task, we only included in the deviant datasets trials that were detected as such, and we only included in the **STD** dataset trials in which no false alarm occurred. The final result of this procedures was 1 dataset per condition, each containing 721 clean trials.

5.6.2 Regions of interest

The regions of interest used for this experiment were spatially identical to the ones used in **Experiment 1** (see 3.6.2). Temporally, the **Fronto-Central ROI** spanned over a 325ms starting at the point of deviance. With respect to word onset, this covered from 325ms to 650ms. This ROI coincided with the region were a MMN response [32, 77, 76] and a P3a response [21, 90, 32] could be expected. The **Parietal** ROI extended temporally from 200ms after the point of deviance, to the end of the epoch. With respect to word onset, this covered from 525ms to 2400ms. This ROI corresponded to the region were a P3b response would be expected [21, 90, 32] (Fig: 3.2).

5.6.3 Statistical analysis

Unless otherwise noted, statistical testing was performed utilizing the procedure described in detail in 2.3. This nonparametric clustering methods offers a straightforward and intuitive solution to the Multiple Comparisons problem.

5.7 Results

5.7.1 Behavioural Results

The behavioural performance of the participants was assessed in 2 different ways. During the blocks in which words were presented, participants reported the occurrence of deviants by performing a mouse click. This provided us an on-line measure of deviant detection. At the end of the experiment, participants performed a forced choice test that allowed us to measure preferences between the different conditions.

On-line deviant detection

Before analysing the on-line detection of deviant, we eliminated from the pool responses occurring before the point of deviance. These early responses represented 2.04% of the total.

Globally, participants had a mean hit rate of 90.61% (σ= 11.07%) and a mean false alarm rate of 2.05% (σ= 5.31%). We eliminated 2 participants from the subsequent analysis due to having a hit rate more than 2 σbelow the mean. Without these 2 participants, the mean hit rate was 92.63% (σ= 6.77%) and the mean false alarm rate was 1.38% (σ= 0.89%), which corresponds to a high discriminability index (*d'* = 3.65).

By evaluating both deviant types individually, we could determine that while **Dv1** had a hit rate of 91.46% (σ= 8%), **Dv2** had a hit rate of 96.14% (σ= 4.39%). To evaluate if this difference in detection rate was significant, we performed a custom surrogate test.

First, given that **Dv1** words were presented more frequently than **Dv2** words (13% and 4% respectively), we drawn form the pool of **Dv1** trials, a random subset matching the number trials in the **Dv2** condition. Next we compared this two sets by means of a surrogate difference of means test with 5000 permutations. On each permutation, the labels of the conditions were shuffled the difference of the means of the shuffled conditions measured, generating in this way a distribution of possible differences of mean under the null hypothesis of no difference between conditions. A p value was calculated as the proportion of difference of means under the null hypothesis that were bigger that the one observed in the original data. We repeated this procedure of drawing a random sample of **Dv1** trials and performing the surrogate test 1000 times, and the final p value was defined as the mean p value of these 1000 iterations. In this way, we could determine that the hit rate of **Dv2** deviants was significantly higher ($p < 0.001$).

Following this, we analysed the reaction time for both deviant types. As the distribution of reaction times was not normal according to a Lilliefors test (D(9886) = 0.07, $p < 0.0001$), median was used as a measure of central tendency for the following analysis. Participants responded with median latencies of 788ms and 734ms after the point of deviance for the **Dv1** and the **Dv2** conditions respectively. In this way participants responded 54ms faster in the **Dv2** condition. To determine if this difference was significant, we followed a statistical procedure similar to the one described in the preceding paragraph, but this time using difference of medians instead of difference of means. The observed difference of median response times resulted to be significant ($p < 0.001$), implying that participants responded faster to the deviants that are presented less frequently (Fig: 5.1).

From these analysis of the on-line deviant detection task we can conclude that, although the performance was close to ceiling for both deviant types, the **Dv2** deviants, which were presented with less frequency, were detected faster and with higher accuracy than the **Dv1** deviants.

Forced choice test

The forced choice test performed at the end of the experiment allowed us to determine if participants had a preference for any condition in terms of considering it as belonging to the words presented during the blocks. All results presented in this section were corrected for multiple comparisons using the Bonferroni-Holm method (See table 5.2).

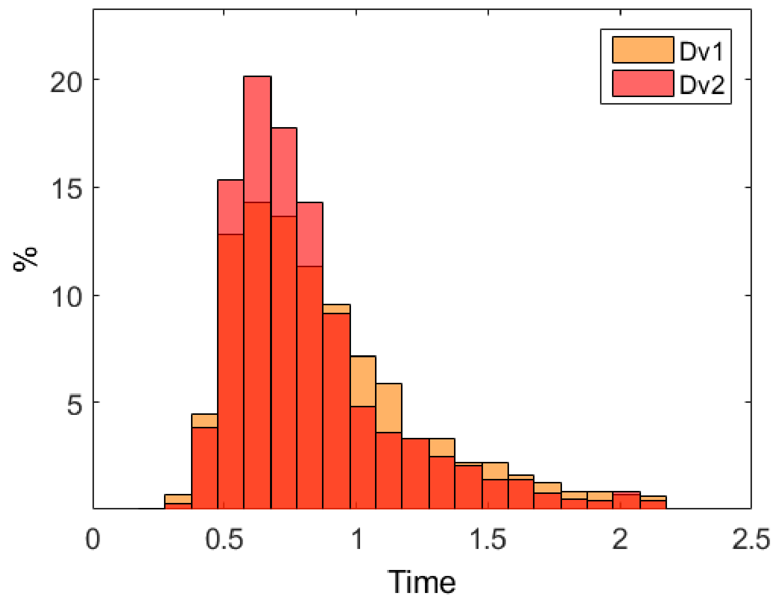

Figure 5.1: Experiment 3. Histogram of response times for **Dv1** and **Dv2** conditions. X axis represents time in seconds after point of deviance. Y axis represents percentage of responses.

Our analysis revealed that participants had a clear preference for **STD** words over **Dv1** deviants (87.13%, $t(33) = 11.81$, $p < 0.0001$) and over **Dv2** deviants (91.18%, $t(33) = 12.4$, $p < 0.0001$).

Besides showing a preference for Standard words over deviant words, participants also displayed a preference for **Dv1** deviants over **Dv2** deviants (59.93%, $t(33) = 3.33$, $p < 0.01$).

Finally we performed a surrogate t test to compare the preference for **STD** words over **Dv1** deviants vs the preference for **STD** words over **Dv2** deviants. This comparison only approached significance ($t(33) = -2.06$, $p = 0.067$).

In brief, from these analysis we can conclude that participants easily discriminate **STD** words from deviants of both types. Additionally, they preferred the most frequent type of deviant (**Dv1** over **Dv2** deviants). These results imply that participants could keep track of the frequency of presentation of all stimuli categories, and that the likelihood of considering a word as one of the words presented during the experiment depended on its frequency of appearance.

Contrast	Pref.	Statistics
STD vs **Dv1**	87.13%	$t(33) = 11.81$ $p < 0.0001$
STD vs **Dv2**	91.18%	$t(33) = 12.40$ $p < 0.0001$
Dv1 vs **Dv2**	59.93%	$t(33) = 3.33$ $p < 0.01$

Table 5.2: Experiment 3 Behavioural Results. A forced choice test was performed contrasting each condition against the other. The mean preference in each contrast was calculated for each participant and a one sample t test was performed at the group level to test against the null hypothesis of no difference from chance. Results were corrected for multiple comparisons using the Bonferroni-Holm method.

5.7.2 EEG Event Related Potentials

In this section we present the results of our event related potential analysis. As in the previous experiments, we will use the point of deviation as time zero.

Fronto-Central region of interest

The analysis of the **Fronto-Central** ROI (5.6.2) revealed that, as in the two preceding experiments, the presentation of a deviant elicited a MMN response. When contrasting the **STD** and **Dv1** conditions, we could determine that the presentation of the **Dv1** deviant elicited a significant negativity starting at 47ms ($t = 28.14$, $p < 0.05$). The topography of the difference between conditions reveals a fronto-central negativity congruent with a MMN response. Following this first negativity, we identified a second one starting at 251ms ($t = 38.88$, $p < 0.05$) (Fig: 5.2).

The results for the comparison between the **STD** and **Dv2** conditions mimic the ones obtained for the **Dv1** condition. The presentation of the **Dv2** deviant elicited a significant negativity starting at 47ms ($t = 118.8$, $p < 0.001$). Once more the topography of the difference between conditions reveals a fronto-central negativity congruent with a MMN response. A second negativity starting at 207ms was also detected ($t = 77.2$, $p < 0.01$) (Fig: 5.3).

As our analysis determined that both deviant conditions elicited a MMN response, we proceeded to compare the deviant conditions to test whether the amplitude of this response was modulated by the frequency of the deviants. We predicted that the least frequent **Dv2** deviant would trigger a stronger MMN response. In

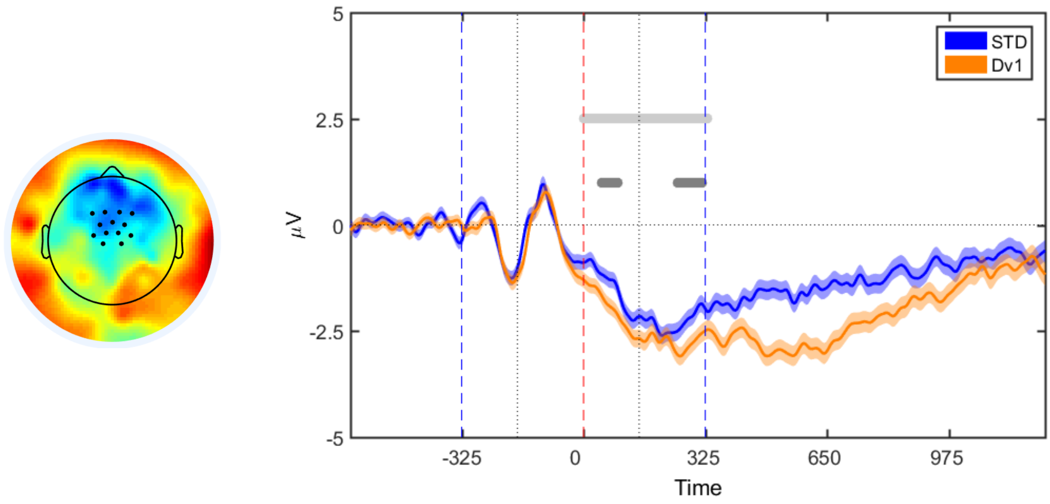

Figure 5.2: Experiment 3. Grand average over **Fronto-Central** ROI. **STD** vs **Dv1**. First and second vertical dashed lines indicate syllable onset times. Third vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Middle grey horizontal line demarks $p < 0.05$. Left: Topography of the difference wave at the time point of maximum difference in the first significant cluster.

contrast with this predictions, no significant difference between the MMN responses elicited by **Dv1** and **Dv2** was detected. However is worth noticing that the MMN response elicited by the **Dv2** condition is numerically more negative and extends longer over time (Fig: 5.4).

Parietal region of interest

The analysis of the **Parietal** ROI revealed that both deviant categories elicit a P3b response. The presentation of a **Dv1** deviant evoked a long lasting positive deflection starting at 247ms and extending till the end of the epoch (t = -2692.8, $p < 0.001$) whose topography corresponds to a P3b response (Fig: 5.5).

In the same way, the **Dv2** deviant evoked a long lasting positivity starting at 315ms and extending till the end of the epoch (t = -2380.1, $p < 0.001$), once more corresponds to a P3b response (Fig: 5.6).

Since our analysis showed that both deviant conditions elicited a P3b response, we compared across them to test if the frequency of the deviants had a modulatory effect. This analysis revealed a modulation, but instead of finding a higher amplitude for the **Dv2** condition as predicted, we found that the P3b response elicited by the

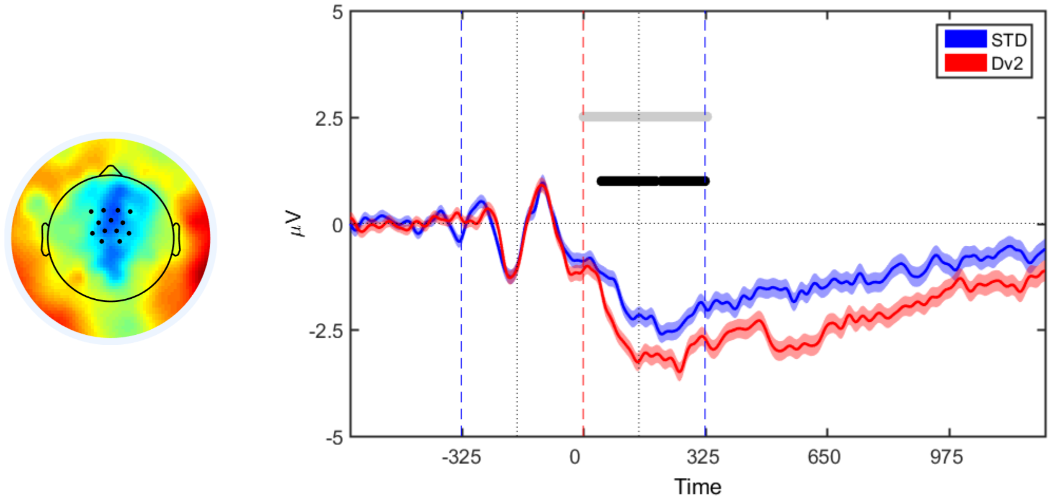

Figure 5.3: Experiment 3. Grand average over **Fronto-Central** ROI. **STD** vs **Dv2**. First and second vertical dashed lines indicate syllable onset times. Third vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line denotes $p < 0.01$. Left: Topography of the difference wave at the time point of maximum difference in the first significant cluster.

Dv1 deviants started earlier than the one corresponding to the **Dv2** deviants ($t = 122.58$, $p < 0.01$). In terms of amplitude, both conditions peaked to the same level (Fig: 5.7).

Next, we compared the responses elicited by both deviant types, but this time, instead of locking to the point of deviance, we locked the signal to the point of response. This allowed us examine the temporal dynamics of the processing of both deviant types before response.

To do this, first we eliminated trials in which the response occurred either earlier than 780ms or later than 2000ms with respect to word onset. We did this to ensure that the remaining trials could be re-segmented from -1000 till 300ms with respect to the time of response. We were able to retain 676 trials per condition (94%) for subsequent analysis.

We could observe that for both deviants, the elicited P3b peaked precisely at the time of response and reached the same final amplitude. As in our comparison of the P3b component locked to the point of deviance, the analysis locked to the time of response showed that even thought both P3b reached the same level, the component elicited by the **Dv1** deviants starts to rise earlier. This can be seen

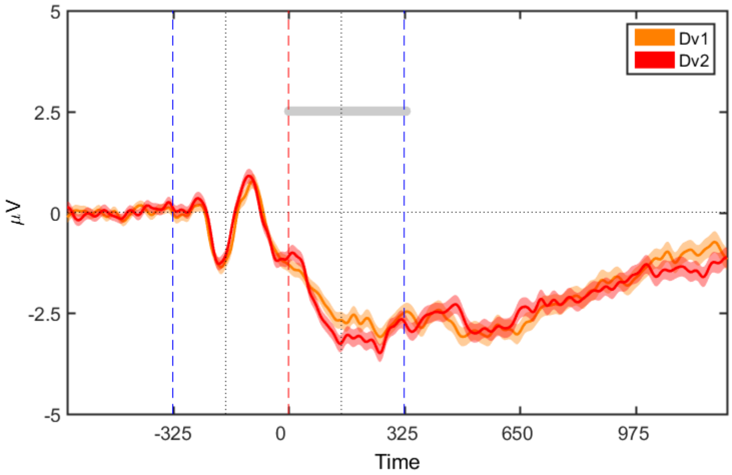

Figure 5.4: Experiment 3. Grand average over **Fronto-Central** ROI. **Dv1** vs **Dv2**. First and second vertical dashed lines indicate syllable onset times. Third vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest.

during 2 consecutive significant temporal clusters, the first one starting 648ms before the response ($t = 104.98$, $p < 0.01$) and the second one starting 468ms before the response ($t = 118.03$, $p < 0.01$) (Fig: 5.8).

5.7.3 Linking prediction error to behaviour

As we anticipated in the introduction of this chapter, one of the goals of **Experiment 3** was to determine if the prediction error signals elicited by the presentation of a deviant composed by an unexpected sequence of phonemes, can have a functional role, by modulating behavioural responses. To answer this question, in this subsection we present a correlation analysis between reaction times and features of the MMN and P3b responses. All results presented in this subsection were corrected for multiple comparisons using the Bonferroni-Holm method.

Fronto-Central region of interest

If as we propose, the MMN response is acting as an error signal that the brain is using to determine if a prediction failed, the amplitude of this component could be correlated with the reaction time in the on-line deviant detection task.

To test for this possibility, we first defined MMN amplitude as the mean am-

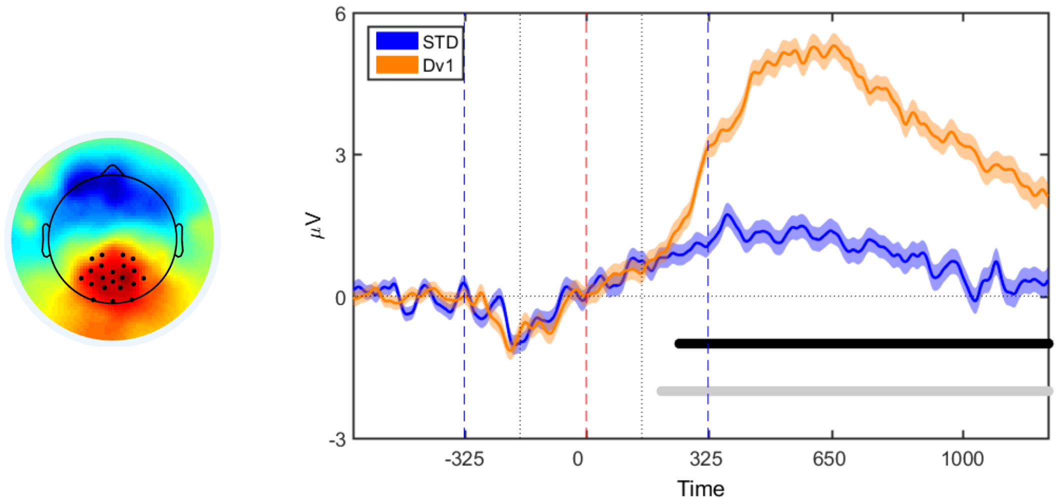

Figure 5.5: Experiment 3. Grand average over **Parietal** ROI. **STD** vs **Dv1**. First and second vertical dashed lines indicate syllable onset times. Third vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line denotes $p < 0.01$. Left: Topography of the difference wave at the time point of the peak of the **Dv1** condition.

plitude in a 148ms time window starting at the beginning of the significant cluster found for both deviant conditions (i.e. 47ms after the point of deviance). Next, we removed from analysis trials that contained a response before the point of deviance, or that had an amplitude 2 σ away from the mean. In this way, we kept 682 (94%) trials for the **Dv1** condition and 685 (95%) trials for the **Dv2** condition. As the distribution of the response times was not normal according to a Lilliefors test (D(9886) = 0.07, $p < 0.0001$) (5.1) we log transformed the reaction time data and we performed a Pearson correlation between response times and MMN amplitude for each condition.

We found a small, yet significant positive correlation for both, the **Dv1** (rho(682) = 0.10 $p < 0.05$) and **Dv2** (rho(685) = 0.10 $p < 0.05$) conditions. More negative (i.e. higher amplitude) MMN responses were associated with shorter reaction times. This implies that the stronger the error signal elicited by a deviant, the faster the participants can classify it as such and produce a response (Figure: 5.9).

Even though these correlations might appear small, it should be considered that individual EEG trials (as the ones used in this analysis) are noisy, as it can be appreciated in the broad range of amplitudes displayed by individual trials. To

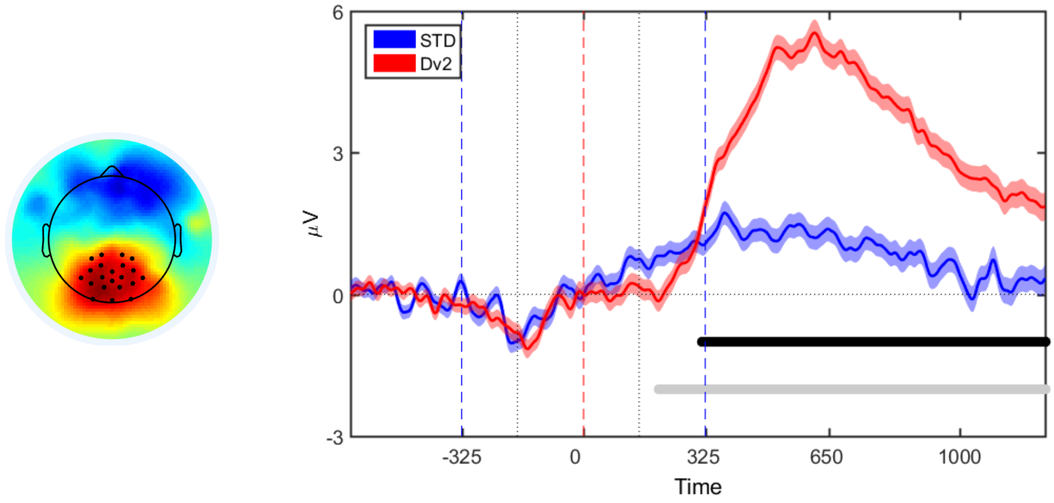

Figure 5.6: Experiment 3. Grand average over **Parietal** ROI. **STD** vs **Dv2**. First and second vertical dashed lines indicate syllable onset times. Third vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line denotes $p < 0.01$. Left: Topography of the difference wave at the time point of maximum difference at the first significant cluster.

reduce this noise, we sorted individual MMN values according to the reaction time in the respective trial, and we divided them in 22 groups of 30 trials each. In this way, MMN values recorded at trials with similar reaction times were kept together. Next, we obtained the mean MMN value and reaction time from each group, and performed a Pearson correlation for this mean values.

Once more, we found a significant positive correlation for both, the **Dv1** (rho(22) = 0.56 $p < 0.05$) and **Dv2** (rho(22) = 0.52 $p < 0.05$). More negative (i.e. higher amplitude) MMN responses were associated with shorter reaction times (Figure: 5.10).

Parietal region of interest

We tested for a correlation between P3b amplitude and reaction time, using the same method described for the analysis of the MMN response. We first defined P3b amplitude as the mean amplitude in a 300ms time window centred at the peak of the P3b responses found in both deviants (467ms to 767ms from the point of deviance).

As before, we removed from analysis trials that contained a response before the point of deviance, or that had an amplitude 2 σ away from the mean. In this way

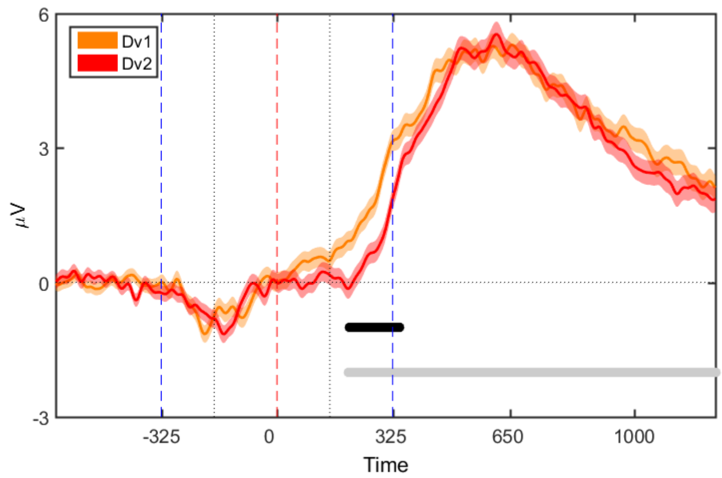

Figure 5.7: Experiment 3. Grand average over **Parietal** ROI. **Dv1** vs **Dv2**. First and second vertical dashed lines indicate syllable onset times. Third vertical dashed line indicates end of the word. Time zero indicates the point at which deviance occur. Error bars denote 1 SEM. Horizontal light grey line delimits time window of interest. Black horizontal line denotes $p < 0.01$.

689 (95%) trials for the **Dv1** condition and 685 (95%) trials for the **Dv2** condition were kept. Once more we log transformed the reaction time data and we performed a Pearson correlation between response times and P3b amplitude for each condition.

This analysis resulted in a highly significant negative correlation for both, **Dv1** (rho(689) = -0.21 $p < 0.0001$) and **Dv2** (rho(685) = -0.21 $p < 0.0001$) conditions. More positive (i.e. higher amplitude) P3b responses were associated with shorter reaction times (Figure: 5.11).

Using the same procedure we followed for the analysis of the MMN responses, in order to reduce noise, we sorted individual P3b values according to the reaction time in the respective trial, and we calculated 22 mean amplitude values.

A significant negative correlation for both, the **Dv1** (rho(22) = -0.62 $p < 0.01$) and **Dv2** (rho(22) = -0.64 $p < 0.01$) conditions was found. More positive (i.e. higher amplitude) P3b responses were associated with shorter reaction times (Figure: 5.12).

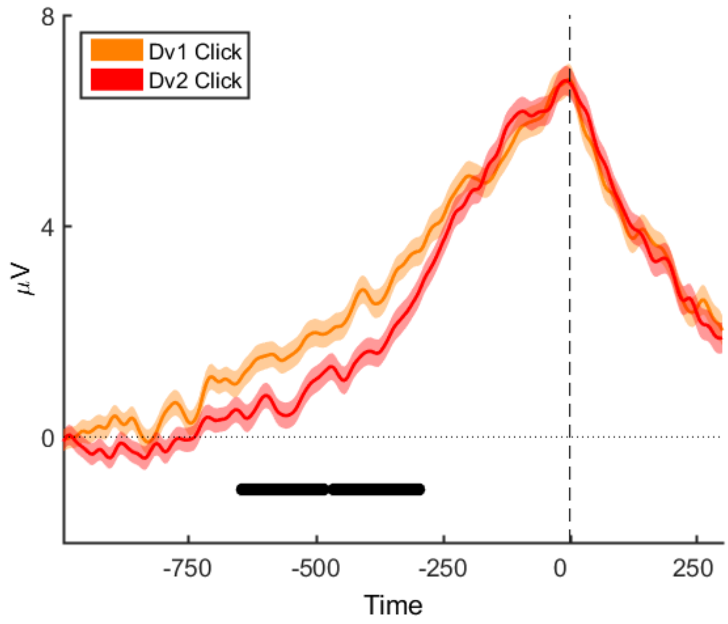

Figure 5.8: Experiment 3. Grand average over **Parietal** ROI. **Dv1** vs **Dv2**. Signal locked to response. Error bars denote 1 SEM. Black horizontal line denotes $p < 0.01$. Black dashed vertical line indicates the moment of response.

5.8 Discussion

5.8.1 Can the phonological predictive system evaluate more than one candidate in parallel?

In the introduction of this chapter, we argued that evidence suggests that prediction in speech can be considered as probabilistic phenomenon. Instead of predicting only the most probable percept, different candidates could be evaluated in parallel and weighted with different strength. For example, it is well established that the amplitude of the N400 response is modulated by the probability of a word given the previous context [70, 71]. This can be taken as evidence of evaluation of different candidate words in parallel. The first goal of the present experiment was to test whether the phonological predictive system can evaluate different phoneme candidates in parallel in a similar fashion.

To test this, we included in our OddBall design, 2 deviant categories that differed in their frequency of presentation. Specifically, we predicted that the **Dv2** deviants, which were presented with a frequency of only 4%, would elicit higher amplitude prediction error signals than the **Dv1** deviants, which were presented in 13% of the

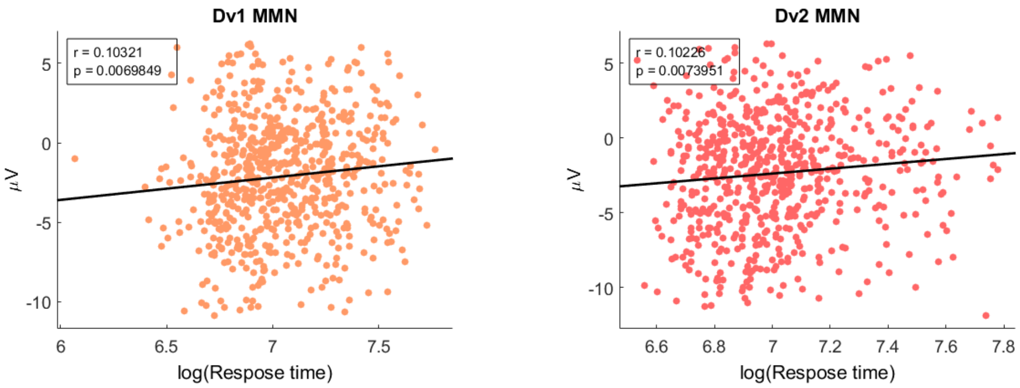

Figure 5.9: Experiment 3. Correlation between MMN amplitude and reaction time for **Dv1** (Left) and **Dv2** (Right) conditions. A significant positive correlation was found for both conditions, implying that higher amplitude MMN responses were associated with faster reaction times.

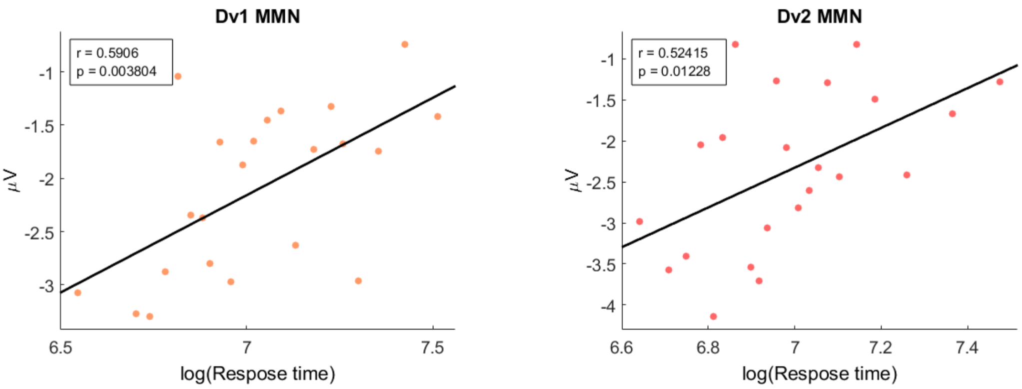

Figure 5.10: Experiment 3. Correlation between MMN amplitude and reaction time for **Dv1** (Left) and **Dv2** (Right) conditions, taking mean MMN amplitudes of blocks of 30 trials with similar reaction time. A significant positive correlation was found for both conditions, implying that higher amplitude MMN responses were associated with faster reaction times.

trials.

Behaviourally, during the on-line detection task, the more infrequent **Dv2** deviants were detected faster and with higher accuracy than the **Dv1** deviants. In line with this, the results of the forced choice test (5.7.1) at the end of the experiment showed that participants considered **Dv1** deviants (i.e. the most frequent deviatn) as more likely to be among the words presented during the experiment, when compared to **Dv2**. This implies that participants were able to keep track of the frequency of presentation of the different deviant conditions. But despite this, although the MMN response elicited by the **Dv2** deviants was numerically more

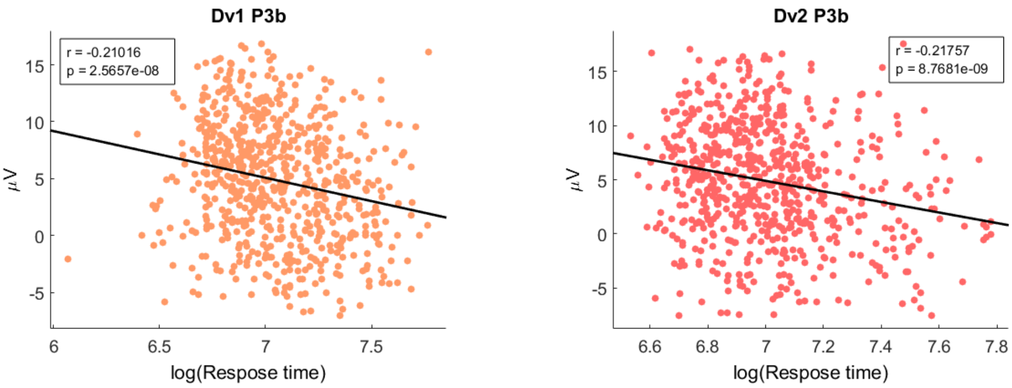

Figure 5.11: Experiment 3. Correlation between P3b amplitude and reaction time for **Dv1** (Left) and **Dv2** (Right) conditions. A significant negative correlation was found for both conditions, implying that higher P3b amplitude were associated with faster reaction times.

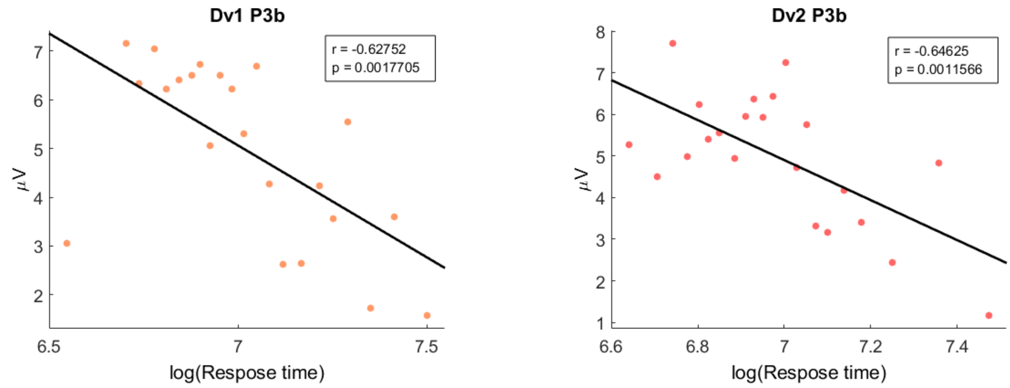

Figure 5.12: Experiment 3. Correlation between P3b amplitude and reaction time for **Dv1** (Left) and **Dv2** (Right) conditions, taking mean P3b amplitudes of blocks of 30 trials with similar reaction time. A significant negative correlation was found for both conditions, implying that higher amplitude P3b responses were associated with faster reaction times.

negative and more spread in time, a comparison of both deviants didn't reach a significant difference. A simple reason for this could be that the difference in the frequencies of presentation used was of only 9%, which might not be enough to generate a difference in MMN amplitude that could be captured in an EEG experiment. A different rate of presentation, such as presenting **STD** words in 65% of the trials, **Dv1** deviants in 30% of the trials and **Dv2** deviants in 5% of the trials could result in a significant modulation by frequency of presentation. Because of this, we consider that the data provided by **Experiment 3** is inconclusive with respect to this matter.

In the case of the P3b response, we did observe a modulation by frequency of presentation, but this modulation was in the opposite direction of our predictions. We expected that the most rare **Dv2** deviant would elicit a higher amplitude P3b response, but instead the data showed that although the peak amplitude was the same for both conditions, this positivity starts earlier in the case of the **Dv1** deviant (5.7). This earlier start of the P3b in the more frequent **Dv1** condition becomes more evident when the signal is locked to the moment of the response (5.8). Considering that we didn't find a significant modulation of the MMN response, and that the difference in frequency between deviant conditions used was of only 9%, the interpretation of this P3b modulation remains elusive.

5.8.2 Functional role of prediction error. Correlations with response time

Given that in chapters 3 and 4 we were able to show that the presentation of words containing an unexpected sequence of phonemes elicit ERPs such as the MMN and P3b responses, which could be interpreted as prediction error signals, one of the goals of the experiment presented in the current chapter was to test whether this prediction error can serve a functional role by modulation behaviour. To test this, we measured reaction times by asking participants to response with a mouse click each time a deviant was presented and we correlated this behavioural measure with the amplitude of the MMN and P3b responses. Specifically, we expected that higher prediction error (i.e. higher ERP amplitude) would lead to shorter reaction times.

The correlation analysis confirmed our predictions, showing that higher amplitudes of both MMN and P3b were associated with shorter reaction times. This is particularly interesting in the case of the MMN response. As we showed before (4.8), this component indexes predictions that are deployed automatically, even if participants are not instructed to perform any task with respect to the deviants. The correlation analysis presented in this chapter suggests that despite this automaticity, the amount of prediction error elicited by an unexpected sequence of phonemes can affect higher level processes such as stimuli classification. By showing that the amplitudes of the MMN and P3b responses are correlated with reaction time, this results support our proposal which states the phonological predictions deployed during the perception of individual words have a functional role.

5.8.3 Phonological prediction in a more complex stimulus set

The stimulus set used in this experiment (5.3) allowed 3 alternatives after the first syllable of each word. This alternatives were presented with different frequencies, with one being the **STD** (83%), one being a more common deviant (**Dv1** 13%) and the last one being a more rare deviant (**Dv2** 4%). In this way, the stimulus set used in this experiment resembled natural speech, in which usually each syllable of a word can be followed by several alternatives.

Even with a more complex stimulus set like the one used for this experiment, when a word with an unexpected sequence of phonemes is presented, this elicits a MMN response that can be interpret as an error signal, product of the violation of a prediction [116]. In this way, the results of the experiment presented in this chapter further supports our proposal which states that in speech processing, predictions are deployed at the phonological level, even if the context does not include higher level information.

Chapter 6

General Discussion

6.1 Introduction

As we argued in the introduction of this thesis 1, one of the strategies that the speech processing system uses to complete the difficult task of handling the variability and noisiness of the speech signal, is to generate predictions [67]. Predictions can be generated at different levels [68], and as extensive communication occurs between levels [34], information from any particular level can facilitate the processing of incoming information at almost any other level [3, 98, 12, 11].

In the case of the phonological level, evidence show that the forthcoming phonemes of a word can be predicted using syntactic [28], semantic [10, 66, 57] and phonotactic [23, 104, 119] information. We proposed that as prediction seems to plays a fundamental role in speech processing [60, 80, 12], phonological prediction might be possible even in the absence of this high level information. To test this, we performed 3 electroencephalography experiments in which only phonological and word form information was available to generate phonological predictions.

6.2 MMN as a marker of phonological prediction

Throughout the 3 experiments described in this thesis, the presentation of an unexpected sequence of phonemes, reliably elicited a MMN response [77, 76]. This event related potential is a well established marker of violation of expectations which can be interpreted as the result of comparing a prediction with the actual bottom-up input. In this way, the MMN response would act as prediction error that the system has to minimize [20, 113, 54, 50].

Importantly, the deviant words used in these experiments were constructed by cross-splicing standard words. Therefore, each syllable in a deviant word was acoustically identical to a syllable in a standard word. This represents an advantage with respect to previous works on the topic of phonological prediction, where the phoneme to be predicted was either omitted [10], or replaced by a non-linguistic sound [66, 57]. In the case of the stimuli used in our experiments, the only feature that defines a phoneme as deviant, was that given the phoneme X_n, instead of the usual phoneme X_{n+1}, the phoneme Y_{n+1}, which belongs to a different word, was presented. In this way, the MMN response registered in these experiments, is not elicited by a change in instantaneous low level auditory features, but by the violation of an abstract rule [84].

The amplitude of this MMN response was modulated by the number of syllables congruent with a standard word presented before the point of deviance, with higher MMN amplitude (i.e. more prediction error) when 2 congruent syllables were presented instead of 1. This implies that phonological predictions can be constructed using information from (at least) the 2 preceding syllables.

Experiments 1 and 2 differed in the attentional set held by the participants. While in **Experiment 1** participants were instructed to count the occurrence of "mistaken" (i.e. deviant) words, in **Experiment 2** they were not informed about the occurrence of deviants and were simply instructed to attend to all the words. Our analysis showed that the MMN responses elicited under these two different attentional conditions were indistinguishable (4.8.1), which suggests that phonological predictions are deployed automatically, even if the task at hand does not require detecting abnormalities in the speech stream.

Finally, besides determining that the presentation of an unexpected sequence of phonemes generates a predictive error signal, in **Experiment 3** we were able to show that this prediction error can influence behaviour, as we found that higher amplitude MMN responses were correlated with faster reaction times in an on-line deviant detection task.

Taken together, the MMN effects found in the 3 experiments presented in this thesis, suggest that even when no high level linguistic information such as syntax and semantics is present, the human speech comprehension system can use phonological information from (at least) the past 2 syllables to deploy phonological predictions in an automatic way

6.3 The P3b. A marker of Phonological prediction beyond the MMN?

The P3b response is a high level and relatively late component which amplitude and latency are sensitive to several experimental parameters. This component has been reported to reflect perceptual stimulus discrimination, memory processing [21, 90], participants' attentional set [90] and conscious access [85], although it can also reflect subliminal processing [97]. Is modulated by task features such as stimulus dissimilarity [31] or target to target interval [55], as well as response features like accumulation of evidence [81, 109]. Considering this, the interpretation of such a multiphasic component can be challenging.

In **Experiment 1**, participants were instructed to count the occurrence of "mistaken" (i.e. deviant) words. Under such instructions, the P3b response amplitude (as in the case of the MMN response) was modulated by the number of syllables congruent with a standard word presented before the point of deviance. This suggests that this component is indeed related to prediction, as its amplitude is enhanced by a previous successful prediction.

A comparison between experiments 1 and 2 (4.8.3) revealed that this component is modulated by attentional set. In the case of **Experiment 2**, where instructions did not differentiate between deviant and standard words, this component was not registered. This could imply that although the P3b response can act as a prediction error signal, it might only be present when predictions are strategic for the task at hand.

But in the case of **Experiment 3**, this component behaved in the opposite direction than the one predicted (5.7.2). We expected that, if this component would act as a prediction error signal, the **Dv2** words, that were presented in only 4% of the trials, should elicit a higher amplitude response than the **Dv1** words, which had a presentation frequency of 13%. Instead, this component reached the same final amplitude for both conditions, and it even ramped faster for the **Dv1**.

Further analysis showed that this component is maximal when the signal is locked to the moment of response, therefore, we can consider that in the case of **Experiment 3**, where participants were requested to respond to the presentation of deviant with a mouse click, it acted as a decision variable that initiates response when a threshold is reached [81, 109]. One possibility is that in this experiment, participants might have use different strategies to respond to the 2 different deviant types. While **Dv2** words were infrequent enough to be treated as deviants, the **Dv1** words could have been frequent enough to be learned, and considered not as deviant, but as a target to which to respond.

As we argued at the beginning of this section, the P3b response is sensitive to several experimental parameters. Even if in **Experiment 1**, where participants were asked to count the occurrence of deviants, this component seems to act as a predictive error signal, the change of response type to mouse click and the different type of deviant manipulation in **Experiment 3** might make impossible a comparison with experiments 1 and 2. A likely scenario is that the P3b component might reflect both aspects of prediction and of the way in which participants are executing the response. Given that the deviant manipulation of **Experiment 1**, which consisted in changing the number of syllables congruent with a standard word presented before the point of deviance, was successful in modulating the P3b response, one possible way to shed light on this topic would be to replicate **Experiment 1**, although using the same type of response as in **Experiment 3**.

6.4 What next?

In this thesis we have shown that it is feasible to study phonological prediction using electroencephalography experiments with an OddBall design. This opens a road of research to approach several open questions on the topic. In this section we will delineate some of them.

6.4.1 How automatic is phonological prediction?

As we have shown, words composed by an unexpected sequence of phonemes automatically elicits prediction error in the form of a MMN response, even if the task at hand does not require deviant detection. A simple way to test the limits of this automaticity would be to perform an experiment as **Experiment 1**, but instructing the subjects to perform a concurrent task such as watching a silent movie. Even thought the main generators of the MMN response can be located in the superior temporal gyrus, a variety of additional generators, including frontal cortices, can be recruited depending on the task at hand [76]. As the MMN elicited in this experiments is the result of the violation of an abstract rule [84] (3.1), if participants' attention is drawn somewhere else, frontal cortices might not be recruited and the MMN response might not be present.

6.4.2 What is being predicted? Just the next phoneme or an entire word form?

We have proposed that the speech comprehension system is able to build phonological predictions where preceding phonemes are used to predict forthcoming ones. In parallel to this, it is possible that preceding phonemes could be use to generate predictions about actual full word form. The increase of prediction error for **XXY** deviants compared to **XYY** deviants reported in experiments 1 and 2 could be the result of a failure of a prediction at the word form level, instead of only quantitative increase of prediction strength at the phonological level.

To test if this type of word form level prediction is possible, deviant words should be constructed in a way in which they would share the same internal statistical structure as standard words, so that each syllable presented could be expected given the previous one, even though their combination at the word form level would be rare.

Endress and Mehler [35] developed a stimulus set with this characteristics, composed by 6 tree-syllabic words and 2 **phantom words** for which there was a word sharing the first and the second syllable, a word sharing the second and third syllable, and a word sharing the first and third syllable (Fig: 6.1).

If this phantom words would be use as deviants in an OddBall experiment, their local transitional probabilities (i.e. from one syllable to the next) would be identical to the ones of standard words, and they would only be defined as deviants at the word form level. Therefore, if the speech processing system is only capable of performing local phonological predictions, the presentation of this type of deviants would not generate prediction error.

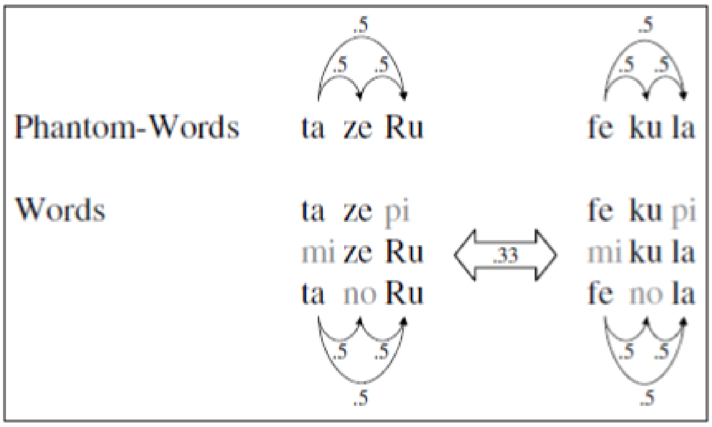

Figure 6.1: Stimulus set used by Endress and Mehler [35]. Stimuli were constructed so that transitional probabilities (TPs) among syllables in words would be identical to TPs among syllables in **phantom-words**. For each of the two phantom-words, there was a word sharing the first and the second syllable, a word sharing the second and third syllable, and a word sharing the first and third syllable. (The syllable that is not shared between a word and the corresponding phantom-word is printed in light gray characters in the figure.) In this way, TPs among adjacent and nonadjacent syllables within words and phantom-words were 0.5. Figure reproduced with permission from the authors.

6.4.3 The eternal language question. Domain general or domain specific?

Speech perception is a high level and very complex cognitive process that is supported by a vast network of brain regions involving temporal, parietal and frontal cortices [47, 48, 49, 62, 99]. As such, it has extensive overlap both functionally and structurally with several other cognitive functions [7]. Because of this, the ability to generate phonological predictions could be language specific, or it could be a domain general process that participates in the construction of complex auditory objects [118, 102, 94]. Given that the MMN response, that acted as a prediction error signal in our experiments, can be elicited with a variety of non linguistic stimulus, and that its main generators are located before processing streams for linguistic and non-linguistic sounds diverge [99, 61, 62, 59], we are inclined to think that this could be the case [116].

To test this, it could be possible to perform an OddBall experiment, in which linguistic and non-linguistic blocks would alternate. Several methods exist to generate non-linguistic stimuli matched in complexity with linguistic ones, such as temporal inversion (i.e. backward speech) [74, 27], spectral rotation [46, 36] and noise vocoding [114, 24]. Such comparison would allow to test if the brain is able to perform the same type of predictions for non-linguistic stimuli.

6.5 Final remarks

The main goal of this thesis was to test whether the human speech processing system is able to construct phonological predictions within a word, using solely phonological and word form information. In 3 different electroencephalography experiments we were able to show how the presentation of a word composed by an unexpected sequence of phonemes, elicits a MMN response. This event related potential is a well established marker of violation of expectations that can be interpreted as prediction error, resultant of the comparison between a prediction and the incoming bottom-up input.

Is worth underlining that in our experiments, participants' brains constructed predictions in an automatic way, with new pseudowords that were learned in a period of minutes and without the aid of higher level information such as syntax or semantics. This implies a formidable capacity and flexibility of the language system to learn the sequences of phonemes composing new words and generate predictions about upcoming phonemes based on the preceding ones. This capacity might play a fundamental role in the difficult task of mapping a complex, variable and noisy signal as speech into meaning.

Bibliography

[1] R. Powers Albert, Megan Kelley, and Philip.R. Corlett. Hallucinations as Top-Down Effects on Perception. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2016.

[2] H. Althen, M. Huotilainen, S. Grimm, and C. Escera. Middle latency response correlates of single and double deviant stimuli in a multi-feature paradigm. *Clinical Neurophysiology*, 127(1):388–396, 2016.

[3] Gerry Altmann and Mark Steedman. Interaction with context during human sentence processing. *Cognition*, 30(3):191–238, 1988.

[4] Matthew A J Apps and Manos Tsakiris. The free-energy self: A predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews*, 41:85–97, 2014.

[5] Stephen R Arnott and Claude Alain. Stepping out of the spotlight: MMN attenuation as a function of distance from the attended location. *Neuroreport*, 13(17):2209–2212, 2002.

[6] Martijn Baart and Arthur G. Samuel. Early processing of auditory lexical predictions revealed by ERPs. *Neuroscience Letters*, 585:98–102, 2015.

[7] Francesco P. Battaglia, Gideon Borensztajn, and Rens Bod. Structured cognition and neural systems: From rats to language. *Neuroscience and Biobehavioral Reviews*, 36(7):1626–1639, 2012.

[8] Tristan a Bekinschtein, Stanislas Dehaene, Benjamin Rohaut, François Tadel, Laurent Cohen, and Lionel Naccache. Neural signature of the conscious pro-

cessing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5):1672–7, feb 2009.

[9] Tristan A Bekinschtein, Stanislas Dehaene, Benjamin Rohaut, François Tadel, Laurent Cohen, and Lionel Naccache. Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5):1672–7, 2009.

[10] Alexandra Bendixen, Mathias Scharinger, Antje Strauß, and Jonas Obleser. Prediction in the service of comprehension: modulated early brain responses to omitted speech segments. *Cortex; a journal devoted to the study of the nervous system and behavior*, 53:9–26, apr 2014.

[11] Idan Blank, Zuzanna Balewski, Kyle Mahowald, and Evelina Fedorenko. Syntactic processing is distributed across the language system. *NeuroImage*, 127:307–323, 2016.

[12] Megan Ann Boudewyn, Philip R Corlett, Karl Friston, Meredith Brown, and Gina R Kuperberg. A Hierarchical Generative Framework of Language Processing: Linking Language Perception, Interpretation, and Production Abnormalities in Schizophrenia. *Front. Hum. Neurosci*, 9(November):1–23, 2015.

[13] David Brainard. The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436, 1997.

[14] Van Den Brink, Colin M Brown, and Peter Hagoort. Electrophysiological Evidence for Early Contextual Influences during Spoken-Word Recognition : N200 Versus N400 Effects. *Journal of Cognitive Neuroscience*, 13(7):967–985, 2001.

[15] Andreja Bubic, D Yves von Cramon, and Ricarda I Schubotz. Prediction, cognition and the brain. *Frontiers in human neuroscience*, 4(March):25, 2010.

[16] Christian Büchel, Stephan Geuter, Christian Sprenger, and Falk Eippert. Placebo analgesia: A predictive coding perspective. *Neuron*, 81(6):1223–1239, 2014.

[17] E T Bullmore, J Suckling, S Overmeyer, S Rabe-Hesketh, E Taylor, and M J Brammer. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE transactions on medical imaging*, 18(1):32–42, 1999.

[18] Maximilien Chaumon, Dorothy V.M. Bishop, and Niko A. Busch. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, 250:47–63, 2015.

[19] Srivas Chennu and Tristan a Bekinschtein. Arousal modulates auditory attention and awareness: Insights from sleep, sedation, and disorders of consciousness, jan 2012.

[20] Srivas Chennu, Valdas Noreika, David Gueorguiev, Alejandro Blenkmann, Silvia Kochen, Agustín Ibáñez, Adrian M Owen, and Tristan A Bekinschtein. Expectation and attention in hierarchical auditory prediction. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(27):11194–205, 2013.

[21] Marco D. Comerchero and John Polich. P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110(1):24–30, 1999.

[22] Toni Cunillera, Estela Càmara, Juan M. Toro, Josep Marco-Pallares, Nuria Sebastián-Galles, Hector Ortiz, Jesús Pujol, and Antoni Rodríguez-Fornells. Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, 48(3):541–553, 2009.

[23] Ghislaine Dahaene-Lambertz, Emmanuel Dupoux, and Ariel Gout. Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, 12(?):635–647, 2000.

[24] Matthew H. Davis and Ingrid S. Johnsrude. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229(1-2):132–147, 2007.

[25] Ruth De Diego-Balaguer and Diana Lopez-Barroso. Cognitive and neural mechanisms sustaining rule learning from speech. *Language Learning*, 60(SUPPL. 2):151–187, 2010.

[26] Stanislas Dehaene and Jean Pierre Changeux. Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, 70(2):200–227, 2011.

[27] G Dehaene-Lambertz, a Montavont, a Jobert, L Allirol, J Dubois, L Hertz-Pannier, and S Dehaene. Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and language*, 114(2):53–65, aug 2010.

[28] Katherine a DeLong, Thomas P Urbach, and Marta Kutas. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, 8(8):1117–1121, 2005.

[29] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, mar 2004.

[30] Hanneke E M Den Ouden, Peter Kok, and Floris P. de Lange. How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, 3(DEC):1–12, 2012.

[31] Christian F Doeller, Bertram Opitz, Axel Mecklinger, Christoph Krick, Wolfgang Reith, and Erich Schröger. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *NeuroImage*, 20(2):1270–82, oct 2003.

[32] Connie C. Duncan, Robert J. Barry, John F. Connolly, Catherine Fischer, Patricia T. Michie, Risto Näätänen, John Polich, Ivar Reinvang, and Cyma Van Petten. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, 120(11):1883–1908, 2009.

[33] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, and O. Van Der Vrecken. The MBROLA project: towards a set of high quality speech\nsynthesizers free of

use for non commercial purposes. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, 3:2–5, 1996.

[34] Jeffrey L University of California San Diego San Diego C A U S Elman, Mary Hare, Bowling Green K Y U S Bowling Green State University, Ken McRae, and Western University of. Cues, Constraints, and Competition in Sentence Processing. *Beyond nature-nurture: Essays in honor of Elizabeth Bates*, pages 111–138, 2005.

[35] Ansgar D. Endress and Jacques Mehler. The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, 60(3):351–367, 2009.

[36] S. Evans, J. S. Kyong, S. Rosen, N. Golestani, J. E. Warren, C. McGettigan, J. Mourao-Miranda, R. J. S. Wise, and S. K. Scott. The Pathways for Intelligible Speech: Multivariate and Univariate Perspectives. *Cerebral Cortex*, 24(9):2350–2361, 2014.

[37] Thomas a Farmer, Morten H Christiansen, and Padraic Monaghan. Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32):12203–12208, 2006.

[38] F Faugeras, B Rohaut, N Weiss, T a Bekinschtein, D Galanaud, L Puybasset, F Bolgert, C Sergent, L Cohen, S Dehaene, and L Naccache. Probing consciousness with event-related potentials in the vegetative state. *Neurology*, 77(3):264–8, jul 2011.

[39] F Ferreira and J R Clifton. The independance of syntactic processing. *Journal of Memory and Language*, 25:348–368, 1986.

[40] Fernanda Ferreira, Kiel Christianson, and Andrew Hollingworth. Misinterpretations of Garden-Path Sentences: Implications for Models of Sentence Processing and Reanalysis. 30(1):3–20, 2001.

[41] Fernanda Ferreira and Nikole D. Patson. The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, 1(1-2):71–83, 2007.

[42] Derek J Fisher, Bryan Grant, Dylan M Smith, and Verner J Knott. Effects of deviant probability on the 'optimal' multi-feature mismatch negativity (MMN) paradigm. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 79(2):311–5, feb 2011.

[43] K. I. Forster. Priming and the Effects of Sentence and Lexical Contexts on Naming Time: Evidence for Autonomous Lexical Processing. *Quarterly Journal of Experimental Psychology*, 33(4):465–495, 1981.

[44] Lyn Fraizer. *On comprehending sentences: Syntactic parsing strategies.* PhD thesis, Indiana University, 1979.

[45] Dominik Freunberger and Dietmar Roehm. Semantic prediction in language comprehension: evidence from brain potentials. *Language, Cognition and Neuroscience*, 3798(July):1–13, 2016.

[46] a. D. Friederici. The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, 91(4):1357–1392, 2011.

[47] Angela D Friederici. Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, 6(2):78–84, feb 2002.

[48] Angela D Friederici. The neural basis of language development and its impairment. *Neuron*, 52(6):941–52, dec 2006.

[49] Angela D Friederici. Pathways to language: fiber tracts in the human brain. *Trends in cognitive sciences*, 13(4):175–81, apr 2009.

[50] Karl Friston. A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1456):815–36, apr 2005.

[51] Karl Friston. The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301, 2009.

[52] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience*, 11(2):127–138, 2010.

[53] Susan M. Garnsey, Neal J. Pearlmutter, Elizabeth Myers, and Melanie a. Lotocky. The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language*, 37(1):58–93, 1997.

[54] Marta I Garrido, James M Kilner, Klaas E Stephan, and Karl J Friston. The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 120(3):453–63, mar 2009.

[55] Craig L Gonsalvez and John Polich. P300 amplitude is determined by target-to-target interval. *Psychophysiology*, 39(3):388–396, 2002.

[56] Jeremy Goslin, Claudia Galluzzi, and Cristina Romani. PhonItalia: a phonological lexicon for Italian. *Behavior research methods*, 46(3):872–86, 2014.

[57] David M. Groppe, Marvin Choi, Tiffany Huang, Joseph Schilz, Ben Topkins, Thomas P. Urbach, and Marta Kutas. The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Research*, 1361:54–66, 2010.

[58] Olaf Hauk. Preface to special issue "prediction in language comprehension and production". *Language, Cognition and Neuroscience*, 31(1):1–3, 2016.

[59] Gregory Hickok. The functional neuroanatomy of language. *Physics of life reviews*, 6(3):121–43, sep 2009.

[60] Gregory Hickok. 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, 2012.

[61] Gregory Hickok and David Poeppel. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2):67–99, 2004.

[62] Gregory Hickok and David Poeppel. The cortical organization of speech processing. *Nature reviews. Neuroscience*, 8(5):393–402, 2007.

[63] J. A. Hobson and K. J. Friston. Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology*, 98(1):82–98, jul 2012.

[64] Kathleen Joos, Annick Gilles, Paul Van de Heyning, Dirk De Ridder, and Sven Vanneste. From sensation to percept: The neural signature of auditory event-related potentials. *Neuroscience & Biobehavioral Reviews*, 42:148–156, 2014.

[65] N. M. Kane, S. H. Curry, S. R. Butler, and B. H. Cummins. Electrophysiological indicator of awakening from coma. *The Lancet*, 1(8846), 193.

[66] Makio Kashino. Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology*, 27(6):318–321, 2006.

[67] Dave Kleinschmidt and T Florian Jaeger. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2):148–203, 2015.

[68] Gina R Kuperberg and T Florian Jaeger. What do we mean by prediction in language comprehension? *Language Cognition & Neuroscience*, 31(1):32–59, 2016.

[69] M Kutas and S A Hillyard. Reading senseless sentences: brain potentials reflect semantic incongruity., 1980.

[70] Marta Kutas and Kara D. Federmeier. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12):463–470, 2000.

[71] Marta Kutas and Kara D Federmeier. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *NIH Public Access*, pages 621–647, 2014.

[72] Gregory A. Light, Lisa E. Williams, Falk Minow, Joyce Sprock, Anthony Rissling, Richard Sharp, Neal R Swerdlow, and David L Braff. Electroencephalography (EEG) and event-related potentials (ERPs) with human participants, 2010.

[73] Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, 2007.

[74] Yasuyo Minagawa-Kawai, Alejandrina Cristià, and Emmanuel Dupoux. Cerebral lateralization and early speech acquisition: A developmental scenario. *Developmental Cognitive Neuroscience*, 1(3):217–232, 2011.

[75] B W Muller, Christiane Achenbach, Robert D Oades, Stefan Bender, Ulrich Schall, Bernhard W Müller, Christiane Achenbach, Robert D Oades, Stefan Bender, and Ulrich Schall. Modulation of mismatch negativity by stimulus deviance and modality of attention. *Neuroreport*, 13(10):1317–20, 2002.

[76] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12):2544–2590, 2007.

[77] Risto Näätänen. Mismatch negativity (MMN): Perspectives for application. *International Journal of Psychophysiology*, 37(1):3–10, 2000.

[78] Risto Näätänen, Satu Pakarinen, Teemu Rinne, and Rika Takegata. The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, 115(1):140–144, 2004.

[79] D Norris, J M McQueen, and a Cutler. Merging information in speech recognition: feedback is never necessary. *The Behavioral and brain sciences*, 23(3):299–325; discussion 325–370, 2000.

[80] Dennis Norris, James M. McQueen, and Anne Cutler. Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, 3798(December 2015):1–15, 2015.

[81] Redmond G O'Connell, Paul M Dockree, and Simon P Kelly. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature neuroscience*, 15(12):1729–35, 2012.

[82] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:156869, jan 2011.

[83] Bertram Opitz, Teemu Rinne, Axel Mecklinger, D Yves von Cramon, and Erich Schröger. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage*, 15(1):167–174, 2002.

[84] Petri Paavilainen. The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, 88(2):109–123, 2013.

[85] Felipe Pegado, Tristan Bekinschtein, Nicolas Chausson, Stanislas Dehaene, Laurent Cohen, and Lionel Naccache. Probing the lifetimes of auditory novelty detection processes. *Neuropsychologia*, 48(10):3145–3154, 2010.

[86] Denis Pelli. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4):437–442, 1997.

[87] H. N. Phillips, A. Blenkmann, L. E. Hughes, T. A. Bekinschtein, and J. B. Rowe. Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions. *Journal of Neuroscience*, 35(25):9255–9264, 2015.

[88] Holly N. Phillips, Alejandro Blenkmann, Laura E. Hughes, Silvia Kochen, Tristan A. Bekinschtein, Cam CAN, and James B. Rowe. Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex*, 2016.

[89] M J Pickering and S Garrod. An integrated theory of language production and comprehension. *The Behavioral and brain sciences*, 36(4):329–347, 2013.

[90] John Polich. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(10):2128–48, oct 2007.

[91] John Polich and Jos?? R. Criado. Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, 60(2):172–185, 2006.

[92] J. Schomaker and M. Meeter. Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience and Biobehavioral Reviews*, 55:268–279, 2015.

[93] Sarah Schuster, Stefan Hawelka, Florian Hutzler, Martin Kronbichler, and Fabio Richlan. Words in Context: The Effects of Length, Frequency, and Predictability on Brain Responses During Natural Reading. *Cerebral Cortex*, (June):1–16, 2016.

[94] Sophie K Scott. Auditory processing–speech, space and auditory objects. *Current opinion in neurobiology*, 15(2):197–201, apr 2005.

[95] Stephen H. Scott. A Functional Taxonomy of Bottom-Up Sensory Feedback Processing for Motor Actions. *Trends in Neurosciences*, xx(8):1–15, 2016.

[96] Claire Sergent, Sylvain Baillet, and Stanislas Dehaene. Timing of the brain events underlying access to consciousness during the attentional blink. *Nature neuroscience*, 8(10):1391–400, oct 2005.

[97] Brian H. Silverstein, Michael Snodgrass, Howard Shevrin, and Ramesh Kushwaha. P3b, consciousness, and complex unconscious processing. *Cortex*, 73:216–227, 2015.

[98] Irina Simanova, Jolien C. Francken, Floris P. de Lange, and Harold Bekkering. Linguistic priors shape categorical perception. *Language, Cognition and Neuroscience*, 3798(January):1–7, 2015.

[99] Karsten Specht. Neuronal basis of speech comprehension. *Hearing Research*, 307(0378):121–135, 2014.

[100] Michael J Spivey-knowlton, John C Trueswell, and Michael K Tanenhaus. Context Effects in Syntactic Ambiguity Resolution : Discourse and Semantic Influences in Parsing Reduced Relative Clauses. *Canadian Journal of Experimental Psychology*, 47(2):276–309, 1993.

[101] Johanna Steinberg, Hubert Truckenbrodt, and Thomas Jacobsen. The role of stimulus cross-splicing in an event-related potentials study. Misleading formant transitions hinder automatic phonological processing. *The Journal of the Acoustical Society of America*, 131(4):3120, 2012.

[102] Mitchell Steinschneider, Kirill V. Nourski, Ariane E. Rhone, Hiroto Kawasaki, Hiroyuki Oya, and Matthew A. Howard. Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. *Frontiers in Neuroscience*, 8(8 JUL):1–13, 2014.

[103] Melanie Strauss, Jacobo D Sitt, Jean-Remi King, Maxime Elbaz, Leila Azizi, Marco Buiatti, Lionel Naccache, Virginie van Wassenhove, and Stanislas Dehaene. Disruption of hierarchical predictive coding during sleep. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11):E1353–62, 2015.

[104] Yue Sun, Maria Giavazzi, Martine Adda-Decker, Leonardo S. Barbosa, Sid Kouider, Anne-Catherine Bachoud-Lévi, Charlotte Jacquemot, and Sharon Peperkamp. Complex linguistic rules modulate early auditory brain responses. *Brain and Language*, 149:55–65, 2015.

[105] Mari Tervaniemi, Anu Castaneda, Monja Knoll, and Maria Uther. Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *Neuroreport*, 17(11):1225–1228, 2006.

[106] Mari Tervaniemi, Mika Rytkönen, Erich Schröger, Risto J. Ilmoniemi, and Risto Näätänen. Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & memory (Cold Spring Harbor, N.Y.)*, 8(5):295–300, 2001.

[107] Juan M. Toro, Scott Sinnett, and Salvador Soto-Faraco. Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2):25–34, 2005.

[108] Matthew J. Traxler. Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, 18(11):605–611, 2014.

[109] Deirdre M. Twomey, Peter R. Murphy, Simon P. Kelly, and Redmond G. O'Connell. The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, 42(1):1636–1643, 2015.

[110] C Van Petten, S Coulson, S Rubin, E Plante, and M Parks. Time course of word identification and semantic integration in spoken language., 1999.

[111] Cyma Van Petten and Barbara J. Luka. Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2):176–190, 2012.

[112] Benjamin T. Vincent. A tutorial on Bayesian models of perception. *Journal of Mathematical Psychology*, 66:103–114, 2014.

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

[114] Conor J. Wild, Matthew H. Davis, and Ingrid S. Johnsrude. Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, 60(2):1490–1502, 2012.

[115] Michael P. Wilson and Susan M. Garnsey. Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, 60(3):368–392, 2009.

[116] István Winkler. Interpreting theMismatch Negativity. *Federation of European Psychophysiology Societies*, 21(1992):147–163, 2007.

[117] István Winkler and István Czigler. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 83(2):132–43, feb 2012.

[118] István Winkler and Erich Schröger. Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain and Language*, 148:1–22, 2015.

[119] Sari Ylinen, Milla Huuskonen, Katri Mikkola, Emma Saure, Tara Sinkkonen, and Petri Paavilainen. Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain and Language*, 162:72–80, 2016.