WHAT PMN? INVESTIGATING PHONOLOGICAL PROCESSING IN SEMANTICS-FREE SPEECH PERCEPTION WITH ERP

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2021

Massimiliano Canzi

School of Arts, Languages and Cultures

I Table of Contents

Word Count *(pp. 16-190)*: **62.124**

II List of Figures

III Abbreviations & Acronyms

AFA: Adaptive Factor Adjustment

AIC: Akaike Information Criterion

AL: Artificial Language

AVE: Average Channel Reference

BIC: Bayesian Information Criterion

CMS: Common Mode Sensor

DRL: Driven Right Leg

EEG: Electro-encephalography

ELAN: Early Left Anterior Negativity

ERAN: Early Right Anterior Negativity

ERP: Event-Related Potential

FDR: False Discovery Rate

ICA: Independent Component Analysis

IFO: Information Flow Order

ISI: Inter Stimulus Interval

LF: Lexical Feedback

LMM: Linear Mixed-Effect Model

MMN: Mismatch Negativity

PMN: Phonological Mismatch / Mapping Negativity

PSO: Post Stimulus Onset

IV Abstract

The Phonological Mapping Negativity (PMN) event-related potential (ERP) component has been, in the past, primarily linked to phonological mapping and mismatch, as part of processes of lexical retrieval, in speech perception and spoken-word recognition. However, an updated theory suggests the PMN serves as a neural marker for the analysis of acoustic input (Newman and Connolly 2003). In addition, a recent review of the limited, existing literature on the PMN has advanced concerns regarding methodological shortcomings and contradictory findings of previous studies (Lewendon et al. 2020).

Recently, despite limited and contrasting evidence in regards to the exact role of the PMN in processes of spoken-word recognition, clinical research has been published that focussed on the elicitation of the PMN as a direct marker of phonological processing ability in patient populations, including Wernicke's aphasia patients (Robson et al. 2017). Considering past methodological limitations, the limited study sample, and inconsistencies among findings in the literature on the PMN, an in-depth investigation into the exact nature of the component has been long overdue. Three novel experiments were carried out to test the elicitation of the PMN in contexts of auditory, phonetic and phonological mismatch in the absence of lexical retrieval, with the aim of testing the sensitivity of the PMN to phonological and pre-lexical mismatch alone.

No PMN was observed across all three experiments. Among the possible explanations, the theory that the PMN reflects phonological mapping only as a function of lexical activation appears to be the most likely. This theory would suggest that the PMN, even though sensitive to phonological mapping, should be categorised primarily as a semantic response or that it should not be considered as a direct marker of phonological mapping. The need for further experimentation into the nature of the PMN and neighbouring components is addressed, as well as limitations inherent to the ERP field are discussed.

V Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning;

VI Copyright Statement

So we all raise a standard **/** *To which the wise and honest soul may repair*

To which a hunter **/** *A hundred years from now, may look and despair*

And see with wonder **/** *The tributes we have left to rust in the parks*

Joanna Newsom

VII Acknowledgements

I am extremely grateful to my supervisors Wendell Kimper and Pat Strycharczuk — as well as to my independent reviewer Thea Cameron-Faulkner — who, in the past four years, have taught, supported, guided and trusted me far beyond the scope of this PhD. I would also like to thank Linnaea Stockall, Vera Hohaus and Ricardo Bermúdez-Otero for their role in the evaluation of this thesis. For their direct impact on the realisation of this work, I would like to extend my gratitude to Ariadne Loutrari and to my research assistants Hui Chen, Lauren Forrest, Chloe Gornall, Tristan Hill, Yuerong Shen, Ellen Symonds, Xinrong Wang, and Ziyun Zhang.

I am also forever indebted to my parents, to whom this thesis is dedicated, for their unconditional love and support.

Finally, I would like to thank Lorenzo Moretti, Andrea, Jenny, and Marco Nini, Kaiyue Xing, Lucia Maltoni, and Edoardo Altamura for the beautiful memories made in the past four years; Tamara Rathcke, Bettina Braun, Diego Frassinelli, Caterina Moruzzi, and Sahra Styger for the warm welcome to Konstanz; Juliette Angot, George Bailey, Hannah Booth, Stefano Coretta, Roisin Cosnahan, Lisa Donlan, Alina Mclellan, Donald Morrison, Anh Khoi Nguyen, and Stephen Nichols for being fantastic colleagues, friends and sources of inspiration throughout this doctoral journey.

Chapter 1

Introduction

1.1 Background

The total number of neuroimaging experiments revolving around the collection of event-related potential (ERP) data for the study of language processing, comprehension and perception has been increasing rapidly in recent decades. Some of the reasons behind this advancement can be attributed to the advent of cheaper and more reliable technology, more streamlined and standardised electro-encephalography (EEG) data processing techniques, increased knowledge of ERP components and their key role in understanding cognitive processing in humans. Another important reason is that new and revised theories of language comprehension and spoken-word recognition often require to be investigated with experimentation that allows a more direct approach to studying cognitive functions, compared to classical behavioural experimental paradigms in linguistics.

Advancements in ERP research allow nowadays for much broader applications of the data collection technique compared to even two or three decades ago. However, there are still limitations in the field and extensive gaps in the literature that, often times, prevent researchers from being able to reliably interpret ERP evidence and weigh it in favour of one theory or another. One of the main reasons for this is that, while some components have received a large amount of experimental attention in the past decades, other components have been flying under the radar. Information that has been collected through experimentation regarding these under-studied ERP components is often inconsistent across the few studies that are present in the literature. For example, the Phonological Mapping / Mismatch Negativity (PMN) component, the focus of the current research project, is among those ERP components that, unlike other well-known responses often associated with language processing such as the N400 (Kutas and Hillyard 1980: Kutas and Hillyard 1984) and P600 (Hagoort et al., 1993), have received very little experimental attention in the past few decades. The PMN component has been linked primarily to phonological (Connolly et al. 1992; ...; D'Arcy et al. 2004) and pre-lexical (Newman et al., 2003) mapping in spoken-word recognition. However, as extensively discussed in *Chapter 2* of this work, findings of existing research are in disagreement when it comes to the ERP component's topographical distribution and function in language processing (Lewendon et al., 2020).

Although this level of uncertainty surrounding the nature of a specific ERP component is not unheard of in ERP research, that very situation becomes extremely dangerous when researchers start employing the PMN in clinical experiments, for instance as a marker of phonological processing abilities (e.g. Robson et al. 2017). The role of the PMN component in language processing has been in question for many years now together with its existence. Indeed, the option that the PMN might in fact simply be another representation of other neighbouring responses such as the N400 and MMN still exists (Lewendon et al., 2020). Studying a patient population's phonological processing abilities based on the elicitation of the PMN could be extremely problematic when interpreting results. The PMN response might in fact not represent phonological processing *alone* but it might be sensitive to some other aspects of language processing, as some have proposed in the past (Connolly and Phillips, 1994). In such scenarios, measuring the PMN response as a direct marker of phonological processing abilities might result in biased and incorrect interpretations of experimental findings.

1.2 ERP in speech perception

Event-related potentials (ERP) are shifts in electrical potential inside the brain, usually time-locked to the processing of external stimuli and sensory information. ERP are measured with electro encephalography (EEG), dedicated equipment consisting of electrodes placed on the scalp of experimental subjects. The electrodes measure electrical potential on the surface of the scalp, originating from the brain, the skin and the surrounding environment, during the presentation of experimental stimuli. Electrodes transfer the electrical signal data to a a receiving device connected to a computer, where it is digitised. *Chapter 3* goes into more detail about the implementation of the EEG methodology and the set-up of equipment.

In the past few decades, the evolution of computational and experimental technology, especially with EEG technology equipment becoming more affordable and widespread, has allowed linguists to study responses of the brain to sensory information in cognitive processes in relation to language. Some of these processes, or groups of processes, include reading and listening comprehension, speech perception, spoken-word recognition and more. Studies on the aforementioned topics have been providing extensive amounts of information about linguistically-related cognitive processes and ERP components which, in turn, help inform and provide evidence for the evaluation of existing theoretical frameworks and architectures of grammar (see Getz and Toscano 2021 for a recent review).

When talking about ERP in relation to language processing and spoken-word recognition, some notable components come to mind such as the N400 (Kutas and Hillyard 1980; Kutas and Hillyard 1984) and P600 (Hagoort et al. 1993; Hagoort and Brown 2000) components, often linked to semantic processing and syntactic processing (among other things) respectively. However, there are other ERP components such as the PMN (Connolly and Phillips, 1994), often associated to phonological mismatch and mapping, that have so far received less experimental attention over the years. The number of studies run in the past on the nature of the N400 component, for instance, have probably reached the hundreds. This has helped to provide a fairly clear picture of what the component represents, what it reacts to, what its scalp distribution is and when it is elicited. On the other hand, studies on components such as the PMN have only been recently hitting two digits. The sum total of the knowledge on this particular component has only reached the surface when it comes to reliably categorise the PMN during speech perception and phonological processing. In fact, only few paradigms have been tested so far and there are still many contexts in which it is unknown whether a PMN response should or should not be present. Furthermore, many of the findings regarding the elicitation of the PMN component show inconsistencies regarding its scalp distribution and its role among processes of speech perception and spoken-word recognition.

Components such as the N400 or P600, especially when compared against components such as the PMN, are very consistent in their scalp distribution, latency and in the type of experimental paradigm that elicits them. Because of their reliability and stability, these components have been used in many more studies. Whenever the N400 and PMN components are elicited in an experimental paradigm, an explanation on the function of the N400 can often be reliably provided building upon decades of research and hundreds of studies covering a wide variety of cases where the N400 is elicited. On the other hand, more assumptions need to be made about certain behaviours of components such as the PMN, considering fewer studies have been run to investigate its nature and role in language processing.

1.3 Aims and Objectives

All past experiments run to study the PMN component and phonological processing in spoken-word recognition generally combine phonological mismatch with the presence of contextual semantic information and processing (e.g. Connolly et al. 1992; Connolly and Phillips 1994; Newman et al. 2003; D'Arcy et al. 2004). Although partially dissociated, just like the N400, the PMN has been found to be sensitive to semantic mapping (Connolly and Phillips 1994 as reported in Lewendon et al. 2020). No experiment has been run to test whether the PMN is sensitive to phonological mapping and mismatch in a context where lexical processing is completely absent from the speech stream. For this particular reason, it is at this time impossible to exclude that, although the PMN certainly appears to be sensitive to phonological information, it might be so only in specific contexts where semantic processing and lexical retrieval are also present in combination with phonological mapping and mismatch. If that was the case, the PMN could be regarded as a lexical ERP component with a phonological sensitivity.

Experiments one to three of this research project test three novel (for PMN research), distinct experimental paradigms of speech-sound and sound processing. The aim is to investigate whether the PMN component can be elicited through the sole presence of phonological mismatch in semantics-free stimuli (Experiments one and two) as well as whether distinct phonological responses can be discovered when comparing the processing of speech-sound and sine-wave tone sequences (Experiments two and three). Experiments one and two also investigate whether the degree of focussed attention, known to modulate amplitude and latency of some ERP responses (e.g. Näätänen et al. 1993) directly affects the observation of the PMN component. This is done with the goal of obtaining a better understanding of the type of response the PMN represents. All three experiments, while aiming to be informative as stand-alone units, were also designed to be highly comparable to examine and correlate findings, directly and indirectly.

1.4 Relevance

Studying how ERP responses behave in uncharted experimental territory is, in its own right, informative and necessary when trying to build an atlas of ERP components. The investigation into an ERP response such as the PMN, however, can also indirectly inform broader research questions that have been the centre of debates in speech perception, language processing and spoken-word recognition in the past decades. Understanding whether the PMN component is simply related to phonological and word-form mapping, or whether it is elicited by semantically-derived processing can — in future research — allow for studies on the investigation of information flow order (IFO) in speech perception and architectures of grammar models. Understanding whether the PMN is consistently elicited, it is directly linked to one specific grammatical "module" or whether, similarly to other components originally thought to be strictly linked to grammatical processing such as the P600, its response can be traced back to non language-specific cognitive processing, can also inform theories of brain architecture, such as modularity of mind (e.g. Fodor 1983) and the extent of feedback on speech perception (Getz and Toscano, 2021).

ERP components can be thought of as building blocks that can be used to investigate theories in language processing, memory, cognitive workload, etc. because of the almost real-time and direct access they grant into brain function and activity. However, findings are only as reliable as the methodology used. Specifically because the elicitation of an ERP component, or lack thereof, can be informative in providing evidence towards one account or another, it is paramount to reliably determine what each ERP component represents in the beginning, how it is distributed and when it is observed. The PMN component, differently from other components related to language processing, has received limited experimental attention in the past three decades. For this reason, its role, topographical distribution and even existence are still not agreed upon. The current research aims at increasing the body of work on the nature of the PMN component in order to, one study at a time, more confidently be able to label the PMN component for what it really is and, in future research, exploit its elicitation for the study of debated topics in linguistics and cognitive sciences such as architectures of grammar and speech perception modelling.

1.5 Thesis structure

The current thesis comprises seven chapters, including the *Introduction* chapter. The *Research Context* chapter introduces the reader to the existing body of literature on the topic of relevant ERP components in speech perception and spoken-word recognition as well as addressing the gaps in the literature that the current project aims to fill. In the *Research Context* chapter specifically, ERP components often mentioned in this work are introduced and their role is compared to that of the PMN component, which remains the central focus of the study. Some of the most well known models of speech perception are also summarised with the aim of presenting the reader with information in regards to the general frameworks in which the majority of the findings of previous research have been contextualised.

Following, the *Methods* chapter summarises the common aspects of methodology among all three experiments, including EEG data collection and equipment, software used, preprocessing and processing steps and statistical analysis. Methodological details specific to each of the experiments are presented in the experiment-specific chapters. A the end of the *Methods* section, the reader can find the links to obtain the code used for all three experiment presentations, as well as processed ERP data, statistical analysis and data visualisation code in **R**.

The *Experiment one*, *Experiment two* and *Experiment three* chapters present each of the experiments as well as comparisons of findings among datasets and discussion of the findings of each of the three studies. Finally, the *General Discussion* and *Conclusion* chapters tie all findings of experiments one to three together with previous research studies to discuss how the PMN component — and early pre-N400 responses — should be interpreted in contexts of speech and sound perception. While the *General Discussion* chapter focuses mostly on one theory that links the elicitation of the PMN to the presence of lexical and semantic mapping during phonological mismatch, the chapter also presents more speculative theories and the implications these would have on current thinking surrounding architectures of grammar and speech perception modelling. Limitations of the studies and suggestions for future research are addressed in the *Conclusion* chapter of this work. Finally, *Appendix A* includes the discussion of additional findings relating to sequence violations and stimulus presentation expectations.

Chapter 2

Research Context

2.1 Introduction

The current chapter introduces the reader to literature on the Phonological Mismatch Negativity or PMN[1] (PMN) ERP component and its role in aspects of speech perception and spoken-word recognition. Together with an overview of the PMN component, limitations and inconsistencies characterising many of the existing studies on the PMN are also discussed. Other ERP components often observed during speech perception and spoken-word recognition are also introduced to the reader. Furthermore, a comparison is drawn between the knowledge present about the PMN component specifically and other neighbouring components of spoken-word and language processing, such as N40 and MMN. *Chapter 2* proceeds to advance an argument in favour of why it is necessary to carry out an in-depth investigation on the nature of components such as the PMN, including collecting more evidence to support recent interactional accounts of speech perception (e.g. Getz and Toscano 2021). Theoretical assumptions are also discussed in the second half of the chapter. The PMN has received much less experimental attention compared to more widely studied components in the field of language processing, such as the N400 (Kutas and Hillyard 1980; Kutas and Hillyard 1984) and P600 (Hagoort et al., 1993) components. Understanding the exact specifics and nature of even

[1] Just recently, the phonological mismatch negativity has been re-labelled as Phonological Mapping Negativity. This name change has been carried out in order to clearly differentiate the PMN from the Mismatch Negativity (MMN), generally thought to be a separate component. However, in this particular work, I mostly refer to the PMN as Phonological Mismatch Negativity, considering this reflects the majority of the use across most of the existing literature on the topic.

one more ERP component could, in future studies, help shed light on decade-long questions and debates in speech perception, such as information flow order and the extent of modularity of mind.

2.2 Event-related potentials

2.2.1 PMN

The Phonological Mismatch Negativity component is commonly maximal between 250 and 300 ms post stimulus onset. It has been mostly observed in frontal and central areas of the scalp[2] (Connolly and Phillips 1994; D'Arcy et al. 2004; Connolly et al. 2001; ...). In the literature, the PMN has mainly been linked to phonological mismatch and processing and it has been most largely elicited in experiments where — similarly to how the N400 is often elicited — a sentence-final highly-probable word is substituted with a word whose first phoneme or syllable does not match the expected, high-probability stimulus, regardless of its meaning (e.g. *the piano is out of pizza*) (Connolly and Phillips 1994; D'Arcy et al. 2004). The PMN and the N400 were thought to reflect similar patterns in lexical and semantic processing (e.g. Connolly and Phillips 1994; Diaz and Swaab 2007). However, it was observed that if the final word of a sentence was substituted with a less likely word beginning with a different phoneme but retaining most of the semantic meaning, a PMN component would be elicited, but a N400 one would not. This example reinforced the idea that the PMN and N400 components reflect, respectively, phonological and lexical processing of language and speech (Connolly et al. 1992; Connolly and Phillips 1994; D'Arcy et al. 2004; Diaz and Swaab 2007; Groppe et al. 2011).

Here, I review some of the main research findings on the PMN and their limitations, both in terms of methodology and shortcomings in discussing the role of the PMN component. Only a dozen papers were published containing research whose focus was directly aimed at discovering new information about the PMN component over the course of thirty years. Even today there is still very little agreement over what the function of the PMN is, what its topographical distribution is and whether the component is more comparable to higher-level, abstract ones such as the N400 or whether it is closer in nature to earlier, less abstract components such as the mismatch negativity (MMN).

[2] Inconsistencies regarding the topographical distribution of the component among existing studies are discussed later on in this chapter.

PMN as phonological marker of lexical retrieval

In 1994, Connolly and Phillips provided a contextualisation of the PMN component in the framework of the Cohort model of word recognition (Marslen-Wilson, 1984). The Cohort model, in brief terms, suggests that a series of three steps (i.e. activation, selection and integration) is responsible for processes of spoken word recognition. According to Marslen-Wilson (1984), both sensory information and context can activate a series of words in memory — *a cohort* — given linguistic and extra linguistic context and previous speech information. Connolly and Phillips (1994) suggest that the PMN falls exactly into the lexical selection stage of the cohort model. Once a cohort of words is activated and the presented stimulus does not match any of the expected outcomes activated by the cohort — because the first phoneme of the unexpected stimulus does not match the first phoneme of the highly likely stimuli in the context — the PMN response can be observed as a result of phonological mismatch during lexical activation. However, it is not clear whether semantic context is necessarily required to trigger the phonological mismatch response of the PMN. In fact, no experiment was ever run to test whether PMN can be elicited with phonological mismatch in semantics-free contexts. More recent studies on the PMN, such as a 2009 study by Newman et al. suggest that the PMN is a marker of acoustic information merging with pre-lexical phonological mapping. This contextualisation of the PMN component eliminates semantic context entirely from the requirements for the elicitation of the PMN.

In Hagoort (2000), an effect comparable to the PMN — but referred to as N250 (Hagoort and Brown, 2000) — was observed as a response, in conjunction with a later, N400-like component, to mismatching sentence-final words in a given context. However, Mismatching sentence-final words in a similar experiment involving the presentation of written rather than auditory stimuli did not evoke a similar response. These findings suggest that, while the PMN (or N250) is explained as indexing mismatch between word form and expectations (Hagoort and Brown, 2000), it is limited to auditorily presented stimuli. One year later, a similar study by Van Den Brink et al. 2001 also obtained, with a design similar to the original studies by Connolly et al. (1992; 1994), an early negative response (dubbed N200) to semantically anomalous stimuli beginning with phonemes that differed from the congruent completion, in sentence-final position. Van Den Brink et al. (2001), similarly to previous studies by Connolly and Phillips (1994) and Hagoort et al. (2000) link the early negativity (until now referred to as either PMN, N200 or N250) to "*lexical selection process[es], where word-form information resulting from an initial phonological analysis and content information derived from the context interact*" (Van

Den Brink et al., 2001). This contextualisation of the PMN component by Connolly and Phillips (1994), Hagoort et al. (2000) and Van Den Brink et al. (2001) fits well within existing theories of spoken-word recognition and is complemented by results on the PMN obtained by experiments employing different methodologies.

In a study by Newman et al. (2003), native English participants were presented with a real spoken English word (e.g. clap). Following the presentation of the stimulus, subjects were required to mentally remove its first phoneme (e.g. lap) and, soon after, to listen to the resulting manipulated stimulus. However, in some target trials, a different stimulus would presented (e.g. dog, map, xap). A PMN effect was observed in response to manipulated stimuli which presented a phonetic mismatch from the participants' expectations, regardless of lexical status. Kujala et al. (2004) similarly presented subjects of their experiment with a written word on a screen (e.g. *map*), followed by a single phoneme (e.g. *t*). Participants were instructed to substitute the new phoneme to the first existing phoneme of the priming stimulus, with the aim of obtaining a new word, which would be later presented to them auditorily. However, and this should now be a familiar paradigm to the reader, a mismatching word / nonce-word was presented across some of the trials instead (e.g. *lap*). Flouting phonological expectations resulted in the elicitation of a PMN component among the resulting effects. Newman and Connolly (2009) later ran another similar experiment where the first stimulus of each trial to be presented could either be a word or a non-word, before participants were asked to delete the first phoneme. In this more recent experiment, similar results were also found where the PMN is observed as a response to the presentation of a mismatching phoneme, regardless of lexical status of the stimulus presented.

D'Arcy et al. (2004) conducted a high-resolution ERP study with the aim of spatially differentiating the N400 response to an earlier, PMN/N250 component to anomalous word-form stimuli in a given sentence context. While the design of the study partially resembles previous methodologies (e.g. Connolly and Phillips 1994, Van Den Brink et al. 2001), D'Arcy et al. (2004) introduced new variables that shed light on different properties of the PMN. Participants of the study were instructed to visually and auditorily investigate pairs of sentences. The first sentence of the pair would be presented visually (e.g. "The man is teaching in the classroom"). The second sentence of the pair would be presented auditorily and either included a congruent or semantically anomalous sentence-final word, given the context presented in the first sentence (e.g. "The man teaches in the... school/barn"). Anomalous sentence-final words were semantically anomalous, highly unlikely given the context and began with mismatching phonemes compared to their congruent counterpart. All anomalous stimuli elicited both an N400 and a PMN response. However, congruent stimuli were coded as being highly probable or unlikely, with less probable items being more difficult to anticipate. While none of the congruent responses elicited an N400 component, less probable stimuli elicited a PMN-like response, spatially separated from a classic N400 response (D'Arcy et al., 2004). This result was suggested to index differences between word-form expectations and the stimulus presented, regardless of lexical status of the stimulus. These findings further support the idea that the PMN, as a response to auditory stimuli, is more likely linked to mismatch in word-form goodness-of-fit rather than to lexical retrieval, and should then be elicited in contexts where lexical activation or lexical retrieval are not necessarily present.

These more recent findings, while collected through methodologies different to previous experiments such as the one by Connolly and Phillips (1992; 1994) inform us that, for instance, syntactic and lexical activation do not appear to be necessary for the elicitation of a component such as the PMN. The main differences between Connolly and Phillips' (1994) and Kujala's (2004) experiment, for instance, is that the former incorporates stimuli in a syntactic context by embedding them in a sentence, while the latter does not. Furthermore, the lexical activation context is also reduced in later experiments, where one word is presented at a time, reducing lexical context influences on expectations. However, lexical activation has always been present in all experimental paradigms tested so far and it still plays the role of a confounding variable when trying to determine the underlying origin of the PMN component. Even in regards to lexical activation, we have seen how – in experiments such as Kujala (2004) and Newman et al. (2009) – as little as presenting a mismatching monosyllabic word is enough to obtain both N400 and PMN responses to stimulus presentation. Similarly, no experiments were ever created that aimed at eliciting the PMN during passive listening tasks. Although it is often said that the PMN is not elicited without the inclusion of active behavioural task, this is mostly to be attributed to lack of trying.

However, some interesting nuances in the results have not been sufficiently explained by Connolly et al. (1994). For example, the PMN response was biggest when the sentence-final word was both semantically and phonologically incongruous with the rest of the sentence. This would suggest that the PMN is sensitive to both phonological and semantic information. If it were the case that the PMN was only sensitive to phonological information, the component's amplitude would have been very similar between the aforementioned condition and the one where only phonological mismatch was present but semantic one was not.

The PMN component has been labelled as sensitive to phonological mismatch. However, it was only observed — across all experiments — in contexts that also presented processes of lexical retrieval and activation in combination with phonological mapping and mismatch. The component does seem to mirror phonological and word-form mismatch — sometimes as part of lexical retrieval — and it appears to mostly behave differently from the N400 component (D'Arcy et al., 2004). All that we currently know, on the other hand, is that the PMN embodies a response to mismatching information, possibly both phonological and lexical during the perception of a word. Most of the times, this response appears to be to unexpected phonological information in a given semantic (and syntactic) context. Whether the component is a response to the perception of a mismatching phoneme or syllable in the specific, whether it is a response to a mismatching phonological representation, or whether it is an extension of the N400 component with its sensitivity to semantic information is yet to be determined with any certainty. Unless there can be a definitive agreement on the nature of the PMN, it would be unreliable to employ the PMN component in experimental paradigms that aim at determining differences between lexical and phonological processing in speech perception. For instance, recent studies (e.g. Robson et al. 2017) investigated differences between phonological and lexical processing abilities in Wernicke's aphasia patients by comparing the elicitation of the PMN and N400. The study claimed that the two components respectively mirror phonological and lexical processing. However, because the research on the PMN hardly comes to one, clear conclusion about the role of the PMN in spoken-word recognition and speech perception, basing research that claims its role is tied to a particular function could lead to misinterpretation of current (and future) findings.

The future of the PMN

The examples I have provided so far highlight some of the shortcomings of the studies run on the PMN, namely that they all contained some level of overlap between lexical and phonological mapping in their experimental paradigm. This overlap between phonological and lexical mapping is detrimental when the aim of a study is to try and determine which of the two levels of processing is mainly responsible for the activation of a specific ERP component. Arguably, even when using nonce words rather than real words — such as in Kujala's (2004) experiment or Newman et al.'s (2003) — if it is the case that nonce-words are used in a context where an active, behavioural word recognition task has been administered, lexical activation and retrieval is most likely among the effects that are observed in speech perception and spoken-word recognition. There are other relevant studies that aim at exploring the PMN in relation to phonological and semantic processing but, as mentioned before, the key studies investigating the component are, as of 2021, less than a dozen. Moreover, findings are highly inconsistent across studies in regards to the component's topographical distribution. Some limitations can also be observed in regards to some methodological choices across most studies. Lewendon et al. (2020) raise serious doubts about the literature on the PMN. Lewendon et al. (2020) mention many of the inconsistencies that have been discussed in this chapter so far, including Connolly and Phillips' (1994) lack of explanation for a PMN component that was more prominent when both lexical and phonological mismatch were present, together with many more confounding factors and methodological limitations across other studies. An addition to their paper is a table, reproduced below, of results and methodological inconsistencies present across a variety of key studies on the PMN. The table highlights some worrying methodological shortcomings and inconsistencies in the interpretation of the findings across studies with similar experimental paradigms.

Most of the papers considered in the brief meta analysis by Lewendon et al. (2020) — 10 out of 11, to be specific — tested a dozen or fewer participants, which is nowadays considered a very small sample size for reliable data collection in ERP experiments, especially when dealing with under-studied components such as the PMN. Some of these experiments (e.g. Connolly and Phillips 1994; D'Arcy et al. 2004; Newman and Connolly 2009) used 40 or fewer trials per participant, others fewer than 60 (i.e. Connolly et al. 2001). Experiments one to three of this research project, for instance, aimed for 20 participants and a minimum of 100 to a maximum of 150 trials for the target condition, per participant[3]. On top of methodological inconsistencies and gaps in the limited, existing literature, some of the main issues arise in the classification of the PMN. While the component has mostly been observed in frontal and central regions of the scalp (Connolly et al. 1990; Connolly et al. 2001; D'Arcy et al. 2004; Newman and Connolly 2009), it was also described as a parietal response (Connolly and Phillips 1994; D'Arcy et al. 2000) and as a component distributed across the midline (Van Petten et al., 1999) or the entire scalp (Van Den Brink et al., 2001).

In summary, what we know about the PMN component so far, despite the aforementioned

[3] Because in experiments like experiment one, for instance, the target condition is only presented 33% of the time, this means that control trials are often more than 300 to 400 per participant.

Table 2.1: Lewendon's (2020) report on PMN results and methodological inconsistencies in existing research, expanded.

Reference	**Topography**	**Stimuli**	**Methodological considerations**
Connolly et al. 1990	Un-subtracted waves: frontocentral; Subtracted (difference) waves: central	(English) sentences that varied in sentential constraint	10 participants (trials per condition unclear)
Connolly et al. 1992	Flat distribution across midline sites	Phonologically correct masking on the electrophysiological responses to terminal words of spoken (English) sentences differing in contextual constraint	Responses not visible in averaged waveforms
Connolly and Phillips 1994	Frontal, central, and parietal	Terminal words of spoken (English) sentences	10 participants (40 trials per condition)
Van Petten et al. 1999	Flat distribution across midline	(English) gated words were used as congruous and incongruous sentence completions	
D'Arcy et al. 2000	Early N2b: parietal; Late N2b: distributed across scalp	standardized reading test that was formatted for computer presentation	
Connolly et al. 2001	Frontal	visual word/non-word that was followed by the brief presentation of a prime letter with the instruction to anticipate the word/non-word formed by replacing the word's first letter with the prime letter	10 participants (min. 60 trials per condition). Conflicting MEG data acknowledged to invalidate PMN results
Hagoort and Brown 2000	Posterior	(English, spoken and written) sentences were presented that contained three different types of grammatical violations	12 participants (60 trials per condition). "N200" response to semantic expectation violations. No isolation of phonological anomaly.

Reference	Topography	Stimuli	Methodological considerations
Van Den Brink et al. 2001	Flat distribution across scalp	(English) spoken sentences that ended with either (a) congruent, (b) semantically anomalous, but beginning with the same initial phonemes as the congruent completion, or (c) semantically anomalous beginning with phonemes that differed from the congruent completion	
Newman et al. 2003	Frontotemporal	To omit the initial phoneme from a word (clap without the /k/) after which they heard a correct (lap) or incorrect (cap, ap, nose) answer	Early onset P300 contamination in phonological expected condition. Authors could not confirm absence of PMN in this condition
D'Arcy et al. 2004	Frontocentral	visualauditory sentence pairs that related within a semantic hierarchy	10 participants (24 trials per condition
Newman and Connolly 2009	Frontal and central	word/nonword prompt with the instruction to delete the initial sound and determine the resulting segment. Following the prompt [...] an aurally presented response that matched/mismatched expectations	13 participants (40 trials per condition)

studies taking place over the course of more than 30 years, is still very limited. What we do know about the PMN is not universally agreed upon when considering previous studies that have been reviewed in this chapter. This is true both in terms of what the component represents as well as in regards to where it is observed and distributed across the scalp and its relationship with lexical and semantic retrieval. Furthermore, thanks to recent advancements in both the fields of electro-physiology and statistics, we now know that most of the studies run to investigate the PMN component would nowadays be considered under powered, both in terms of participants tested and number of trials per participant. The aim of this particular section of *Chapter 2* is to highlight the need of further experimentation on the function of the PMN component in the role of speech perception and language processing in order to provide an updated account of the component and, possibly, through experiments one and two, further evidence of what does or does not elicit the component itself. The focus of this and future work is eventually to accurately be able to define the PMN as either a response to phonological mismatch, lexical mismatch or a combination of the two. In the following sections, other ERP components often observed in language perception and processing experiments are reviewed and their link to the PMN is discussed in detail.

2.2.2 N400

The N400 component (Kutas and Hillyard, 1980) is a negative-going component traditionally observed at around 400 ms post stimulus onset, although it can extend between the 200 and 500 ms range. It has been, across the decades, linked to lexical and semantic processing and mismatch (e.g. Kutas and Hillyard 1980; Kutas and Hillyard 1984; ...; Rabovsky and McRae 2014; Erlbeck et al. 2014). The amplitude of the N400 component is maximal in central and parietal regions of the scalp (Kutas and Hillyard, 1984). Over the course of the past forty years, the N400 component has been one of the most studied and investigated ERP components in the field of language processing. The N400 has been observed during processes of semantic mismatch and processing in language as well as in contexts that do not involve language or linguistic information. For instance, a visual N400 can be observed when two pictures, whose content is semantically unrelated, are presented during experimentation (e.g. Nigam et al. 1992). The elicitation of the N400 component is not only limited to processing of semantically unrelated items in language and pictures, but its presence is often observed across a variety of sensory information, including olfactory stimuli (Invitto et al. 2018). The N400 component has is also observed when semantic mismatch happens across different sensory modalities, e.g. when the meaning of a word (e.g. *pizza*) does not directly match the flavour of a particular food (e.g. *orange*) (Skrandies and Reuther, 2008).

Although the N400 component has been observed across a variety of modalities and responses to different sensory information, it has been most extensively connected to semantic mismatch in language, in studies of both speech and language perception and comprehension. The N400 component is very reliable in regards to its amplitude, latency and topographical distribution. Because of of this, experiments were designed that aimed to elicit the N400 component in relation to other cognitive processes in the brain. Some of the examples include memory or attention experiments (i.e. if a word was perceived as semantically unrelated in a specific context, it could also be hypothesised that participants were able to access memory information regarding word meaning). Kutas herself, one of the researchers behind the discovery of the N400, describes the component as a very effective dependent variable *"for examining almost every aspect of language processing"* (Kutas and Federmeier, 2011) because of its "remarkably constant" latency (Kutas and Federmeier, 2011).

In the past few decades, many and diverse studies have been run that either directly investigated or indirectly employed the N400 for the study of language. By now, we have a very solid understanding of the nature of the N400 component. The N400 is usually observed at around 400 ms post stimulus onset and it responds to some level of lexical activation and semantic processing and retrieval (Kutas and Hillyard 1980; Kutas and Hillyard 1984; ...). For this reason, a very general statement can be made to say that as early as 400 ms following the perception of a speech stimulus, some process of lexical activation and semantic mapping can be observed in healthy adults. It is thanks to many and many different studies that we can, one piece at a time, draw a clearer picture of what an ERP component looks like as a response to a range of different stimuli. Because there are endless possibilities when it comes to processing information and what type of mismatch can be induced, there are near endless paradigms that can implemented to investigate the behaviour of ERP components during specific cognitive processes.

For instance, even in the case of the N400 that, as I have mentioned, is one of the most studied ERP components, it is still unclear whether it should be interpreted as a response directly to semantic processing (e.g. Kutas and Hillyard 1984) or whether it should also linked to some level of syntactic processing to a smaller extent (e.g. Weber and Lavric 2008). If it is the case that the N400 is sensitive to both semantic and syntactic processing and mapping, it could also be the case that the N400 represents some more general level of cognitive processing, which in itself, would also include semantic mapping. Only by running more and more studies, can more evidence surface that connects existing ERP responses to *unexpected* cognitive processes, allowing for a clearer picture of the role and function of ERP responses.

2.2.3 MMN

The N400 component has been observed as a response to sensory information not limited to linguistic and language data. It is primarily considered a *meaning* component. As previously discussed, the N400 is often elicited as a result of semantic activation and retrieval. Similarly we can say that the PMN has mainly been elicited through some level of lexically-induced phonological mismatch. Differently from the N400 component, we have very little evidence of the PMN being activated outside of a linguistic specific contexts. On the other hand, there is a number of ERP components that are not language or speech specific necessarily but that are important nonetheless in contexts of speech and sound perception. Most of the following ERP components are going to be making an appearance among the experimental findings later on in this research project.

The mismatch negativity (MMN) is a negative-going cross-sensorial ERP component often observed in frontocentral regions of the scalp between 150 and 250 ms post stimulus onset (Näätänen et al. 1993; Näätänen and Alho 1995). The MMN component is commonly studied in the auditory (Näätänen et al. 1993; Näätänen and Alho 1995) and visual domains (Pazo-Alvarez et al., 2003). The mismatch negativity reflects the perception of a deviant stimulus in a sequence of standard stimuli (Garrido et al., 2009). In the auditory domain, a deviant stimulus can be identified by differences in pitch, duration, stress and frequency range (Erlbeck et al., 2014). In the visual domain, a deviant stimulus could be a picture of a red square among a series of pictures of blue squares, as an example (Pazo-Alvarez et al., 2003). The MMN has been elicited in conditions where no active task was present and it requires very little attention to stimulus presentation on the subject's part (Phillips et al., 2000).

The MMN has been linked to the perception of deviant stimuli in both speech and non speech stimulus types in auditory mismatch (Näätänen, 2001). Concerning the perception of speech-specific stimuli, such as syllable or single phonemes for instance, the MMN reflects the perception of changes in the stimulus along a series of different parameters, including duration of the stimulus, frequency distribution, timbre and so on (Näätänen, 2001). The MMN is often elicited by placing a deviant stimulus — a stimulus that, as mentioned, presents one or several different parameters from the norm — in a repetitive sequence of standard, or non deviant, stimuli (e.g. Näätänen et al. 1993; Näätänen and Alho 1995). Because of the consistency of the MMN (or its magnetic counterpart, the MMNm) in terms of latency and topographical distribution, countless experiments employ methodologies that aim to elicit this component in order to investigate many different aspects of speech and speech-sound perception and discrimination (e.g. Näätänen 2001; Pulvermüller 2001; Froyen et al. 2008; ...), similarly to what happens with the N400.

One of the main reasons why the MMN is such a reliable marker of the perception of deviant auditory stimuli is that the component has been elicited regardless of participants attention to stimulus presentation (e.g. Näätänen et al. 1993; Näätänen and Alho 1995). However, many studies have confirmed that focussed attention to stimulus presentation does in fact impact the amplitude of the MMN component (e.g. Woldorff et al. 1991; Woldorff et al. 1993. The amplitude of the MMN is also affected by how big the difference is between the deviant and the standard stimuli. For example, a study by Näätänen 1991 demonstrates how, the bigger the frequency difference between the deviant and standard stimuli, the bigger the peak difference between the two curves at the time the MMN is elicited.

When comparing the MMN to the PMN, one of the main differences would be that the MMN is mostly thought to represent auditory mismatch while the PMN is primarily connected to phonological mapping in the specific. Whether the PMN is mostly a response to the auditory mismatch portion of the phonological form is yet to be determined. However, there has been a series of research studies that linked the MMN component to more language-specific aspects of sound mismatch. For instance, Shtyrov et al. (2003) noticed that the MMN was affected by grammatical changes in syntactic structure of phrases perceived by unattentive participants. These findings suggest that MMN does not only react to just sound mismatch (Shtyrov et al. 2003; Lewendon et al. 2020) and also that syntactic mapping and its influence can be observed earlier than it is usually expected (i.e. around 600 ms with the elicitation of the P600) and without necessarily the need for participants attention to stimulus presentation.

Other studies have also linked the MMN to phonological (Pulvermüller, 2001), semantic and acoustic changes (Weber et al., 2004) in speech and language processing. These examples shift the idea that we have of the MMN as a simple response to sound mismatch to a more complex picture of what the component really represents. This debate also shows how even with very well studied components such as the MMN there is still doubt regarding the precise function or nature of the component itself. If the MMN does represent some level of phonological mapping and mismatch, as well as sound mismatch, with a similar topographical distribution as the PMN, the idea that these two components are somewhat linked to one other, possibly at opposite ends of a spectrum (Lewendon et al., 2020) appears to be always more and more likely.

2.2.4 Why focus on the PMN?

When it comes to PMN research specifically, there are confounding factors in the paradigms of previous studies on the component that raise further questions and highlight the need for additional experimentation to determine what the nature of the PMN really is. Although the PMN has been labelled as a phonological component, it has only been elicited whenever lexical activation was also present. However, phonological processing of speech sounds does not necessarily require semantic mapping by definition. As pointed out, the limitations of previous research are not only limited to experimental confounding factors and variables but also to resulting inconsistencies in the findings of existing studies.

The experiments in this thesis aim to separate semantic and phonological processing in speech perception to try and investigate how the PMN component behaves in contexts where phonological mismatch is present but no lexical or syntactic context directly mediates phonological processing. In experiment three, which takes a different turn than experiments one and two, sound perception is studied where no linguistic information whatsoever is present. The exploratory nature of this work aims at creating environments that have not been looked at before in terms of phonological and sound perception and to try and determine which ERP components, with a particular focus on the PMN and MMN, are elicited when different types of stimuli and different types of task are employed in the design of experiments. In experiments one, specifically, separation of lexical and phonological information is done by the creation of an extremely miniaturised artificial language (discussed in more detail in *Chapter 4*.

Some have argued that the PMN represents phonological mismatch during lexical retrieval (e.g. Connolly and Phillips 1994), placing the PMN component closer to components such as the N400 on a spectrum that goes from sound processing to the processing of meaning. However, another possible direction is that the PMN is an extension of neighbouring pre-N400 components such as the MMN or that the two components share similar origins. Although the MMN has been often linked with lower-level sound mismatch, evidence was also provided that connects the MMN to phonological mismatch, as previously discussed. The similarities between the two components are striking and, while the MMN has been proven to be elicited regardless of participant's attention, which has never been a focus for PMN studies, it is not to exclude that there is an underlying, common origin for the two components. Dissociating phonological processing from semantic and syntactic processing should allow access to insights in speech-sound processing in semantics-free contexts. This, in turn, could also help determine the amount of sensitivity of both components to phonological or semantic information.

2.2.5 Learning by dissociation

Dissociation can be a very important keyword when it comes to learning more about ERP components and their function in a particular series of processes, such as speech perception in this particular case. Dissociation is a simple concept to understand but, especially when it comes to ERP research, a much more difficult one to implement and examine.

Many studies have managed to separate (i.e. dissociate) and have presented differences between the N400 and P600 components by providing evidence that presents the response to semantic (N400) and syntactic (P600) stimuli as different (e.g. Guillem et al. 1995; Brouwer et al. 2017; Brouwer and Hoeks 2013; ...). A dissociation can be created by, for instance, determining that different types of stimuli generate responses that differ in amplitude, latency, scalp distribution and so on. However, once we have determined that semantic and syntactic processing point to the activation of different processes in the brain, this does not automatically prove that the P600, for instance, is only and specifically linked to syntactic processing. At the same time, it is only possible to claim that the N400 and P600 are dissociated in the particular environment that they have been tested in. This, in turn, would also mean that unless all possible environments concerning the activation of either the N400 or P600 have been explored, there is always a remote but not completely unlikely possibility that these two components represent similar processes in specific environments and that they are not fully dissociated.

Another way ERP research conducts investigations into the role of specific ERP components is, rather than studying multiple processes and determining what responses they cause, by eliminating every other process but the one that is being investigated. This is done with the aim of determining what type of response that specific paradigm creates without the influence of external variables. In more specific terms, the aim of this research project is to investigate whether the PMN component's response is to phonological information specifically rather than to lexical and phonological at the same time, or simply lexical. What can be done to achieve this is to separate the two dimensions of lexical and phonological processing and to test only one of them at the time. However, this is not always possible.

Let us try to imagine how one could test semantic mismatch in language without activating phonological processing. In order to convey a certain meaning, we do need to do so through a linked, specific form. What can be done, however, and what is currently devised in this doctoral project, is to test phonological processing without the activation of lexical information. This, on the other hand, can be easily achieved by presenting participants with nonce words or sequences of syllables that do not amount to a specific word in the target language. Separating components to learn more about a specific process or, conversely, separating processing to learn more about specific components is the key to providing clear and reliable evidence to build knowledge and an accurate atlas of ERP components.

Once a very in-depth understanding and grasp of all the ERP components that make up a specific process has been obtained, experiments aiming at elicit specific ERP components can be devised to explore more general, theoretical questions. For instance, in this very case, having a clear understanding of what every component in speech perception represents would allow us to be able to devise experiments that reliably tested more general and overarching questions and hypotheses in the realm of speech perception theories and models. The following sections aim at introducing different theoretical frameworks regarding speech perception and information flow order. Although the question addressed by this work remains much more specific, dealing with the nature of a small ERP component, all of the research conducted aims at, in the future, building a better picture of what speech perception and spoken-word recognition look like as a whole when it comes to what processes are activated in the brain.

2.2.6 ERP experimental paradigms

Like in the majority of experimental fields across science and social sciences, a number of different paradigms can be used in ERP experiments when investigating language processing, comprehension and perception. In particular, a number of decisions need to be made when designing an experimental paradigm so that it is able to capture and elicit the ERP components that are the focus of the study. For instance, because the focus of experiments one to three is to determine whether sound and speech-sound processing elicit phonological mismatch regardless of semantic mapping, the decision was taken to only include semantics-free stimuli as part of all experiments, in order to avoid confounding variables. In this section in particular, I want to address the difference between the implementation of *passive listening* (Experiment one) and *active listening* (Experiment two and three) experiments, why the two differ, what are the relevant differences and why it is necessary to take both types of paradigm into account when investigating the nature and behaviour of specific ERP components such as the PMN.

Passive listening tasks (or experiments) during EEG data collection are characterised by participants being exposed to stimuli without having to pay direct attention to the their presentation and without the inclusion of behavioural tasks. On the other hand, active listening is often characterised by the implementation, during ERP measurements, of active behavioural tasks that prompt participants to directly pay attention to the stimuli presented. The main advantage of a passive-listening task is that motor activity, which could create further noise and interference in the data collected, is reduced throughout the experiment as no behavioural responses have to be provided by the participants.

Active tasks make up the majority of experiments in ERP research. Some examples, in speech perception specifically, include all of the experiments on the PMN discussed above (e.g. Connolly and Phillips 1994; Kujala et al. 2004; D'Arcy et al. 2004), where participants were instructed to either pay attention by simply being told to listen to the stimuli presented (Connolly and Phillips, 1994) or where behavioural tasks were included so that subjects had to modify or recognise incoming stimuli or press buttons according to whether they thought a stimulus belonged to a specific category (Kujala et al., 2004). Direct attention to stimulus presentation during auditory ERP experiments can be achieved with devices as simple as telling participants to pay attention to stimulus presentation and does not necessarily require the inclusion of complicated, behavioural tasks. The main advantage of using active tasks in an ERP experiments is that the amplitude of many components is directly mediated by attention to stimulus presentation, with many ERP responses being easier to observe when participants are directly paying attention to each of the stimuli. Some of these ERP components include the MMN (Näätänen and Alho, 1995), N400 (Kutas and Hillyard, 1984) and P600 (Verhees et al., 2015). Some ERP components have never been observed in contexts where participants were not directly observing stimulus presentation, such as the PMN. On the other hand, some components are generally present in contexts where participants are not expecting the presentation of a stimulus, such as the N1 (Näätänen and Picton, 1987), which is in fact suppressed if the participant can predict an unexpected stimulus is going to be presented (Grau et al., 2007).

In passive (listening) experiments, on the other hand, participants are instructed to ignore the presentation of stimuli during data collection. This is done either by simply instructing subjects not to pay attention to the experiment or by instructing participants to, in case of auditory listening experiments, watch a mute film or mute video on a computer screen to make sure that attention is diverted away from auditory stimulus presentation. Low-level components that do not require direct attention to stimulus presentation and that are usually a direct response to stimulus perception, often found in the N1 and pre-N1 range, can be observed without direct attention to stimulus presentation. Components that follow the N1 can also be observed but, as mentioned in the paragraph above, amplitude tends to be lower for all ERP responses in general when compared to responses paired with active behavioural tasks. Passive listening tasks do not have many benefits over active experiments. However, the main benefit is that ERP responses are generally smaller, which can be considered a positive side effect if components that are not the focus of the study want to be limited. For instance, in a study on language processing, the elicitation of a non-language specific component such as the P3 might be considered an issue more than an asset. For this reason, implementing a passive paradigm might reduce the overall amplitude of components such as the P3 which, in turn, would reduce the level of contamination on neighbouring trends. However, the majority of experiments that include passive tasks do so in order to directly compare the responses collected to those of corresponding active experiments, in order to determine whether a specific ERP component can be observed regardless of participant's attention or whether certain cognitive processes take place regardless of an active implemented task.

A study by Oades and Zerbin (1995) combined active and passive tasks to study topographical component distribution during an oddball paradigm experiments to determine whether temporal attention to stimulus presentation directly impacted component origin. This study discovered that, while pre-P2 components (e.g. N1) were characterised by similar latency and topographical distribution regardless of focused-attention, post-P2 components varied both in terms of latency and topographical distribution for target odball stimuli, suggesting that multiple origin points for the same component are present depending on the level of focused-attention to stimulus presentation (Oades and Dittmann-Balcar, 1995). In the experiments of this research project, both passive and active experiments are implemented. The aim is to explore all possible contexts in which the PMN can be elicited in order to study whether focused-attention to stimulus presentation is a requirement for the component's elicitation or whether it is not.

2.3 Theoretical assumptions

When investigating ERP in language processing and comprehension, the current thinking has been that of linking specific ERP responses to matching grammatical modules. While this is not true for every ERP component, especially because many of the ERP responses observed during language comprehension are not directly linked to language specifically (i.e. N1, P3, etc), it is more the case for components such as the N400, PMN, P600 and, to a smaller extent, components such as the MMN. Sometimes, it can be hypothesised that there is a one-to-one relationship between a specific ERP component and a corresponding grammatical module, like the N400 being primarily linked to semantic processing. Other times, ERP components can be observed as responses to multiple levels of processing, with some of these related to grammatical modules. This is the case for responses like the MMN, which is found both as a response to mismatching stimuli in a sequence as well as being influenced by phonological processing.

Being able to link language-specific responses to specific grammatical modules allows researchers to run ERP experiments in order to investigate general questions about architectures of grammar, by observing which ERP responses are activated and when they are activated, for instance, presenting participants with ambiguous information and mismatching linguistic stimuli during experimentation. However, in order to be able to accurately investigate grammatical architectures by the study of ERP components, it is necessary to have a very clear understanding of what each of these ERP components indexes when it comes to language and speech processing specifically.

The current thesis aims to answer one main research question: whether the PMN can be observed as a response to some degree of phonological mismatch in the absence of lexical activation of any of the stimuli. In particular, all stimuli used across the three experiments in this collection are not lexically activated. While previous literature (e.g. Connolly et al. 2003, Kujala et al. 2004, Connolly et al. 2009) showed that the PMN is found as a response to both lexically-activated and non lexically-activated stimuli in a non-sentence context, some forms of lexical activation and lexical expectations were always present through the implementation of behavioural tasks, or through the presentation of lexically-activated priming stimuli, presented at the beginning of each trial. For this reason, a confound is present so that it is not possible to determine whether the PMN is acting as a response of the presentation of a mismatching word form (phonological information) of a non lexically-activated stimulus or whether it is a response to a mismatch which derives from a "failed" lexical retrieval task. In this section, I aim to investigate theoretical assumptions in regards to the nature of phonological and lexical processing and storage during speech perception, in order to clearly and explicitly define the context in which the three current experiments have been designed. Furthermore, the aim of the current section is to provide a brief introduction to current issues in modelling speech perception, which a component like the PMN — and ERP studies in more general terms — could help address once a solid foundation has been built on the role of ERP components in language processing.

2.3.1 Information flow order

Across the decades, many theories of architecture of grammar and models of speech perception have tried to shape the relationship between different levels of language processing (e.g. phonological and lexical) as well as trying to define those processing levels as one or another series of cognitive processes. Some of the original research on the PMN (e.g. Connolly et al. 1992; Connolly and Phillips 1994) was grounded in the cohort model of speech perception (Marslen-Wilson, 1984). The Cohort model is the first speech perception model of word recognition, proposed in 1984, two years prior the formulation of the TRACE model (McClelland and Elman, 1986). It is the first model to take full account of the temporal and transient nature of the speech signal. The Cohort model suggests that word recognition comprises three separate stages: *activation*, *selection* and *integration*. Phonetic information activates a cohort of possible competing lexical entries that, the more information is provided, the more are inhibited until one item is selected and integrated (Marslen-Wilson 1984; Weber and Scharenborg 2012). In particular, in the Cohort model, only lexical entries that share the initial phonemes with the presented stimulus are activated, roughly after around 150-200 post stimulus onset (Marslen-Wilson, 1984). The more phonemes are presented, the more mismatching lexical entries are then eliminated from the cohort until hopefully only one candidate remains.

Different speech perception models and difference theories behind architectures of grammar provide very different assumptions. This is true both regarding the direction of the information processing chain, as well as completely different takes on the type of mental representations and the way information is accessed during cognitive processing of language and speech, i.e. whether lexical feedback is present or not. However, experiments one to three aim of the current thesis aim to elicit the PMN component with experimental designs primarily based on previous, published methodologies used to elicit the PMN, rather than on specific theoretical assumptions. In fact, not only does this work not commit to a specific framework of speech perception and architecture of grammar, but the main point behind studying the role of the PMN is so that it later can be exploited in designing experiments to investigate theories of speech perception and grammar.

Models of speech perception such as the Cohort model Marslen-Wilson (1984) and the TRACE (McClelland and Elman, 1986) model, to pick two of the most influential models over the past decades, differ in the way they model information being accessed and processed during processes of speech perception. Unlike the cohort model, TRACE is an interactional model of speech perception. Unlike theories of modularity of mind and feed-forward perception and processing models (e.g. Fodor 1983), the foundation of the TRACE model rests on a structure called "the Trace", which is a network of units that work for both perception and memory in the brain (McClelland and Elman, 1986). One key difference between the two models is that the TRACE model also accounts for lexical feedback, which the cohort model does not.

Lexical and pre-lexical feedback is the idea that information can travel from more abstract processing units to less abstract ones during processing and perception, in order to facilitate spoken-word recognition (see Getz and Toscano 2021 for a review of recent findings). for example, contextual syntactic and semantic information might mediate the perception of low-level phonetic or acoustic processing. A few examples of very well studied phenomena related to the existence of lexical feedback are the Ganong effect (Ganong, 1980) and lexical feedback in compensation for co-articulation (e.g. Samuel 1996; Samuel and Pitt 2003; Magnuson et al. 2003.

Recent findings in cognitive sciences, with both evidence from fMRI and EEG, strongly support an interactive, top-down model of speech perception, with lexical and linguistic information influencing early stages of sound perception (Getz and Toscano 2021). For example, evidence from MEG suggests that areas of the brain often associated with high-level language processing appear to be activated during acoustic and phonetic processing, too (Gow Jr and

Olson, 2015). More evidence also supports a more complex, adaptable and flexible set of cognitive processes in speech perception, supported by parallel processing streams (Scott, 2017). In a recent model summarised by Hickok and Poeppel 2016, speech perception is explained by two parallel, processing streams (i.e. ventral and dorsal streams) that process speech signals for comprehension and acoustic signals respectively (Hickok and Poeppel, 2016). However, while the current agreement in speech perception theory is that interaction is present and feedback mediates spoken-word recognition, the exact extent to which top-down influences and interaction mediate sound perception is yet to be determined.

In an experimental design such as the majority of the ones on the PMN, where lexical and pre-lexical mismatch are both present, whether contextual lexical information can mediate pre-lexical processing could matter and could potentially explain some of the findings on the PMN. In that particular case, what type of model is chosen to frame the assumptions and the findings could change the discussion on the PMN entirely.

However, in the experimental designs used across this thesis, no lexical activation or lexical retrieval / manipulation task is present at all across the entirety of the study. In experiment one, the priming used to create mismatch is only explained by transitional probabilities, and no abstract, linguistic information such as lexical or syntactic information is present. Any mismatch found in relation to the PMN could be safely attributed to pre-lexical mapping, including word-form, phonetic or acoustic mismatch. In experiments two and three, mismatch is created by the behavioural task, rather than by expectations caused by lexical activation or retrieval (e.g. Connolly et al. 1992; Connolly and Phillips 1994; D'Arcy et al. 2004). For experiment one, the only general assumption is that expectations can be caused by transitional probabilities, which is widely accepted by all models of speech perception, including cohort and TRACE (Marslen-Wilson 1984; McClelland and Elman 1986). For experiments two and three, where the behavioural task does not entail any use of lexical activation or processing, information flow order between less and more abstract layers of language processing should not influence mapping and mismatch of phonetic information.

One of the key elements of this work is that, not only are theoretical assumptions limited when it comes to models of speech perception, but also that the findings of this work, whether the PMN exists in response to non lexically-active phonological mismatch, can also later be used to explore the extent to which lexical and pre-lexical feedback influences earlier level of speech perception. While current theories of speech perception drive away from a completely modular account (e.g. Scott 2017; Getz and Toscano 2021, the extent to which feedback and interaction in language processing work requires further investigation, which could be aided by newly acquired knowledge on components such as the PMN.

2.3.2 Mental representations

Another major discrepancy among different models of speech perception and architectures of grammar is that of mental representations. Especially when talking about the PMN indexing phonological and lexical processing, different models assume that this information is linked and retrieved differently in the brain. In this section, two main competing accounts are presented and a point is made that, regardless of the underlying assumption, the current studies of the PMN should be able to collect evidence regarding the nature of the PMN to investigate its link to phonological and word-form mapping specifically.

In particular, most recent studies on the PMN define the component as as goodness-of-fit response between word-form information and speech input, regardless of lexical status of the stimulus (e.g. Newman and Connolly 2009). This suggests that word-form information of an expected stimulus is pitted against word-form information of the speech input and, when the two do not match, a PMN component is obtained. This happens regardless of lexical activation of the target stimulus, as shown by previous literature (e.g. Newman et al. 2003; Newman and Connolly 2009). However, it is still unclear what aspects of word-form information the PMN is sensitive to. In all studies on the PMN, the phonological / word-form mismatch was obtained with a segmental mismatch, where at least the first phoneme of the target stimulus was presented as different than that of the priming stimulus (e.g. Newman and Connolly 2009) or from the expectations created by contextual information (e.g. Connolly et al. 1992; Connolly and Phillips 1994).

Different models propose that the "best-fit" word-form (and lexical) information can be stored and retrieved in different ways. For this particular work, we will focus on abstract representations and exemplar accounts. All models of speech perception mentioned until now (e.g. Cohort and TRACE models) assume that, for each lemma, there is an abstract entry and representation of that particular item, linked to an abstract representation of its word form and meaning (Marslen-Wilson 1984; McClelland and Elman 1986; Norris 1994; Scharenborg and Boves 2010). The Minerva2 model is, on the other hand, an exemplar model of memory (Hintzman, 1986).

The Minerva2 model was used to simulate spoken word recognition through episodic memory, explaining that acoustic information and differences across utterances are relevant and they are directly matched with all episodic representations of the instance of a particular item, every time it is uttered. With Minerva2 there is no need for feedback, as there is no middle layer between input and lexical representations, since each input is directly mapped onto its episodic counterpart (Goldinger, 1998). However, it's been argued that a direct mapping of form and meaning for each input is much less optimal and it is computationally expensive. Breaking the process of word-recognition down in separate cognitive steps — as it has been described by other models we have previously discussed — is probably preferred and consistent with priming and cognitive experiments.

To summarise, most speech perception models assume that an abstract, lexical representation — linked to one phonological word-form representation — is stored in memory and, depending on the model, accessed and selected through different cognitive processes that might or might not include interaction and feedback. The Minerva2 and other exemplar models of memory and speech perception, on the other hand, assume that lexical and phonological (and acoustic, and contextual, etc.) information are stored together in one exemplar, for each observation of a given exemplar. These exemplars are then stored in proximity with similar other exemplars and the speech input is compared to the existing, stored exemplar clouds in order to activate and select the correct item.

However, both previous literature on the PMN and the current study define the PMN as a mismatch between speech input and the selected, expected word-form of an incoming stimulus, whether the stimulus contains lexical information or not. For this reason, the only underlying assumption necessary for the mismatch causing the PMN to occur is that some phonological, word-form information is selected to be matched onto the incoming speech input. All models of speech perception assume that some item and its corresponding form information is selected after the process of retrieval, regardless of whether this happens through the selection among a cohort of mental representations or whether the speech input is compared to clouds of exemplars stored in memory. As long as some word-form information is selected at the end of that process and as long as that information does not match the speech input the PMN should be elicited.

Regarding stimulus lexical information, absent across all three experiments presented in the current thesis, the type of representation or how the information is stored should also not affect the elicitation of the PMN component. Whether lexical representations are selected from a cohort of options, or whether stimulus lexical information is present in the exemplar together with matching word-form, episodic and acoustic information should not play a role in the comparison of expected word-form information with the speech input, as that information is retrieved either prior or simultaneously to lexical information.

The phonological information that the PMN indexes is broadly associated to word-form information in most of the existing literature on the PMN. Future and more specialised studies on the PMN could also help inform whether the component is mostly sensitive to some aspects of word-form information more than others. As mentioned above, all pasts experiments elicit the PMN with mismatch in segmental word-form information (i.e. mismatching vowels, consonants) but none focus on other aspects, such as intonation and other supra-segmental features. In addition, lexical information, both contextual and specific to the stimulus, does not seem to be necessary for the elicitation of the PMN, as the component was evoked as a response to mismatching stimuli in minimal settings as seen in Newman et al. ()2003; 2009). However, identifying whether word-form mismatch alone can trigger the PMN might suggest that later accounts of the nature of the PMN are more likely to be correct, compared to original discussions on the component that placed it as a direct response to mismatch in the lexical retrieval stage of spoken-word recognition.

2.4 Where do we stand?

The use of electro-encephalography is very promising in the context of providing more evidence of the extent of bottom-up and top-down processing of speech perception, primarily because of its great temporal resolution. Understanding more about when lexical activation happens and its mediation of phonetic processing (e.g. Getz and Toscano 2021) can provide valuable benefits to understanding information flow order and cognitive processing in speech and sound perception. This endeavor, however, does not come without its drawbacks.

The PMN component stands out from other well-known ERP components in the realm of speech perception and spoken word recognition. This is because of the limited amount of studies run to investigate its nature and because of its dubious role as a response to word-form processing in relation to lexical activation (Lewendon et al., 2020). Being able to precisely pinpoint the role of the PMN component, whether strictly phonological, lexical, a combination of these two or a response to anything else, by separating specific aspects of speech perception with dedicated experimental paradigms, can be informative in multiple ways. Knowing more about the PMN component and its role in speech perception can be an extremely advantageous tool for the study of speech perception, modularity of mind and grammatical architectures.

Chapter 3

Methods

3.1 Introduction

Electro-encephalography (EEG) is a non-invasive electrophysiological technique used to record brain activity. By placing electrodes on the scalp of experimental subjects, EEG allows for the recording of shifts in electrical activity in the brain. In experimental paradigms using EEG, electrical potential is often measured in response to the presentation of specific time-locked stimuli or events. These shifts in electrical potential, used to investigate cognitive processing in response to external events, are measured in the brain as a responses to experimental stimuli and they are called event-related potential (ERP) components. Compared to other imaging techniques, such as functional magnetic resonance imaging (fMRI), the main advantage of EEG and ERP studies is their temporal resolution. EEG equipment is often times capable of measuring shifts in electrical potential with a sampling frequency of up to 2048 Hz (2048 measurements per second). The temporal resolution that this method provides makes it a primary choice when investigating the response timing of the brain during the perception and processing of any type of information. On the other hand, imaging techniques like fMRI have the advantage of a much more precise level of topographical imaging that is just not available with EEG. One of the reasons behind this differences is that the electrodes are placed, for obvious reasons, on the outside of the participant's scalp. Because of this, they do not allow accurate (or true three-dimensional) measurements when it comes to ERP component origin and topographical distribution.

The current chapter describes the equipment and the procedures employed for the collection, pre-processing and analysis of EEG and ERP data, across all experiments in this thesis. The hardware and software used is presented in detail and the pre-processing steps — used to turn raw, continuous EEG data into filtered, baseline-corrected and subject-averaged ERP data sets — are explained and justified to the reader. Information regarding experimental design and stimulus presentation and synthesis can be found in the chapters dedicated to each experiment. Whenever possible, consistency in collecting, processing and analysing data across all three experiments — together with the development of similar experimental paradigms throughout the study — was a priority during the creation of protocol and the choice of software, hardware and methods for the realisation of all experiments.

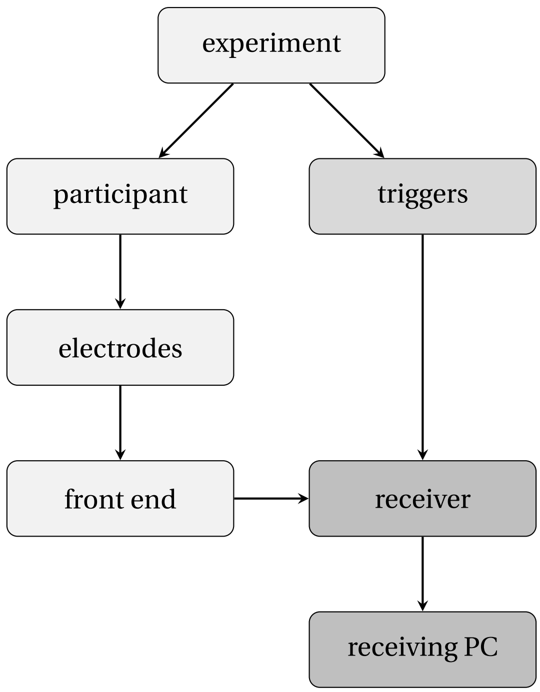
3.2 Location

The Psycholinguistics Laboratory at The University of Manchester was used as primary location for data collection and storage throughout all experiments. The laboratory is equipped with a sound attenuated (34 dB attenuation) experimental booth, with one experimental position dedicated to EEG data collection. The EEG position in the sound attenuated booth contains a desk, housing two computer monitors, a set of speakers and a chair, facing the desk and placed 1.5 metres from the main experimental-presentation screens. The speakers are placed on the sides of the monitor, facing the participant at equal distance. Across all three experiments, the monitor was only used to present a fixation cross and messages indicating the start and end points of the rest blocks. All stimuli were presented auditorily only. The volume was kept consistent for all experiments, at a comfortable level that allowed both clear listening and maximum comfort for the entire duration of the experiment. The average dB level, measured from the participant's perspective, was of an average 60 dB. A computer keyboard and a mouse were placed on the desk, but no behavioural task that required their use was implemented while recording EEG data. Because no input from the participant was recorded, there was no need for using specialised behavioural analogue boxes or gaming controllers. Keyboard and mouse were only used to skip rest blocks and start the experiment. Electrodes are connected to a front-end device in the experimental booth, placed on a second desk behind the participant. Both the receiving and experiment presentation computers are located outside the booth to avoid alternate current (AC) interference. No other electrical device was plugged in inside the booth during the data collection phase. No artificial (or AC) lighting was used to avoid further interference. The experimental booth is naturally lit by a window on its side.

3.3 The equipment

This section aims to describe all hardware and software used for the execution of the experiments and the processing of the data.

Figure 3.1: Signal flow from stimulus presentation to EEG data capture



- **Electrodes:** 64 BioSemi pin-type active electrodes (A+B) for the scalp and 6 flat-type active electrodes (EX1-6) for the eyes were used across all three experiments of this research study. Ag-AgCI electrode tips provide low noise and low offset voltages. Pin-type electrodes were applied onto a 64-channel BioSemi head cap with a 10:20 electrode distribution system as pictured in figure 3.2
- **Front-end:** *BioSemi Active-Two AD-box* is a low-power, galvanically isolated front end with the possibility of connecting up to 240 electrodes. The box was placed near the participant during data collection and it was connected to the electrode bundles originating from the cap. The box is DC powered with a battery to avoid AC-related interference affecting the data. The front end is itself connected to the receiver through an optic fibre cable.
- **Receiver:** The *BioSemi USB2 Receiver* converts the signal from optic to digital and it transfers

the data to the receiving computer, where it is stored. The receiver also collects data sent as analogue triggers (i.e. markers that temporally align the presentation of the stimuli with the continuous EEG data, necessary for the computation of average ERP).

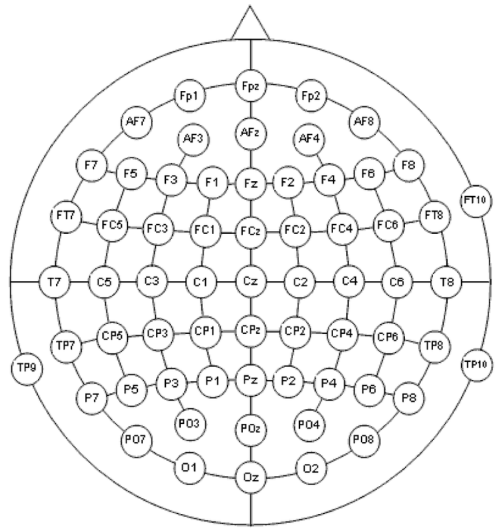
- **Software:** All experiments' stimulus presentation was coded in *Neurobehavioral Systems*' **Presentation** (Systems 2004; v. 21.1). For the experiment structure, two proprietary languages were used for the coding: `PCP` and `SPL`. The data were monitored during collection and they were saved to disk with the software suite **ActiView** by **BioSemi** (v. 7.07). EEG data preprocessing was carried out in **Matlab** (R2019a) and its extensions **EEGLAB** and **ERPLAB** (Delorme and Makeig 2004; Lopez-Calderon and Luck 2014; version 2019.1 and 7.0.0 respectively). Averaged ERP data were later processed, analysed and visualised with `R` (R Core Team 2020; version 4.0.3) in **RStudio** (RStudio Team 2020; version 1.4.11). R packages used for the processing, visualisation, significance testing and modelling of ERP data include the `ERP` (Causeur et al., 2018), the `lme4` (Bates et al., 2007), `lmerTest` (Kuznetsova et al., 2017) and `pbkrtest` (Halekoh et al., 2017) (imported through `lmerTest`) packages as well as packages from the `tidyverse` (Wickham et al., 2019) suite. Post-hoc pairwise comparisons were carried out in `R` with the use of the `emmeans` (Lenth et al., 2019) package. *Type III* ANOVA tables with adapted F-tests for linear mixed effect models (Kenward-Roger approximation; Halekoh et al. 2017) were created using the `anova(type = "III", ddf = "Kenward-Roger")` function, modified by the `lmerTest` and `pbkrtest` to work specifically with `lme4` objects. In case of non-significant interactions, the best model fit was chosen by AIC/BIC comparison using the `anova()` function from base `R`. Topographical scalp maps were created using cubic spline interpolation methods from the `akima` R package (Akima et al., 2016). All analyses, visualisation and modelling were *knit* from R Markdown into commented html files for improved readability. All output files are freely available to download from GitHub links as detailed in the final section of the current chapter.

- **Audio/video:** The computer monitor used for the presentation of experimental stimuli and instructions is a Dell 23" computer monitor with a resolution of 1900x1080 and an aspect ratio of 16:10. The audio presentation device is a pair of Logitech stereo speakers.

3.4 Protocol

Participants were recruited among the undergraduate and postgraduate student population at the University of Manchester and they were compensated č20 for their time. Each exper-

Figure 3.2: Channel locations for the 64-electrode 10-20 BioSemi system.



imental session, from start to finish, lasted two full hours on average. Undergraduate experimental assistants were recruited to help with experimental and equipment set-up. One researcher was always present during data collection to ensure the protocol was followed correctly.

3.4.1 A solid foundation

The chapters focussing on each of the experiments (*Chapters 4*, *5* and *6*) deal in detail with how the stimuli were created and how they were arranged when presented. The aim of this section, on the other hand, is to introduce a paper, by Astheimer and Sanders (2011), whose influence was crucial in determining the nature of the stimuli used in experiments one and two. Furthermore, some methodological choices used in the study by Astheimer and Sanders (2011) were also employed in the experiment of the current research project, working as a baseline for the creation of the general experimental paradigms for all three experiments. This is particularly true for the creation of the experimental paradigm of experiment one. Astheimer and Sanders (2011) created partwords by combining 11 synthesised CV syllables and they then trained participants to learn to recognise the stimuli through computerised training and testing experimental blocks. This was done with the aim of testing whether presentation of word-initial syllables, in continuous speech and during passive listening, elicited a greater response compared to word-medial and -final syllables. ERP evidence can be used to show that focussed attention is shifted to a greater extent to the presentation of word-initial syllables, since they are assumed to be more informative in spoken-word recognition (e.g. Connine et al. 1993). The results of the ERP study confirmed that word-initial responses elicited a greater N1 response linked to attention and unexpected stimuli (Näätänen and Picton, 1987), compared to word-medial and -final syllables after training (Astheimer and Sanders, 2011).

The results of Astheimer and Sanders' (2011) experiment are directly relevant to the aims of this thesis, as they inform us that focussed attention is stronger to the presentation of word-initial syllables because of the higher amount of information they carry during spoken-word recognition. However, what is even more relevant is the level of detail and planning applied to the realisation of the methodology and the choice of the stimuli. All part-words created were balanced to resemble real speech in terms of phonemic and syllabic transitional probability ranges. Moreover, ulterior exploratory and behavioural checks were run to make sure that none of the part-word stimuli stood out from the pool of stimuli for any particular reason nor that any of the items resembled a real word in English. For this specific reason, the same blueprint of the stimuli used in Astheimer and Sanders' (2011) was applied for the creation of stimuli of experiments one and two. In addition, our stimuli in experiments one and two were controlled to an even stricter extent with syllable durations being exactly the same for every syllable. In Astheimer and Sanders (2011), syllable duration ranged between 190 and 310 ms. This extended range could have played a confounding factor in the current experiments, where the focus is on testing the response to phonological and phonemic differences between syllables.

Other methodological components from Astheimer and Sanders' (2011) paper are present in the current experiments. For instance, a similar experimental protocol was adopted, with the order of presentation of the experimental blocks and the choice of having a baseline EEG recording being extremely alike in both experiments. The baseline recording — before participants had learnt the stimuli — was followed by training and testing behavioural tasks, before EEG is later recorded again once participants had learnt the nonce words devised for the experiments. Overall, drawing on stimuli and experimental methodology choices that have been successfully tested in other experiments improves the validity of the methodology of experiments one, two and three. As an added benefit, similar methodological choices between experiments add an extra layer of comparability between new and old research that might allow for direct and indirect comparisons of findings, working as partial replication of older studies in the field.

3.4.2 Equipment set-up

Participants were seated on a chair and their cranial circumference measurements were taken in order to choose one of two sizes of head-caps (Small, 52-56 cm; Medium, 56-60 cm). Then, once the cap had been applied, the nasion-to-inion distance was measured. To allow for topographical normalisation, the nasion-to-inion distance was used to be able to normalise across different size heads (inter-participant). The Cz electrode was adjusted to be found at the midpoint of the scalp for every participant across all three experiments. Once in place, the pin-type scalp electrodes were applied to the cap, using an electrolyte solution (*Signa Gel*) for better conductivity and stable contact between the scalp and the electrode tip. The gel was applied to each hole of the cap with plastic syringes, prior to the application of each electrode. Six face electrodes were subsequently applied to monitor noise caused by eye movements and blinks and activity caused by nearby face muscles. Two electrodes were placed on the side of each eye to capture lateral eye movements. Two electrodes were placed at the top and two more at the bottom of each eye to identify blinks. These electrodes are generally referred to as EX1 to EX6 and they were applied to participants' faces with the use of a double-sided tape (specifically designed for this application) and a drop of electrolyte solution for best signal-to-noise ratio measurements

3.4.3 Dealing with noise

After all electrodes were in place, the quality of the signal was visually monitored in ActiView. Electrodes that showed an abnormal impedance measurement were corrected with the addition of more electrolyte gel to increase signal-to-noise ratio and, therefore, reduce impedance. Although, in most cases, this one step was sufficient to restore the channel to an usable state, the signal from some of the electrodes remained noisy at times. In these specific cases, the noise was corrected during the pre-processing phase (after data collection, before re-referencing to the average of all electrodes) either through spherical interpolation or by directly removing the electrode channel from the dataset entirely. In the vast majority of cases, however, applying a notch filter at 50 Hz was enough to get rid of the majority of noise across all participants' data.

3.4.4 Instructions

A different set of instructions was read to participants for each experiment and specific details are discussed in the respective chapters. The average data collection phase lasted around one full hour, including breaks. Among the instructions, participants were told to keep their gaze on a fixation cross, placed in the middle of the screen during the duration of the entire experiment, with the exclusion of rest blocks. Participants were also instructed to moderate and/or control blinks whenever possible, blinking especially only when stimuli were not being presented, during the inter-stimulus interval (ISI). Generally, participants were given a 30-second rest block every about 3 minutes of experimental trials.

To avoid excessive fatigue, the possibility of taking a longer breaks during data-collection blocks was offered to all participants. However, no participant took advantage of this offer in any of the three experiments. Only programmed rest blocks were used with no exceptions. Before the start of an experiment, participants were shown live EEG data recordings of their own brain from a secondary screen — placed in the booth and turned off during the duration of the experiment — to show the negative effects and the amount of noise created by sudden movements or excessive blinking.

3.4.5 Debriefing

Upon conclusion of the experiments, the head cap was removed from the participants' scalps and so were the face electrode. A short debriefing followed. Compensation of GBP 20 participant was provided and participants were then dismissed. Lastly, electrodes and head cap were cleaned with lukewarm water and disinfected with the use of isopropyl alcohol after every experiment. Equipment was then placed onto dedicated drying racks to dry until the next experimental session.

3.5 Data pre-processing

A fixed protocol was not only followed for the experimental procedure but it was also extended to data pre-processing and, in particular, to all the steps necessary for converting raw, continuous EEG data into averaged ERP data. The data collected consisted of one single file per participant. The file contained measurements of continuous electrical activity over 70 electrode sites (i.e. 64 scalp + 6 face electrodes), recorded at a rate of 2048 kHz and at a resolution of 32 bits.

3.5.1 Preliminary steps

Each BioSemi .bdf file was imported into Matlab and EEGLab and it was turned into an EEGLab dataset. Channel locations were added to each of the 64 scalp channels for future topographical analyses on the data. The experimental triggers sent during data collection from the experimental presentation software were paired with corresponding labels used to disambiguate which trigger (or code) was representative of which type of event (i.e. presentation of a matching or mismatching stimulus). Before filtering, channel data was briefly visually analysed for particularly noisy channels or issues with data collection, alignment of experimental triggers and so on. EEG data were also re-sampled from 2048 to 512 Hz, from the starting 2048 Hz. Resampling allows for faster pre-processing and smaller-size datasets. For reliable measurements, sampling rate should be at least double the highest frequency of interest (Nyquist frequency). In this particular case, 512 / 2 Hz is still a very high frequency considering all ERP experiments linked to language processing are usually characterised by frequencies between 5 and 15 Hz.

3.5.2 Filters

The recordings from all electrodes were notch filtered at 50 Hz to remove AC interference. The notch filter used is based on the Parks-McClellan notch filtering algorithm implemented as a function in Matlab's Signal Processing Toolbox. The resulting data were then band-pass filtered between 0.01 and 40 Hz to remove DC offset (on the low end) and possible other sources of noise (on the high end) that might reduce the signal-to-noise ratio in the dataset. An IIR Butterworth band-pass filter was used for this pre-processing step. Filtering data on the frequency domain can often create artefacts and imprecisions in the time domain. However, the very conservative range of the band pass thresholds used in the pre-processing chain of these experiments should not affect any components of interest to our study, usually characterised by frequencies of 10 Hz or lower. Once filtering was completed, the dataset was briefly visually analysed with a particular focus on electrode channels that might still present sources of noise that filtering was not able to remove. It only happened twice that an electrode showed abnormal levels of noise, possibly due to poor connection or the electrode falling out of place during data collection. In this rare cases, data was spherically interpolated at the electrode site. Before referencing, Independent Component Analysis (ICA) was run. ICA was used primarily to remove blinks only, in order to avoid deteriorating the data by removing components that might have been informative. In most cases, artefact rejection was enough to eliminate trials contaminated with blinks.

3.5.3 Re-referencing

Electrodes from BioSemi are referenced online to the Common Mode Sensor (CMS) electrode during data collection. THE CMS electrode injects a small amount of noise in the data, supposedly reducing the amount of external noise picked up by the electrodes. When data is imported into EEGLAB it needs to be re-referenced to one or multiple electrodes among the scalp or face electrodes. When the data was imported, all channels were temporarily referenced to the average of temporal electrodes T7 and T8. However, after the data had been cleaned with filters and channel interpolation, average channel reference (AVE) was applied, reverting to the previous T7-T8 reference. Average reference of all (scalp) electrodes is considered among the best possible references for reduction of noise and for topographical analysis of the data. However, data was not referenced to the average of all electrodes on import simply because some of the channels could have been particularly noisy, for some subjects. For this reason, AVE re-referencing following pre-processing data-cleaning steps guarantees a much cleaner dataset.

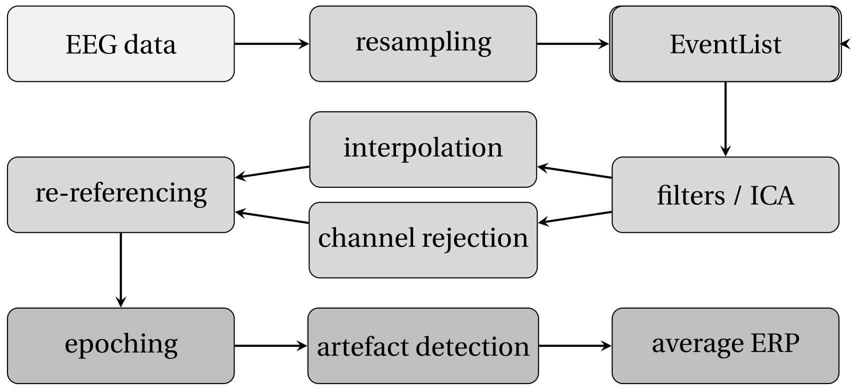
3.5.4 Epochs

Once the data had been filtered and bad channels had been removed, the data was divided into epochs starting 200 ms before the presentation of a stimulus, with a length of 1000 ms. Epoching the data reduces the size of the dataset, by removing any unnecessary recorded signal. However, it is only done at a later step since processes like filtering (or ICA) work better on continuous data. Each epoch is baseline corrected to the period preceding the presentation of the stimulus (-200 ms to 0). Artefact detection was performed with a moving average window across all 64 scalp electrodes. The threshold was set to 50 micro Volts. Finally, average ERP components were calculated for all experimental conditions once noisy epochs had been removed.

3.6 Statistical analysis

ERP data tend to be very noisy and unpredictable, even with a well-controlled methodology and properly set-up equipment. Noise in EEG data has different sources: **1)** sources external to the brain, such as eye blinks, electrical skin conduction and room electrical interference (or AC

Figure 3.3: Pre-processing protocol from continuous EEG data to average ERPs.



hum) are among the biggest causes of noise and artefacts. On top of that, **2)** having fewer trials or participants than required, investigating effects that are lesser known or harder to elicit can definitely contribute to the quality of the signal that, once processed and cleaned, is used for statistical analyses. A methodological paper by Luck and Gaspelin (2017) shows that, by dividing their ERP dataset into two groups (or conditions) by simply splitting the data points at random between either the first or second condition, researchers were able to find significant differences, using otherwise standard significance testing and data modelling techniques, between the two groups. This was certainly due to the amount of noise and variability in ERP data. These significant differences were not justified as the data was split randomly across the two conditions and there was no real trend to account for the differences.

The different analyses carried out on the datasets of experiments one to three can be split in two main groups: channel-level analyses and mean amplitude modelling. In mean amplitude modelling analyses, cubic-spline interpolation scalp maps were generated in `R` (R Core Team, 2020) displaying means amplitude measurements collected throughout several time windows with the `akima` (Akima et al., 2016) package. After describing visual trends in the data, differences between conditions, scalp region and hemispheres for mean amplitude values measured across specific time windows — where trends were previously discovered with a more exploratory analysis using the multivariate adaptive factor adjustment method — were tested using linear mixed effect regression models (LMM) in R with the `lme4` (Bates et al., 2007) and `lmerTest` (Kuznetsova et al., 2017) packages. Together with the main effect of experimental condition and its interaction with region and hemisphere, intercepts were allowed for experimental subjects in the random structure of the model[1]. Whenever the interaction of exper-

[1] Because of the nature of the data and the low amount of subject groups, LMM resulted in *singular fits*

imental condition and region or experimental condition and hemisphere were reported as significant, post-hoc pairwise-wise comparisons and contrast tables were produced using the `emmeans` package in `R`. While LMM allow to test whether average amplitude values in a specific time range are significantly different across different conditions and scalp regions, the AFA method (Sheu et al., 2016), used in this work as a mean of exploratory analysis of the data, is generally used for channel-level analysis to determine the extent of the differences at each of the electrode sites at any given point in time throughout a grand-average epoch.

For the statistical testing of significant differences between experimental conditions at a channel level, an adaptive factor adjustment (AFA) procedure (Sheu et al. 2016) was employed throughout all three experiments as an exploratory method to inform a more accurate use of LMM models for the analysis of mean amplitude over time, overall a superior measure of ERP activity compared to peak amplitude. The statistical method is suggested specifically for the analysis of ERP data, characterised by a high degree of complex dependence pattern over time (Sheu et al. 2016). Because ERP components are often weak and rare, mass univariate analysis of the data is often challenging (Groppe et al. 2011; Sheu et al. 2016). Furthermore, their short duration and inter-subject variability decreases the overall signal-to-noise ratio. ERP data are often characterised by an elevated number of measurements, especially when high sampling rates are used. When testing over a large number of measurements, it is extremely important to correct for false positives (Woolrich et al., 2009), while maintaining "*reasonable power for correct detection*" (Sheu et al., 2016). The adaptive factor adjustment procedure was compared to other common multiple testing methods often employed in ERP data analysis, such as the BY (Benjamini and Yekutieli, 2001), SVA (Leek and Storey, 2008) and LEAPP (Sun et al., 2012) procedures. The AFA method was found to perform better at false-discovery rate (FDR) compared to all other aforementioned methods (Sheu et al., 2016), while maintaining enough power to correctly identify small, significant differences. The AFA method overcomes the challenge of reliable and effective statistics of ERP components with a joint modelling of signal and noise processes, including input where signal is completely absent, such as the baseline section of the data (Sheu et al., 2016).

The combination of mean amplitude modelling and channel-level exploratory analyses

in few occasion. This is known to happen when the variance of a factor, usually a random factor in package `lme4`, is very close to zero. While this can be problematic, it often is expected in cases where mean values are measured within subjects. For this reason, random intercepts for experimental subjects were not excluded from the model in those particular instances, considering only one intercept was allowed in the first place and that all models always reached convergence.

provides a much clearer, complete picture of the observed effects and trends in the data.. The decision to combine two types of analyses and, moreover, to take on a more descriptive approach when it comes to contextualising the findings of all experiments and to discuss all general trends is primarily a consequence of the extremely exploratory nature of experiments one to three. In particular, the paradigms used in this research project have not been used before to investigate sound and speech-sound perception specifically. For this reason, expectations on which components would be elicited, the size of the difference and topographical distribution is not as clear cut as in many other existing studies.

3.7 Reproducibility

The `R` output (*knit* to `html` files) which includes all of the code used to process, plot, interpolate and model the ERP data of all three experiments, as well as *.csv* files of all processed ERP data are available to download on GitHub at the following repository: `mcanzi/phd_codedata`

The `R` code written for the processing, visualisation and modelling of all findings is very comparable across all three experiments. The code includes multiple custom functions specifically devised for the analysis, visualisation and modelling of ERP data in long-format. The aim is to be able to provide readers access to these functions by releasing a free `R` package to the community soon after completion of the doctoral cycle. The thesis document, including source code, bibliography files and all high-resolution figures is also available on GitHub at **`mcanzi/phd_thesis`** for download.

Note: LMM model summaries, together with *type III* ANOVA tables (from `lmerTest`) have not been included in the main text of this work. The same is true for `emmeans` least square means and contrast tables (Lenth et al., 2019), which would have amounted to pages and pages of text. All summaries are readily available through the GitHub link provided above. For the reader's convenience, the code output, model summaries and contrast tables of all models of each experiment have been *knit* to one `html` output file, named `model-summaries.html` and located in the experiment-specific sections of the aforementioned GitHub repository.

Chapter 4

Experiment 1

4.1 Introduction

The aim of experiment one is to test whether the Phonological Mismatch Negativity (PMN) component can be elicited in a passive listening experiment which introduces a context where phonological mismatch is accounted for but no lexical or syntactic mapping is present. This experiment establishes an important baseline for my research. This is done by exploring speech-like perception and passive listening, where the speech signal has been stripped of any lexical and syntactic information. In particular, the aim of experiment one is to investigate how the brain reacts when phonetic and phonological expectations are not met in language perception and processing, with one caveat. In this particular paradigm, expectations are neither created by semantic nor syntactic elements of language, but they are delivered through stimulus presentation order and rate. There have been quite a few experiments that employ ERP in the study of speech perception that have tried to determine which kind of mismatch elicits which ERP component in language perception and processing. The PMN component has, for instance, been linked to phonological mismatch in speech perception and spoken-word recognition. Other components, such as the N400 and P600[1], have been linked to semantic and syntactic mismatch and mapping respectively (Kutas and Hillyard 1980; Connolly and Phillips 1994; Gunter et al. 1997; Hagoort et al. 1999).

[1]Although the P600 component has been linked and correlated with syntactic mapping and processing in language comprehension, it takes on a much wider role of response to rule violations in contexts that are external to that of language and speech perception.

However, as addressed in *Chapter 2*, often times theories regarding the role of specific ERP components, originally linked to one particular event or type of processing, can often be updated and completely turned around when new evidence is introduced (Hagoort et al. 1999). Confounding factors in experimental paradigms, non-reproducible methodologies and overall inconsistent results across studies over the decades keep the role and the existence of the PMN in a very fragile state. While the component has been linked to phonological mismatch and mapping, there is still quite a lot that needs to be researched in order to draw a clearer picture of how this specific ERP component behaves across different experimental environments. The reason why it is not uncommon for the role of ERP components to be updated so frequently is mainly connected to how young the overall field is. The more we learn about ERP and ERP research, the better researchers become at finding out new aspects of their function of ERP components.

Another example is the MMN component which, on the other hand, was extensively and consistently linked with the perception of deviant sounds and stimuli in a sequence (e.g. Näätänen 1991; Näätänen et al. 1993; Näätänen and Alho 1995). However, new evidence was more recently introduced that linked the component to other types of processes in speech and language perception, namely more abstract processes such as phonological, semantic and syntactic processing (Shtyrov et al. 2003; Pulvermüller 2001; Weber et al. 2004). These two examples, the PMN and MMN, exemplify that new theories are often introduced by revisiting older research or by carrying out new experiments to increase the sum-total knowledge that we have in regards to specific ERP components or particular cognitive processes.

4.1.1 Background

The way research in linguistics and cognitive sciences has been linking ERP effects with specific types of reactions to mismatch / processing in the brain is, often times, by trying to isolate specific layers of language processing and perception. Taking a look at one of the few studies carried out on the differences between the PMN and N400 (Connolly and Phillips, 1994), thought at some point in time to be representative of the same ERP component (Diaz and Swaab, 2007), we find a very well known common experimental paradigm for the realisation of these types of studies. Several sentences are presented to participants making sure that the final word in the sentence is, from the context, highly predictable[2] (e.g. the piano is out of *tune*). Sometimes, however, a different word is presented in sentence-final position and, it has

[2] This is often referred to as *cloze probability* (Connolly and Phillips, 1994).

been found that **1)** if the word shares the first phoneme or syllable with the predictable word in that context but the words do not semantically match, e.g. *tuna*, a greater N400 component is observed but no PMN (Connolly and Phillips 1994; Diaz and Swaab 2007). However, if **2)** the word is semantically acceptable in the context but it does not share any phonological properties with the expected word, the opposite scenario is found (Connolly and Phillips 1994; Diaz and Swaab 2007). When the word is **3)** semantically unrelated and it also does not share any phonological properties, e.g. the piano is out of *pizza*, both effects are found. It has also been observed that in the case where both semantic and phonological mismatch occur, the N400 component is observed later than in conditions **1)** and **2)**. For this reason, it has been theorised that while the N400 is a response found to be greater when semantic mismatch is present (Kutas and Hillyard, 1984), the PMN, originally named phonological mismatch negativity because of this particular scenario, must be connected with phonological mismatch and processing in speech perception (Connolly and Phillips 1994; Diaz and Swaab 2007).

Connolly and Phillips (1994) proposed — by observing the behaviour of the PMN and N400 components across all conditions — that the PMN component appears to be sensitive to phonological mismatch specifically while the N400 component is elicited through mismatch and mapping of semantic information. In other terms, Connolly and Phillips (1994) proposed that the two ERPs were dissociated from one another. They also linked the two components to two separate levels of processing in spoken-word recognition and speech perception. However, as discussed in more detail in *Chapter 2*, what Connolly and Phillips (1994) failed to overtly observe is how the PMN effect was strongest when both semantic and phonological mismatch were present. This stands to indicate that, although the PMN appears to mostly be sensitive to phonological mapping, it is also somehow mediated by the presence of semantic mismatch. In future studies, Connolly et al. (1995; 2001) followed up on previous findings to build on the foundation that they had laid down regarding the nature of the PMN and its relationship with phonological mapping and mismatch.

The reason why we seem to agree that the PMN reflects phonological processing is simple. Research in the past uncovered ERP components that are (also) linked to semantic and syntactic mismatch, such as the N400 and P600 (Kutas and Hillyard 1984; Gunter et al. 1997) following 400 ms post stimulus onset. We also observed earlier ERP (pre-200 ms) that seem to be connected with stimulus and sound processing, mismatch, evaluation and perception at a lower level, such as auditory MMN and N1 (e.g. Nyman et al. 1990). The PMN - whose latency has been found to be in between components like the MMN and N400 (Connolly and Phillips, 1994) - has been mostly observed when the first phoneme of a highly probable word ended up being different from the one participants expected. Because of this, it seems intuitive to think that the PMN reflects some sort of higher-level mismatch in phonetic and phonological processing (i.e. phonological mismatch driven by semantic and possibly syntactic expectations).

Since this first definition, the hypothesised role of the PMN, just recently undergoing relabelling and often being referred to as the Phonological Mapping Negativity because of its sensitivity not limited to mismatch, has been updated multiple times. At the same time, ERP components are rarely only connected to one type of mismatch or event in the brain. It is becoming more arguable at this point whether we even have "language specific" ERP components and whether it even makes any sense to talk about a "phonological" or "semantic" component when discussing, for instance, the N400, which has been observed across experiments using visual (Proverbio and Riva, 2009), auditory/linguistic (Kutas and Hillyard 1980; Kutas and Hillyard 1984), and even olfactory stimuli (Invitto et al., 2018). In addition, not only are single ERP components found to be linked to multiple levels of processing and mismatch, but they have been found to interact in the same context (e.g. the N400 and P3 components have been found interact in the perception of semantically unexpected and physically deviant words by Arbel et al. 2011). In addition, responses in the brain usually cause what are referred to as *complexes*, which is a combination of multiple components being elicited as a group due to the perception of a specific stimulus. Across the three experiments of this work, complexes like the N1-P2 complex are often mentioned in relation to the perception of speech sounds and mismatching stimuli.

4.1.2 How to look for the PMN

In electro-physiological research, isolating a single process or a single ERP component to determine its nature, even when we design an extremely well-controlled experiment, is not a trivial task. Stripping language of all the layers that could play a confounding role in language processing, such as semantic processing, syntactic processing, pragmatics, etc., while at the same time investigating phonological mismatch requires complex methodologies. Unfortunately, complex methodologies often do not work well when paired with ERP data collection methods. These methods often require fairly simple designs. On top of that, even when stripping language down to only speech sounds, for instance, it is still almost impossible to determine whether the response of an ERP component to phonological mismatch is connected to auditory mismatch or to mismatch of phonological representations in a context where auditory stimuli are present.

It would be unreasonable, given the evidence, to outright deny that the PMN has a connection with some level of phonological mismatch and mapping. However, ERP components are often linked to multiple aspects of speech perception and processing. For this reason, it is particularly relevant to the study of language processing and comprehension to be able to determine exactly what the PMN component represents. Whether it mirrors a lower-level sound mismatch, similarly to earlier components or a higher-level, semantically informed, phonological mismatch effect has yet to be determined. Experiments like that of Kujala et al. (2004) or Newman et al. (2003), while removing some of the possible confounding factors (such as syntactic activation), introduce new biases such as, for instance, using a mix of audio-visual methods (Kujala et al., 2004). Experiment one of this thesis builds on the structure of experiments like Newman et al. (2003) and others (see Newman and Connolly 2009; Astheimer and Sanders 2011) to study whether the PMN component — and other early auditory perception-related components — can be observed during speech-like perception (i.e. semantic- and syntactic-free speech signal).

In experiment one, I aimed to remove semantic and syntactic expectations in speech perception by creating a streamlined speech signal resembling a miniaturised *artificial language*. Participants were trained to recognise and remember two pairs of trisyllabic nonce words. The nonce-word pairs were structured so that the second item of each pair was always going to follow the first one 100% of the times. This created the expectation that, once having heard item A, item B would definitely follow. In the main experimental block, item B of each pair was mismatching 33% of the trials, so that the first syllable of the nonce word would not match the stimulus previously learnt by participants, failing to meet the subject's expectations on what word they were going to hear next. These expectations, however, differently from what usually happens with real languages, were not created by semantic information or sentence structure but they were externally manipulated.

4.1.3 Artificial languages

In *Section 4.1.2*, I mentioned how experiment one of this research project makes use of an artificial language (AL) for the investigation of the PMN. More specifically, in this particular case we are dealing with an extremely miniaturised semantics-free and syntactic-free version of an artificial language. An AL is explained as a series of word items that, just like in natural languages, can be used in different combinations and, especially in terms of transitional probabilities, have a certain chance of appearing, in a given a context, next in combination with one of the other lemmas in the same language. In experiment one, the simplest possible artificial language was created. Four trisyllabic CVCVCV *words* only were available as vocabulary, to which no meaning was assigned. Furthermore, these four items were strictly paired in only two possible combinations. Participants were trained to learn the four words and the combinations in which they were presented in a process of computerised learning.

The main reason behind the implementation of a miniature artificial language rather than an existing — unknown to all participants — natural language was that in an artificial language, phoneme transitional probabilities and syllable transitional probabilities can be systematically controlled from the outside, together with the ability to hand-pick the exact phonemes that make up words in the language. However, there is one main argument against using artificial languages instead of a natural language in experiments dealing with the investigation of language-specific ERP components such as the PMN. The argument revolves around the somewhat unknown extent of similarity between processing natural languages and artificial languages in the brain. Existing research on the processing of artificial languages with ERP has shown how, following participants' training in the learning of a miniaturised artificial language, similar ERP responses were found when processing artificial languages compared to processing a participant's native natural language (Friederici et al., 2002). In particular, Friederici et al (2002), demonstrated how, training in learning an artificial language, made all the difference between eliciting language-related ERP responses (with training) and not eliciting them in the absence of specific language training. The aim of Friederici et al.'s (2002) paper was, ulteriorly, to demonstrate that languages learnt outside of the critical period would elicit responses during comprehension that mirrored those found when processing a participant's native language.

Another study by Tabullo et al. (2013) investigated ERP responses to mismatch in semantics-free artificial grammars taught to participants and to mismatch in the participant's native language (Spanish). They discovered that, although not all ERP components were shared between the two conditions, artificial grammar sequences elicited N400 and P600 components among the subjects' responses. The elicitation of a P600 or posterior late positivity response in the mismatch condition would suggest that, syntactic sequences in a known language and sequences in an (artificial) languages exhibit similar cognitive mapping and processing by the brain. More specifically, the authors suggest that both N400 and P600 can be considered responses to unfulfilled stimulus expectations in a sequence (Tabullo et al., 2013).

In existing literature in linguistics and cognitive sciences, artificial languages have been preferred over the use of natural languages in a number of occasions. Primarily, the use of artificial languages have been employed to study infants' and adults' language acquisition learning procedures (e.g. Braine et al. 1990; Gómez and Gerken 2000; Folia et al. 2010) and to test hypotheses such as the critical period (Friederici et al., 2002). Other examples include the use of artificial language learning for clinical investigations into semantic and syntactic abilities of aphasic patients (Glass et al., 1973), feature-based generalisations (Finley and Badecker, 2009) and, among many other, studies on language evolution (e.g. Christiansen 2001).

The artificial language employed in experiment one was, from a complexity classification stand-point, extremely simple. No lexical mapping was introduced for any of the four words in the AL. Furthermore, during passive listening tasks, it is common to present participants with a mute film in order to distract them from auditory stimulus presentation. It was thought to be best not to implement this feature in order to avoid participants possibly semantically connecting specific words to images on the screen. No syntactic relationships among the items were introduced either, but it was made clear that the four words were strictly grouped into two pairs, with transitional probabilities between the two items of each pair being 1. The reason why only four words — and two pairs — were created for the experiment was that four seems to be the maximum amount of stimuli participants can comfortably hold in short memory (Cowan, 2001). Considering the training phase was done directly prior to the presentation of the stimuli and that making sure participants knew the stimuli perfectly was such an important feature of the study, only few stimuli were used to maximise learning outcome.

4.2 Methods

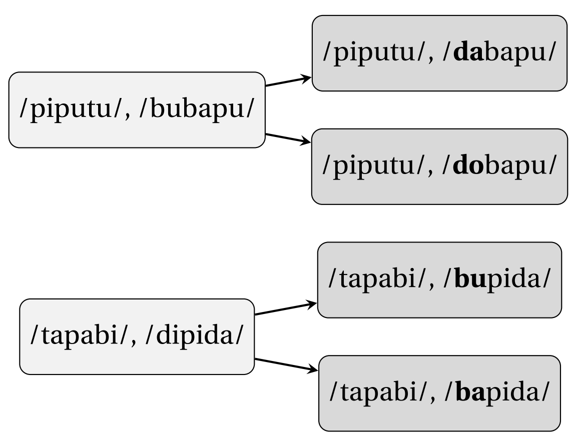
4.2.1 Stimuli

Six pairs of trisyllabic (CVCVCV) nonce-words (e.g. *tapabi*) were synthesised in Praat and used as the set of stimuli for experiment one. Two pairs of nonce words are henceforth called *standard pairs*. These two nonce-word pairs were presented to participants in the training session, as highlighted in the experimental protocol below. Subjects were instructed to memorise and recognise the four nonce words forming the two standard pairs. The other four nonce-words pairs are referred to as *manipulated pairs* from this point onwards. While almost exactly the same as the two standard pairs, they differed only by the first syllable of the second nonce word of each pair. Two manipulated versions of each standard pair were created. Each nonce word measured 700 ms in length and each pair would, in total, measure 1900 ms, including a 500 ms pause between the two items of the pair. Pitch contour, vowel length, syllable and word duration were controlled among standard and manipulated pairs too, in order to reduce the number of confounding differences between the stimuli.

Each CV syllable was synthesised using Mac OSX text-to-speech software and it was later modified in Praat to control for vowel length, syllable length and pitch contour. For consonantal sounds, only plosive consonants /p/, /b/, /d/, /t/ and /k/ were used. As for the vowel sounds, only /a/, /i/, /o/ and /u/ were employed. Standard nonce words and pairs were created so that phonemic transitional probability was controlled for. The stimuli used were previously planned and employed by Astheimer and Sanders (2011) who checked for stimulus bias and controlled for several other factors such as phonemic transitional probability, resemblance of nonce words to English words, etc (as detailed in *Chapter 3*). The transitional probability between the two items of each pair — for each of the standard pairs — was 1. This meant that the probability of a specific nonce word being in position two, after hearing the first word of each pair, was 100%.

The diagram below presents the nonce words, part of standard and manipulated pairs, employed for the experiment.

Figure 4.1: standard (left) and manipulated (right) nonce-word pairs used as experimental stimuli.



4.2.2 Behavioural pilot

Before running experiment one in its full form, I decided to run a behavioural pilot with 5 participants based on sections two and three of the experiment (i.e. *training* and *testing*) of the experiment. This was done in order to test **1)** whether any of the nonce words used as stimuli where too similar to each other or too similar to existing English words and **2)** whether the computerised training was as effective in practice as it was in theory for participants to recognise and distinguish standard pairs to their manipulated counterpart. All 5 participants were native speakers of British English. They rated the stimuli as *Not at all similar to existing English speaking words*. No stimulus was rated as standing out from the others for any particular reason and all participants completed the testing phase of the experiment on the first try after a short (2 - 5 minutes) training phase. Because of these reassuring results, the stimuli used for the behavioural pilot were kept unchanged for the experiment. No participant that took part to the behavioural pilot also took part to the final experiment. No compensation was given for the participation to the pilot.

4.2.3 Participants

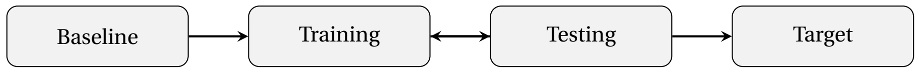
Twenty two (F = 12) undergraduate students at The University of Manchester participated in this experiment. All participants were native speakers of English, right-handed adults and they all reported normal hearing and normal or corrected-to-normal vision. No participants disclosed use of psycho-active medications or the presence of any known neurological conditions. Written consent was taken from each participant before the beginning of the experiment and compensation of č20 was given upon completion of the final experimental block. Two participants did not complete the experiment due to technical difficulties during experimentation and the data from two more participants were discarded because of an elevated number of artefacts and noise. In total, data from 18 participants (F = 10) was included in the analysis of the results.

4.2.4 Design and instructions

Experiment one comprised four distinct experimental blocks: *baseline*, *training*, *testing* and *target* blocks. Each block and its function is now presented in detail.

Baseline: During this first phase, EEG data was recorded from participants for an average of 10 minutes, in which both standard and manipulated pairs were presented 66% and 33% of

Figure 4.2: Order of experimental blocks for the procedure of experiment one.



the time respectively. A break between each pair of 1 second was also introduced. The goal of the baseline section was to record participants' reaction to standard and manipulated speech sequences without any prior active learning of the stimuli themselves. No behavioural task was present in the baseline section and no rest blocks were provided. The baseline section, as already mentioned, lasted only 10 minutes on average. While this is not a great deal of time to collect clean ERP data from subjects, extending the collection time in this first phase would have meant keeping participants in the lab for a much longer time.

Training: Following the baseline block, participants were instructed to learn the two standard nonce-word pairs through computerised training. Each pair was presented auditorily to the experimental subjects. Participants were instructed to press <enter> on the keyboard to move on to the next pair of stimuli. Participants were given as much time as they needed to confidently memorise and recognise standard stimulus pairs. The average training phase lasted approximately 5 minutes (2 - 8 minute range).

Testing: Then, participants undertook a computerised behavioural testing block. Both standard and manipulated pairs were presented auditorily at equal rates and subjects had to disambiguate between pairs they had previously learnt in the training phase and those that they had not. Participants were instructed to press <enter> on the keyboard if the stimulus heard was exactly as they had learned in the training session, while <tab> if they thought it had been changed in any way. In order to pass the test, participants had to provide five correct answers in a row. The test was repeated a total of three times for each participant. If a participant failed the test twice in a row, subjects were presented the training block again. All participants passed the testing phase on the first try except one participant, who passed it on the second try after brief retraining. The data from that participant ended up being discarded because of a noisy recording. No EEG data were collected neither in the training nor in the testing blocks of the experiment.

Target: The main data collection block was administered to all participants. Subjects were instructed to listen to the pairs of nonce words they had previously learnt in a passive listening task. However, unbeknownst to them, 33% of the time, one of the manipulated pairs would be presented instead (mismatch condition). During this phase, EEG data were recorded from 64 flat-type active electrodes on the scalp and 30 second rest blocks were provided every three minutes of trials. In particular, ERP responses were measured whenever the second nonce word of each pair would be presented, as the first word of each pair, regardless of the fact that it was a standard or manipulated stimulus, would not change across conditions. The target block lasted an average of one hour per participant. Participants were offered a choice of warm or cold drink to consume during the breaks. Although presenting mismatching stimuli at a lower rate compared to matching ones could have introduced unwanted responses to stimulus presentation rates (i.e. so called *odball* paradigm), the decision was a compromise to avoid habituation to the presentation of manipulated items had the presentation rate been 0.5.

4.2.5 Expected effects

For the dataset collected in experiment one, two and three, I decided — beforehand — to focus on a subset of electrodes when testing for differences (at a channel level) between conditions. In particular, I ran statistical analyses on 18 electrodes, located in prefrontal, anterior, frontal (AF and F), central (FC and C) and parietal (P) regions of the scalp, where effects of interest to this study, such as N1, MMN, PMN, P3 and N400 are most likely to be located (Kutas and Hillyard 1980; Connolly and Phillips (1994); Kujala et al. 2004; Garrido et al. 2009). Electrode sites were included from multiple regions of the scalp with different distributions. For instance, only three parietal sites were included in the analysis since only some evidence was found that suggested the PMN has a parietal distribution (e.g. Connolly and Phillips 1994). However, most of the scalp sites are distributed in anterior, frontal and central areas of the scalp, where N1, MMN and PMN have most often been observed.

Experiment one is the first exploratory study of this research project. For this particular reason, all effects are expected based on previous literature and the type of experimental design that was chosen. For experiments two and three, which employed a similar design to experiment one, predictions made are also influenced by the results of the current experiment. In particular, N1 and MMN components — linked to unexpected stimulus perception and stimulus mismatch (Näätänen and Picton 1987; Näätänen et al. 1993: Näätänen and Alho 1995) — are characterised by negative peaks between 80-120 and 150-250 ms respectively in frontocentral regions of the scalp (Garrido et al. 2009). The N1, specifically, is linked to unexpected sound perception in the absence of task demands (Näätänen and Picton 1987), which is the case for experiment one, considering participants have been instructed not to actively listen to the stimuli being presented. The MMN, also elicited regardless of participant's focussed attention to stimulus presentation, has also been linked to the perception of deviant stimuli in a series of standard stimuli. This is, to some extent, represented by the inclusion of mismatching phonemes in the manipulated nonce-word pairs.

The PMN has mostly been observed with peaks negatively at around 250-300 ms post-stimulus onset in frontal and frontocentral scalp sites [3], which is where and when a difference between the match (standard pairs) and mismatch (manipulated pairs) conditions is expected should the PMN reflect phonological mismatch (Connolly and Phillips 1994; D'Arcy et al. 2004). The P3 (as P3a) or P300, another component linked to stimulus categorisation and mismatch (Garrido et al., 2009) as suggested by its name, generally peaks positively at around or after 300 (range 250-500) ms post stimulus onset and it is strongest in parietal regions of the scalp. (Garrido et al., 2009).

4.2.6 Hypotheses

As informed by more recent literature on the PMN, which places the component as a goodness-of-fit index between expected word-form information and incoming speech input, regardless of the role of lexical activation, one main hypothesis can be put forward for experiment one.

H1: The PMN component can be observed as a response to phonologically mismatching stimuli in the absence of any form of lexical activation.

4.2.7 Statistical testing

As described in more detail in *Chapter 3*, the statistical analysis of the data was carried out on two levels mainly, namely the mean amplitude and channel level. An adaptive factor adjustment procedure (AFA) put forward by Sheu et al. (2016) was implemented for the channel-level analysis. The procedure works better than other types of multivariate analysis at "discovering interesting ERP features" (Sheu et al., 2016) while correcting for multiple comparisons and time-series data correlation. The AFA method was implemented through a software package in R, called ERP (Causeur et al. 2018). Because of the exploratory focus of the study, particular

[3]While, as discussed in *Chapter 2*, the PMN has also been observed in parietal (Connolly and Phillips, 1994) and other sites rather than frontal and central (e.g. Van Petten et al. 1999), the frontocentral distribution is the one most commonly observed one.

attention is also given to scalp distribution and a descriptive approach to general trends and effects across the scalp.

For mean amplitude modelling, linear mixed effect models (LMM) were fit to mean amplitude data, measured across time windows of interest. Experimental condition, region and hemisphere were added as main effects as well as a three-way interaction. The best model fit was obtained by methods of model comparison. Random intercepts were allowed for experimental subject. Cubic-spline interpolation scalp maps were also created to display mean amplitude measurements. LMM were run based on previous predictions, visual exploration and channel-level analyses.

4.3 Results

In this section, ERP results are presented from both target and baseline experimental sections. ERP grand-average curves have been computed and statistical results are described. *Mismatch minus match* difference curves are also presented to highlight differences between the two experimental conditions. Topographical maps and mixed models have been devised to focus on component origin analysis and topographical distribution. The discussion section of the chapter, at a later stage, contextualises the findings of this experiment with existing theories and past research.

4.3.1 Target block

Figures 4.3 and 4.4 present the grand-average response across all participants to matching and mismatching pairs by electrode site, time-locked to the presentation of the second nonce word of each pair. As described in the methodology section of the experiment, the first item of every pair, whether matching or mismatching, remains the same across conditions. Only the first syllable of the second nonce word was swapped in the mismatch version of the stimuli. ERP responses have been captured starting 200 ms before the presentation of the target syllable, until 800 ms post syllable onset. However, because the focus of experiment one is on components usually observed earlier than 400 ms post stimulus onset, the figures included in this chapter often show data between -100 and 500 ms post stimulus onset. This allows us to visualise more details and provide a clearer picture of the patterns highlighted by the data. Furthermore, difference curves (mismatch condition minus match condition) have also been included to help visualise direction of differences between the curves.

Responses are highlighted for channel locations in the anterior-frontal (AF), frontal (F), frontocentral (FC), central (C) and parietal (P) regions of the scalp.

Figure 4.3: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 500 ms PSO for prefrontal, anterior-frontal and frontal sites in the experimental *target* block.

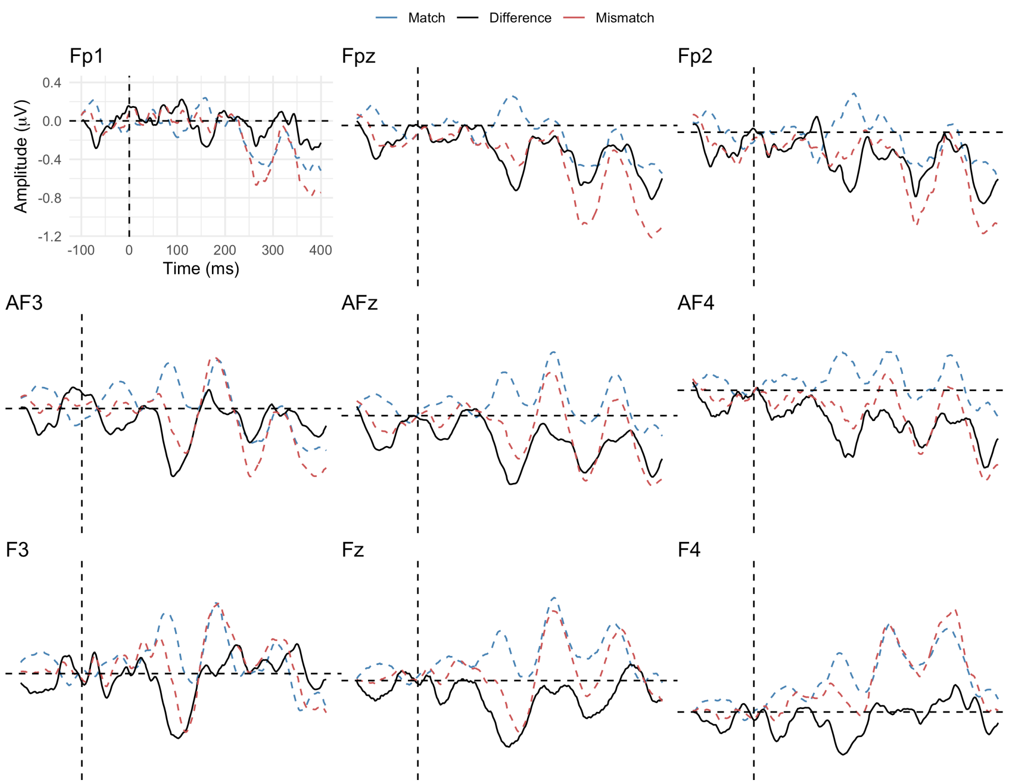


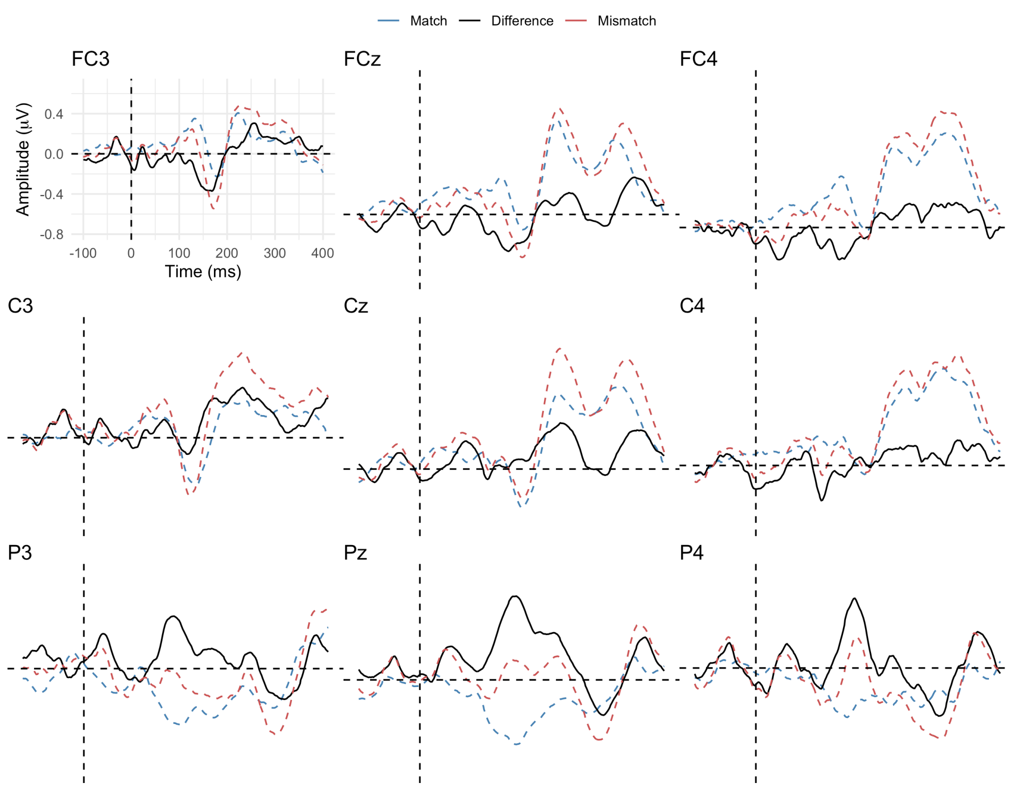
Figure 4.3 highlights responses in anterior and frontal (FP, AF, F) scalp sites[4]. A negative trend is quite visible in the anterior and frontal region of the scalp, strongest across the midline, between 150 and 200 ms and maximal at around 160 ms at AFz and Fpz for the mismatch condition. This negative-going deflection is most likely linked to the elicitation of the MMN component. It is often observed in this area of the scalp, often found in the 150 to 250 ms range post stimulus onset (Näätänen et al., 1993).

Following the elicitation of a negative trend between 150 and 200 ms, there seem to be multiple negative-going polarities, respectively in the 250-300 ms range and the 350-400 ms range. The

[4]Right away, it can be observed that the data could be described as noisy when looking at the pre-stimulus interval. This, however, is not caused by particular noise in the data, but it is a by-product of small effects due to the passive listening nature of the task. What this means is that, while signal-to-noise ratio is not particularly high, it is most likely not due to noise being particularly high, but rather to signal being particularly low. This is one of the limitations of passive listening experiments.

difference between the curves is maximal across prefrontal and anterior-frontal scalp sites. Later in this section, it is shown how what appears to be a pair of negative-going effects for the mismatch condition in frontal areas is in fact a by-product of central positive components between 250 and 450 ms. These trends are most likely correlated with the elicitation of a component referred here as P2 (because of its latency) and the P3(a) component to the presentation of novelty, oddball stimuli.

Figure 4.4: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites in the experimental *target* block.

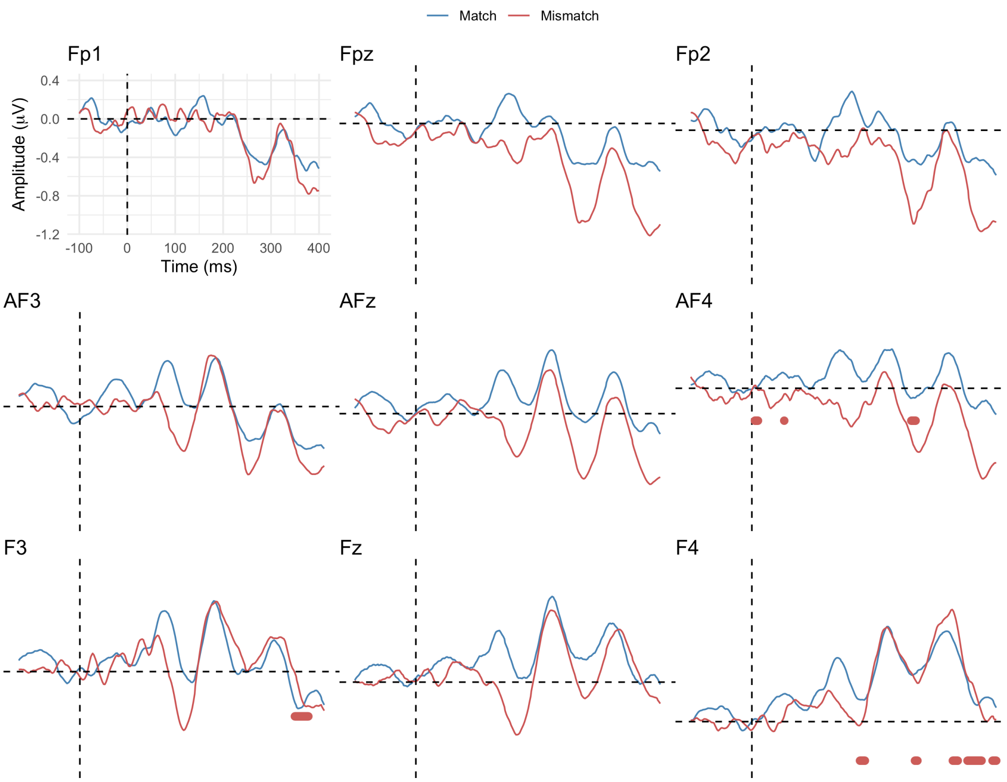


In figure 4.4, moving from the front of the scalp towards the back of the scalp, the difference between the two curves between the 100 and 200 ms marks is reduced in amplitude until polarity is reversed. A similar trend can be observed for the two following deflections that now present a positive trend for the mismatch condition, maximal across the midline at frontocentral and central electrode sites. Many of these deflections in the difference curve seem to share amplitude levels comparable to pre-stimulus interval noise. This would suggest that their status is probably not statistically significant. On the other hand, their propagation across the scalp helps visualise general average trends that seem to be shared across most participants. With most single-subject curves being enclosed between the -3 and +3 μV and grand average curves of all participants (figures reported in the chapters) showing effects often between -0.5 and 0.5 μV on the scale, statistical power is not at its highest.

The adaptive factor adjustment method (Sheu et al., 2016) was applied to the ERP curves from the mismatch and match conditions in order to investigate which of these trends observed in figures 4.3 and 4.4 is statistically significant at a channel level. The AFA method is a conservative method of ERP and time-series data multivariate analysis that promises to capture nuances and differences that are small while, at the same time, providing a conservative approach when it comes to noisy data (Sheu et al., 2016). The unpredictable and generally noisy nature of ERP data is naturally bound to false positives. In order to balance that out, the AFA method was tested on 18 electrodes that were hand-picked before the data was even collected. These electrode sites would show — based on previous literature — the strongest effects for the components that are the focus of this work. The same 18 electrodes are analysed throughout all three experiments. By testing a limited number of electrodes the number of false positive is limited to some extent.

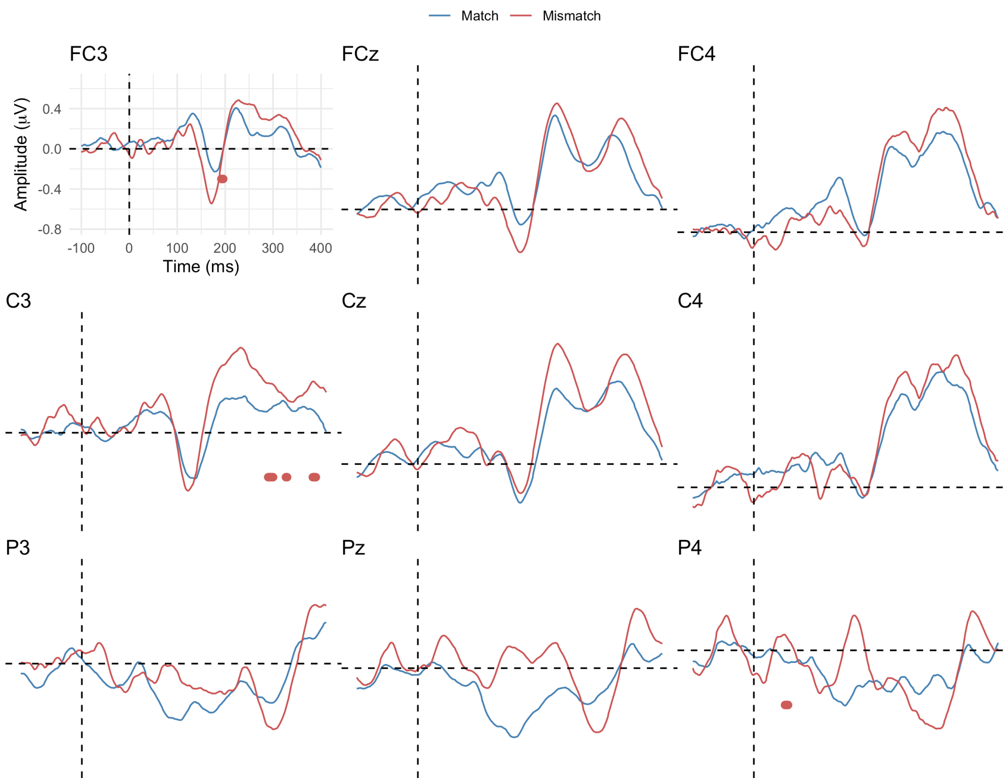
Figures 4.5 and 4.6 present the ERP curves for the mismatch and match conditions. The red segments signal to the reader where the differences between the two conditions are statistically significant (on the x axis) with $p < 0.05$ at each of the scalp site.

Figure 4.5: Grand-average ERP curves for match / mismatch experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites in the experimental *target* block.



Most of the statistically significant differences at a channel level, found between match and mismatch condition in prefrontal, anterior and frontal scalp sites can be observed in the right hemisphere. Three time windows of significance are mainly displayed in the findings: the first one is found before and around 100 ms post stimulus onset (Fp2). The second peaks negatively for the mismatch condition before 200 ms (AF4 and F4). The third window is a positive trend for the mismatch condition between 300 and 400 ms post stimulus onset (F4). This distinction is clearest in the right frontal scalp site F4, where small but significant differences are found surrounding the 200 ms mark and, later, in the P3 range. Figure 4.5 is a good example of the conservative nature of the AFA procedure, highlighting only few differences across 9 scalp sites, although some other trends are visible that have not been highlighted in other regions. Statistical differences presented in the data often show minimal size difference, in the order of one tenth of a μV.

Figure 4.6: Grand-average ERP curves for match / mismatch experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for frontocentral, central and parietal sites in the experimental *target* block.



Moving on to the results of the AFA procedure on frontocentral, central and parietal scalp sites, most of the effects are visible in frontocentral and central scalp sites only. On the other hand, minimal effects can be observed in the parietal region of the scalp and only some small differences are elicited in central sections. When it comes to frontocentral scalp sites, mainly two effects are visible. The first effect, maximal in left frontocentral electrode FC3, is a significant negativity for the mismatch condition between 150 and 200 ms, with a peak at 170 ms post stimulus onset. Other significant effects, across midline and right frontocentral scalp sites, as well as left central scalp site C3, display differences in the 250-400 ms range, characterised by an overall positive deflection, maximal over FC ad C electrode sites, likely connected to novelty P3 and P3a-b related positivity. The topographical analysis, presented in the next section, highlights the distribution of the aforementioned components more clearly.

The focus of the analysis is primarily on the first 350 ms starting from the presentation of the stimulus. All effects linked to phonological, phonemic and auditory mismatch are usually elicited before the observation of the N400 component. In experiment one, no N400 was observed. In theory, no N400 was expected among the findings of this particular experimental paradigm. This is considering that neither lexical nor syntactic processing or mapping were introduced. However, the topographical analysis will reveal a later positivity in central and parietal sites that is usually linked with the elicitation of the P600 component, often connected to syntactic violation but generally related to rule violation in speech, music and general sequences. Besides differences linked to the elicitation of the MMN component and those correlated to the observation of P3 and P3-related effects, no negativity between 250-300 ms could be observed in relation to the elicitation of the PMN component. Looking specifically at the scalp sites that have been often connected with the elicitation of the PMN, both with grand-average curves and subtracted curves, no clear-cut PMN can be observed.

4.3.2 Modelling

Besides being one of the most consistent reference methods available for the study of ERP at a channel level, the offline average reference (AVE) is very well suited for topographical analyses of ERP components. For this reason, I have included cubic-spline interpolation scalp maps showing the mean amplitude measurements (in both conditions) measured across time windows where significant differences were earlier captured using the AFA procedure or where general trends were visually identified. Furthermore, I have tested differences in mean amplitude measurements for the selected time windows using linear mixed effect model regressions (LMM) with the `lme4` (Bates et al., 2007) and `lmerTest` (Kuznetsova et al., 2017) package in `R` (R Core Team, 2020). I have explored pairwise contrasts of condition across scalp region, when necessary, using the emmeans package (Lenth et al., 2019). Experimental condition, scalp region and hemisphere were added as fixed factors in each model, as well as their three-way interaction. Intercepts for experimental subjects were allowed in the random structure of the model. *Type III* ANOVA tables were generated with approximated degrees of freedom using the Kenward-Roger method (Kuznetsova et al., 2017). Figure 4.7 displays average amplitude values calculated between 150 and 200 ms post stimulus onset for both mismatch and match experimental conditions.

Figure 4.7: Grand-Average mean amplitude intervals (between 150 and 200 ms PSO) for mismatch / match experimental conditions in the experimental *target* block.

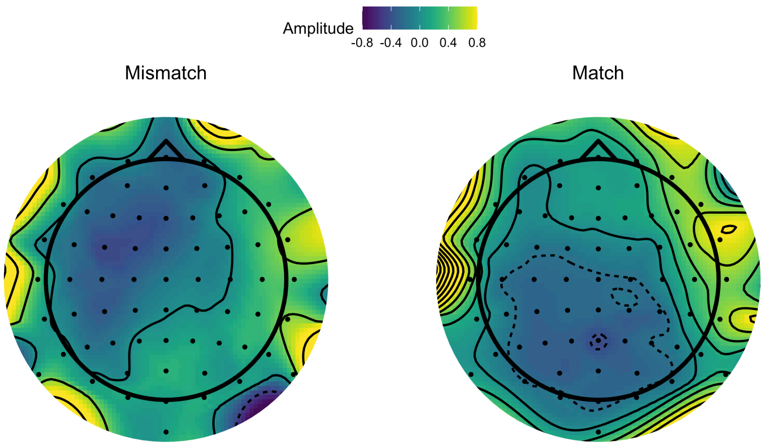


Figure 4.8: Grand-Average mean amplitude intervals (between 300 and 350 ms PSO) for mismatch / match experimental conditions in the experimental *target* block.

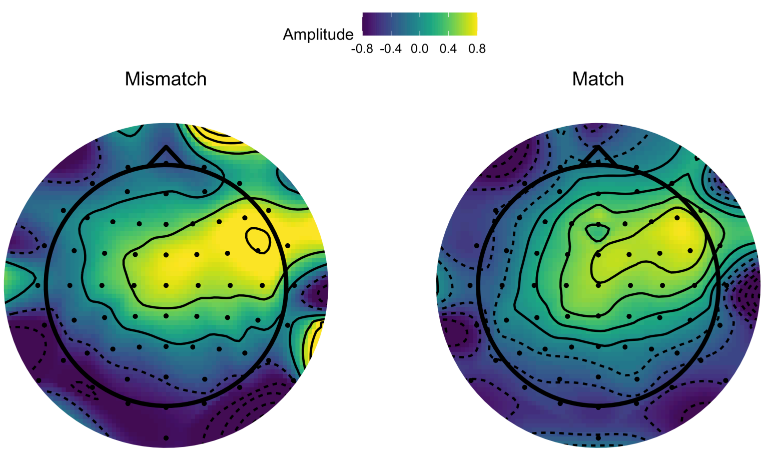


Figure 4.9: Grand-Average mean amplitude intervals (between 250 and 300 ms PSO) for mismatch / match experimental conditions in the experimental *target* block.

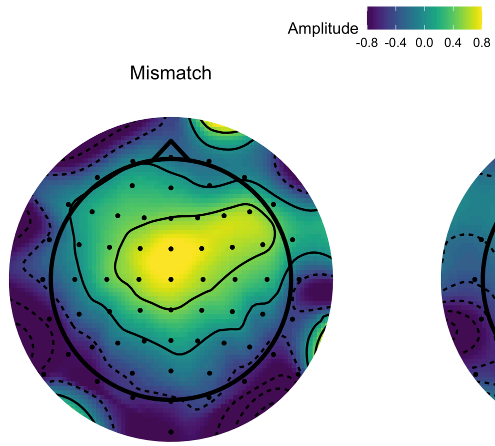
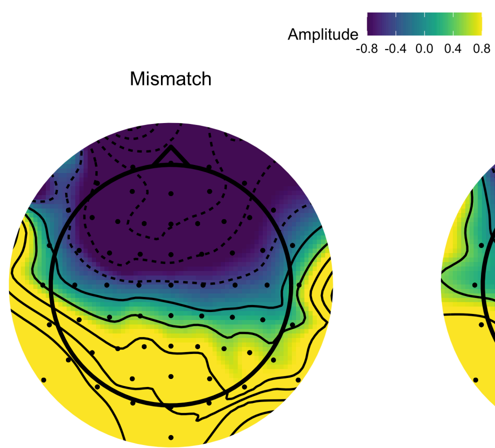


Figure 4.10: Grand-Average mean amplitude intervals (between 575 and 625 ms PSO) for mismatch / match experimental conditions in the experimental *target* block.



A negative deflection is highlighted for the mismatch condition over frontocentral and frontal scalp sites, mostly distributed over the midline with a slight left hemisphere prevalence. This difference was previously captured by the AFA method and a significant difference was displayed between the two conditions at the left frontocentral scalp site. The MMN, which originates from the anterior frontal gyrus (Näätänen et al., 1993), is the most likely candidate for the effect observed in the 150 to 200 ms time window. The LMM fitted to mean amplitude data summarised between 150 and 200 ms reported a significant main effect of scalp region

[$F_{(10,1797)}$ = 6.723, $p < .0001$] and hemisphere [$F_{(1,1797)}$ = 44.384, $p < .0001$], as well as a significant interaction between experimental condition and scalp region [$F_{(10,1797)}$ = 3.784, $p < .0001$]. Post-hoc pairwise comparisons for experimental condition within scalp regions reported a significant main effect of experimental condition at anterior-frontal [$t_{(1797)}$ = -2.244, Est = -0.27 μV, SE = 0.12 μV, $p < .01$)[5] and frontal [$t_{(1797)}$ = - 2.720, Est = -0.23 μV, SE = 0.08 μV, $p < .001$) scalp sites over both hemispheres.

Following, figure 4.8 focusses on the statistically significant difference, maximal between 300 and 350 ms post stimulus onset, most likely linked to the elicitation of a P3-like component or, in more general terms, positivity linked to stimulus attention and evaluation processes, common of the 300 to 400 time window. Looking back at figures 4.5 and 4.6, while most significant differences were observed in the 300 to 350 ms range, another component or general trend is visible between the 300 ms and the MMN component. This peak is, similarly to the difference highlighted in figure 4.8, negative at the front and positive at the centre. Because the 250 to 300 ms range is a time window critical for both stimulus evaluation and the elicitation of the PMN, mean amplitude modelling can provide evidence towards whether the differences observed are statistically significant as well as provide information on the distribution of the effect.

Both in figures 4.8 and 4.9, a central positivity is observed throughout both time windows (250-300 ms, 300-350 ms). As presented by the findings of the AFA procedure, most of the significant differences between the two curves at a channel level are found for the second positive deflection, rather than for the first. These findings suggest that the deflection found between 300 and 350 ms is most likely linked to the elicitation of an oddball and novelty component, such as the P3. The previous deflection, found to be similarly distributed in terms of region and average amplitude across both conditions is most likely linked to some general and common process of stimulus evaluation and possibly accompanying the elicitation of a preceding N1 component. The original channel analysis of the 250 ms to 350 ms time window does not suggest the origin of a negative trend for the mismatch condition that could be linked to the PMN.

In the 250-300 ms range, a LMM fitted to mean amplitude values reported only significant main effects of region [$F_{(10,1797)}$ = 17.57, $p < .0001$] and hemisphere [$F_{(1,1797)}$ = 13.50, $p < .001$], but no significant main effect of experimental condition or its interaction with scalp region.

[5] A negative estimate suggests that mean amplitude was more negative for the mismatch condition compared to the match condition.

Similarly, a second model fitted to mean amplitude data summarised between 300 and 350 ms post stimulus onset suggested only significant main effects of scalp region [$F(_{10,1797}) = 13.58$, $p < .0001$] and hemisphere [$F(_{1,1797}) = 57.11$, $p < .0001$]. There does not appear to be a significant main effect of condition, nor its significant interaction with region or hemisphere, throughout both time windows in the P3 (and PMN) range.

Finally, although mainly out of the scope of this study, a late posterior positivity can be observed for the mismatch condition when compared to the match experimental condition (Figure 4.10). This positivity, which starts from as early as 500 ms post stimulus onset is maximal at around 600 ms in posterior regions of the scalp. Both the temporal and topographical distribution of the observed component suggests the elicitation of a P600-like response. As detailed in *Chapter 2*, the P600 is often connected with rule violation in syntactic contexts (Hagoort et al., 1993). The component is observed in experiment one despite no syntactic context being present in the experimental design. It could be suggested that, although no syntactic mapping is present, the presentation of mismatching, incongruous stimuli acts as a violation of an existing sequence, created by the alternation of the presentation of standard stimuli. This effect is reinforced, later on in *Chapter 5* and *Chapter 6*, by similar findings within the data of experiment two and three.

The LMM fitted to mean amplitude data between 575 and 625 ms post-stimulus onset reported a significant main effect of scalp region [$F(_{10,1797}) = 77.253$, $p < .0001$], hemisphere [$F(_{1,1797}) = 10.990$, $p < .0001$], as well as a significant interaction between experimental condition and scalp region [$F(_{10,1797}) = 7.217$, $p < .0001$]. Pairwise comparisons of experimental condition within each scalp region, across hemispheres, report a significant main effect of experimental condition with a positive direction for the mismatch condition in parietal [$t(_{1797}) = -4.34$, Est = 0.63 μV, SE = 0.14 μV, $p < .0001$)] , parieto-occipital [$t(_{1797}) = -3.100$, Est = 0.71 μV, SE = 0.23 μV, $p < .01$)] and occipital [$t(_{1797}) = -2.353$, Est = 0.83 μV, SE = 0.326 μV, $p < .05$)] scalp regions.

4.3.3 Baseline

In the baseline block of the experiment, participants listened to matching and mismatching nonce-word pairs before training during the EEG data collection. For this reason, no particular effect driven by sound / phoneme expectations caused by the difference between matching and mismatching stimuli should be observed between the two conditions. At the same time, some degree of mismatch, due to short-term statistical learning and the fact that mismatching stimuli are presented at a lower rate, is to be expected to a smaller extent. For instance, the N1 component, linked to the perception of unexpected sound stimuli, could be elicited regardless of the fact that participant had not been trained to recognise matching and mismatch stimuli.

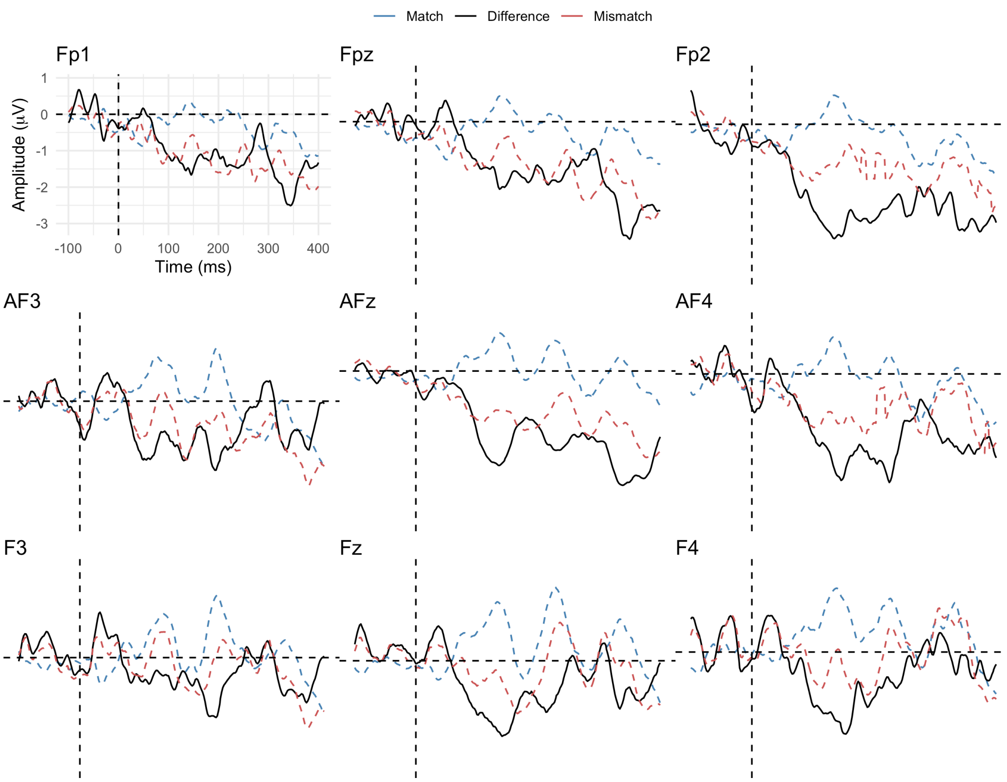
When it comes to components such as the MMN and P3, the situation becomes interesting. In theory, no MMN or P3 should be expected. In the baseline block, participants did not actively listen to the stimuli being presented and, because no training had been administered up until that point, participants had not previously learnt the differences between matching and mismatching pairs. However, the number of stimuli used is fairly low and each stimulus is repeated a high number of times in the experimental block. This could provide the participants with enough time and repetition during trials to learn to recognise a stimulus through sheer repetition. This level of statistical learning could result in a mismatch response, in the form of MMN or P3, when a mismatching stimulus is presented.

The baseline block is much shorter than the target block of the experiment. Because of this, the results from this section of the experiment are only presented as a way to provide a brief visual comparison between the two experimental blocks. The number of trials was kept to a minimum, a necessary compromise to avoid making the study too long. The target block of the experiment certainly remains the focus of the study. However, adding data from a baseline block could potentially provide access to further insights and ideas that could be explored further in future experiments. The idea behind the implementation of a baseline experimental block was inspired by a study from Astheimer and Sanders (2011), described in great detail in *Chapter 3*.

The results presented in figures 4.11 and 4.12 introduce a situation with a particularly low signal-to-noise ratio (SNR) in the data. Low SNR can be observed thanks to the high pre-stimulus noise threshold level. Before the 0 ms mark, no stimulus is presented during the trial. For this reason, all the activity collected before that time, in each epoch, should amount to random noise and, for that reason, should average to zero across trials. However, as it can be noticed, there is a level of activity as high in the pre-stimulus section as in the post-stimulus onset section[6].

[6] Although pre-stimulus noise is certainly higher in the baseline block than it is in the target block, the noise is hardly higher than 0.5 μV, which is a very low threshold. The reason why the data appear to be particularly noisy is most likely connected to the very small size of the effects.

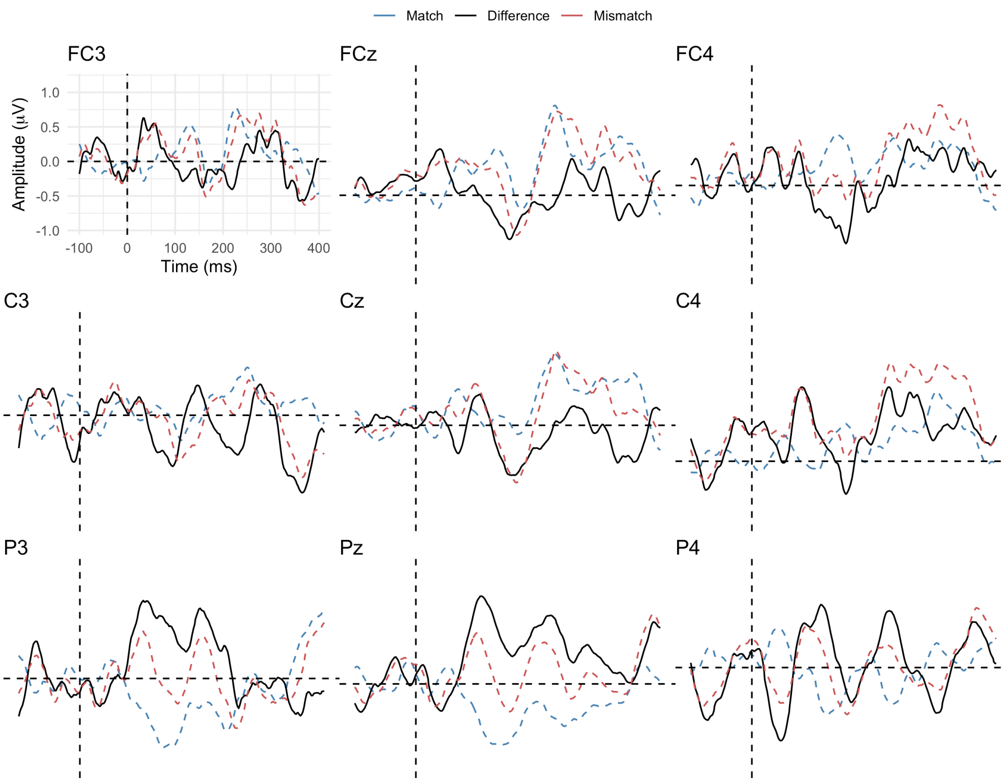
Figure 4.11: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 500 ms PSO for prefrontal, anterior-frontal and frontal sites in the experimental *baseline* block.



The AFA procedure was run on the data from the baseline block of the experiment. However, no statistically significant differences between the match and mismatch conditions were observed. This lack of statistically significant findings could be attributed to the low signal-to-noise ratio in the data for the baseline section. Low SNR does not usually allow for an adequately powered statistical analysis. On the other hand, the lack of findings could also be attributed to the fact that, because participants had not previously learnt the differences between matching and mismatching pairs, no effect was elicited when presenting stimuli from one or the other group.

No clear trends or significant effects were isolated through the implementation of topographical component origin analysis. An LMM model run to test differences for mean amplitude measurements, between match and mismatch conditions, in the baseline condition also reported no statistically significant effects across multiple time windows. In particular, no significant pre-200 effect was discovered (between 150 and 200 ms) at any scalp site ($t(19.7)$ =

Figure 4.12: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites in the experimental *baseline* block.



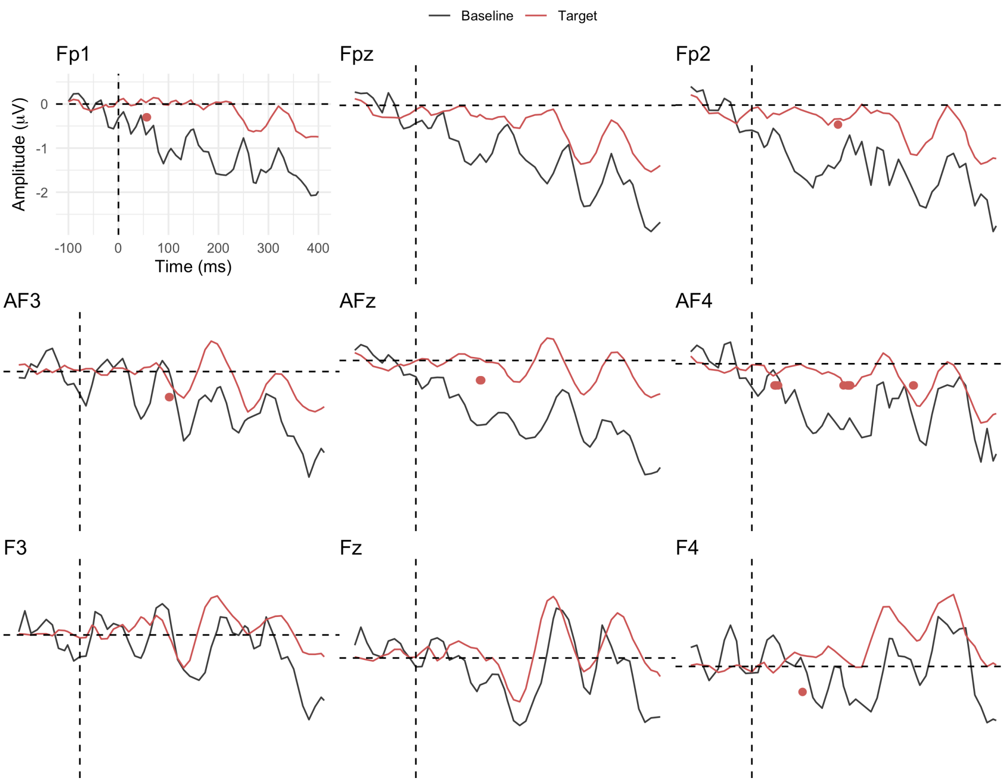
-1.20, Est = -0.10, SE = 0.14, $p = 0.1726$), nor any subsequent effects including the late P600-like component ($t(54.2) = -1.8$, Est = -0.40, SE = 0.22, $p = 0.0666$).

Responses to mismatching stimuli in the baseline block and responses to mismatching stimuli in the target block were compared. These differences and similarities are highlighted in figures 4.13 and 4.14 below.

Figures 4.13 and 4.14 display a clear difference in overall amplitude range between the two experimental blocks across the whole time window. This is, once again, due to the fact that the baseline block contains averages of much fewer trials when compared to the target block. The more averages are computed, the more noise is reduced, resulting in overall smaller amplitude peaks. Conversely, more trials results in more precise measurements and a clearer representation of the overarching effects.

The AFA procedure only captures very few differences between the two curves across the ma-

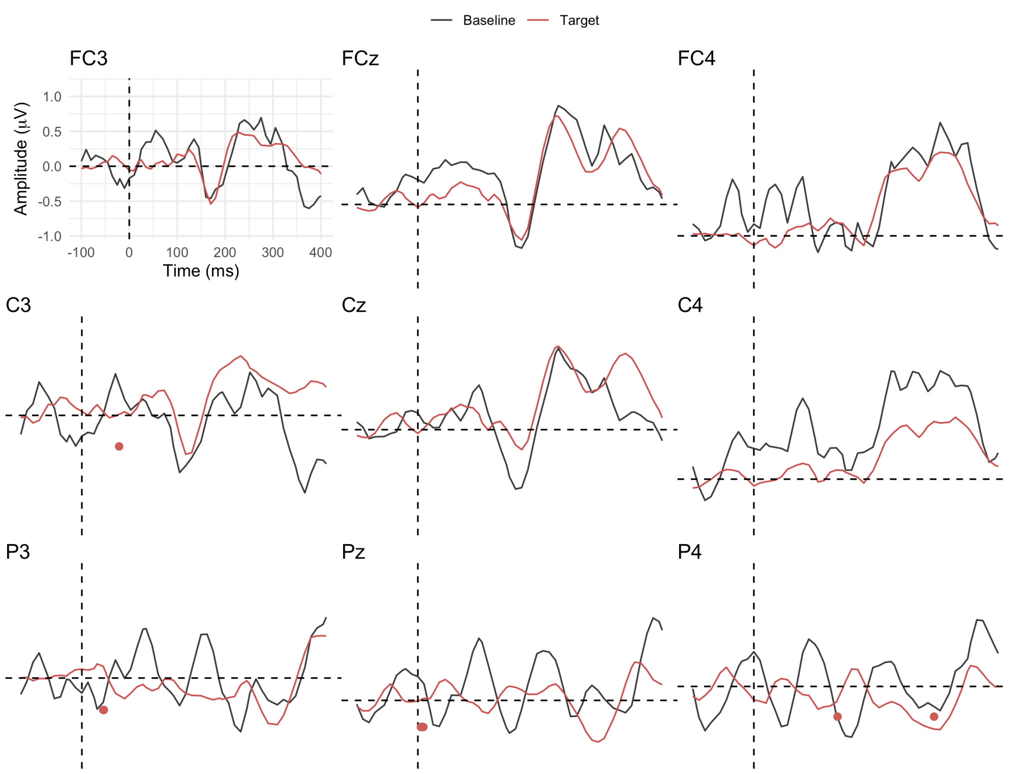
Figure 4.13: Grand-average ERP curves for mismatch experimental conditions and AFA procedure results calculated between -100 and 500 ms PSO for prefrontal, anterior-frontal and frontal sites in the experimental *baseline* and *target* blocks



jority of electrodes sites. However, setting aside differences in amplitude and amplitude peaks, which cannot be reliably measured at this point, some similarities are present between the two conditions. This is especially true in regards to component (or trend) latency of the visible ERP complexes across the time axis. For instance, when focussing especially on central scalp sites, as pictured in figure 4.14, the latency of ERP complexes such as the P2-N2-P3 is very comparable. In some particular cases, amplitude peaks are also fairly comparable between the two blocks, as it is visible for the P3 peak in right frontal, central and parietal scalp sites. This comparison, although not the focus of the study, provides further insights on the behaviour of ERP responses to known and unknown syllable sequences.

Overall, the comparison between baseline and target data highlights some very general similarities between the two experimental blocks. These similarities could suggest that, for instance, some level of mismatch and expectations are present in the baseline block as they are in the target block. This is particularly visible with comparable P3 peaks. However, trends in the baseline condition were not inferentially supported neither by the AFA method nor by the implementation of an LMER model run on mean amplitude values. It is possible that some level of mismatch was also present in the baseline block but that not enough power was available, considering the limited number of trials, to statistically confirm the visual trends. On the other hand, it could also be suggested that trends visualised in figures 4.13 and 4.14, for the baseline condition, are mostly result of noise in the data. The discussion section aims at reconciling the results obtained in experiment one with the literature on ERP research. The focus is mostly on the data collected in the target block, which remains the central focus of experiment one, but brief comparisons with baseline data are present.

Figure 4.14: Grand-average ERP curves for mismatch experimental conditions and AFA procedure results calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites in the experimental *baseline* and *target* blocks.



4.4 Discussion

In the results section of this chapter, I have highlighted the similarities and differences between the ERP data in response to the perception of mismatching and matching nonce-word pairs in the target experimental block. No PMN was component was observed across all conditions and time-windows. Significant effects were observed mostly in frontal, central and frontocentral sites between 100 and 200 ms and past 300 ms after the presentation of mismatching and matching stimuli. Differences preceding 200 ms and following 200 ms were characterised respectively by an increased negativity and positivity for the mismatch condition over the match condition. The two main effects were linked with two expected ERP components, namely the mismatch negativity (MMN) and the P3, correlated with sound mismatch and the presentation of deviant, oddball stimuli across trials. The absence of a PMN could be attributed to P3 contamination, discussed in more detail later on in this section. No higher-level component related to the processing of lexical and semantic meaning, such as the N400, was observed. However, a P600-like positivity was elicited in posterior regions for the mismatch condition following the 500 ms mark.

4.4.1 PMN

The PMN is not one of the effects that were observed among the findings of experiment one. For this reason, *H1* has to be rejected. There are different possible explanations behind the absence of a PMN response. Significant differences between the mismatch and match condition were observed in relation to components such as the MMN and P3. Because the MMN is often elicited by sound mismatch and the P3 is thought to be a response to the presentation of oddball stimuli, it could be concluded that participants did recognise mismatching stimuli as such. Drawing from this, it could also be suggested that, because participants were aware of the difference between the stimuli in the two conditions, a reaction to phonological mismatch — which is how the stimuli in the two conditions were differentiated — should have been observed.

Previous research studies were often found to disagree with regards to the topographical distribution of the PMN component, with some studies placing its elicitation in frontal/central regions (Connolly et al. 1990; Connolly and Phillips 1994; Connolly et al. 2001; D'Arcy et al. 2004), some posterior (Hagoort et al., 1999) and some evenly distributed across the scalp (Van Petten et al. 1999; Connolly et al. 1992 D'Arcy et al. 2000). However, mostly, there was agreement in placing the PMN as generally maximal around or before 300 ms post stimulus onset among most influential studies on the topic (e.g. D'Arcy et al. 2004; Connolly and Phillips 1994; ...). However, neither the AFA analysis nor the observations made through topographical origin analyses indicated a negativity similar to that described in past PMN research across all scalp sites. Furthermore, considering that the PMN has been observed as originating across many different regions of the scalp, origin analysis is consequently less reliable. In addition, the PMN has never been elicited in passive-listening experimental paradigms. Experiment one of this research project is the first experiment to directly tackle the question of whether the PMN can be elicited through phonological mismatch in a passive listening experimental design. Considering how the PMN and MMN share similar temporal distribution, it is not to exclude that a PMN component, elicited in past experiments characterised by a passive listening experimental methodology, might have been mislabelled as MMN, a more well known component.

In experiment one, the one difference between matching and mismatching pairs was based on phonetic and phonological properties alone, i.e. a different CV combination for the first syllable compared to the matching counterpart. No lexical activation, syntactic context or any other type of linguistically charged information was responsible for the difference between the mismatch and match conditions. Should the PMN be representative of phonological mismatch and mapping by itself, its elicitation would have been expected in the context of experiment one. An elicited PMN component would have been observed through the presence of negativity between 200 and 300 ms, with a negative peak for the manipulated condition at around 250-280 ms, reflecting phonological mismatch and mapping processes linked to the manipulation of highly likely stimuli.

PMN and phonological mapping

One of the possible explanations why no PMN component has been elicited for the mismatch condition is that the PMN does not in fact reflect word-form mismatch or that, at least, it does not reflect phonological mapping by itself, but rather as a lexical-activation response. The PMN component has often been linked to phonological mismatch whenever phonological mismatch was accompanied by some form of lexical activation, retrieval or processing in experiments with cloze probability sentences (e.g. Connolly and Phillips 1994), phoneme substitution (e.g. Kujala et al. 2004; Newman et al. 2003) and so on.

The PMN was elicited in previous experiments by changes in phonological mapping and mismatch, semantic mapping or a combination of both. However, the component was always observed in contexts where external linguistic information and lexical context were always present. It is possible that contextual information and semantic activity might have had an impact on the elicitation of components such as the PMN, outside of the control of the researcher. For instance, the fact that in the experiment by Connolly and Phillips (1994) the PMN was found to be greater in amplitude when both semantic and phonological mismatch were both present — compared to when only phonological mismatch was present — should already be an indicator that the PMN could be sensitive to information in language processing not limited to phonological mismatch. In addition, there were no concrete examples in which the PMN was elicited in contexts where no lexical activation was present at all. This, in turn, would suggest that it cannot be excluded that the PMN is in fact a by-product of lexical activation more than it is of phonological mapping.

For instance, the PMN could be speculated to be some form of earlier lexical phonological marker, compared to the N400, activated by lexical retrieval in speech perception and spoken-word recognition. This theory would suggest that, although the component appears to be mostly sensitive to phonological mapping and mismatch, it responses to phonological mapping but only in contexts where semantic and lexical activity are present. In the specific cases where semantic or lexical components are removed from the speech(-like) signal, there is no need for the phonological mapping linked to the elicitation of the PMN to take place. On the other hand, if it was the case that the PMN was sensitive to phonological and pre-lexical mapping, a PMN should have been observed regardless of lexical activation in experiment one.

PMN and methodological decisions

Another possible explanation for the missing PMN relies on the fact that the PMN is a fairly small effect, compared to more reliable effects such as the P3, which could have by itself contaminated the observation of the PMN component. Furthermore, passive listening paradigms such as the one carried out in this experiment — without the implementation of a behavioural task — tend to generate smaller ERP responses to begin with (Astheimer and Sanders, 2011). On the other hand, when active tasks are incorporated in a behavioural design, components such as the P3 also tend to be bigger. If the P3 component had been bigger than what it was in our experiment, chances are it would have completely obscured any other smaller or neighbouring component in its proximity, which is still what might have happened for the PMN. Eliminating behavioural tasks has at least helped ensure that the P3 component would be of manageable amplitude, while, unfortunately, reducing overall signal-to-noise ratio for both conditions. No previous research on the PMN ever used passive listening paradigms. For this specific reason, it is not possible to directly compare the findings of this experiment with previous findings employing a similar methodology. It should also be mentioned that some ERP components are not elicited in passive tasks and this might as well be the case for the PMN. However, this can not be determined from the interpretation of the findings of experiment one alone, reason why experiment two was also devised.

PMN and MMN

Interestingly, the MMN component covered a similar role to that of the PMN in its sensitivity to complex acoustic and possibly phonological changes in the presentation for mismatching stimuli. While this is not in itself evidence to directly link the MMN and PMN components, the similarities in behaviour, topographical and temporal distribution definitely suggest that the two components might in fact share a common origin. This hypothesis is later discussed in much more detail in the discussion section of experiment two, in *Chapter 5*.

4.4.2 N1 and MMN

The auditory N1 is a negative-going ERP component often elicited by the presentation of an unexpected stimulus in the absence of task demands (Näätänen and Picton 1987; Näätänen 1991). As extensively reviewed in *Chapter 2*, the component is often part of an ERP complex, named the N1-P2 complex. In the N1-P2 complex, as the name suggests, the N1 component is often followed by a positive-going P2 peak. In experiment one of this research project, an N1 component was expected for a series of reasons. First of all, the N1 component is a low level response which is sensitive to the presentation of unexpected auditory stimuli, whether they carry any level of linguistic information or, for that matter, any information at all. In particular, the N1 is usually considered a response to direct sensory (e.g. acoustic) information rather than to abstract features. It has also been generally thought not to be influenced by external and contextual information, given its early nature as a response to stimulus presentation. Differently from the MMN, although the N1 appears to represent a direct response to auditory stimuli, it is instead often correlated with shifts in auditory temporal orienting as a response to stimulus presentation.

On the other hand, the mismatch negativity component is often observed between 150 and 250 ms post stimulus onset in frontocentral regions of the scalp (Näätänen et al., 1993). The MMN component usually responds to the presentation of a deviant auditory stimulus in a sequence of standard, matching auditory stimuli. The component was originally thought to be a direct response to the perception of deviant sounds (Näätänen et al., 1993) and changes in frequency and duration. However, it has more recently been linked to the perception of phonological and, in more general terms, linguistic information as well as to direct frequency and duration changes (e.g. Shtyrov et al. 2003).

In experiment one, a negative-going effect is observed for the mismatch condition compared to the match condition between 150 and 200 ms with maximal peaks between 160 and 170 in frontal and frontocentral regions of the scalp. This effect has been associated to the presence of an MMN component, considering that the presentation of mismatching nonce words can be considered equivalent to the presentation of deviant stimuli in a sequence of standard stimuli. The response of the MMN, although most often connected with more trivial changes in much more simple stimuli (e.g. the presentation of a 2000 Hz tone following the presentation of multiple 1000 Hz tones) has been in the past observed as a response to phonological mapping (e.g. Pulvermüller 2001). This observation, in turn, could be further evidence of the sensitivity of the MMN component to higher levels of abstractness in the perception of deviant stimuli, where differences between matching and mismatching stimuli are not as obvious as doubling the frequency of a tone, but more subtle in terms of changing place of articulation of a consonant, for instance, as done in experiment one of this research study. Interestingly, the MMN takes in this particular experiment on the role that I previously theorised the PMN would take on. Coincidently, the PMN has also not been observed. Moreover, although the MMN is modulated by temporal attention to stimulus presentation (Woldorff et al., 1991), it generally does not require attention to be elicited, which appears to be the case for the findings of this experiment.

4.4.3 P3

Components that can either be linked to the elicitation of the P3a and P3b (see *Appendix A* for more information) have been observed and differences between match and mismatch condition tested significant at some of the central scalp sites. Unfortunately, the P3 (and overall, effects between the 200 and 400 ms range linked to stimulus evaluation, attentional shifts, odball stimuli) are quite problematic when it comes to trying to observe a component such as the PMN, with which it shares a very similar distribution in the temporal domain. Normally, components such as the P3 present much higher amplitudes whenever active tasks are present (Wronka et al., 2008) as a response to a stimulus that is also the target of the behavioural task. Because of this, a passive listening task was chosen for the first experiment of the project, in order to simultaneously test whether the PMN could be elicited in the absence of direct attention to stimulus presentation and in a condition where effects from the P3 component would be minimised.

Furthermore, the PMN should, in theory, be a direct response to phonological mismatch. In experiment one, no lexical or syntactic mapping was present and the main technique used to create a mismatching and matching cohort of stimuli during stimulus presentation consisted in the presentation of manipulated stimuli at a much lower rate. This was done in order to avoid participants getting used to mismatching stimuli, which in turn would have caused habituation from the participants' perspective. Unfortunately, while P3 contamination was not welcome, it was part of a conscious design choice and a compromise in experimental design. As mentioned above, had matching and mismatching stimuli been presented at the same rate, while reducing the risk of a P3, the risk of mismatching stimuli becoming more and more familiar throughout the experiment would have most likely increased. Especially because of the requirements of ERP experiments to be characterised by a fairly simple design, methodological compromises often need to be made and should be accounted for.

4.4.4 P600

The elicitation of a late positivity in posterior regions of the scalp, often associated with the observation of the P600 component (see *Appendix A* for more information), has been noted in previous sections of this chapter. These findings are discussed in more details in the next two chapters, where similar patterns are discovered in experiments two and three. The elicitation of the P600 component in contexts where no syntactic information is present opens up interesting discussions on the way participants perceive stimulus presentation sequences. Furthermore, the observation of a P600-like response opens up a discussion on how mismatch in the phonetic / acoustic realm can create, in speech as well as in music (as mentioned in *Chapter 2*), syntactic and rule violation responses in sound perception.

4.4.5 What's next

The absence of the PMN component among the findings of experiment one opens up more interesting conversations than its anticipated presence. As suggested in previous sections of this chapter, there is a chance that the absence of the PMN component is linked to the experimental design and, in particular, to the passive nature of the experiment. However, other theories that could consistently explain the absence of a PMN component in response to phonological mismatch are just as likely. In the next two chapters of this work, further experiments are presented specifically to investigate differences in speech and sound perception. Understanding the exact role of pre-N400 components such as the PMN during spoken-word recognition might be the key, in the future, to answering decade-old questions regarding the architecture of grammar and speech perception.

The second experiment of this doctoral research was devised in tandem with experiment one and it employs behavioural tasks as devices for attention during EEG data collection. At the same time, experiment two retains a very comparable methodology to that of experiment one. The findings of experiment one and two are informative on their own. However, when combined, they are far more effective at determining whether active tasks and a higher degree of focussed attention are required to elicit the PMN or whether its elicitation (or lack thereof) is independent of attention to stimulus presentation but derived on other factors such as the absence of semantic processing and lexical retrieval.

Chapter 5

Experiment 2

5.1 Introduction

In the current chapter, experiment two is discussed in great detail. Moreover, results from experiment two are also compared to those of experiment one. Experiments one and two were both designed to test the effects of — and the ERP responses to — phonological mismatch in speech-like semantics-free perception. The main difference between the two experiments lies in the way subjects interact with the experimental stimuli. In experiment one, participants were instructed to passively listen to auditory stimuli presented during the EEG recording session. On the other hand, participants of experiment two were required to interact with the stimuli by the implementation of a behavioural task. The two experiments share many features, including the set of stimuli used and part of the methodological procedure. These parallelisms have been implemented specifically to increase comparability between the findings of the two experiments.

5.1.1 Overview

In *Chapter 4* I discussed how the PMN component has been primarily linked to reflect processes of phonological and pre-lexical mapping, according to the limited literature (e.g. Connolly and Phillips 1994; D'Arcy et al. 2004; Newman and Connolly 2009). Had this been the case, that the PMN is elicited with word-form mismatch in the absence of lexical activation, its presence should have been observed among the findings of experiment one. However, this does not appear to be the case. A set of two possible theories was advanced to provide a logical explanation for this particular phenomenon: the main theory — extensively summarised in the discussion portion of the previous chapter — suggests that the PMN might not in fact reflect phonological mapping as previously suggested. The Phonological Mismatch Negativity was elicited with the implementation of a few different methodologies over the years. However, inconsistencies in the findings and methodological limitations restrain the reliability and consistency of the interpretation of the PMN as primarily a phonological component. It could be theorised, for instance, that the PMN reflects some phonological mismatch and mapping under the condition of lexical activation and semantic retrieval processes, rather than representing mapping and phonological mapping alone.

The second possible theory that can be advanced to explain the absence of the PMN component in the data that was previously collected is rooted in the methodology of the experiment. Although the PMN might reflect phonological mismatch in speech perception, the absence of active, behavioural tasks limited the extent to which the PMN component could be elicited. This phenomenon could be explained by one of two reasons: methodological limitations or specific temporal attention requirements of the component itself that were not met in experiment one. This is where experiment two becomes relevant.

Experiment two was devised, together with experiment one, to test whether behavioural tasks and, in more general terms, the degree of focussed attention to stimulus presentation had a direct impact on the elicitation of the PMN component. Investigating whether the PMN is observed in an environment where no lexical activation is present but, differently from experiment one, behavioural tasks have been implemented to focus participants' temporal attention on stimulus presentation, is helpful in delimiting the nature of the PMN component. This is done by testing multiple, similar contexts in which the PMN component should or could be elicited. Should the PMN component be observed among the findings of experiment two, it would be considered evidence to the account that the PMN requires a higher degree of focussed attention to stimulus presentation, provided by the implementation of behavioural tasks, to be observed. At the same time, the results would point in the direction that the PMN also does not require lexical processing — since no lexical activation is present in experiment two as well as in experiment one — to be elicited.

On the other hand, should the PMN component not be elicited by the paradigm created for experiment two, the data would begin to suggest that, regardless of the degree of attention to stimulus presentation, no PMN component can be elicited without the introduction of lexical activation or, in more general terms, some form of semantic mapping.

Active vs passive listening in ERP experiments

In *Chapter 2*, I addressed how different types of experimental paradigms, with a focus on active vs passive stimulus presentation, can have an effect on what ERP components are elicited and to what extent their observation differs. Furthermore, a point was made of how passive listening tasks can help reduce noise in the data by eliminating motor responses in participants. In this section, I aim to provide more specific information regarding ERP components that are often associated with experiments in speech perception and processing, including N1, MMN, P3 and the PMN.

Specifically in speech perception, ERP components behave differently depending on the degree of focussed attention to stimulus presentation, predictability and the inclusion of task demands and behavioural experimental components. Generally, amplitude is bigger for components activated when active tasks or direct attention to stimulus presentation are present. Some components can be elicited with no attention (and no consciousness to a smaller extent). Other require attention to stimulus presentation for their elicitation. The (auditory) N1 component, for instance, linked to the perception of an unexpected sound in absence of task demands, has been elicited regardless of temporal attention (Budd et al., 1998) and its amplitude is usually greater the shorter the inter-stimulus interval (ISI) (Wang et al., 2008).

An auditory — but also not linguistically specific — component that is often associated with research when participants are not paying direct attention to stimuli is the P3 (P3a more specifically) component to novelty stimuli. Not only is the P3 elicited in contexts where only passive tasks are present (Polich and McIsaac, 1994), but the component has also been extensively studied in contexts where participants were *unaware* of the study. The component, known best as the oddball component, was elicited by auditory oddball paradigms in coma patients (Gott et al., 1991). However, interestingly, it was not observed in patients under the effects of anaesthesia (Plourde and Boylan, 1991). Entering in the specifics of coma and anaesthesia research is not in the scope of this study at all. However, I am relying on this information to provide tangible examples of extreme contexts in which some ERP components and effects do not require focused attention to stimulus presentation, nor the presence of active tasks, to be elicited.

In a study on both the auditory MMN and N400 components across different types of tasks (i.e. active tasks, ignoring stimuli), it was found that the N400 component completely disappears in the passive *ignore* condition while it is present in normal active task conditions (Erlbeck et al., 2014) when semantic mismatch occurs. The MMN component, while being characterised by a bigger amplitude in active conditions does not completely disappear but rather becomes smaller in passive, ignore task conditions (Erlbeck et al., 2014). Across all of the studies mentioned in this work about the PMN component, all included active tasks (e.g. Connolly and Phillips 1994; D'Arcy et al. 2004; Newman et al. 2003; Kujala et al. 2004; Diaz and Swaab 2007; Newman and Connolly 2009). Testing for the PMN in both active and passive task environments could help determine whether the PMN component can be elicited in contexts where no semantic mapping is present and, at the same time, also to investigate the extent to which focussed attention to stimulus presentation mediates the amplitude, latency and topographical distribution of the component.

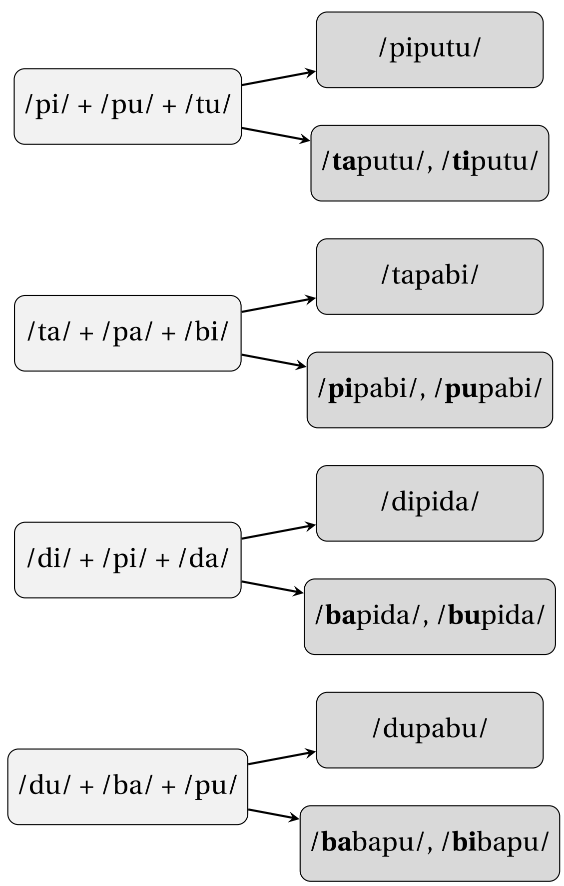
5.2 Methods

Experiment two takes a different approach to that of experiment one by introducing behavioural tasks. At the same time, the experiment mostly employs the same stimuli and a similar methodology for maximum comparability between the results of the two experiments. The protocol was overall more straightforward, consisting of one experimental section only, and participants performed behavioural tasks during the main experimental block. The core of experiment two consisted of participants listening to a sequence of three syllables, presented auditorily and separated by a short pause. Once all three syllables had been presented, subjects were instructed that they had four seconds to mentally concatenate the three syllables in the sequence to form a (nonce) word. Participants were also instructed not to alter the order in which the syllables were presented. Following the four-second pause in which participants completed the behavioural task, subjects would hear a nonce word that either matched the nonce word made up of the three syllables (66% of total trials) or a mismatch nonce word that presented a different first syllable (33% of total trials). Both the match and mismatch stimuli would present the same trisyllabic structure and the only differences would be the choice of CV combination in the first syllable.

5.2.1 Stimuli

The stimuli employed in experiment two match those used in experiment one. However, they were arranged and combined differently and in different contexts. The four match nonce words — that comprised three syllables each — were the four nonce words used as *match stimuli* for experiment one. In experiment two, the nonce words were presented one syllable at a time to participants, e.g. A+B+C, with a short 0.5 s pause in between each syllable. They were then presented after a four-second pause as either *match stimuli*, e.g. ABC, or *mismatch stimuli*, such as A'BC or A"BC. In this context, A' and A" were different CV syllables from the matching A syllable. However, mismatch syllables maintained a similar structure to the original CV, match syllable. Figure 5.1 highlights all possible combinations of match and mismatch stimuli in experiment two.

Figure 5.1: Stimuli



No behavioural pilot was run for experiment two to assess whether the syllables and/or nonce words resembled words in English since the same stimuli used in experiment two were already implemented in experiment one. As described in *Chapter 4*, a behavioural pilot was run as part of experiment one to test for bias or confounding factors in the stimulus pool, including resemblance to real words. Furthermore, because the same stimuli had been created and used in experiments (e.g. Astheimer and Sanders 2011) prior to experiments one and two of this research project, most of the safety checks, such as correcting for phonemic transitional probability, were previously run by other researchers.

5.2.2 Participants

20 (F = 13) undergraduate students at The University of Manchester took part in experiment two. All of the twenty participants were native speakers of British English, right-handed adults who reported normal or corrected-to-normal hearing and vision. One of the participants who took part to experiment two also took part to experiment one. However, data collection for the two experiments was carried out more than six months apart so there is very little ground to advance the suggestion of stimulus information retention from the previous experiment. No participants disclosed use of psycho-active medications or reported any known neurological condition. Written consent was taken from each participant before the start of the experiment and č20 compensation was offered upon completion of the experiment. All participants completed the experimental procedure and no dataset was discarded from the grand average of data, as no recording exceeded a number of rejected trial above the 10% mark of all trials.

5.2.3 Design and instructions

The experimental protocol comprised 10 experimental blocks with 30 trials per block, totalling to 300 trials, divided as 200 match and 100 mismatch trials. Between each block, a 30-second rest block was implemented. Participants of the study were told that the option of a longer break was available should they have requested it. No subject took advantage of this. From the beginning of trial 1 to the end of trial 300 the average running time for experiment two was of one hour and ten minutes including rest blocks.

Each trial had a duration of 10 seconds. Three syllables were presented auditorily to each subject. Each syllable had an average duration of 200 ms and a pause of 500 ms was placed to separate the first syllable to the second and the second from the third. After the third syllable had been presented, participants were given four seconds to combine the three syllables together, mentally, without rearranging their order. Participants were instructed to combine the syllables in order to form a (nonce) word. They were told that after four seconds, they would have to be hearing that particular word being presented — auditorily — back to them. However, the matching nonce word was only presented, randomly, to participants only 66% of trials and, as mismatch, 33% of the total trials.

Participants were instructed to perform all tasks mentally. They were instructed not to press any buttons on the keyboard, speak or repeat the syllable or the words out loud, as that would create noise in the data. Participants were presented a set of three training trials at the beginning of the experimental block to present the task and to test everything was functioning correctly in the experimental booth. All three training trials were match trials.

5.2.4 Expected effects

The same 18 electrodes that were isolated for running significant testing analyses on data for experiment one were selected for experiment two, as well, since similar responses were expected across the two experiments. Mean amplitude modelling using LMM was carried out on all scalp sites. Employing the same stimuli across experiment one and experiment two, together with the use of a similar methodology, meant that comparable effects should be expected throughout both experiments. For this reason, comparisons between the results from both experiments likely offered insights on the differences between the two designs. Because stimuli of experiment two did not introduce semantic mapping in any way, no N400 effect was expected for the presentation of mismatch stimuli, compared to matching stimuli, similarly to experiment one. On the other hand, effects like N1 and MMN were expected due to the mismatch in sound perception and the same goes for effects like P3, connected to stimulus categorisation and oddball stimuli presentation. Finally — and arguably most importantly — a PMN component was expected, connected to phonological mismatch in speech perception. Because the tasks used in experiment two were active, effects were generally greater in amplitude and this allowed for more definition and a higher signal-to-noise ratio in the data compared to previous results from experiment one.

5.2.5 Hypotheses

Based on the design of experiment two, and in light of the findings of experiment one, a similar hypothesis can be put forward as in our previous experiment.

H1: The PMN component is expected to be observed as a response to word-form mismatch, primed by the presentation of a behavioural task, in the absence of lexical status of the stimuli.

5.2.6 Statistical testing

Similarly to experiment one, all significant testing was carried out at the channel level with the use of an adaptive factor adjustment (AFA) statistical method (Sheu et al. 2016) and with the implementation of linear mixed effect regression models (LMM) for the modelling of mean amplitude. In addition, mean amplitude cubic spline interpolation topographical maps were created to visually investigate component origin across the scalp as well as to provide visual aid to the presentation of statistical models.

5.3 Results

The results section provides detailed comparisons of responses from multiple conditions in experiment two, together with a comparison of results of experiment two and experiment one. The main focus of this study is the difference between the subjects' response to mismatching vs matching nonce-word stimuli and whether a clear PMN component is elicited in a context that is, at the same time, stripped of any lexical activation and where behavioural tasks are present. However, some portions of the results section of the current chapter also deal with supporting results such as ERP responses to the presentation of each syllable in the sequence and a comparison of responses to syllables and mismatch nonce words when appropriate.

5.3.1 Match vs mismatch

Figures 5.2 and 5.3 present a comparison of ERP responses for the mismatch and match conditions across sets of anterior (AF), frontal (F), frontocentral (FC), central (C), and parietal (P) scalp sites, where all of the expected effects are likely to be observed, according to existing literature and previous studies. At the same time, in black, a difference curve displaying the *mismatch minus match* condition was included. AFA results are also discussed later on this particular section of the chapter. However, general trends are also observed and interpreted regardless of their effect size and significance status. Later, mean amplitude modelling is carried out in order to try and determine the size and distribution of effects and trends across the scalp, across different time windows following the presentation of matching and mismatching stimuli.

Figure 5.2: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.

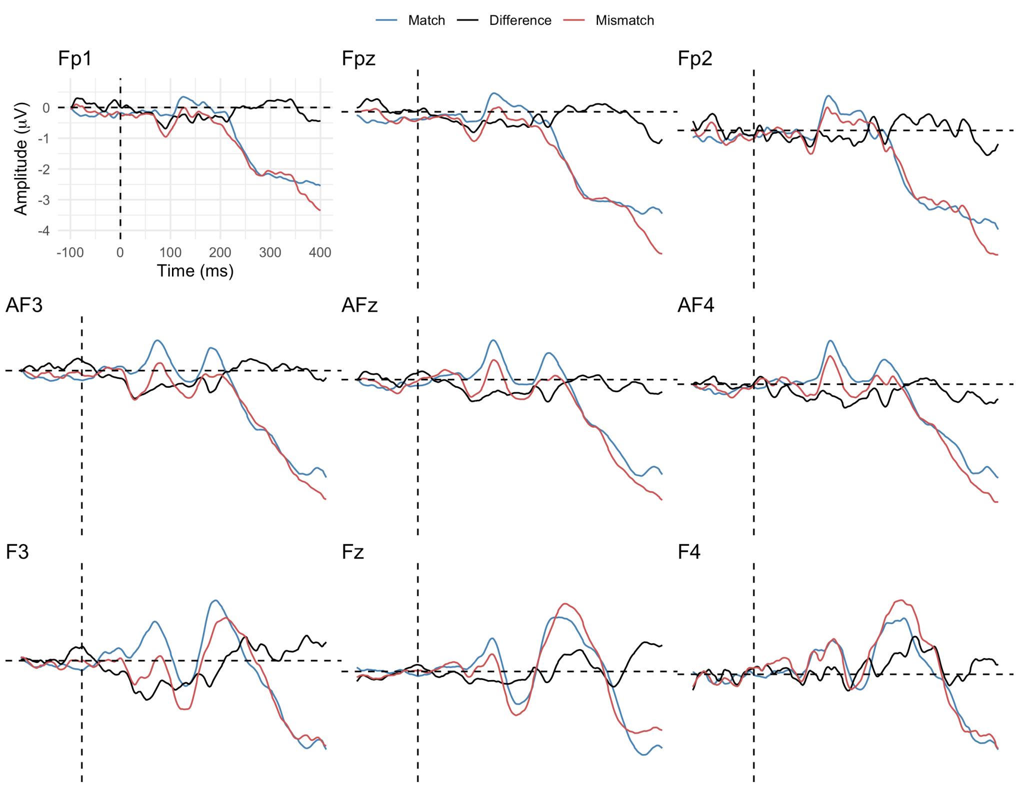
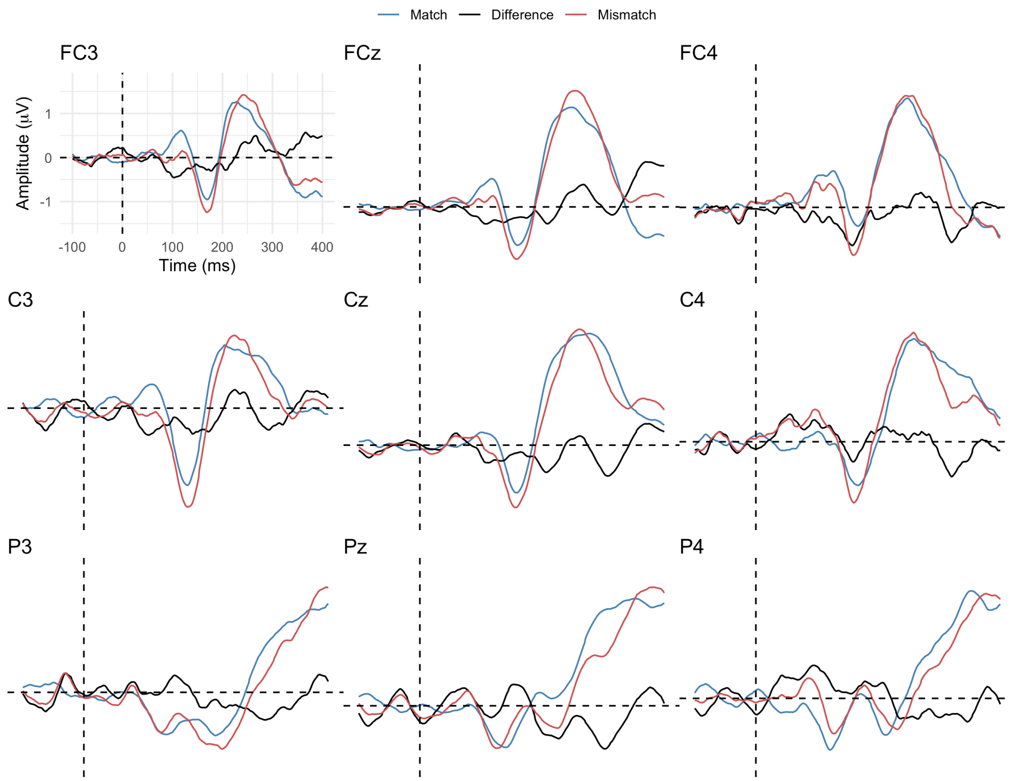


Figure 5.2 displays visible trends in frontal, anterior and prefrontal scalp sites across the mismatch and match conditions. Particularly, the difference curve of the two conditions highlights a negative-going deflection, maximal in left prefrontal and anterior-frontal scalp sites (Fp1 and AF3) with peaks overlapping with the 100 ms mark. The deflection is followed by a negativity for the mismatch condition that persists until 200 ms post stimulus onset. The negative peak and the following negative deflection could be most likely associated with the presence of N1-like and MMN responses to the presentation of mismatching stimuli. These two (and other) differences between the curves are later tested with the implementation of the AFA procedure (Sheu et al., 2016). What appears as another negative deflection peaking at 200 ms for the mismatch condition, especially when focussing on the difference curve, is most likely connected with a positivity in central sites, linked to the elicitation of the P3 component and discussed in more detail in the following paragraph. A positive deflection can also be ob-

served for the mismatch condition peaking around the 400 ms mark primarily across frontal scalp sites (Fz and F4).

Figure 5.3: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 400 ms PSO for frontocentral, central and parietal sites.

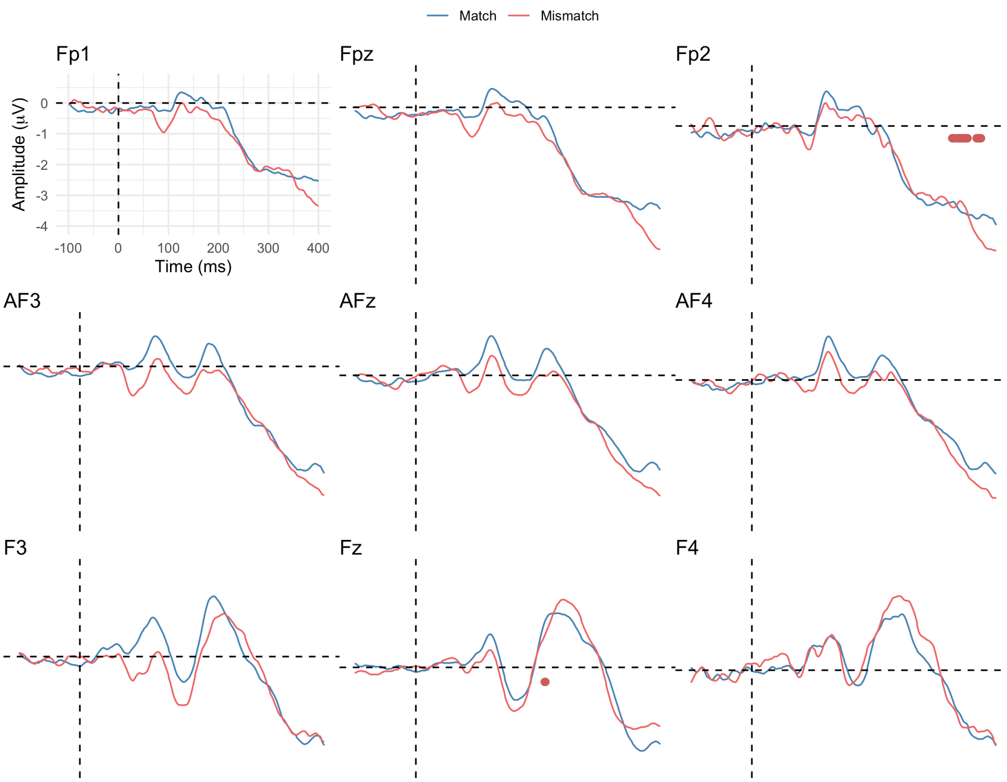


In midline frontocentral and central scalp sites (FCz and Cz), where they are maximal, positive deflections can also be observed for the mismatch condition around 250 and 350 ms post stimulus onset. These effects, similarly to the findings in experiment one, are most likely linked to effects of stimulus categorisation, presentation rates and the processing of oddball stimuli. However, the most visible effect is found in central and parietal sites, where the difference curve highlights what appears to be a negative deflection peaking at the 300 ms mark in midline and right central and parietal sites (Cz, Pz, C4, P4). The component's amplitude appear to be bigger than the pre-stimulus noise. This suggests, to an extent, that the trend is not necessarily an effect of noise in the data. The negative deflection peaking at 300 ms post stimulus onset could be associated with the elicitation of a PMN component, considering they both share a similar distribution in the time domain, including peak latency. However, its topographical distribution is very limited to a few central scalp sites and, in previous literature, many of the findings suggest that the PMN is often elicited in frontal and central scalp sites or with a flat distribution across the whole scalp and midline (Lewendon et al., 2020). When presenting AFA results and topographical distributions of the observed effects, more is said about the elicitation of this negative-going deflection that, regardless of its association with a PMN component, was not initially observed in the results of experiment one.

Figures 5.4 and 5.5 display significant differences between the two experimental conditions, according to the AFA procedure method (Sheu et al. 2016). The statistically significant differences — with an alpha threshold set to .05 — are represented with red dots and lines for the significantly different portions of the curve in the time domain. It can be observed that statistical differences are mostly present around 300 ms post stimulus onset in figure 5.4 while more effects are found in the time window through scalp sites of figure 5.5.

Focussing on anterior and frontal scalp sites, no particularly relevant effects were captured with the implementation of the AFA procedure. The only differences highlighted have been observed around the 300 ms mark (between 270 and 300 ms depending on the scalp site) and they often coincide with very small positive deflections — $< 0.5\ \mu V$ — for the mismatch condition. The data and topographical distribution of the effects suggest that these differences are a by-product of a negative polarity that is observed in central and parietal scalp sites for the mismatch condition, which in turn results into a negativity in frontal and occipital scalp sites alike. Interestingly, although differences at the 100 and 160-170 ms mark can be spotted especially in prefrontal and frontal scalp sites respectively, the general trend is not assessed as significant by the application of the AFA method. However, considering that the average amplitude of the effects surpasses, in all cases, the average noise in the pre-stimulus interval, it could be hypothesised that the lack of statistically significant effects in the curves in that specific time window is mainly due to a lack of power or an extreme conservativeness in the implementation of the AFA procedure.

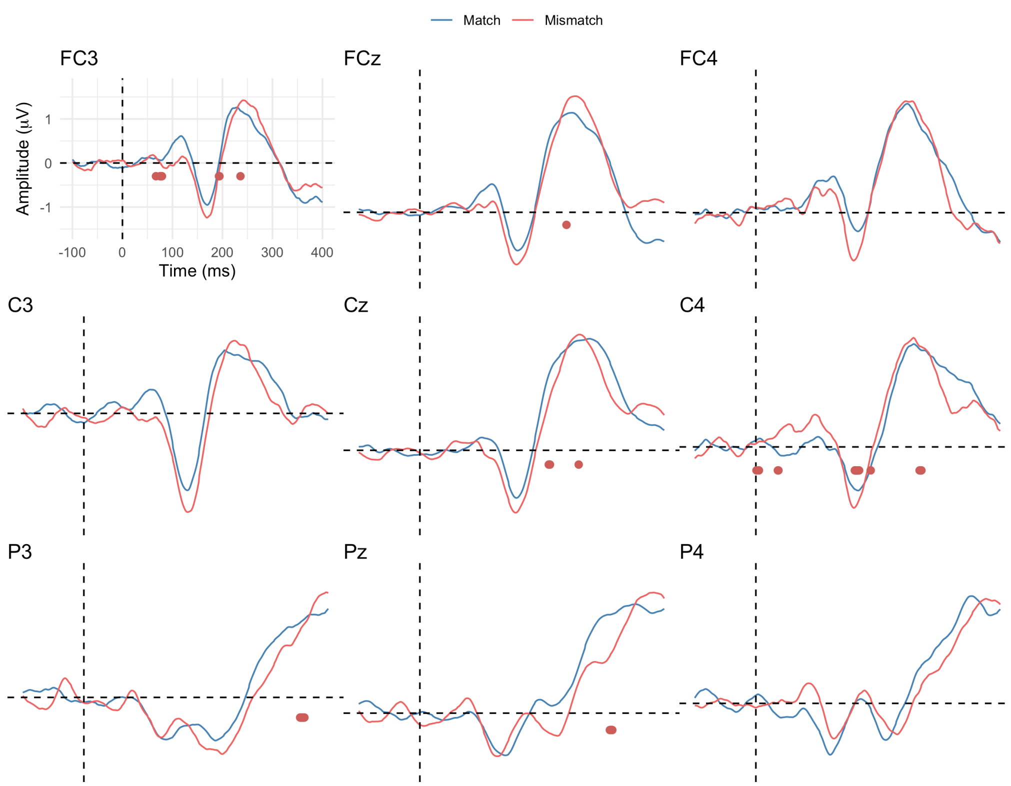
Figure 5.4: Grand-average ERP curves for match / mismatch experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.



However, the situation is particularly different once results of the AFA method on frontocentral, central and parietal scalp sites are considered. At FC3, three distinct sections of the curve are statistically different, at around 100, 250 and 350 ms respectively. The first two differences are most likely linked with N1 and P3-like effects, linked to attentional shifts in stimulus presentation and the processing of unexpected, low-rate stimuli. The third effect is, on the other hand, likely linked to this small difference between the two curves observed in the previous set of electrodes on the right side of the scalp. However, the amplitude difference in the time-window of significance is extremely small — 0.06 μV — which could suggest that the difference might be due to the elicitation a false-positive effect. These are rare with the implementation of the AFA procedure, due to its conservativeness and its specific ability to separate data and noise, but never to be excluded when dealing with ERP and EEG data.

The most insightful trends, when it comes to frontal, central and parietal regions of the scalp, are clustered around the central and parietal scalp sites, with a particular focus on midline and right parietal scalp sites (Pz and P4) where the observed effects are maximal. The AFA procedure highlights a significant difference at the 300 ms mark for parietal scalps sites Pz and P4, followed by a general trend of negativity in both electrodes (significant in P4) that continues after the 400 ms mark. While there certainly is P3 contamination in parietal scalp sites with what looks like a negative peak shy of the 250 ms mark (by-product of a central positivity), the negativity highlighted for the mismatch condition, compared to the match one, following that 250 ms mark and continuing on until and after 400 ms, seems to be representative of its own separate component, as highlighted later with topographical representation of the observed results.

Figure 5.5: Grand-average ERP curves for match / mismatch experimental conditions and AFA procedure results calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites.



In order to try and determine whether the negativity observed in central and parietal regions of the scalp, following the 300 mark, was mainly caused by P3 contamination or whether it was possibly originated by phonological or lexical mismatch[1], I have compared the findings

[1]No lexical activation is present, however, the N400 has been mainly observed as a centroparietal negative deflection between the 250 and 500 ms mark, with peaks at around 400 ms (Kutas and Hillyard

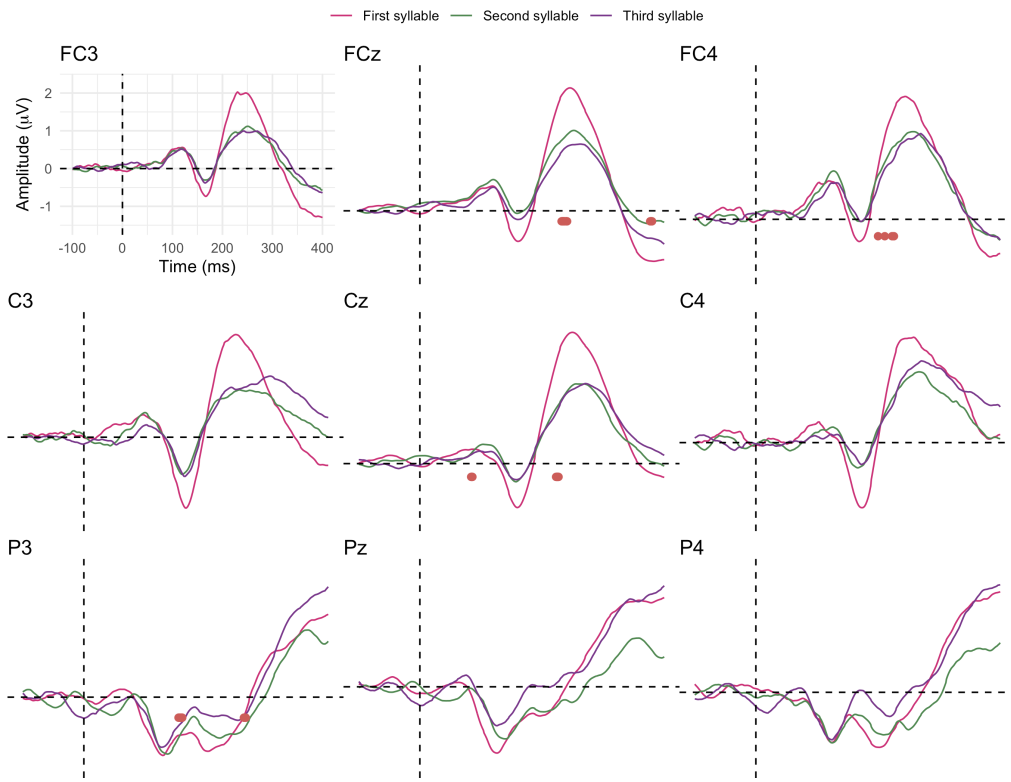
presented above with the responses to the presentation of each of the three syllable stimuli in the sequence played to participants in the first half of each trial.

In particular, considering that each of the four syllable sequences is unique and also considering that every sequence-onset syllable is also unique, in that no syllable is used twice in the same position across all trisyllabic sequences, there is a 25% chance, at the beginning of each trial, that a participant will hear one of the four sequence-onset syllables. However, following the presentation of a specific syllable in the first position, the second and third syllable have a 100% presentation rate and, for that reason, should be expected by participants, granted that they had learnt the syllable sequences through passive statistical learning. It is not unlikely that the participants had learnt — although no training session was directly administered with behavioural tasks — the sequences considering that only four syllable sequences were used throughout the whole experiment and that they were each presented more than 150 times during experimentation.

When considering the expected response to the presentation of each of the syllables in a sequence, the main difference between the presentation of the first and second syllables in a sequence is determined by the rate of stimulus presentation and not by any type of task-related phonological or lexical mismatch. For this specific reason, effects like N1 and P3 are to be expected to a small extent. However, no PMN (or N400 for that matter) should be observed as a response to the presentation of the first syllable in a sequence. If a similar pattern of negativity, following the negative peak at the 250 ms mark in parietal scalp sites — by product of a central P3 — is not observed in the perception of single syllable in the first half of each trial, it can be argued that the effects found to differentiate the mismatch and match curve analysed before are possibly due to the elicitation of a component possibly representing some level of linguistic mismatch in the processing of mismatching stimuli. Moreover, responses to mismatching stimuli are those mostly susceptible to intra-subject variation, considering that they are the ones averaged from the least amount of trials (33%), followed by responses to matching stimuli (66%) and, finally, by the response to each of the syllable in a sequence (100%).

1980; Kutas and Hillyard 1984.), similarly to what has been elicited in right central parietal electrodes in this data set.

Figure 5.6: Grand-average ERP curves for first / second / third syllable experimental conditions and AFA procedure results calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites.



What was predicted as one of the options related to the origin of the negative deflection at 300 ms post-stimulus onset for the mismatch condition can be more clearly observed in figure 5.6. First of all, although three curves have been included in the grand-average plots, corresponding to the perception of each of three syllables in the sequence, the marks indicating significant differences tested with the AFA method are in regards to only differences between the perception of the first and second syllable of each syllable sequence. This is because the main difference should be expected between the first and second syllables rather than from the second to the third. It is assumed that a similar difference should be observed between the first and the second and the first and the third syllables. To improve plot legibility, only comparisons between the perception of the first and the second syllables in the sequence have been presented.

Across all frontocentral scalp sites (FC3, FCz and FC4) statistically significant differences have been observed between the conditions representing the perception of the first and second syllable of each sequence between 150 and 200 ms, where an MMN effect has been found to be maximal. However, when focussing on parietal scalp sites specifically, there seems to be no significant negative trend for the perception of the first syllable sequence compared to the perception of the second. The only difference between the two conditions is determined by statistical rate of stimulus presentation in the absence of phonological mismatch, while at the same time maintaining an active task and direct focussed attention to stimulus presentation. It could be hypothesised that the negative deflection characterising parietal scalp sites — P2 specifically in figure 5.3 — could be associated with a specifically phonological responses or, in more general terms, a difference not associated with a P3-like response.

The 300 ms mark is an extremely relevant time point for this series of three experiments, considering it is often when the PMN is maximal for the experimental target condition whenever phonological mapping or mismatch is present. However, and this is discussed in much more detail in the following discussion section, the effect that has been observed in this particular context is only observed as significant over one parietal scalp site. Overall, the trend that can be observed over few central and parietal electrodes is characterised by minimal differences. The literature on the PMN is not clear on its placement in terms of topographical distribution. The component has been observed in a variety of locations, including central sites. However, the PMN is often observed in frontal / central sites or across the entire scalp, which is not like the findings of this particular experiment.

At the same time, the PMN has only ever been observed in settings where both lexical and phonological processing were active and, in this particular case, only phonological mismatch was present. Only experiments where both semantic and phonological mapping were present simultaneously were carried out while investigating the PMN. For this reason, using previously observed topographical distributions of the component to identify whether the effect discovered in the current experiment is similar to the PMN as originally observed does not constitute a fair and direct comparison. The smaller distribution of the component and its reduced effect might be, in fact, linked to the idea that the component was elicited in a context where no semantic mapping was present.

Furthermore, the absence of a post-P3 negativity was not observed for the *first* - *second* syllable condition comparison. However, it cannot be ruled out that this effect is also a by-product of stimulus categorisation and oddball stimuli presentation rate effects. For this particular experiment, these effects might have been more prominent in the match and mismatch conditions, taking into consideration that the odball stimuli in question were also the target of the active, behavioural tasks included in implementation of the study. For this particular reason, very careful assertions are made on the nature of the observed 300 ms negativity. Experiment three implements the same exact methodology of experiment two but without the inclusion of linguistic stimuli. If a 300-ms negativity, similar to the one observed in central and parietal scalp sites for the mismatch condition of experiment two, is not found among the results of the mismatch-match comparison in experiment three it could be suggested that the difference was determined by the presence of phonological mismatch. On the other hand, because presentation rates of stimuli in experiment two and three are exactly the same, if a similar effect is observed in both experiment two and three, its nature can be linked to processes of stimulus categorisation and presentation rates. It could, then, also be suggested that its lack of elicitation in experiment one is linked to the lack of behavioural tasks and temporal attention to stimulus presentation.

5.3.2 Modelling

Similarly to the presentation of results for experiment one, cubic spline interpolation topographical maps have been created and mean amplitude has been modelled with linear mixed-effect models (LMM) to investigate ERP component distribution across the scalp for the time windows of interest to the study, where either significant effects or overall trends were highlighted in channel-level analysis. Experimental condition, scalp region and hemisphere were all included in the model as main effects as well as a three-way interaction. When using LMM, random intercepts were allowed for experimental subjects. Model structures were later reduced through step-wise regression to reach best model fit through AIC / BIC comparisons. Results are reported from *type-III* ANOVA tables often used in place of model summaries for multilevel multi-factorial designs. Five time windows are analysed here, related to the elicitation of four to five plausible ERP components, namely the N1, MMN, PMN, P3 and, as a particularly unexpected event, the P600.

Figure 5.7: Grand-Average mean amplitude intervals (between 80 and 100 ms PSO) for mismatch / match experimental conditions

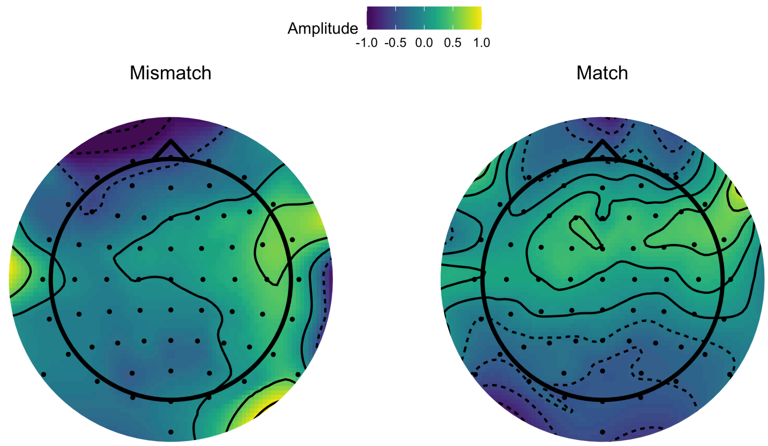


Figure 5.8: Grand-Average mean amplitude intervals (between 160 and 210 ms PSO) for mismatch / match experimental conditions

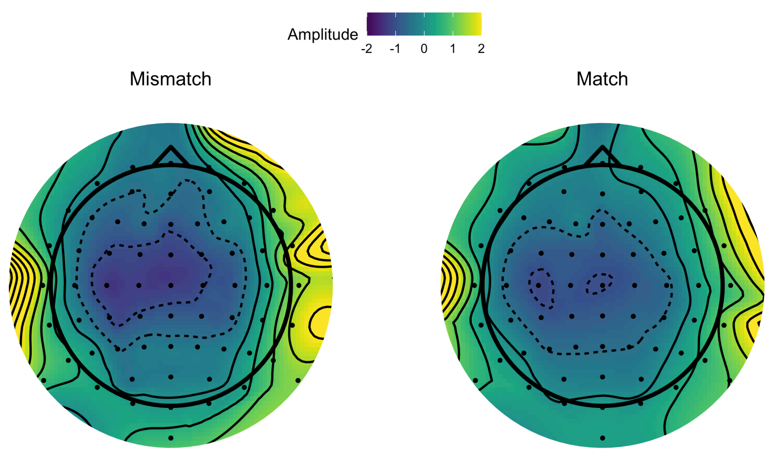


Figure 5.9: Grand-Average mean amplitude intervals (between 240 and 280 ms PSO) for mismatch / match experimental conditions

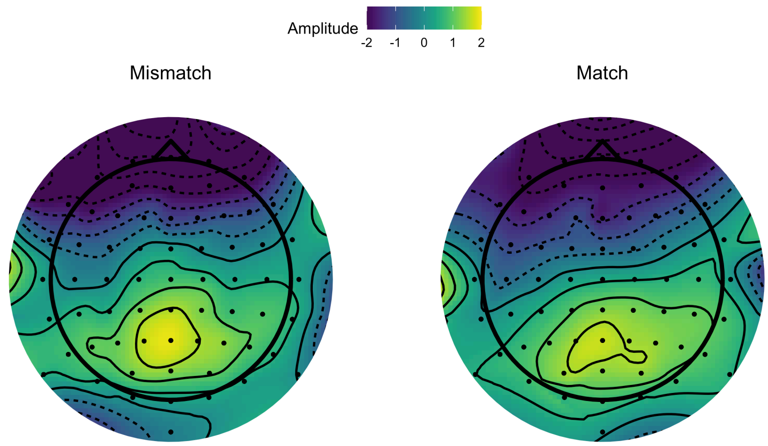


Figure 5.10: Grand-Average mean amplitude intervals (between 290 and 310 ms PSO) for mismatch / match experimental conditions

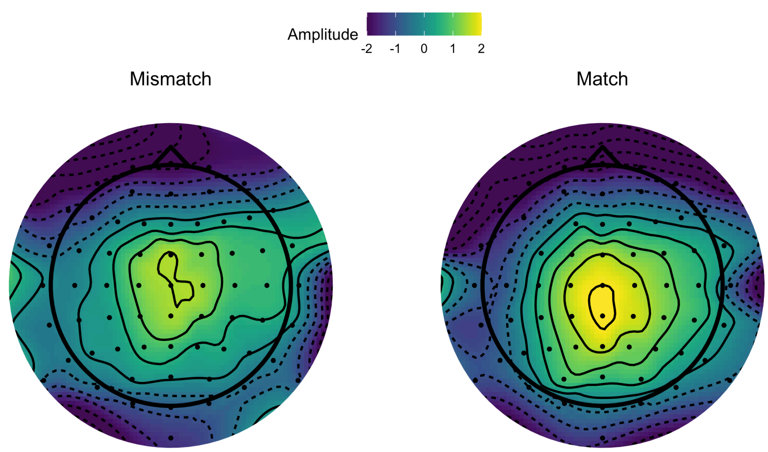


Figure 5.11: Grand-Average mean amplitude intervals (between 575 and 625 ms PSO) for mismatch / match experimental conditions

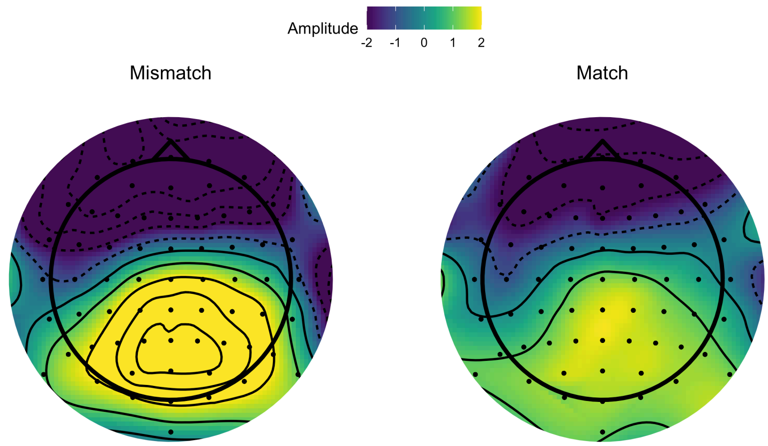


Figure 5.7 displays what appears to be a negative trend for the mismatch condition between 75 and 100 ms post stimulus onset in left frontal and prefrontal scalp sites. The negativity is most likely connected to one of the components of the N1 ERP response, often elicited across different sources in the scalp (Näätänen and Picton, 1987). The effect, however, did not appear to be significant at the channel level by the implementation of the AFA procedure across most scalp sites and the size of the differences appears to be minimal when compared to other trends and differences displayed across following time windows. A LMM was fitted to the mean measurements of amplitude between 75 and 100 ms post stimulus onset. There was a main effect of experimental condition [$F(_{1,1966.7}) = 6.52$, $p < .05$] scalp region [$F(_{10,1949.7}) = 16.85$, $p < .0001$], hemisphere [$F(_{1,1949.2}) = 44.66$, $p < .0001$], as well as a significant interaction between experimental condition and region [$F(_{10,1949.2}) = 8.70$, $p < .0001$][2]. In order to explore pairwise comparisons of the effect of experimental condition within each scalp region — in each hemisphere, if required — pairwise contrasts were calculated with the `emmeans` package (Lenth et al., 2019). Post-hoc pairwise comparisons reported a significant effect of experimental condition for prefrontal, anterior-frontal, frontal and parietal scalp regions over both hemispheres.

[2] F statistics were calculated with *Kenward-Roger* approximation for degrees of freedom in models with random effects. anova() function as per package lmerTest (Kuznetsova et al., 2017).

Table 5.1: Least-squared means and pairwise contrasts of match and mismatch experimental conditions. Mean amplitude (in μV) between 75-100 ms PSO

Region	Match: $\bar{x}$	Mismatch: $\bar{x}$	Estimate	DF	***t***	***p***
prefrontal	-0.27 μV	-0.73 μV	-0.45 μV	1951	-3.64	<.001
anterior-frontal	-0.15 μV	-0.55 μV	-0.39 μV	1953	-4.45	<.001
frontal	0.06 μV	-0.24 μV	-0.31 μV	1953	-5.03	<.0001
parietal	-0.08 μV	0.14 μV	0.22 μV	1958	2.14	<.05

Following the negative-going trend observed around the 75-100 ms mark, a negativity-going deflection can also be observed for both mismatch and match experimental conditions with peaks between 160 and 170 ms post stimulus onset in figure 5.8. This negativity, which follows the prefrontal and frontal N1-like response, also appears to be mostly distributed in anterior and frontal scalp sites and it appears to be marginally more negative for the mismatch condition. The effect appears to be maximal in frontal scalp sites Fz and F1. Fitting a LMM on mean amplitude measurements between 160 and 210 ms post stimulus onset, there was a significant main effect of scalp region [$F(_{10,1958.0})$ = 12.42, $p < .0001$] and hemisphere [$F(_{1,1958.0})$ = 60.32, $p < .0001$], as well as a significant interaction between experimental condition and hemisphere [$F(_{1,1958.0})$ = 4.26, $p < .05$]. However, no significant interaction between experimental condition and scalp region was reported. Post-hoc comparisons on the interaction between hemisphere and experimental condition report a significant effect of experimental condition over the left hemisphere [$t(_{1961})$ = -2.64, Est = 0.19 μV, SE = 0.06 μV, $p < .01$]. Because no significant interaction between experimental condition and scalp region was discovered, pairwise comparisons for experimental condition within scalp regions are not recommended. Because of this, the more pronounced negativity in the left hemisphere for the mismatch condition cannot be directly attributed to frontal regions specifically. When discussing a similar effect in the results section of experiment one, a doubt was raised regarding the nature of the component observed at the 150-200 ms mark, insinuating that the effect was either mirroring a late N1 effect or some other component in the range usually reserved to the MMN in auditory mismatch studies. However, a clearer N1-like effect can be observed in experiment two which would suggest that this negative deflection, which does not appear to be significant in a whole-scalp analysis, is probably linked to the elicitation of later components, possibly in the range of an MMN to stimulus presentation.

Between 240 and 280 ms, where a positivity is displayed by difference curves in figures 5.2 and 5.3, fitting an LMM to mean-amplitude measurements reveals significant main effects of scalp region [$F_{(10,1949.2)}$ = 88.32, $p < .0001$], hemisphere [$F_{(1,1949.0)}$ = 118.82, $p < .0001$] and a significant interaction between scalp region and experimental condition are reported [$F_{(10,1949.0)}$ = 3.60, $p < .0001$]. When exploring pairwise contrasts of experimental condition within scalp regions, a main effect of condition, in the form of a negative trend for the mismatch condition, was only found at occipital [$t_{(1961)}$ = -2.04, Est = -0.64 μV, SE = 0.30 μV, $p < 0.01$] and parietal [$t_{(1961)}$ = -2.94, Est = -0.39 μV, SE = 0.13 μV, $p < .01$] scalp sites. This effect does not appear to be related to the elicitation of a P3-like response.

Figure 5.10 highlights topographical differences between the mismatch and match condition between the time window of 280 and 320 ms post stimulus onset. The map displays a small negative-going deflection in central scalp sites over the midline and in the right hemisphere for the mismatch condition, when compared to the response to matching, control stimuli. When referring back to Figure 5.5, a difference in latency between the match and mismatch peaks between 200 and 250 ms can be seen, with the mismatching condition negative peak being maximal 17 ms after the peak in the match conditions at scalp site P4. Whether the negative-going deflection that derives from this peak difference is a by-product of this difference in latency or whether the difference in latency is a by-product of the negativity is not possible to ascertain at this point. For this reason, this difference between the two conditions is treated as either a source of contamination from the elicitation of P3-like component (although, as previously discussed, not visible for the separate syllable condition) or an effect of phonological mismatch connected to the elicitation of the PMN or an equivalent component located in this specific region of the scalp. Results of LMM, fitted to mean amplitude between 280 and 320 ms post stimulus onset reported significant main effects of scalp region [$F_{(10,1939.9)}$ = 94.48, $p < .0001$] and hemisphere [$F_{(10,1939.2)}$ = 74.34, $p < .0001$], as well as significant two-way interactions of experimental condition and region [$F_{(10,1939.2)}$ = 2.95, $p = .001$] and region and hemisphere [$F_{(10,1939.2)}$ = 3.83, $p < .0001$].

Table 5.2: Least-squared means and pairwise contrasts of match and mismatch experimental conditions. Mean amplitude (in μV) between 280-320 ms PSO

Region	Match: $\bar{x}$	Mismatch: $\bar{x}$	Estimate	DF	***t***	***p***
parieto-occipital	0.51 μV	-0.04 μV	-0.36 μV	1943	-2.71	<.01
frontotemporal	-0.86 μV	-0.29 μV	-0.39 μV	1941	-1.97	<.05
frontal	0.84 μV	-0.56 μV	-0.31 μV	1946	-0.14	<.01

However, pairwise comparisons of experimental condition by scalp region only revealed a negative shift between match and mismatch condition at parieto-occipital scalp sites, likely in response to an increased positivity for the mismatch condition over frontotemporal and frontal scalp sites.

Finally, a late positivity can be visually observed for the mismatch condition with peaks around 600 ms post stimulus onset and maximal in parietal scalp sites. Results of linear mixed effect regression modelling on mean amplitude measurements between 575 and 625 ms post stimulus onset suggest significant main effects of experimental condition [$F(_{1,1966.6}) = 9.67$, $p = .001$], scalp region [$F(_{10,1949.2}) = 285.10$, $p < .0001$] and hemisphere [$F(_{1,1949.2}) = 32.79$, $p < .001$]., as well as a significant interaction between experimental condition and scalp region [$F(_{10,1949.2})$ $= 26.16$, $p < .001$]. Post-hoc pairwise contrasts suggest a main effect of experimental condition, with positive estimates from match to mismatch conditions, across centroparietal, parietal and occipital scalp sites.

Table 5.3: Least-squared means and pairwise contrasts of match and mismatch experimental conditions. Mean amplitude (in μV) between 575-625 ms PSO

Region	Match: $\bar{x}$	Mismatch: $\bar{x}$	Estimate	DF	***t***	***p***
centroparietal	1.64 μV	2.94 μV	1.23 μV	1946	5.27	<.0001
parietal	1.36 μV	2.51 μV	1.15 μV	1946	6.35	<.0001
occipital	1.33 μV	2.20 μV	0.86 μV	1946	2.13	<.01

The effect has been linked with the elicitation of a P600 component as a response the presentation of mismatching stimuli. Although the P600 component is often associated with syntactic processing, it is also generally connected to rule violation and it is possible that, given the nature of the task where syllables needed to be combined together to form a single item, mismatch in the processing of a part of this item would result in a syntactic-like breach in the processing of a sound sequence (Tabullo et al., 2013). This pattern, found across the findings of all three experiments, is discussed in more detail in *Chapter 7* of this work. Because the focus of the study is on pre-N400 components, a general, in-depth discussion is only presented in the final stages of this work.

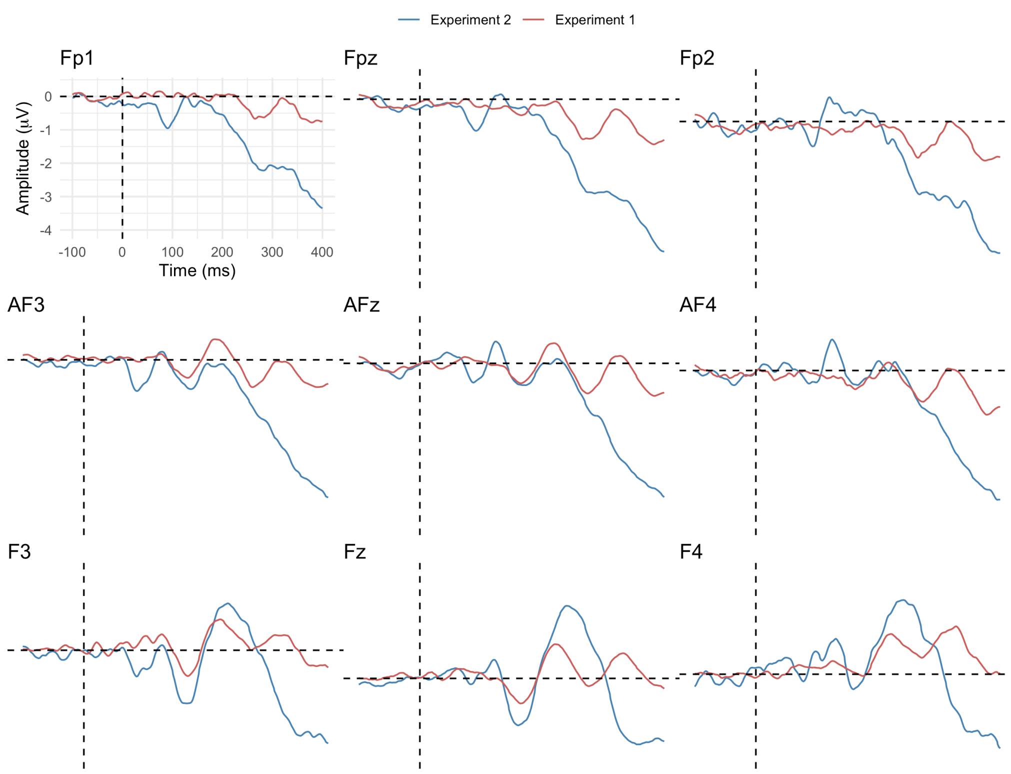
5.3.3 Experiment one vs two

Having seen the results from experiments one and two, let us compare the perception of mismatch stimuli across the two experiments. The scope of this section is provide a visual investigative comparison between the sets of mismatch responses in both experiments, considering they both share similar underlying methodological choices and stimuli. However, statistical differences were not tested considering that the two experiments are characterised by varying range of amplitude values in their responses as an effect of the type of task included and other experimental paradigm-related choices. However, the findings from both experiments and their implications are later compared and discussed in extensive detail in the general discussion Chapter of this research thesis.

The first difference that can be noticed between the results of the two experiments is linked to the overall amplitude range of the data in experiment two, compared to experiment one. Figures 5.12 and 5.13 highlight the responses to the presentation of mismatching stimuli only in both experiments. Because of the implementation of active behavioural tasks in the creation of the design of experiment two, the overall amplitude range of the ERP curves is bigger than data collected from passive listening tasks, which tend to generate smaller ERP curves (e.g. Astheimer and Sanders 2011). In figure 5.12, an N1 response can be observed for the experiment two condition around 100 ms, while this is not visible for the experiment one condition. Whether it is the case that the responses was not elicited in experiment one or whether effects were generally too small to surface, it cannot be determined. Generally, the N1 response is obtained regardless of attention to stimulus presentation and is often present specifically in the absence of direct attention.

The MMN-like effect observed across both conditions, while being generally and expectedly larger in the experiment two condition, is also characterised in that specific condition by a shift in latency of about +15 ms at its maximal peak (AFz). However, regardless of the difference, the peak is in the range of a common MMN appearance regardless of the difference in peak latency, which is probably due to the difference in task nature, passive in the first and active in the second experiment.

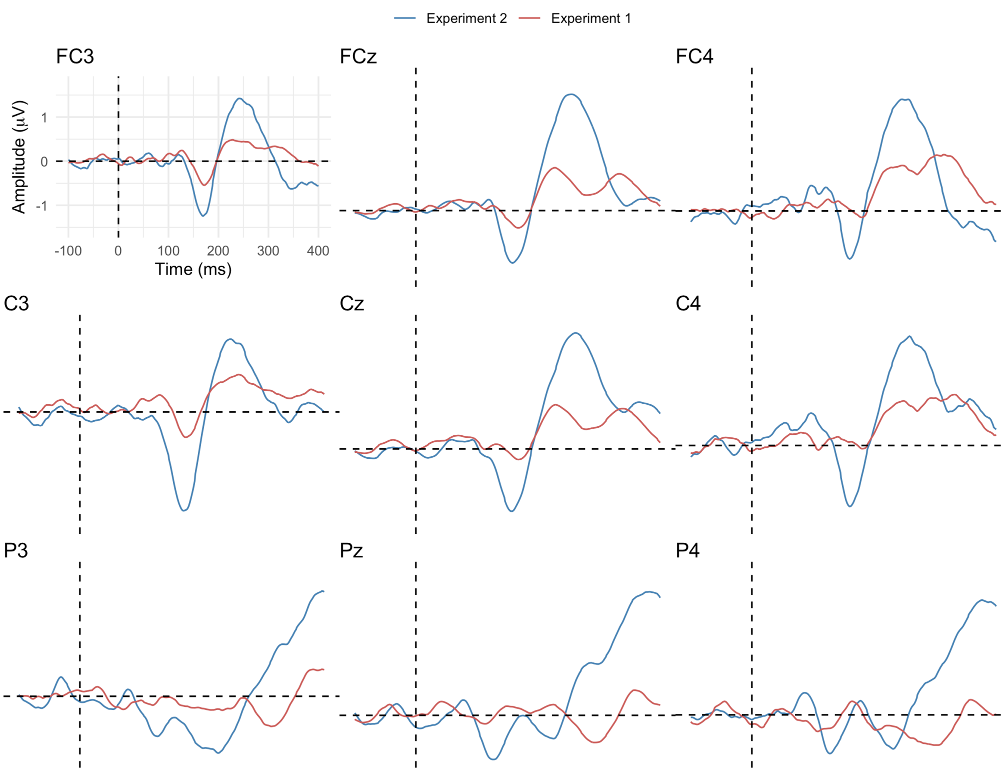
Figure 5.12: Grand-average ERP curves for mismatching stimuli in experiment two vs experiment one and AFA procedure results calculated between -100 and 500 ms PSO for prefrontal, anterior-frontal and frontal sites.



In central and parietal scalp sites (Figure 5.13), the main difference between the results of the two experiment lies between 200 and 400 ms post stimulus onset. The first positive peak in central scalp sites is much larger in amplitude for experiment two than experiment one, which is not unexpected considering amplitude differences between the two studies. An unusual finding, however, is that the following positive peak, between 350 and 400 ms appears to be very comparable across the two conditions. Because of the differences in amplitude, it would not be reliable to directly compare the results of the two experiments beyond a certain point.

Specifically, no inferential statistics were run on the difference between the two experiments, considering most of the differences in amplitude would probably be due to the general amplitude differences compared to specific, comparable components.

Figure 5.13: Grand-average ERP curves for mismatching stimuli in experiment two vs experiment one and AFA procedure results calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites.



5.4 Discussion

5.4.1 PMN

Contrary to expectations, experiment two did not elicit a clear-cut, significant PMN effect for the mismatch condition, which also means that *H1* has to be rejected. The findings of experiments two are consistent with those of experiment one, where similar ERP trends were observed but no clear PMN was found. However, as mentioned above, a negative-going deflection with a very small effect size was observed in central and parietal scalp sites over the midline and in the right hemisphere. The trend was focussed on electrodes such as Pz and P4 where some significant differences were reported between the two experimental conditions, although minimal both in terms of distribution in the time domain and effect size in the amplitude domain.

According to the data collected in experiment one, I proposed two possible explanation for the lack of observation of the PMN component. The main theory suggested that the PMN does not in fact represent phonological mismatch as a general cognitive process but that it might reflect some specific aspect of phonological mismatch, where the presence of lexical activation and retrieval is necessary. This was originally suggested by Connolly and Phillips 1994 who, adapting the framework of the Cohort model (Marslen-Wilson 1984), suggested that the PMN is elicited during the lexical retrieval phase, as a result of phonological mismatch. The other possible explanation to the lack of PMN elicitation in experiment one, was that because no active task was introduced in the experimental design, the PMN component was either too small to be detected, it was shadowed by bigger nearby components, such as the P3, or it was simply not elicited because it might require higher-level processing, not guaranteed in passive tasks where participants do not have to pay direct attention to stimuli. Experiment two drew from this second explanation to directly test the impact of behavioural tasks on the elicitation of the PMN component.

However, despite experiment two employing active tasks, no clear cut PMN component — especially when comparing the results of this study to previously existing findings in the literature — was elicited. If the absence of the PMN component in the previous experiment was due to smaller effects because of passive tasks or if the lack of a PMN component was due to its requirement of a device for attention towards the stimuli, the current experiment should have been able to cover both of those requirements thanks to its design. At the same time, there was the observation of the aforementioned negative-going deflection for the mismatch condition that was located, throughout a very limited topographical distribution, that could possibly provide some interesting insights and reflections on the nature of a phonological mismatch component. Here, I investigate some of the possible options in regards to the elicitation of what could be interpreted for a limited, very small PMN response.

5.4.2 Other findings

A series of trends and effects were observed when comparing mismatch and match conditions of experiment two. In particular, a prefrontal and anterior negative-going deflection was elicited maximal at around 100 ms for the mismatch condition, linked with the elicitation of one of the N1 frontal components. The negativity was followed by another negative-going polarity, peaking at 160 ms post stimulus onset in anterior and frontal scalp sites, associated with the observation of a mismatch negativity response to the presentation of mismatching stimuli. Following, positive deflections could be observed for both conditions, with a few sections of significant difference between the curves, especially between 250 ms in central and parietal scalp sites. These responses are most likely connected to P3 contamination and the presentation of oddball stimuli.

At 300 ms, the mismatch condition presents a negative-going trend followed by an overall small negative section in parietal scalp sites continuing until or just past the 400 ms mark depending on the specific scalp site, originally though to be linked to phonological mismatch. There is no way to safely determine that the effect is not linked to P3 contamination or to artefacts in the data, especially when considering that the overall extent of the difference was more limited when compared to other statistically significant effects across the data. Furthermore, this trend was only visible at two scalp sites and was not picked up by the topographical analysis and the implementation of the LMM. However, the negative-going deflection was not observed as a response to the perception of the three syllables in the syllable sequence presented in the first half of each trial. Because the syllables themselves also differed in presentation rate just as much as mismatching an matching stimuli did, P3 contamination was also expected (and it was, in fact, elicited) but phonological mismatch was not. The fact that the 300 ms negative-going deflection was not found in that particular instance could suggest that that very effect is possibly not linked to P3 contamination and elicitation but rather to either of the remaining options, namely phonological processing or artefacts in the data.

At the end of the previous chapter, we hypothesised that the lack of PMN component was either derived by the fact that either the PMN did not directly respond to phonological mismatch in all context as we previously thought it did, or that the passive methodology of the experiment was not able to elicit such component because of requirements of temporal attention during stimulus presentation. The current discussion portion of this chapter aims at contextualising the current results with previous and existing hypotheses on the elicitation of the PMN as well as that of other components.

5.4.3 N1, MMN and P3

Effects like the N1, MMN(-like) and P3 or similar trends have been observed throughout the data set and amplitude was generally greater for the presentation of mismatch stimuli, compared to that of match stimuli. Some of these effects were also visible when investigating differences between the perception of the first, second and third syllable in a syllable sequence. These results, together with the findings presented in the previous chapter of this work, provide further evidence that, in this particular case as well, participants recognised mismatch stimuli as such, since mismatch and oddball components have been elicited by stimuli presented in the mismatch condition. When it comes to components such as the N1, the reason why it was not observed in experiment one but it an N1-like response was observed in experiment two (although weak) was probably connected to an overall low signal-to-noise ratio in the data set of the first experiment of this research chapter, as clearly showed in the findings presented in figures 5.12 and 5.13.

Comparing results from both experiments, especially when it comes to mismatch pairs and mismatch stimuli, highlighted very similar response patterns on both sides, stressing how the use of similar methodologies and stimuli — with the main difference between the experiments lying mainly in the absence of active task demands in experiment one — can provide an extensive ground for comparison. According to existing literature, the observation of all the ERP components mentioned above should not be considered unexpected or particularly insightful in either experiment one nor experiment two. However, being able to replicate some of the findings of papers and experiments in the literature provides an extra layer of evidence in favour of the choice of experimental paradigm. it can also serve as methodological validation, informing us that experiment two managed to elicit findings that were expected because of previous research when it comes to components such as the N1, MMN and P3, components that have been studied to a much larger extent than smaller, often less reliable components such as the phonological mismatch negativity.

P3 contamination and latency differences

The most likely explanation to the observation of a negative-going polarity between 300 and 400 ms, maximal at 300 ms at parietal scalp sites Pz and P4 is that the difference was caused by the contamination of P3 and odball response elicitations, both in the amplitude and latency dimensions. For instance, this is clearest at P4 where, the mismatch condition presents a delayed negative-going peak in the 200-250 ms range, compared to the match condition, consequently presenting a following negative-going difference between the two conditions, possibly caused by the latency difference at the previous peak. However, the opposite is also an option, that the shift in peak latency was created by the presence of mismatch and that the negativity and latency difference is due to the mismatch response in the first place.

As investigated in the results section of the current chapter, this negative-going polarity difference is not present when observing responses to the presentation of the first and second syllable of the syllable sequence presented in the first half of each trial. There were differences in stimulus presentation rates among those stimuli that should have prompted similar P3 response and contamination to that of the mismatch-match conditions. These findings suggest that the differences observed between match-mismatch conditions were not prompted by P3 contamination but rather by a genuine phonological mismatch response. However, it is also important to take into account that, while there are similarities in presentation-rate differences when comparing first-second syllables and match-mismatch conditions, there are also differences in the type of task that accompanies the presentation of each type of stimulus. For instance, when syllables were presented, participants were only instructed to listen and remember the sequence. However, in the presentation of either matching and mismatching stimuli, subjects were told to check that the matching stimulus corresponded to the memorised version created by concatenating previous syllables. The effect of task difference on the elicitation of ERP components is unpredictable and, for this reason, it is impossible to comfortably determine where this difference was caused by one or the other factor.

Phonological mismatch

Let's, for one moment, set aside the possibility that the component observed maximal at 300 ms post stimulus onset in parietal electrodes, followed by negative polarity reaching the 400 ms mark in the right hemisphere, might be caused by P3 contamination. In this section, I want to discuss the eventuality that this negative-going trend is in fact connected to mismatch in stimulus presentation caused by phonological differences. Specifically by comparing the elicited component to previous literature on the PMN, which is the component that should have been observed in this specific context given the methodology used for the creation of the experiment and the overall distribution of the component in the time domain.

First of all, the negative component was observed mainly in central and parietal midline scalp sites, maximal at 300 ms post stimulus onset. While the PMN has been observed as distributed across the midline (Connolly et al., 1992), its midline distribution was however not limited to central and parietal scalp sites only. Other studies that did manage to elicit the PMN in central and parietal scalp sites, on the other hand, were also able to locate it in both hemispheres and with a flat distribution (Van Den Brink et al. 2001; D'Arcy et al. 2004). The effect observed in experiment two shares the most similarities with the PMN observed by Desroches et al in their paper about processing of phonological similarities (Desroches et al., 2009). In that specific paper, the PMN was found, accompanied by the N400 effect, in midline and right central and parietal regions of the scalp, similarly to the findings of experiment two around the 300 ms mark. However, in the paper by Desroches et al. (2009) the PMN was also somewhat visible across frontal scalp sites, which is not the case in the findings of experiment two. Moreover, the overall size of the difference of the PMN as reported in the literature is much more prominent than the weak effect observed in the mismatch-match conditions of experiment two of this research project.

Another point that should be made to further speculate on the elicitation of this PMN-like negativity is that no PMN was elicited in the previous literature in contexts that presented no lexical retrieval tasks or related cognitive processes[3]. For this reason, every time the PMN was elicited in the past, it was elicited in a context where confounding phonological and lexical processes were both employed in the process of speech perception and recognition. It could be determined that the smaller effect and limited topographical distribution of the observed component might be linked to the fact that it was elicited in a context where only phonological mismatch was present, but without the inclusion of lexical processing. Furthermore, previous literature refers to phonological processing and lexical processing as direct, generalised, one dimensional processes in speech perception. The complex nature of these series of processes and the many different levels of mismatch that are available whenever processing words and speech might, to begin with, be the cause of many inconsistent results that have been found when studying the PMN. This is true mainly because, what is thought as one response to phonological mapping, could consist of many, smaller, different responses to more specific aspects of phonological and lexical mapping that were erroneously linked to one single ERP component.

[3]Some experiments did elicit the PMN through the perception of nonce words (e.g. Newman et al. 2009). However, the nonce words were embedded in lexical retrieval tasks that would have caused a lexical retrieval error response in the case of mismatching meaning, regardless of the nature of the (nonce) word.

Some similarities, especially in the time domain, are present between the observed component and the PMN as described in the literature. However, previous literature and the extent to which the component was observed in experiment two, does not warrant enough evidence to be able to link the effect / trend discovered in experiment two to the elicitation of an actual PMN component.

5.4.4 Experiment-level discussion

The findings of experiment two suggest that, even accounting for the presence of active behavioural tasks, no clear PMN component could be observed. The absence of the PMN is thought to be linked to the absence of lexical selection and semantic activity in processes of speech(-like) perception, which remains the one main difference between experiments one and two and prior research. The fact that, even while changing the design to include active tasks, no distinctive PMN component was found could suggest that the PMN component might reflect some level of phonological mapping but to a smaller extent and in specific contexts only, specifically in contexts that also employ semantic mapping and processing as displayed in previous research. It is possible that the PMN does not directly or solely represent phonological mismatch as a whole, but that it is representative of a either some smaller level of phonological mapping or of the interaction between phonological mapping and lexical retrieval.

In the overview to experiment one, I discussed how experiments like Darcy (2004) and Connolly and Phillips (1994) employed, in their design, syntactic context and lexical likelihood to create phonological mismatch in sentence-final spoken-word perception contexts. To specify, the reason why participants expect 'tune' at the end of the sentence '*the piano is out of tune*' is primarily because of its meaning. Whenever a mismatching word is used instead of the highly likely option, a word such as 'pineapple', phonological mismatch is observed but, arguably, it is in its own right driven by mismatch that originates from lexical and syntactic contexts. In the experiment by Kujala et al. (2004) or in that by Newman et al. (2003), syntactic context was eliminated from the pool of existing variables by creating a different methodology. In Kujala et al. (2004), participants were presented a real word on the screen (e.g *cap*), followed by a letter (e.g. *t*). Subjects were then asked to substitute the first letter of the word with the letter presented and they would be later be presented the resulting word but, similarly to the methodology of this experiment, they were sometimes presented mismatching words — or nonce words — instead, causing PMN. However, while syntactic activation was excluded from the experimental design, some form of lexical activation and retrieval was still present throughout the behavioural task.

While a response different from the N400 to the presentation of *map* instead of *tap*, for instance, could certainly be categorised as a result of phonological mismatch, this phonological mismatch could also, once again, be originally created by semantic context and processes of lexical retrieval. The target stimulus is, without a doubt in this case, phonologically different from the expected, highly probable, word. However, this difference is caused by lexical retrieval expectations determined by semantic contexts and expectations. The same is true if the stimulus presented is a nonce word (e.g. *xap*). Certainly, phonological mismatch is present from the missed presentation of the high-likelihood stimulus, but there are also semantic and lexical levels of mismatch (i.e. not having a meaning at all is certainly a different representation of *meaning* than that of the original stimulus).

When analysing components such as the N400, connected to lexical processing and mismatch in language processing, if we were to remove language entirely (i.e. using images instead, removing all other features that are connected to language besides meaning activation and retrieval) an N400 component, which has been found to be elicited with an array of different sensorial information (e.g. Invitto et al. 2018), is still elicited with the presentation of images containing mismatching or semantically unrelated subjects. Similarly, components such as the P600 are found when syntactic-like rules are floated in contexts such as reading music (Patel et al., 1998) and arithmetic tasks (Núñez-Peña and Honrubia-Serrano, 2004). This evidence suggests that the N400 and P600 are related to some sort of processing type — namely semantic for the N400 and syntactic for P600 — that is not language specific. If we were to, instead, remove lexical activation no PMN is observed with only phonological mismatch present, which could suggest that the PMN component is linked more closely to lexical retrieval than it is to phonological mismatch, almost suggesting that phonological mismatch is not entirely *phonological* unless lexical retrieval and/or semantic mapping are also present. This could, in its own right, support the idea that sounds and speech sounds are treated the same way during processing unless lexical information is also added, separating then processing such as sound perception to speech perception and spoken word recognition.

Finally, even in the situation that the methodology of these two experiments was not able to, in some way, elicit the PMN component — which at this point after the results of two experiment being consistent is a very small chance — we should not consider these results a failure. We should, instead, consider the current findings as further proof that such small, understudied components, hard to elicit and inconsistent across methodologies, perhaps are not a good enough marker by which to measure the extent of other, more general processes when it comes to speech perception, phonological mapping and spoken-word recognition. In the next chapter, experiment three deals with sound perception, where no linguistic information is at all presented in the acoustic signal, to provide a comparable baseline for experiment two. Experiment three, whose findings are relevant to this work on their own but even more relevant when analysed together with those of experiment two, focus on the differences between sound perception and processing with and without phonological processing. This is done with the aim of investigating what responses and to what extent incorporate phonological processing and mismatch, in a context where no PMN component was elicited.

Chapter 6

Experiment 3

6.1 Introduction

The distinct function of the PMN component, often hypothesised to be located somewhere on a spectrum between pre-lexical (acoustic, phonological) and semantic (lexical) mapping in speech perception and spoken-word recognition, is still under evaluation. This is true especially considering the limited existing literature published on this specific topic, that is also characterised by results that are not consistent when it comes to the component's topographical distribution and role in speech perception. Experiment one and two aimed at separating the PMN from the N400 component — or phonological and semantic processing — by presenting subjects with semantics- and syntax-free mismatching speech sounds and sound sequences. The main goal of the two experiments was to investigate whether lexical activation was a necessary process for the elicitation of the PMN component or whether phonological mismatch was sufficient where expectations were dictated by linguistically-external factors, such as the experimental paradigm itself.

Experiment three aims to replicate the methodology of experiment two while substituting each syllable with a sine-wave tone, removing any linguistic information from the sound signal. By doing this, the goal is to highlight ERP responses that are specific to sound processing, in experiment three. Hopefully, by comparing results to the findings of experiment two, the combination of the two experiments should help isolate specific components that are particularly sensitive to phonological processing and mapping. Experiments two and three work very well together for the realisation of this purpose because, in experiment two, stimuli are both auditory and linguistic. For this reason. responses to stimulus presentation might be both in response to auditory / acoustic properties of the stimulus and linguistic ones. In experiment three, on the other hand, stimuli are auditory but not linguistic, which means that responses are primarily to acoustic properties of stimuli and, every response found in experiment two but not in experiment three could be isolated as a linguistically-specific ERP component.

6.1.1 PMN and MMN: phonological processing

It appears that the PMN shares similarities with the N400 component and, as it has been observed in the past (e.g. Connolly and Phillips 1994), the PMN is also sensitive to semantic processing to a smaller extent. At the same time, increasing number of similarities have been drawn between the PMN and the mismatch negativity (MMN) component. These similarities have surfaced thanks to previous research aimed at discovering different nuances of the MMN component, specifically in relation to its response to linguistic information. The MMN, originally primarily thought to be a response to deviant stimuli (Näätänen et al. 1993; Näätänen and Alho 1995) has been discovered to be sensitive to linguistic information, including syntactic (Shtyrov et al., 2003) and phonological (Pulvermüller, 2001) processing. The aim of experiment three is to build on the foundation laid by experiment two and one and in order to explore the differences between speech-like perception (carried out in experiment two) and sound perception, central focus of experiment three.

By removing speech sounds and presenting participants with simple sine-wave tones, while keeping the methodologies between experiments two and three unchanged, experiment three aims at isolating ERP responses to non-speech sound perception in an experimental environment that is as comparable as possible to that of experiment two. The aim is to, at a later stage, be able to compare responses to the presentation of speech sounds to those following the presentation of (non-speech) sine-wave tones. The comparison allows access to insights on phonologically-specific responses in experiment two that, because of the nature of the experimental stimuli, should not be present among the components of experiment three. However, by analysing the results of experiment two alone, it is not possible to determine with certainty whether a pre-N400 effect is specifically in response to phonological processing, sound and auditory processing or stimulus categorisation.

For instance, no PMN component was elicited in experiment two. However, a negative-going polarity was observed for the mismatch condition and for the presentation of the first syllable in a sequence with a peak between 150 and 200 ms, a time window in the range of the MMN effect, usually ranging from 150 to 250 ms post stimulus onset (Näätänen et al. 1993; Näätänen and Alho 1995). At a channel level, the effect was only found to be statistically significant in comparing the responses of the first syllable to that of the second syllable in the sequence, but not when comparing mismatch to match condition. While this could suggest that the effect is a response to stimulus categorisation more than it is to phonological mismatch, the state of significance could also be explained by the increased number of trials (and, therefore, less noise) in the first and second syllable sequence condition. Similar results were observed among the findings of experiment one, where no behavioural tasks were present. The MMN component is often observed whenever a mismatching, deviant (auditory) stimulus is presented interrupting a sequence of standard, non deviant stimuli. In particular, the MMN has been mostly linked to changes in frequency or duration of the stimulus, which, in linguistic contexts could be exemplified by presenting a long vowel in a series of short vowels or an /s/ sound in a sequence of /d/ sounds (Näätänen 1991; Näätänen et al. 1993; Näätänen and Alho 1995; Erlbeck et al. 2014).

In the literature, the MMN was primarily regarded as a low-level response to deviant (sound) stimuli, with no particular link to the processing of linguistic information (Näätänen and Alho, 1995). However, previous research on the mismatch negativity component linked the MMN to aspects of linguistic processing, at levels such as syntactic (Shtyrov et al., 2003), phonological and acoustic (Pulvermüller 2001; Weber et al. 2004). For example, Pulvermüller et al. (2001) presented subjects with a spoken syllable either completing a Finnish word or a pseudo word and measured MMN responses, while asking participants to ignore stimulus presentation[1]. The findings displayed a clear MMN effect around the 150 ms mark whenever a Finnish word was presented, compared to the presentation of a pseudo word. A control group of subjects who did not speak Finnish showed no MMN effect for either condition.

Although this particular example is not directly related to the findings of this work, it suggests that the MMN is also sensitive to processing in phonological — and linguistic — mapping rather than simply being a marker of acoustic or auditory mismatch. This sensitivity to phonetic and phonological mapping of pre-N400 components such as the MMN is likely to have

[1] Often, the only cited difference between the PMN and the MMN is that the PMN was never elicited in the absence of focussed attention to stimulus presentation.

been observed among the findings of experiments one and two. No PMN was observed in either experiment. However, phonological mismatch could have been incorporated in the observation of an early mismatch and categorisation component, such as the N1 and MMN. If the MMN component — or any other component or response observed in experiment two — were responding directly to phonological processing, it should be different in experiment three than previously found in earlier experiments — considering only non-linguistic information is provided and sine-wave tones are presented.

To summarise, the PMN has been found to share similar behaviours to the N400 and to the MMN, including sharing topographical distribution[2] with both responses. Furthermore, the MMN has been found in the past to be sensitive to both phonological and syntactic levels of processing in speech perception. If this theory was further supported by more experimental evidence, the MMN and PMN would both likely share a (mainly) anterior-frontal distribution, a peak between the 150 and the 300 ms mark and a sensitivity to phonological processing, mapping and mismatch. The main difference between the two components is that, experimentally, the PMN was never elicited without focussed attention to stimulus presentation, such as in passive listening tasks. This was confirmed to be the case for the MMN many times over (e.g. Näätänen et al. 1993; Näätänen and Alho 1995). Until the realisation of experiment one of this research project, no PMN was ever elicited in passive experiments because no experiment ever tested this specific hypothesis. With the advancement of the field and the realisation of more research it is possible that the similarities between the two components will be even more striking the more evidence is collected.

In addition, it is possible that the PMN was already elicited in passive listening tasks as part of experimental paradigms that were focussed on testing other hypothesis, but that the component was mistaken for many of the neighbouring and similarly activated responses of language processing. More specifically, because the MMN, PMN, N2 and other ERP components (e.g. ELAN) all share a similar topographical and temporal distributions and, if we consider the fact that the PMN is possibly the least known of all the aforementioned components[3], it is not unlikely that particular instance of MMN, N2 or ELAN being elicited in previous language experiments were, in fact, a PMN in disguise. The reason why no PMN was observed in passive-

[2] Chapter 2 describes examples where the PMN has been observed in frontal, central, parietal and posterior regions of the scalp across several experiments.

[3] the PMN was never listed in any editions of Luck and Kapperman's *Handbook of Event-Related Potential components* (Luck and Kappenman, 2011), despite many studies on the component being available on the topic by the time the first edition of the book was released.

listening tasks is mainly linked to the fact that no research ever looked into the elicitation of the PMN in passive listening tasks.

Based on both previous literature and the findings of the two previous experiments carried out in this research project, the main hypothesis of experiment three focusses on the elicitation of pre-PMN components. If any of the ERP responses observed in experiments one and two were primarily sensitive to phonological mapping — rather than mainly responding to auditory mismatch and stimulus categorisation — the amplitude of those effects should be reduced in experiment three compared to previous experiments. For instance, a reduced MMN or N1 component is not described as such by the overall amplitude of the negative-going peak, but by the difference between the peak observed in the mismatch and the match condition of the experiments. While overall amplitude responses can vary across experiments for a variety of reasons, the important feature is the relative difference between the curves of the conditions analysed in the experiment. However, if it was the case that the MMN component was mostly a response to auditory mismatch rather than to phonological mapping, the effect between match and mismatch conditions of experiment three should be either similar or more pronounced than the effect observed between match and mismatch condition of previous experiments.

The focus of this experiment is on the MMN and pre-300 ms components, as their similarities with the PMN appear to be more striking the more evidence is collected on the two components. However, there are a handful of ERP components that are usually observed both during speech and sound processing. One of these components is the N1 which, because it reacts to external stimuli, does not differentiate between linguistic and non linguistic information. However, while it has been proven in the past the the N1 can be affected by external factors (Curio et al., 2000), it is most likely too low-level of a response to be directly linked to phonological mapping and mismatch. For this reason, no particular N1 difference is expected between the two experiments.

6.1.2 Sound, speech and music

In order to create sound sequences containing different combinations of sounds while, at the same time, removing linguistic information from speech sounds, sine-wave tones varying in pitch were used as stimuli for experiment three. Section **6.2** goes into detail in explaining what stimuli were used and how they were created. However, one of the issues to address when presenting the methodology of the current experiment is that, when employing sine-wave tones (or any combination of tones) as stimuli in place of syllables, the experiment is now tackling aspects of cognition that are common of both language and musical processing. Two or more tones presented one after the other and varying in pitch may very well be perceived as a musical tune. Let us first consider the similarities and differences in speech and music perception. While speech sounds and musical notes carry very different representations and amounts of information, there are several ERP components that are often observed when both processing speech and processing music, in particular when it comes to syntactic violations and the P600 component.

The P600 ERP component (see *Appendix A* for more information) is often associated in language studies with processing and mismatch in the syntactic realm (Kaan et al., 2000). However, studies in both language and musical processing have found that the P600 may reflect a general — non-language specific — type of response to rule violations. For instance, a P600 component is elicited whenever a note that does not belong in a particular scale, or key, is perceived in a musical context (Patel et al. 1998; Schön et al. 2004). Because the P600 reflects a sensory non-specific type of rule (syntactic in language) processing, it is also not limited to just music and language. In a recent study, the P600 component was linked to rule violation in arithmetic tasks (Núñez-Peña and Honrubia-Serrano, 2004).

6.1.3 ERAN

The Early Right-Anterior Negativity (ERAN) is often associated with harmonically inappropriate chords and notes (Koelsch et al., 2001) and the processing of music-syntactic information (Koelsch, 2009) in the perception of music (Koelsch et al., 2001). The reason why the component is mentioned here is because it can be sometimes confused with the MMN component, especially since both of the two ERP responses peak at and around 200 ms post stimulus onset and that their distribution is mainly anterior[4]. Because of the similarities between the two ERP responses, I have provided a brief explanation to why, despite the similar distribution to the MMN and its connection with *inappropriate* — which could also be seen as mismatching — notes, differences between the two responses should be clear enough to be able to distinguish whether an elicited component resembles the ERAN more than it does the MMN in this specific experimental paradigm.

[4]Early Left-Anterior Negativity (ELAN), on the other hand, is usually associated with early rule violation in linguistically informed stimuli, for instance *The pizza is the in oven* (Friederici and Kotz, 2003).

First of all, the ERAN component is usually elicited by music-syntactic processing in mismatch in relation to long term memory (Koelsch, 2009). The stimuli that have been used in experiment three are sine-wave tone sequences that are learnt by participants, mainly through repetition, during the experimental procedure. Furthermore, no tone sequence resembles a specific musical tune and, considering there are only three tones in each sequence, it is unlikely that long-term memory plays a role in the processing of syntactic information in experiment three, which would result in an ERAN component possibly contaminating an MMN response. Finally, Koelsch noticed how the ERAN component usually presents a delayed peak when compared to both frequency MMN and *abstract features* MMN which could also be helpful in determining whether a response observed in the 150-250 ms range is related to one or the other component.

6.1.4 Stimulus Complexity

A factor that might play a confounding role in the analysis of experiment three data, specifically when directly compared them to experiment two data, is that stimuli in experiment three are, besides not being linguistically informative, less complex than those in experiment two. To be more specific, sine wave tones are simple acoustic sine waves and they, arguably, carry much less information than speech syllables. The first complexity layer difference is of course determined by the absence of linguistic information embedded in experiment three stimuli and, as far as that difference is concerned, this level of stimulus complexity difference between the two experiments is accounted for as it reflects the aim of the study. More specifically, stimuli of experiment three were, on purpose, stripped of speech characteristics to eliminate any phonological or phonemic layer of processing. However, it could also be argued that, because syllables in experiment two are composed of multiple phonemes, which, in themselves, are inherently more complex acoustic and auditory stimuli than simple sine-wave tones of fixed duration and frequency, these stimuli are by definition also more complex than the stimuli used in experiment three, regardless of their linguistic status.

Stimulus complexity has been proven, in the past, to play a role in the mediation of the elicitation of ERP components, with responses to simple stimuli being different than those to more complex stimuli of the same medium. For instance, research on auditory stimuli suggested that the perception of complex sound probes elicited, in participants, processes linked to cognitive workload while simple sounds did not (Dyke et al., 2015). More specifically, the presence of cognitive workload was interpreted as a robust orienting response from the presentation of complex stimuli. The orienting response was indexed by a stronger P3a response, not observed among the components elicited by simple, monotonic stimuli (Dyke et al., 2015). A study by Behroozmand et al. (2011) also showed how, through the realisation of a passive listening task, where complex voice and non-voice stimuli were both manipulated with the introduction of pitch-shifted sections, neural response was larger to the presentation of pitch-shifted sections of the stimulus for both the voice and non voice conditions. However, what is most relevant to our study, is that, while no difference was found between complex voice and non-voice stimuli, the neural response was generally greater as a response to voice and non-voice stimuli than it was to simple tones (Behroozmand et al., 2011).

Furthermore, research on complex stimuli was also carried out in the realm of vision studies with ERP (e.g. Verbaten et al. 1986; ...; Luck 2012), which suggests that the differences between the perception of complex and simple stimuli are not limited to the presentation of auditory stimuli. Unfortunately, although some evidence has been collected that suggests that complexity of the stimulus might play a role in the level of activation of cognitive workload and neural responses — regardless of the amount of linguistic information carried by the stimulus — not enough evidence is present to be able to precisely predict the extent to which this variable is going to affect the findings of the study, when comparing the data indirectly to the findings of experiment two. For this reason, if there appear to be among the results of experiment three signs that indicate that stimulus complexity has played a specific role in the type or amount of potential response difference between the two experiments, this is addressed in due course in the chapter.

6.1.5 Expected effects

The ERP component focus of experiment three is the mismatch negativity component, a negative-going deflection often observed between 100 and 200 ms post stimulus onset, with peaks at around 200 ms in anterior and frontal regions of the scalp. No PMN is necessarily expected in experiment three. However, the 18 electrodes that were selected for channel-level significance testing also included most of the scalp sites most commonly associated with the topographical distribution of the PMN. Furthermore, components like the N1, a response to attentional shifts to stimulus presentation and the P3, sensitive to stimulus presentation rate and cognitive workload activation, are to be expected. A P600 component is expected to some extent, since breaking expectations connected to learnt musical sequences could cause responses related to music-syntactic rule violation, to which the P600 is sensitive to in this particular context.

6.2 Methods

The methodologies of experiment two and three are identical, except for the nature of the stimuli. The current method section deals in more detail with the main differences between the two designs: namely the set of stimuli used in the realisation of the experiment. Just as in experiment two, a sequence of three sounds was presented to participants with a 0.5 seconds pause between the first and second and second and third items of the sequence. Differently from experiment two, however, each of the sounds was a sine-wave tone (in place of a syllable) varying in pitch. After the presentation of the third sine-wave tone in the sequence, participants were instructed to mentally concatenate the three tones in a single tune, without altering the order in which the tones were presented. After a four second pause in which participants completed the behavioural task, subjects were auditorily presented with the tune combining the three sine-wave tones together. However, in 33% of trials, the first item of each tune was substituted with a deviant sine-wave tone with varying pitch. Participants were simply instructed to listen to the tune and to mentally check whether the sine-wave tones in the tune matched the three tones they had previously heard in the sequence, played once at the beginning of every trial.

6.2.1 Stimuli

Twelve sine-wave tone sequences were created using a set of synthesised musical notes, spanning the range of three octaves (from octave two to octave four). While each note was originally of 1000 ms in length, tones used in each of the sequences were shortened to 200 ms each using Praat. In addition, a fade-in and fade-out function were added at 0.01 and at 199.9 ms respectively to smooth the abrupt beginnings and ends of sine-wave tones, which are known to cause clipping in speakers. The fade-in and fade-out functions were applied using the Vocal Toolkit (Corretge, 2012) in Praat.

Tones for the creation of the *matching* stimulus sequences were pseudo-randomly generated with a set of constraints applied to the generation of sine-wave tone triplets, at the base of the creation of each sequence[5]. Following, a list of the constraints used in the pseudo-randomised creation of stimuli.

[5] Experimental stimuli can be downloaded at the following GitHub repository: **mcanzi/ERP_tones**.

- All tones in matching sequences followed tonality and were all selected by a C major / A minor scale. Sequences from the scale were used to avoid tonality violations, which could consequently result in the elicitation of ERP components contaminating experimental data, since early components such as the ERAN tend to be elicited by harmonically deviant musical notes (Koelsch et al., 2001).
- In a given sequence, no two tones were present that were less than one tone apart from one another. This was devised so that no sequences would contain tones that either sounded too similar to each other or that were too dissonant[6]. The goal was to have fairly neutral base sequences and that no particular sequence would stand out either because of dissonance or because of two or more tones being particularly similar to one another.
- In a given sequence, not more than one octave difference between the highest-frequency and lowest-frequency tones was allowed. This constraint was included so that, similarly to the aim for adding the first constraint, no particular sequence stood out from the rest for containing a relative distance between tones that would either be too high frequency (or too low frequency), considering components such as the MMN, for instance, do modulate depending on the frequency difference in deviant stimuli. This constraint was also applied to the creation of mismatching stimuli.
- Across all four base sequences (and the remaining eight mismatching sequences) no two combinations of the same two tones were allowed in order to avoid similarities between different sequences that would cause issues when calculating the probability of a certain sound to be played in a given context. Similarly, all twelve sequences were created so that they all started with different tones from one another, again, to avoid repetition across some of the sequences.

Figure 6.1 displays the four base tone-sequence combinations that were generated pseudo randomly with the application of the constraints described above[7]. The boxes on the left show the combination of three tones that form the base sequence. On the right of the figure, the upper box connected to each sequence represents the matching sequence — by definition the

[6]Musical notes that are half a step of distance on the harmonic scale are usually very dissonant when played next to each other.

[7]Experiment three was devised so that half of the mismatching tunes would include a note that violates tonality constraints. This was done to be able to latter compare responses to mismatching stimuli not violating tonality and those who were, in order to observe differences in P600. Because of the small amount of data collected and, considering how this feature was not paramount for the main goal of the study, all mismatching data was analysed as a whole.

exact same sequence as the base tune — while, the box below, displays both manipulations of the matching sequence with mismatching tones highlighted in bold. The naming convention used in the figure reflects the international convention for musical notes where the name of each note (A to G) is accompanied by the number representing the octave in which the note is placed (in this case, only octave 2 through 4 were used). The # symbol is conventionally read as *sharp* in musical theory.

Figure 6.1: Sine-wave tone stimuli sequences and manipulated combinations with name of musical note (A to G) and octave (2 to 4)

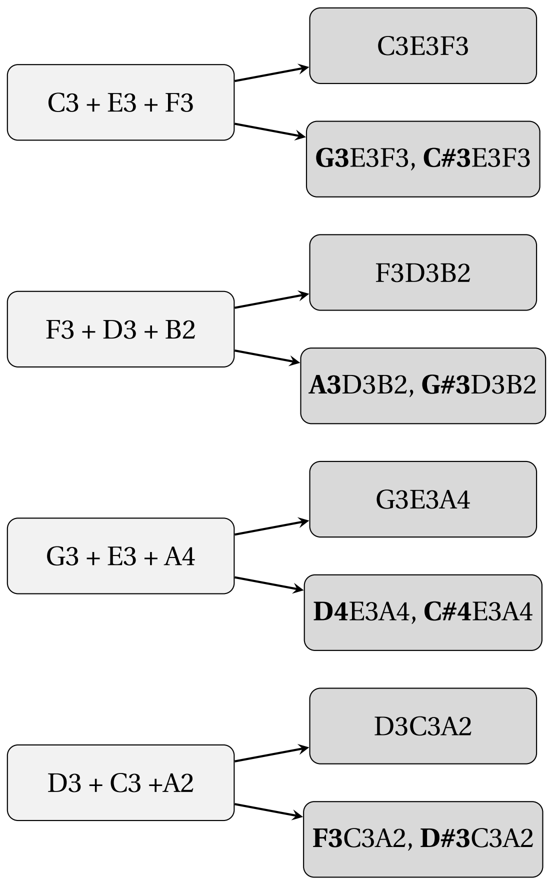
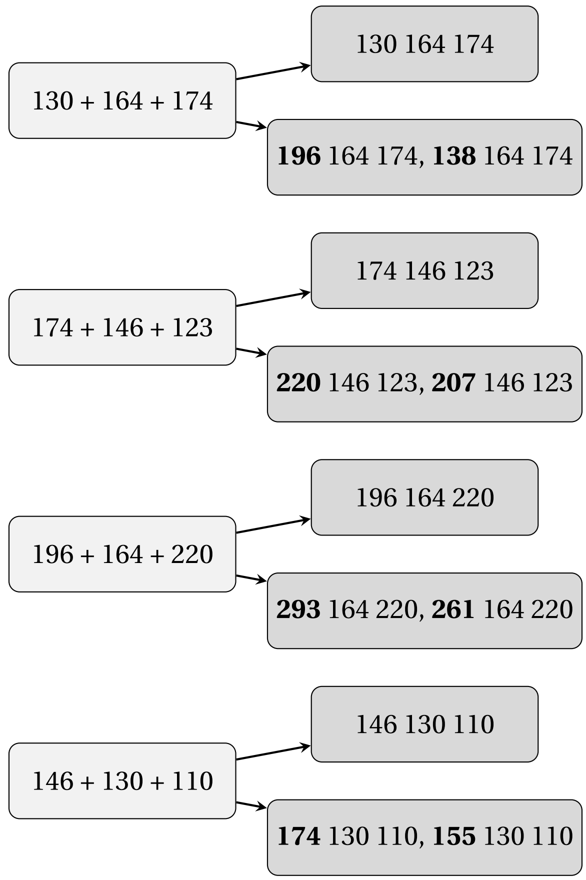


Figure 6.2 below presents the same stimuli that were displayed in Figure 6.1. However, this time, instead of using musical naming conventions, only the frequency of each tone is presented in Hz, providing a clearer picture of acoustics differences between tones in each sequence and, in more specific terms, differences between matching and mismatching pairs.

Figure 6.2: Sine-wave tone stimuli sequences and manipulated combinations in Hz



6.2.2 Participants

12 (F = 8)[8] undergraduate students at The University of Manchester took part in experiment three. The age range of participants was recorded between 19 and 25 years of age (M = 21). All 12 participants were right-handed healthy adults who reported normal hearing and normal or corrected to normal vision. 9 participants reported English as their native language, while the other four reported Castilian Spanish (N = 2) and European French (N = 2) as their native language.

Participants were native speakers of different languages but none of the language was a tonal language. In tonal languages, word meanings are affected by tone and minimal pairs exist where the two items are segmentally the same and the only differences lies on the tone. Avoid-

[8]Data collection of experiment three was interrupted in February/March 2020 by the COVID-19 global pandemic. For this reason, only data from 12 participants — out of the 20 participants scheduled for the experiment — were collected. Because of the the study being underpowered, data used for the statistical analyses were overall less reliable and smaller effects were more difficult to observe.

ing speakers of languages that employ lexical tones was an added criterion since it has been demonstrated that speakers of tonal languages have increased ability when it comes to pitch discrimination in both language and music (Stevens et al., 2013). For similar reasons, none of the participants that were selected for the experiment had any extensive musical training. This was done in order to avoid confounding variables and biases in pitch discrimination abilities caused by musical training. All participants were compensated with č20 in cash upon completion of the experimental procedure. The data collected from all 12 participants were analysed and included in the final analysis, since no data set contained more than 10% invalid or discarded data points.

6.2.3 Procedure

The procedure for the realisation of experiment three followed the exact experimental protocol for experiment two. At the beginning of each experimental session, all participant were presented with a Participant Information Sheet (PIS), containing a general description of the experiment and details about the experimental procedure. Participants were also given the chance to ask questions or express any doubt that they had about the experiment before signing their consent on a provided Consent Form (CF). After taking consent, all 64 electrodes were applied to the participants' scalp following the procedure highlighted in Chapter 2. After all electrodes had been placed on the subject's scalp, participants were asked to sit in the experimental booth where the entirety of the experiment took place.

The whole data collection procedure was composed of one single experimental session, of an average length of one hour. 400 trials were presented, divided in 20 blocks of 20 trials per block. Base tone sequences were presented in full-random order and, whether mismatching or matching sequences were paired with the presentation of base sequences was also decided randomly. Mismatching sequences were programmed to, on average, be presented only 33% of trials. 30-second rest blocks were included in the stimulus presentation program and they were placed between each trial block. At the end of each experimental session, participants were debriefed and compensated for their participation to the experiment. Hot beverages, including beverages containing caffeine, were offered to participants during rest blocks.

6.2.4 Statistical analysis

Significant effects were measured at the channel level using the AFA method (Sheu et al., 2016) for multivariate testing of time-series data between multiple conditions across a set of 18 electrodes that were chosen before the collection of data of the first experiment of this research project. Electrodes are distributed across a variety of scalp regions that are often associated with components of interest to this study, including N1, MMN, PMN, N400 and P600. Additionally, topographical analyses were carried out to identify the origin of the observed components. Linear mixed effect regression models were run on average amplitude measured across specific time windows, to determine statistical differences across different regions of the scalp.

6.3 Results

In this section, grand-average ERP data, difference curves, AFA procedure results, topographical maps and LMM results are presented for all experimental conditions and several comparisons with data from experiment two are drawn.

6.3.1 Mismatch vs Match

Figures 6.3 and 6.4 present an overview of the ERP responses for the match and mismatch conditions of experiment three, with a particular focus on the time window between 0 and 400 ms, where most early sound mismatch and processing ERP effects are expected. Similarly to all ERP grand-average figures included so far in this work, a total of 18 electrodes have been selected as sample for the statistical analyses, since they are considered representative of the scalp regions where most ERP effects elicited in this experiment are observed[9].

Looking at the grand-average ERP curves and at the difference curves for the two conditions in anterior and frontal scalp sites in Figure 6.3, a general positive trend for the mismatch condition, highlighted by the difference curve (in black), can be observed spread across the majority of scalp sites in figure 6.3 between 200 and 280 ms, where components such as the P3a are often observed. The difference at the peak seem to be maximal in prefrontal and anterior-frontal scalp sites, both across the midline and in left and right hemispheres. Secondly, another trend of difference between the two curves is displayed preceding the positive-going polarity, maximal at 170 ms, in right frontal scalp site F4 only. Besides these two particular trends, the curves are fairly comparable and no specific other component or recognisable trend is visible.

[9]It is worth mentioning again that the electrode sites chosen are **1)** consistent across all three experiments and **2)** they had been selected prior to data collection in order to avoid any bias caused by visualising data and trends before deciding where to focus the statistical analysis.

In experiment three, the general amplitude of the grand-average curves is larger. This is most likely connected to the limited number of trials used to compute the average, considering the number of participants included in the analysis is about half of that of previous experiment two or one. Because of this, pre-stimulus interval noise is more pronounced throughout all findings in experiment three, which could also suggest that smaller trends observed across the dataset originated or were partially affected by noise in the data.

Figure 6.3: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.

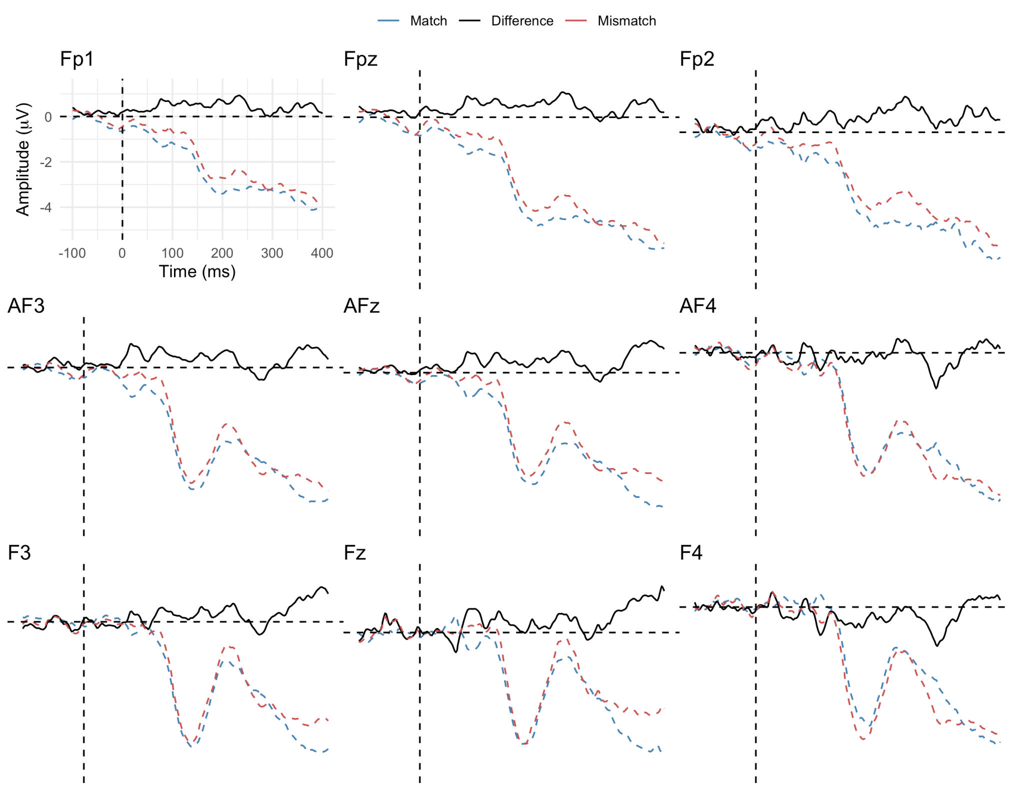
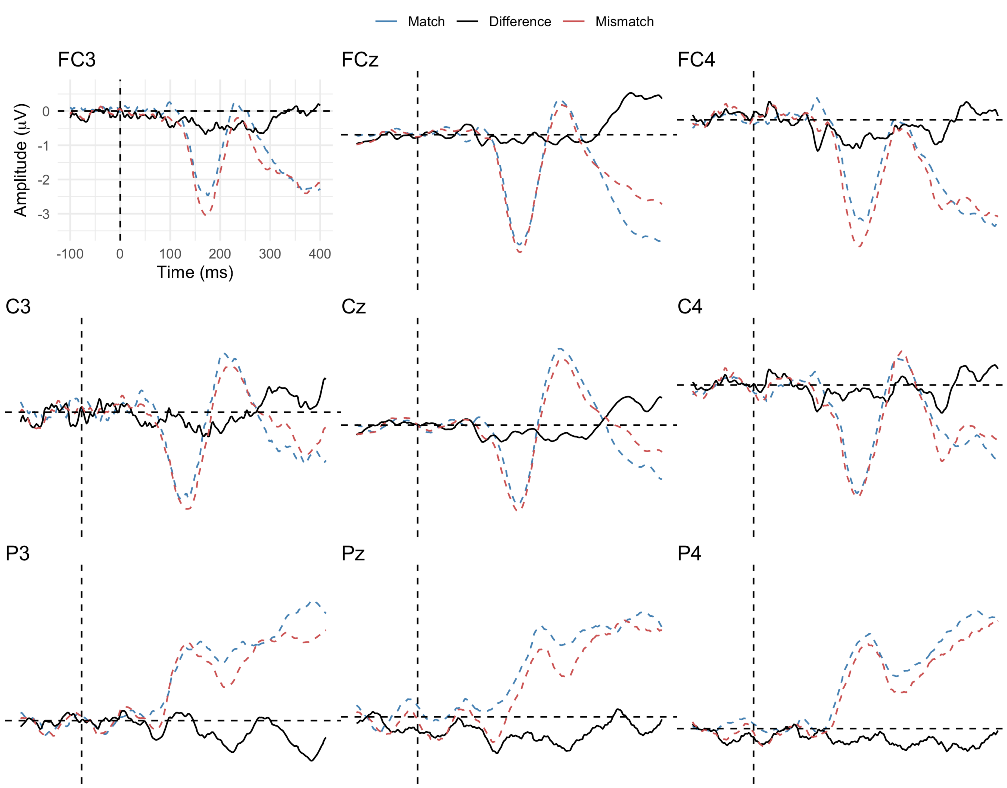


Figure 6.4 displays average ERP curves and difference curves for frontocentral, central and parietal scalp sites. Focussing on the difference curves in particular, only few trends can be observed throughout this set of electrodes. Differences between 100 and 200 ms, which have been observed in one frontal electrode and more prominently in previous experiments are not highlighted in figure 6.4, with very small differences the size of the noise in the pre-stimulus interval, foreshadowing a non-significant difference later on when tested with the use of the AFA procedure. A negative-going polarity in parietal scalp sites (Pz) between 200 and 300 ms

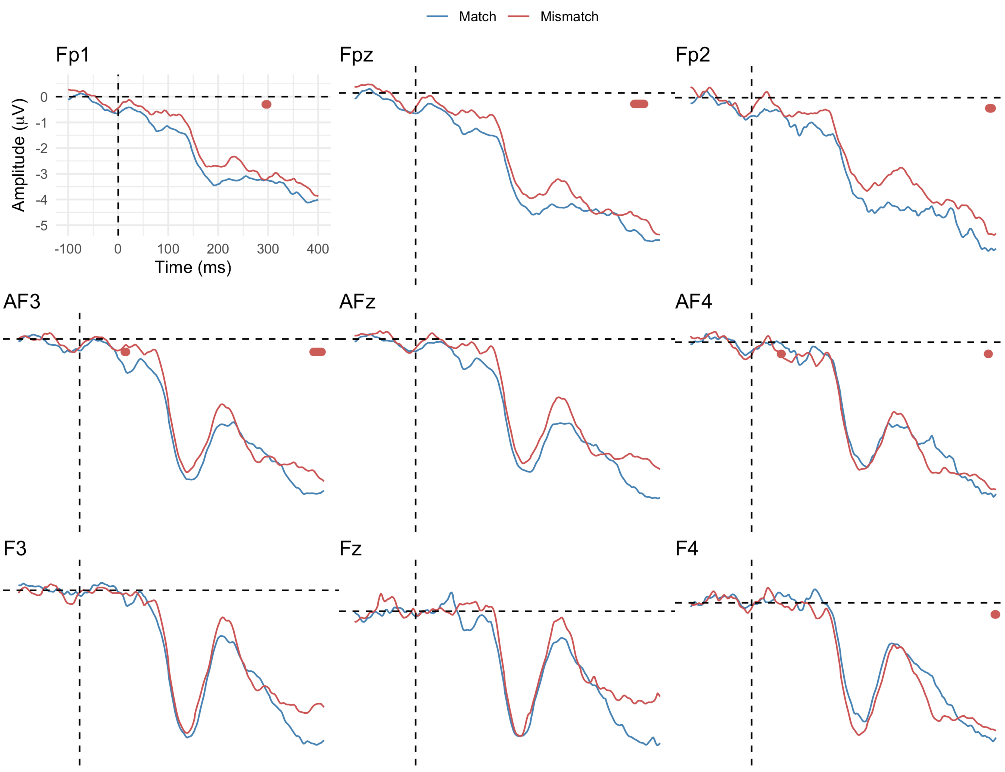
post stimulus onset can be linked to the corresponding positivity in central and frontal sites, previously connected to the elicitation of a P3a component. A trend that is maximal in fronto-central midline scalp site FCz is that of positivity for the mismatch condition between 300 and 400 ms, possibly linked to the elicitation of a P3b component in response to the categorisation of target, odball stimuli.

Figure 6.4: Grand-average ERP curves for match / mismatch experimental conditions and *mismatch minus match* difference curves calculated between -100 and 400 ms PSO for frontocentral, central and parietal sites.



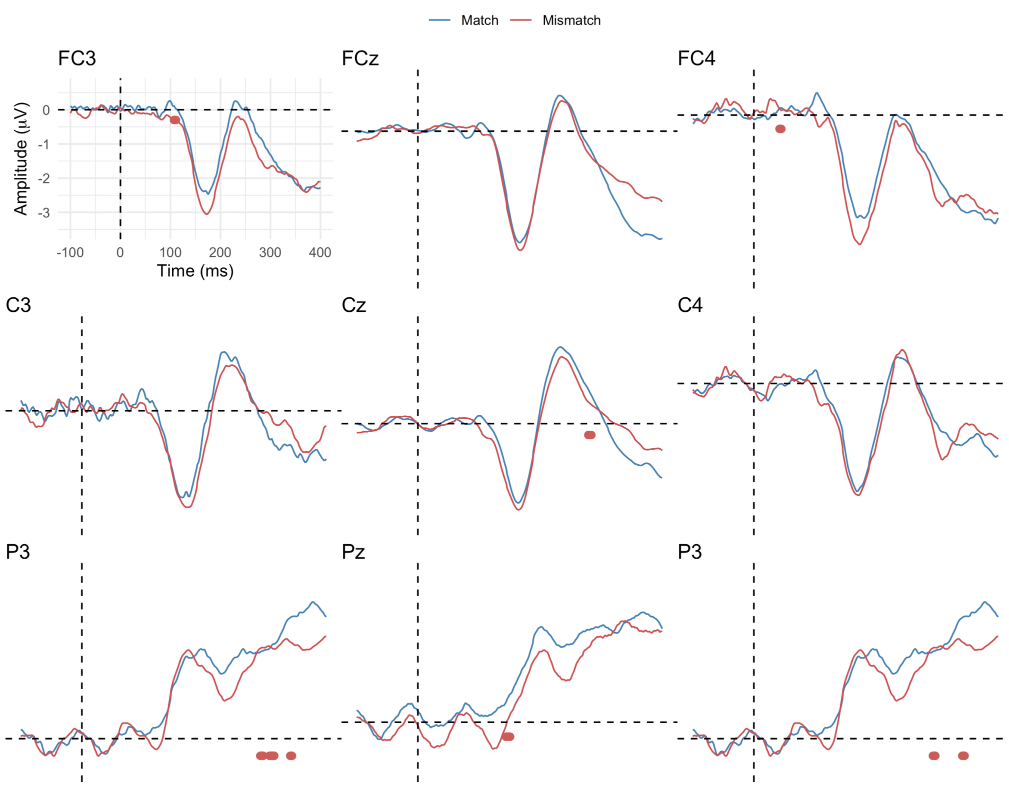
The results of the AFA procedure on anterior and frontal scalp sites (Figure 6.5) display very few differences between the two conditions, with almost no statistically significant differences for the most part throughout the time window. The only statistically significant difference found across many of the anterior and frontal scalp sites is between the 300 and 400 ms mark, where a positivity is observed for the mismatch condition compared to the matching one. The size of the difference, across all electrodes, is very small especially if compared against the overall amplitude of the curves and the pre-stimulus interval noise threshold. However, the consistency across multiple electrode sites suggests that the difference between the curves is caused by a general, underlying component rather than by random noise.

Figure 6.5: Grand-average ERP curves for match / mismatch experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.



Similarly, in central and parietal electrodes, significant differences have been observed short of the 400 ms mark in midline central electrode Cz, with a trend similar to that observed in anterior and frontal scalp sites. However, no other trends have been observed that suggest the elicitation of a different or novel component to the perception of mismatching and matching sine-wave tone sequences. The only other significant difference between the two curves is observed at right parietal electrode P3, at the 250 ms mark, in relation to the corresponding positivity in frontal and prefrontal electrodes. Finally, while differences in peaks can be visually observed across other electrodes, such as C4 and FC4 for instance, around 300 ms, these differences generally appear to be much smaller than the overall distribution in the pre-stimulus interval, suggesting that such differences are smaller than those derived by noise alone in the data. For this reason, none of these differences have been highlighted as significant by the use of the AFA procedure. It could also be the case that these and many other differences between the two curves in the data were not affected by noise in the data specifically but that the low power of the study made it difficult to isolate noise and genuine effects in the data.

Figure 6.6: Grand-average ERP curves for match / mismatch experimental conditions and AFA procedure results calculated between -100 and 500 ms PSO for frontocentral, central and parietal sites.



6.3.2 Modelling

Looking at the difference curves and the grand-average line plots displayed in the previous section of this chapter, only few differences in amplitude, peak and component latency were visually observed across the two main experimental conditions. A similar trend is presented through the modelling of mean amplitude data, where almost no effects of experimental condition are found across multiple time windows for experiment three data. Similarly to results reported for previous experiments, LMM were fitted to model the effect of experimental condition, scalp region and hemisphere, as well as their three-way interaction in R (R Core Team, 2020). Type-III ANOVA tables were generated from the `car` (Fox et al., 2012) package and pairwise contrasts were measured with the `emmeans` (Lenth et al., 2019) package.

Figure 6.7: Grand-Average mean amplitude intervals (between 140 and 180 ms PSO) for mismatch / match experimental condition

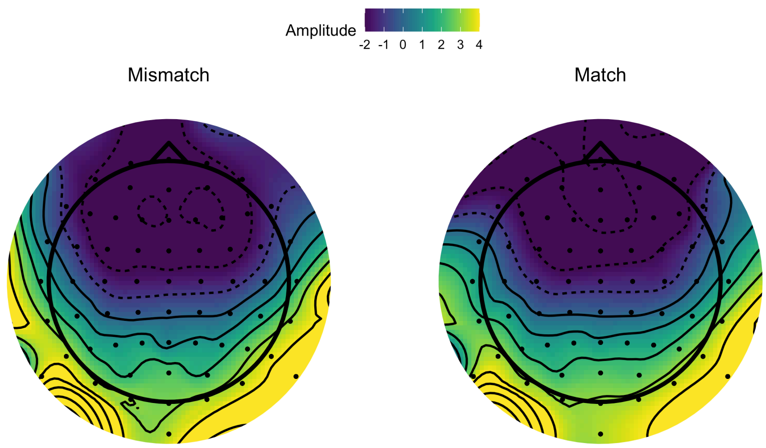


Figure 6.8: Grand-Average mean amplitude intervals (between 290 and 310 ms PSO) for mismatch / match experimental condition

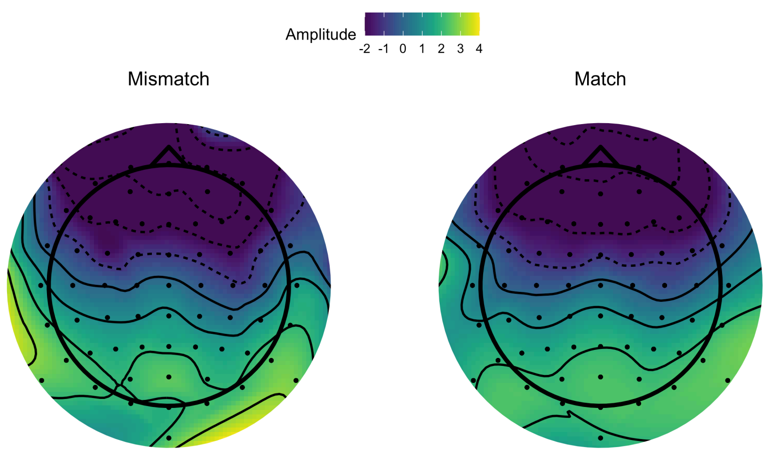


Figure 6.9: Grand-Average mean amplitude intervals (between 575 and 625 ms PSO) for mismatch / match experimental condition

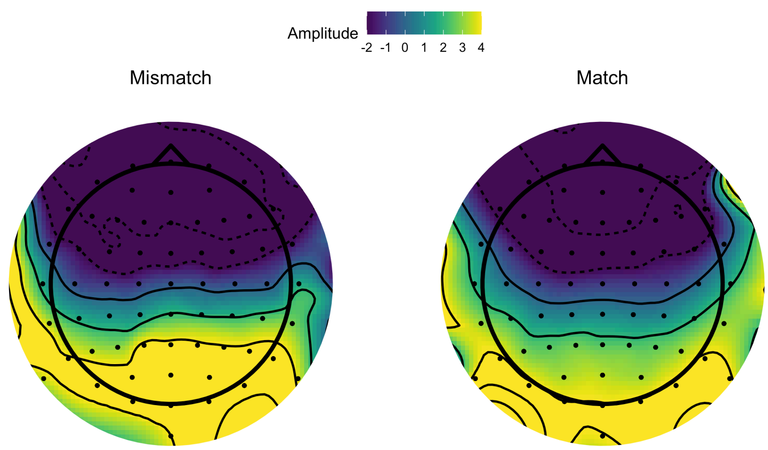


Figure 6.7 displays mean amplitude cubic spline interpolation topographical scalp maps for both the mismatch and match experimental conditions of experiment three, with mean amplitude values measured between 140 and 180 ms post stimulus onset, where an MMN-like trend was observed — despite its significance and status being far from confirmed — in the results of experiments one and two. Inspecting difference curves of experiment three, in Figures 6.3 and 6.4, a negative trend is visible for the mismatch condition across frontal and central scalp sites. A graphical 64-point interpolation presents no distinguishable visual differences between the mismatch and match condition in the time window where the MMN-like effect has been observed to be maximal in previous experiments and, where, in experiment three as well, the negativity trend is maximal scalp sites such as FC3. The results of the LMM suggest no main effect of condition [$F(_{1,1155}) = 0.010$, $p > .5$]. Only a main effect of scalp region was reported as significant [$F(_{10,1155}) = 160.428$, $p < .0001$].

No trend can be visually observed when comparing the cubic spline interpolation scalp maps of mean amplitudes measured in the time window between 290 and 310 ms (6.8), where a very small, possible PMN-like trend was investigated in in experiment two. Figure 6.8 presents very comparable mean amplitude distributions across the scalp for both the mismatch and match condition with no specific negativity trend for one over or the other. Despite an overall increase in grand-average amplitude range in experiment three, overall, compared to previous experiments, experimental-condition specific ERP components or, in more general terms, effects and trends visualised by difference curves, appear to be much smaller. The results of the

LMM suggest no main effect of experimental condition [$F_{(1,1155)} = 0.0002$, $p > .5$] and no significant interaction between experimental condition and scalp region [$F_{(10,1155)} = 0.325$, $p > .5$].

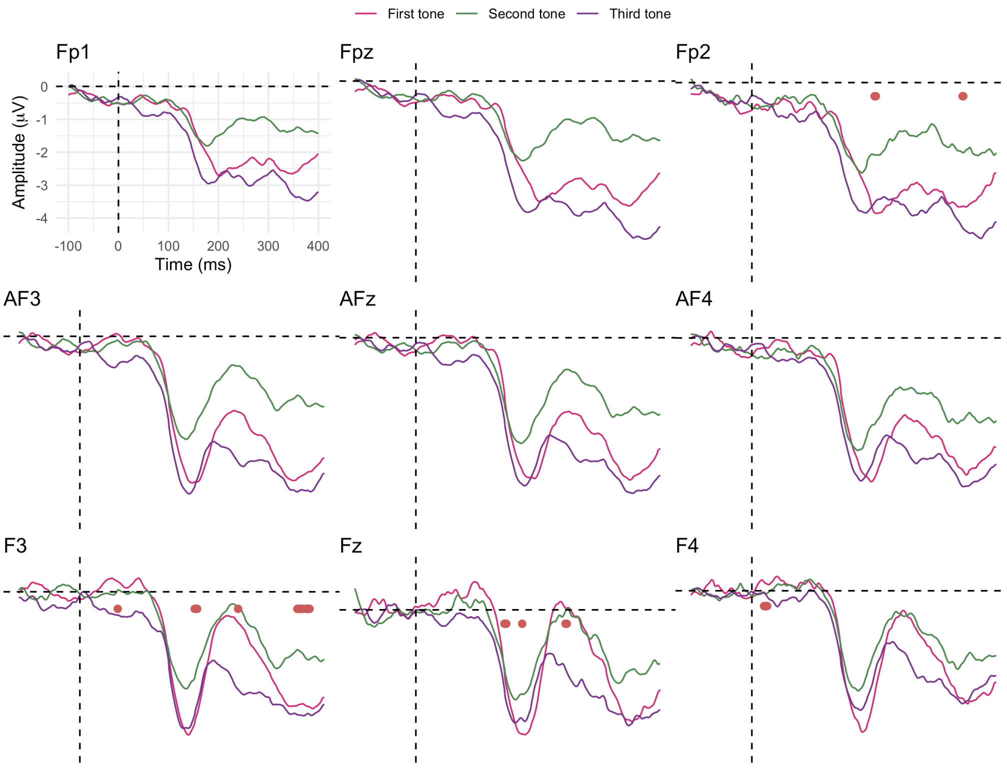
A positive trend is visually observed for the mismatch condition in parietal and posterior scalp sites over the match condition (Figure 6.9). This effect has been linked to the elicitation of a P600 component in all of the previous experiments, although no syntactic activation was present in the study of speech-like perception through experiment one and two. In experiment three, because the focus is musical sequences, a P600 component as a response to the violation of a given sequence — often a musical scale or tonality — was expected. Results of the regression model reports no significant main effect of experimental condition [$F_{(1,1155)} = 1.450$, $p < .05$], but there was a significant main effect of scalp region [$F_{(10,1155)} = 222.274$, $p < .0001$], as well as the significant interaction between experimental condition and scalp region [$F_{(10,1155)} = 2.979$, $p < 0.001$]. post-hoc pairwise contrasts report a main effect of experimental condition, with a positive trend for the mismatch condition, at parieto-occipital [$t_{(1961)} = 2.75$, Est = 1.16 μV, SE = 0.51 μV, $p < .05$] scalp sites.

6.3.3 Tone Sequence

In this particular section, ERP responses to the presentation to the presentation of each of the three sine-wave tones in the first half of each trial are compared. Much like in experiment two, participants were instructed to listen to each one of the three tones, presented separately, and to combine them into one single tune, without changing the order they had previously been presented in. ERP responses to the processing of each tone in the sequence show us differences in processing of sounds and the impact statistical probability in the perception of these sounds, since the first tone of each sequence has a 25% chance of being reproduced, considering there are four tone sequences. However, because each sequence is unique, each following tone (two and three) are supposedly expected by participants, given the first one, on the grounds of statistical learning affecting expectations. Since only four tone sequences were used in experiment three and considering each sequence was presented more than a 100 times to each participant, it can be suggested[10] that participants recognise and remember the sequences used in the experiment after a few trials.

[10]This assumption is also based on the performance of participants in experiment one where, when asked to learn multiple sequences of syllables, it took subjects an average of five minutes — no more than 10 repetitions per stimulus — to remember and recognise nonce-word pairs.

Figure 6.10: Grand-average ERP curves for first / second / third tone experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.



Each curve represents the grand-average response to one of the three tone in a sequence. The red dots, highlighting significant differences between the two curves at $p < .05$, have been implemented on the differences between the perception of the first and second tones only, since it's where the biggest differences are expected and considering visualising differences among each pair of conditions would have cluttered Figures 6.10 and 6.11.

It can be observed from the data in Figure 6.10 that, although not many trends have tested as significant with the implementation of the AFA procedure, some differences between the two conditions have been found specifically at electrode site F3 and Fz. In particular, there is an accentuated negativity at around 180 ms for the perception of the first tone, compared to a smaller response to the presentation of the second tone of each sequence. Another significant segment is observed in electrodes F3 and F4 between 300 and 350 ms post stimulus onset which presents a positive-going deflection for the first-tone condition compared to the presentation of the second tone, once again linked to P3-like components reacting to the processing and categorisation of odball or less probable stimuli.

Similarly to how ERP responses were displayed in experiment two when comparing responses to the perception of single syllables, the first tone of each sequence is the most unexpected and arguably the most informative. Previous findings suggest that word-initial syllables tend to evoke bigger ERP effects compared to word-medial and word-final syllables in speech and speech-like perception (Astheimer and Sanders, 2011), as they are considered to be more informative than following syllables (Connine et al., 1993). The same can be assumed specifically for the first tone in a sequence of tones, where the probability for the first tone to be played is 25% of all possible options. The main reason why the ERP reaction is usually bigger to word-initial syllables (and sequence-initial tones, in this case) is because, since the item is particularly informative, temporal attention is shifted to stimulus presentation, resulting in increased overall amplitude of components. Attention then shifts away from stimulus presentation — once the stimulus has most likely been recognised and identified— which is mirrored by less and less prominent ERP responses, as exemplified negative-going peak observed in the data.

Other than the main difference between 100 and 200 ms, where most auditory mismatch components are located (with distribution in anterior and frontal scalp sites for the most part) the curves of the three conditions are largely comparable throughout. While the number of participants is limited for experiment three, noise in this particular condition is reduced to the minimum because of the very high number of times the first half of each trial is consistently presented to all participants. Every trial in experiment three has a split sequence formed by three separate tones, while only 33% of trials are part of the mismatch condition.

Figure 6.11: Grand-average ERP curves for first / second / third tone experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for frontocentral, central and parietal sites.

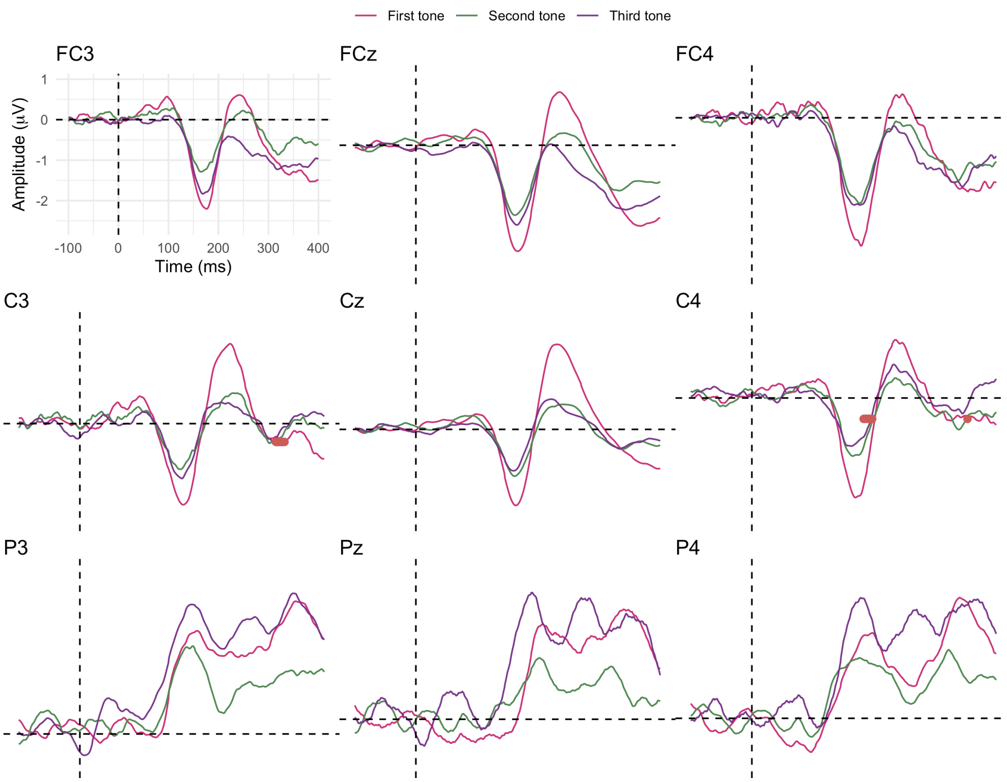


Figure 6.11 displays similar behaviours for ERP curves in response to the three sine-waves tones in the sequence among frontocentral, central and parietal scalp sites. The only visible, recognisable and statistically significant effect is located at electrodes sites C3, FC4 and P4, with a positivity for the second tone at around 300 ms (between 280 and 340 ms), linked to the elicitation of a P3 component in response to a lower presentation rate of the first stimulus compared to the second one. The effect appears to be maximal over parietal scalp sites.

The most visible trend between 150 and 200 ms is a negative-going potential in frontocentral and central scalp sites. However, among central and parietal scalp sites, the effect was not statistically significant at any scalp site included in the pool for the implementation of the statistical analysis. Besides these two specific effects between 150 and 200 ms and between 300 and 350 ms, no other trend, among the three curves, is visible from the data presented. The AFA procedure did not highlight any unexpected significant differences between the ERP responses to the presentation of the first and second tones in the sequence for anterior and frontal scalp sites. While differences at the 300 ms mark, related to the elicitation of the P3, have been found to be significant across a number of electrodes (5/18), the MMN-like effect was only significant at two scalp sites, FZ and F3 which suggests that the pre-200 effect is limited in distribution over the scalp.

Tones vs Syllable in a sequence

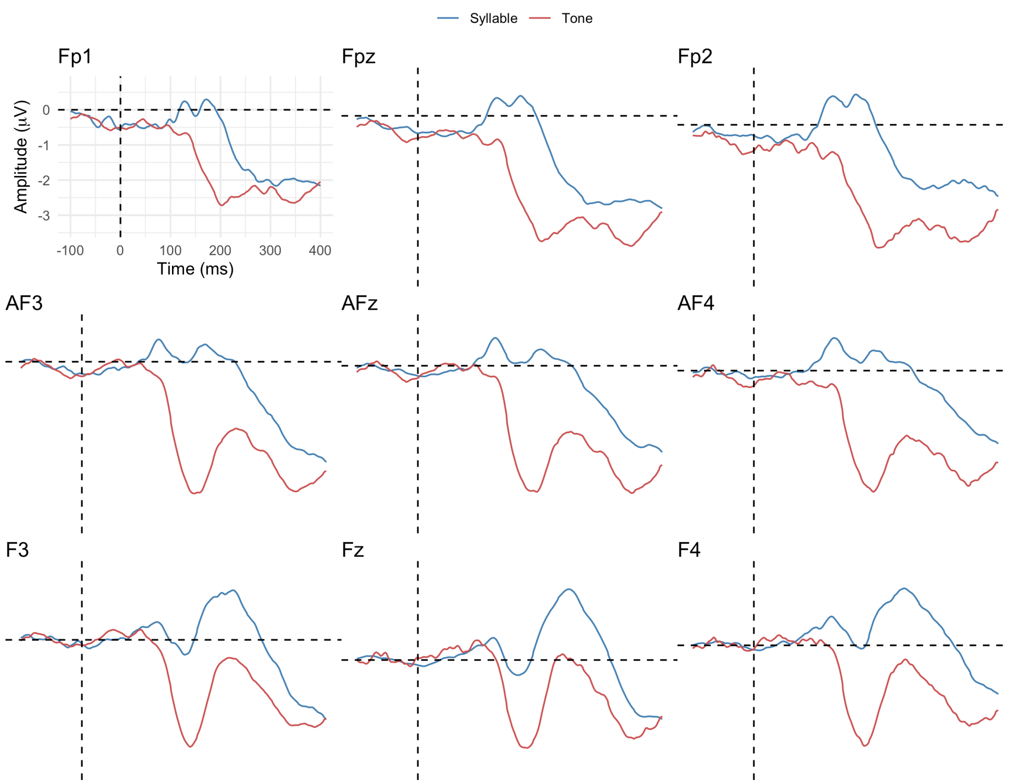
The methodologies of experiments two and three are very alike. The only difference between the two experiments sits in the type of stimuli used (syllables in experiment two, sine-wave tones in experiment three) while everything remained unchanged. To maximise comparability, the same code used to present experiment two was used for experiment three as well, simply substituting sound files used for the stimuli.

In the previous chapter, responses to the perception of each of the three syllables presented in the first half of each trial were presented. In this chapter, the same procedure was carried out for the analysis of the responses to the perception of each sine-wave tone. Similar patterns, to those observed for syllable perception, were found. Here, the goal is to compare the perception of the first syllable of each sequence in experiment two to that of the first tone of each sine-wave tone. The chance of hearing a specific first sine-wave tone — out of the four possible combinations — is that of 25%, which is directly comparable to that of being presented a specific syllable experiment two, since the same number of sequences were used. The sequences were also characterised by similar inter-sequence and intra-sequence transitional probabilities.

The differences observed between the two curves should represent the difference in response to the perception of a simple sine-wave tone, that does not carry any linguistic information and the perception of a linguistically-charged speech sound, in the same context. The context is that of no lexical, semantic or pragmatic meaning influencing expectations. Differently from the match-mismatch condition of the experiment, where expectations to stimulus perception are set by performing an active, behavioural task following the perception of stimulus sequences (focus, on the other hand, of this section), expectations in the perception of syllable / tone sequences are dictated by stimulus transitional probabilities alone. Although the focus here is on the first half of each trial, throughout both experiments, where only stimulus presentation rate varies but no mismatch is present through the administration of a behavioural task, the comparison is still very insightful when it comes to exploring and investigating different trends between sound and speech-sound perception.

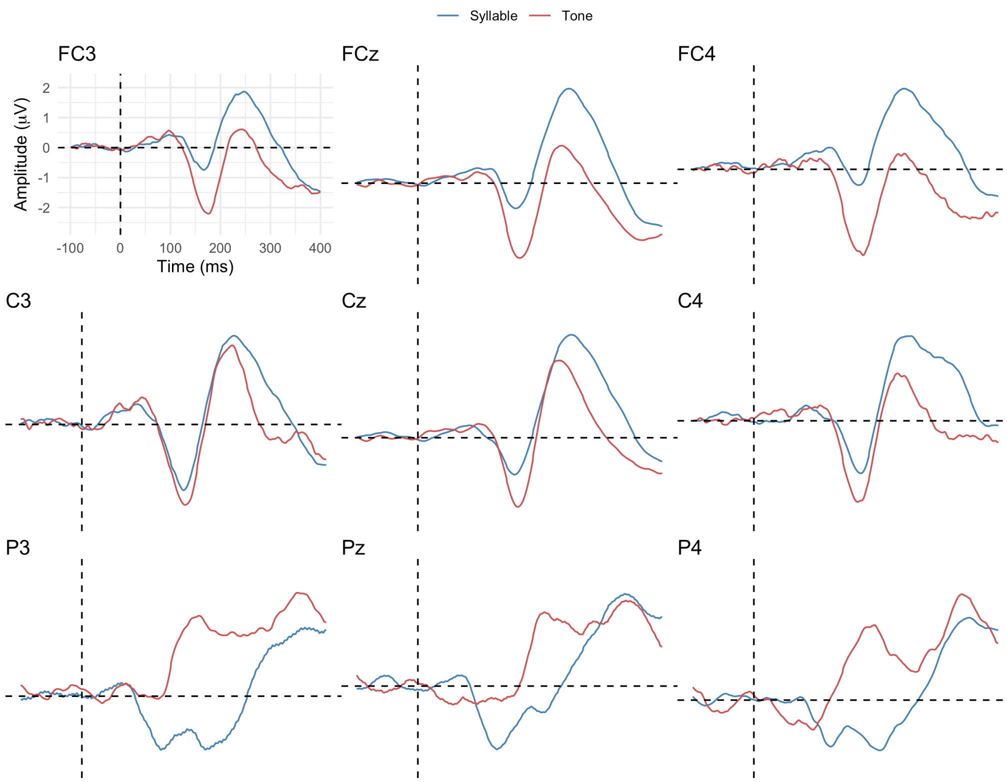
Figure 6.12 displays the differences and similarities between the perception of the first sine-wave and the first syllable of the stimulus sequence for anterior and frontal scalp sites, between -100 and 400 ms. Very noticeable differences across all scalp sites are displayed, generally concerning overall amplitude differences between the two curves rather than latency differences, although small latency differences are still present. First of all, the sine-wave tone condition (red) is characterised by increased negativity in the N1-P2 complex area, especially in anterior and frontal frontal scalp sites, while the same complex is more positively-going accentuated for the syllable condition (blue), with a higher P2 peak and reduced N1. The main difference between the two curves is certainly characterised by the N2 / MMN peak which is much more prominent for the sine-wave tone condition than it is for the syllable condition, with up to 6 μV peak differences between the two conditions. The negative peak is found to be slightly earlier in the syllable condition than it is in the sine-wave tone condition, although the difference is small, with the highest latency difference of 20 ms for anterior and frontal scalp sites.

Figure 6.12: Grand-average ERP curves for first tone (Experiment three) and first syllable (Experiment two) experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.



A similar distribution can be observed in frontocentral, central and parietal scalp sites (Figure 6.13). However, the further away from frontal and central scalp sites, the smaller the overall amplitude, with negative peaks for P3 and P4 electrode sites just short of 3 μV , compared to almost double the size for frontal sites. While effects are generally smaller in lateral sites, what seems like an important difference is visible at around the N1 peak for the tone condition, characterised by increased negativity when compared to the syllable condition. Besides this specific differences, results displayed in this figure are perfectly comparable with those described from Figure 6.12.

Figure 6.13: Grand-average ERP curves for first tone (Experiment three) and first syllable (Experiment two) experimental conditions and AFA procedure results calculated between -100 and 400 ms PSO for frontocentral, central and parietal sites.

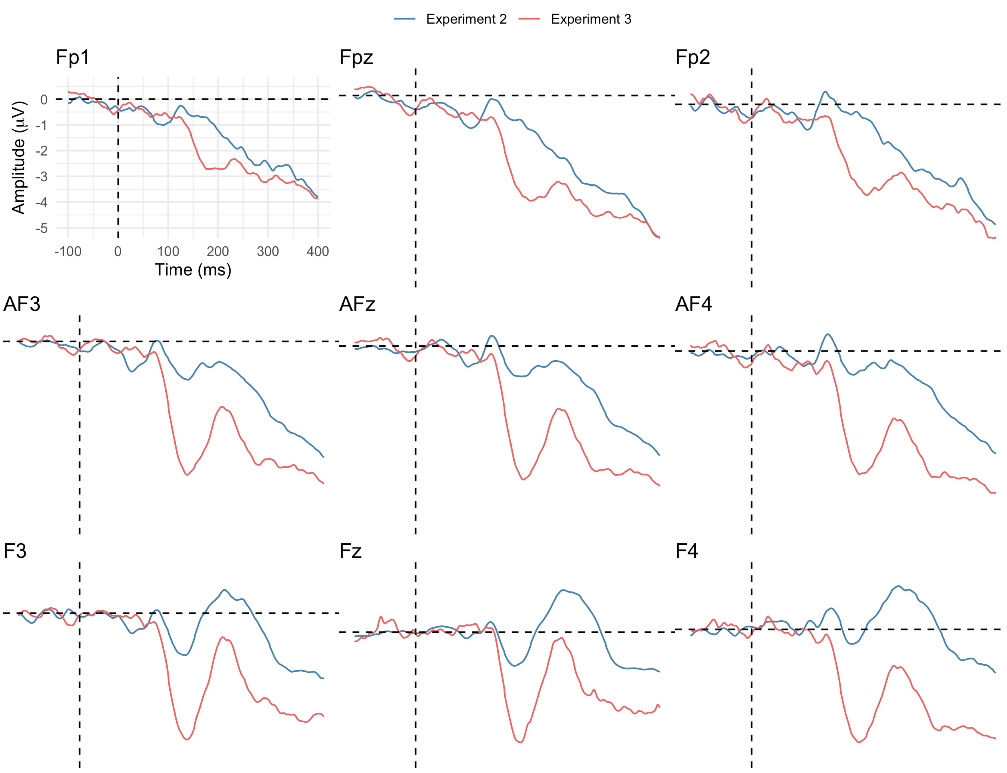


6.3.4 Mismatch in experiment 2 vs experiment 3

The aim of this section of the chapter is to compare the ERP responses to the perception of mismatching nonce words in experiment two and mismatching sine-wave tunes in experiment three. Investigating results within experiment three remains the focus of this experiment. However, comparing the findings of experiments two and experiment three offers access to further insights in the differences of speech-sound and sound perception. Any difference observed could suggest that linguistic and, in particular, phonemic and phonological processing, although not present in the form of a PMN in experiment two, might have been incorporated as part of neighbouring components in experiment three.

Previously, when discussing Figures 6.5 and 6.6, it was pointed out that only very few and very minimal differences were present when comparing the ERP response to the perception of matching and mismatching sine-wave tunes. However, when comparing the mismatch condition of experiment three to that of experiment two, major differences are visible right away, when observing figure 6.14, with significant differences displayed in frontal electrodes F3 and F4 in the N1 and MMN time window respectively[11]. A pronounced positive peak, possibly connected to the elicitation of the N1-P2 complex is clearly displayed in right anterior-frontal scalp sites (AF4 and F4) where differences between the amplitude of the N1 and P2 peaks reach 1.5 μV for the syllable condition. When comparing this with the response to sine-wave tones, we can see that while N1 peaks are generally comparable across both conditions, P2 peaks are almost absent in the latter. While P2 peaks are clearly larger and present higher amplitudes, the latency between N1 and P2 peaks for both conditions is practically the same.

Figure 6.14: Grand-average ERP curves for mismatching stimuli in experiment three vs experiment two and AFA procedure results calculated between -100 and 400 ms PSO for prefrontal, anterior-frontal and frontal sites.

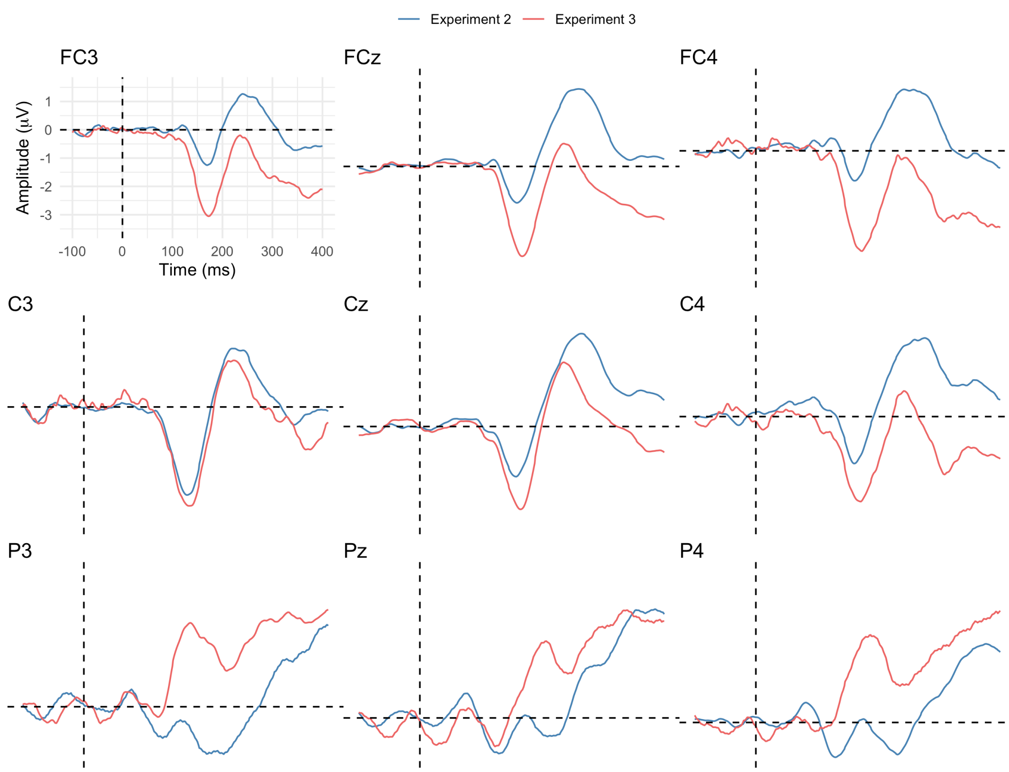


Another striking difference between the two conditions, highlighted previously when comparing the perception of the first tone and syllable of the stimulus sequences, is that the negative-going peak in the tone appeared to be more negative than that in the syllable condition (ex-

[11] Considering how similar differences are displayed across other electrodes sites and how the effect size of the difference seems to be particularly big, especially when compared to results analysed in previous experiments, the absence of significant segments is probably due to high threshold of noise that characterises average data in Experiment 3.

periment 2) at around 150 and 200 ms post stimulus onset, depending on the scalp site. This trend is visible across all scalp sites included in the analysis but the amplitude of the peak seems to be the highest for the sine-wave tone condition in the left anterior-frontal scalp sites, where the effect size difference between the two conditions is the highest. Finally, P3 peak amplitudes seem to be larger for tone condition compared to the syllable condition[12]. Direct differences in amplitude ranges between the grand-averages of the two conditions could be directly caused by the number of subjects averaged, smaller in experiment three than it is for experiment two. For this reason, the focus is on recognising similar or different component patterns among the results of the two experiments, rather than directly comparing amplitude values between the curves with a statistical test.

Figure 6.15: Grand-average ERP curves for mismatching stimuli in experiment three vs experiment two and AFA procedure results calculated between -100 and 400 ms PSO for frontocentral, central and parietal sites.



When comparing figure 6.15 to figure 6.14, similar patterns can be observed in frontotemporal, frontal, central and parietal scalp sites. In particular, P2 peaks are highest in right frontocentral

[12]The difference was calculated as average amplitude between N2 and P3 peaks for both conditions.

and central scalp sites (FC4 and F4), while negative peaks between 150 and 200 ms appear to be less pronounced for both conditions the further away from anterior-frontal scalp sites, which suggests that the origin of the effect is to be attributed to frontal and anterior areas of the scalp and that the component is consistent throughout both experiments. Besides these two observable trends, patterns of ERP responses between the two curves are fairly consistent and comparable in both sets of electrodes.

6.4 Discussion

In experiment three, the perception of mismatching sine-wave tones in a sequence, where expectations were created directly through the design of the experimental task, was tested and analysed using EEG. The results of experiment three are discussed on two separate levels. First of all, the results of this experiment in themselves give us insights in presenting the differences that are present whenever expectations are flouted in sound perception, where no linguistic information is present. At the same time, the results of experiment three are useful because they are directly comparable to those of experiment two, providing the ability to explore responses to mismatch in speech-sound perception to those of sound perception alone.

When comparing participants' ERP responses to matching and mismatching sine-wave tones in their respective sequences, only minimal statistically significant differences were observed between the two conditions in the pre-N400 range, where most of auditory-related and sound processing ERP components are usually observed. Comparing mismatch and match conditions at a channel level, only differences related to the elicitation of a P3 component were elicited and observed across multiple scalp sites. Significant differences earlier than 200 ms were only found at one scalp site and, even non-significant but observable trends were very limited (no N1-like effect).

Differences between the two curves, when comparing mismatching and matching responses, are displayed between the 500 and 700 ms marks. In this time window, a positive-going component can be observed for the mismatch condition in parietal and occipital scalp sites, as displayed through the topographical scalp analysis. The positive deflection is likely linked to elicitation of the P600 ERP response, with which it shares its onset, often at around 500 ms for the P600 (Hagoort et al., 1999), and peak latency at around 600 ms post stimulus onset. The P600 component is sensitive to (syntactic) rule violation and, in the specific case of experiment three, music-syntactic rule violation whenever a mismatching sine-wave stimulus is presented. The presence of the P600 component for the presentation of mismatching stimuli works, in this case, to suggest that participants did recognise mismatching stimuli as such. Moreover, if it was the case that rule violation was observed by participants, any other effect correlated with the presentation of a mismatching auditory stimulus should also have been elicited through this specific experimental paradigm.

Sine wave tones are much simpler, when compared to speech sounds such as single phonemes or whole syllables, both in terms of acoustic construction and in the amount of information they carry. Because of these differences, there is an expectation that responses to these two different types of stimuli are going to also be different. At the same time, because both stimuli were presented with very comparable methodologies, including auditory presentation in both cases, some similarities are also expected in terms of ERP response. Some examples include the P600 (see *Appendix A* for an in-depth discussion of P600 findings across all three experiments) signaling rule violation and a possible N1 and / or P3 effects, in relation to stimulus evaluation. The ERP response to the presentation of mismatching stimuli in experiment three were directly compared to the ERP response to of mismatching stimuli in experiment two. Some differences could be visually observed, especially concerning the overall amplitude levels of the N1-P2 complex and the overall negative-going distribution for the findings of experiment three. Although the MMN effect was not significant in experiment two for the mismatch condition at a channel level, it was isolated through topographical exploration.

However, when comparing the differences between match and mismatch conditions for experiment three to those of experiment two, no particular differences can be observed. This is especially true when focussing only on the results of the AFA procedure. The bigger effects are only those after the 250 ms mark, possibly linked to P3 elicitation, rather than any phonological or speech-specific response. The biggest difference before the 200 ms mark in experiment three can be observed when investigating the perception of the three separate sine-wave stimuli in the first half of each trial, rather than in the differences between mismatch and match conditions in the second half of all trials. In particular, a pronounced negative-going deflection can be observed for the perception of the first stimulus in the sequence compared to the second and third, only in two scalp sites compared to six in the same condition for experiment two. It could be assumed that the reason behind the difference in amplitude is a response to increased attention to the presentation of the first tone as it is highly informative. However, if this was the case, then more differences should be noticed for other components, including the N1 mark as well, considering that the component is sensitive to shifts in temporal attention (Näätänen and Picton, 1987) and that word onsets have been found to present greater N1 responses than word-medial and -final syllables (e.g. Astheimer and Sanders 2011). A detailed discussion and comparison of the negativity associated with auditory and phonological processing across all three experiments is carried out in *Chapter 7*.

When it comes to the 150-200 ms range, a plausible explanation to why a difference is present between the response to mismatching tones and that to the first tone of each base sequence, is that the type of mismatch is different in the two conditions. In the case of the base sequence, the type of response to the first syllable is that of *mismatch* caused by the low probability (25%) of a specific tone appearing in a given context. However, the response to mismatching tunes is directly caused by the result of the behavioural task itself. More specifically, the first tone participants hear in each trial has a 25% change of being presented, which caused increased attention and an unexpected response to stimulus presentation. In the mismatch condition, the unexpected factor is not simply caused by presentation probabilities but by the behavioural task i.e. identifying whether the incoming tune is identical to the one that was previously presented or not.

Participants were, during the second half of each trial, arguably at their most attentive to stimulus presentation. They had previously been instructed to compare the tones that they had previously listened to — and combined into a tune — to the upcoming stimulus sequence. At first, the differences between the two scenarios do not appear to be particularly relevant or well defined. However, it is known that attention can — also indirectly, through working memory — mediate the size and latency of specific components as reviewed in *Chapter 4*. More specifically, while most components are impacted by producing larger amplitudes the higher the level of attention to stimulus presentation, some ERP components — such as the P2, linked to stimulus evaluation — are usually characterised by reduced peaks in the present of attention (e.g. Kanske et al. 2011). The uncertainty of the origin of the difference between the two condition is, as particularly common in ERP experiments, due to the amount of unknown information and, in this specific case, the exploratory nature of the study.

6.4.1 Relevance to PMN?

Similarly to the outcome of experiment two (and, for that matter, experiment one as well) no PMN component was elicited in experiment three. However, this outcome was somewhat expected for this paradigm, considering no phonological mapping had been implemented across the stimuli of the experiment. However, the findings can help shape the boundaries of the nature of the PMN component, despite the experiment not being directly targeted to its elicitation. First of all, the lack of a classic PMN as a response to mismatch in auditory (but not phonological) perception reinforces the idea that the PMN is most likely **not** a response to sound mismatch. While phonological mismatch and sound mismatch are not synonyms, as there is much more to phonological mismatch than there is to sound mismatch and, considering that phonological mismatch does not necessarily have to happen in combination with auditory mismatch (i.e. reading tasks), there is certainly a component of overlap between auditory and phonological mismatch. Until the two are separated, it is not straightforward to be able to determine whether a component works as a response to phonological or auditory mismatch specifically or, alternatively, a combination of both as it seems to be the case for the MMN.

Although a PMN response was not necessarily expected in relation to the absence of phonological mapping throughout experiment three, the absence of an MMN response could suggest methodological limitations particularly tied to lack of power. The MMN is sensitive to phonological information but its primary role is that of response to auditory and acoustic mismatch, which is present in experiment three. At the same time, the elicitation of a clear P600 response would suggest that the study was appropriately powered and that participants perceived mismatching stimuli differently than match. The different interpretations, matched with methodological limitations, do not allow for a clear-cut interpretation of findings.

For this particular reason, the results of experiment three, specifically because no difference was found between the match and mismatch condition in the time range either an MMN or PMN would be expected — in an experimental context where phonological mismatch is not present — are not informative about the nature of the PMN on their own, besides reinforcing the idea that the PMN is not an auditory only response. The following chapter contextualises the findings of all three experiments and it combines existing theories on the elicitation of the PMN with current findings and plausible explanations for the non-elicitation of the PMN.

Chapter 7

General Discussion

The general discussion chapter presents a summary of all relevant experimental findings from experiments one, two, and three. It also ties all collected findings together to previous research with the aim of painting the bigger picture when it comes to the PMN and its function and role in speech and sound perception and processing. Furthermore, the chapter introduces thoughts and implications of the absence of a clear-cut PMN component across all three experiments. Hypotheses are also advanced on the general role of phonological processing in semantics-free spoken-word perception. Finally, the role of ERP components less relevant to the direct focus of the study, such as the P600 found as a response to sequence violation in syntax-free contexts in experimental paradigms involving both language and musical processing, is discussed in relation to the findings of all three experiments.

7.1 Summary of findings

7.1.1 Match and mismatch

No clear-cut PMN response, specific to phonological mismatch and mapping between the 250 and 300 ms mark post stimulus onset, has been observed across the three experiments carried out in this research project. The findings of experiment two exhibited a significant negative-going deflection between 280 and 320 ms post stimulus onset, maximal at 300 ms in parieto-occipital scalp sites only between match and mismatch condition. Although it was originally assessed that the negative trend might have been linked to the elicitation of PMN-like negative polarity, no comparable effect was observed across either of the remaining two experiments. This trend was not significant with the implementation of the AFA procedure (Sheu et al., 2016) at a channel level. However, as previously mentioned, LMM fitted to mean amplitude data reported a very limited significant effect in parieto-occipital scalp sites only. For this reason, it can be suggested that the highlighted trend is not linked to the observation of the PMN. The negative polarity observed for the mismatch condition of experiment two was most likely a by-product of P3 contamination and a shift in P3 latency, observed through the implementation of peak latency measurements for both the mismatch and match conditions in the 250 to 300 ms time range.

Among other trends and effects observed in the findings of experiments one and two, an MMN-like trend, maximal at 160 to 170 ms post stimulus onset, was observed for the mismatch condition (over the experimental match condition). This effect was statistically significant over frontocentral scalp sites among the findings of experiments one. The trend was significant over the left hemisphere, between match and mismatch experimental conditions, among the findings of experiment two. No trend was visible at all in experiment three. Similar patterns were highlighted in experiment two by channel-level analysis and the implementation of the AFA procedure. However, none of the trends could be reliably isolated as one clear-cut specific component. The differences between the match and mismatch conditions in experiment two were much smaller and less consistently captured by the AFA than those of experiment one. In all experiments, although not recognisable as commonly shared trends across all experiments and only significantly recognised at channel level mostly, P3-like trends were present for the perception and processing of target, oddball stimuli being presented across the experiment. An N1-like effect was only recorded in frontal and prefrontal regions — with statistical differences only observed in frontal electrodes — for the mismatch condition of experiment two but not for that of experiments one and three. Finally, the only effect that was consistently significant with the implementation of an LMM model across all three experiments was a late P600-like positivity (see *Appendix A* for an in-depth discussion of all P600 findings) maximal in parietal and parieto-occipital electrodes.

To summarise, some ERP responses to the presentation of mismatching stimuli are present across the findings of experiments one, two and three, with some components shared among either all or two experiments out of three. However, there is no clear PMN component that can be isolated specifically as a response to phonological processing. In addition, besides the consistent elicitation of a P600-like response across all three experiments, all other components could not be reliably labelled and linked to the mapping of a specific level of processing.

7.1.2 Contextualisation of other findings

Other findings that were investigated that did not belong to the main match and mismatch experimental conditions of experiments one to three were not directly impactful in determining the status of the PMN and its elicitation in response to phonological mismatch. They did, however, provide access to insights and ERP responses in speech perception that helped validate the experimental methodologies used. These trends were consistent with the effects expected in those specific contexts — based on existing literature — where transitional probability differences in the presentation of stimuli caused mismatch without the implementation of external tasks or linguistically-derived mismatch, e.g. phonological or lexical. Observing effects that were predicted for those specific conditions in the experiment helped in validating the efficacy of the methodological choices made for the realisation of the experiments.

For example, in the baseline experimental block of experiment one, no statistically significant effects were discovered through channel level or topographical component origin analysis. This was to be expected considering that participants were listening to stimuli they had not heard or learnt prior to the collection of ERP responses, in a context where no active behavioural task was implemented. Although no statistically significant effects were discovered, some general trends were observed that appeared to mirror similar responses to those elicited in the target block of experiment one, although to a much smaller extent. This could potentially suggest that some level of statistical learning of the stimuli implemented in the baseline section could have taken place despite the reduced length of the baseline experimental block.

In experiment two and three, when comparing responses to the first and second stimuli in each of the three-stimulus sequences presented in the first half of each trial of both experiments, effects linked to both stimulus presentation rates (such as P3 and similar effects) and to sound mismatch (such as a N1 and MMN-like negativities) were sometimes observed. The focus of experiments one, two and three was to investigate whether a PMN component could be observed across the chosen variety of experimental paradigms tested in this research project. Although no PMN was expected when investigating responses to the single syllable / tone stimuli in the first sequence of each trial, the findings were insightful in that they were able to display what type of responses were collected in a setting where mismatch was only caused by stimulus transitional probabilities rather than when transitional probabilities were combined with task-specific mismatch in the mismatch vs match conditions of experiments two and three.

7.2 Phonological processing

In *Chapter 2* I have summarised some of the most relevant papers that link the existence of the PMN component to phonological and word-form processing in speech perception and spoken-word recognition. At the same time, I have highlighted the inconsistencies when comparing the findings of many existing experiments run to determine the nature of the PMN component and its topographical distribution. In particular, one of the main limitations of previous research in experiments that aimed at eliciting the PMN component in processes of spoken-word recognition and speech perception was that phonological mapping was always associated, in the chosen experimental paradigms, with lexical activation and lexical retrieval to some extent (e.g. Connolly and Phillips 1994; Newman et al. 2003; D'Arcy et al. 2004; Kujala et al. 2004). This situation created a set of confounding variables which does not allow for a clear-cut dissociation of phonological and semantic processing.

Because of this specific existing gap in the literature, the idea to only use semantics-free stimuli in experiments aimed at eliciting the PMN component was advanced, in order to thoroughly investigate the existence and the specific role of the PMN component in a context where phonological mismatch and mapping were present but no other layer of language processing (e.g. morphological, semantic, syntactic, or pragmatic) had been introduced. This decision was made to further confirm the hypothesis that lexical status of the presented stimuli was not influential to the elicitation of the PMN. Furthermore, some of the existing findings place the PMN in a similar position — both temporal, topographical, and functional — to that of the MMN (e.g. D'Arcy et al. 2004; Newman and Connolly 2009). Because of the similarity between the two components, I have also tested ERP responses to the perception of semantics-free non-speech stimuli, in experiment three, in order to compare the findings with previous responses to the presentation of semantics-free speech stimuli. The aim of the above comparison was that of dissociating phonological processing from auditory mismatch in the elicitation of pre-N400 responses.

The general discussion is focussed on two main aspects of phonological mismatch. The first is the absence of the PMN among the findings of all three experiments, specifically when focussing on experiments one and two. Secondly, in light of the absence of a PMN response, the same section presents hypotheses on whether phonological processing has been observed as embedded in the elicitation of other ERP components, tested by comparing the findings of experiments two and three. Both experiments introduced acoustic mismatch in the presentation of incoming stimuli, with experiment two incorporating phonetic and phonological mismatch with the implementation of speech sounds compared to non-speech sounds of experiment three. An explanation for the above is that, although the PMN was not observed, phonological processing and mismatch could have been incorporated in increased amplitude or latency changes of neighbouring pre-N400 components, for instance, such as the MMN.

7.2.1 PMN

No clear-cut PMN component was observed in any of the three experiments carried out in this research project. The three experimental paradigms, that were created from the beginning specifically to work both independently and in combination, were not capable of eliciting a clear response linked to phonological mapping and mismatch in speech perception. In particular, experiments one and two combined show that, regardless of the level of focussed attention to stimulus presentation, phonological mismatch as an effect of the presentation of mismatching semantics-free stimuli did not contribute to a PMN-like response from participants. At the same time, experiments two and three combined show that, regardless of the (non-) linguistic nature of the stimuli, no specific effects can be observed that appear to be directly linked to the elicitation of a phonological mismatch response in experiment two.

Over the decades, the PMN has received different contextualisations and its role in speech perception and spoken-word recognition has been updated multiple times. Connolly and Phillips (1994), for instance, originally linked the PMN to phonological mismatch during processes of spoken-word recognition, contextualising the component as part of the *lexical selection* stage of the Cohort model (Marslen-Wilson, 1984) specifically (Connolly and Phillips, 1994). If we were to describe the role of the PMN component according to the theory advanced by Connolly and Phillips (1994), the PMN could be interpreted as a mismatch between expectations caused by pre-existing (lexical, syntactic, phonological) contextual information during lexical retrieval of incoming stimuli and mismatching (phonological) information. Being able to dissociate an N400 response from a PMN — although not entirely in previous research (e.g. Connolly and Phillips 1994 as discussed by Lewendon et al. 2020) — is what prompted researchers to suggest that the PMN response was directly linked to aspects of phonological mismatch of incoming stimuli, rather than semantic or syntactic.

Similarly, in a study carried out almost one decade later, Newman et al. (2003) propose an updated theory of the PMN suggesting that the "PMN serves as a neural marker for the analysis of acoustic input merging with pre-lexical phonemic expectations" (Newman et al., 2003). When visualising the *expectations* vs *reality* scenario in Connolly and Phillips' (1994) theory of the PMN, there would be semantic, syntactic, para-linguistic information on one side (*expectations*) and mismatching semantically-driven phonological information (*reality*) on the other. When both factors are present, mismatch is registered and a PMN response is observed. However, when contextualising the explanation provided in the research paper by Newman et al. (2003), the interpretation is that the role of the PMN has shifted to that of a much more lower-level response, linking the PMN component to aspects of language processing such as *acoustic* and *pre-lexical phonemic* information (Newman et al., 2003). In this case, the PMN acts as a goodness-of-fit indicator between expected word-form information and speech input.

These two views of the PMN, that link it directly to lexical retrieval or to pre-lexical processing also shift the importance that theoretical assumptions, highlighted in *Chapter 2*, have on the successful elicitation of the component. In the more recent contextualisation of the PMN, linked to the word-form goodness-of-fit processing, differences in aspects of speech perception models such as information flow order and access to information should not affect the elicitation of the component in the current experiments. However, if the PMN were to be embedded in a context of lexical retrieval and if it were to derive from lexical retrieval processes primarily, assumptions would play a more important role in the overall design of experiments on the PMN.

This more recent view of the PMN, backed up by more than a decade of research, almost entirely discards the role of lexical activation altogether. If it were the case that the PMN was a response to pre-lexical phonological information and a neural marker for acoustic input processing, experiments one and two (and perhaps even three) should have been able to elicit the PMN component. This is specifically true if we consider the experimental paradigms were designed specifically to elicit components that would be reacting to pre-lexical phonological and acoustic processing and mismatch in speech perception, regardless of the semantic status of the stimuli used.

Over the years, the PMN went from being interpreted as a semantically-involved phonological component to a more lower-level response, linked to the processing of acoustic merging and pre-lexical phonological information. Justifying this shift is evidence that presents the PMN component as a response to both word and nonce word stimuli (e.g. Connolly and Phillips 1994; Newman et al. 2003; D'Arcy et al. 2004). The PMN was observed as a response to semantically activated and semantically non-activated stimuli in spoken-word recognition. For this reason, it was suggested that semantic status of the stimulus — i.e. whether the word or item presented had an associated meaning whether physical or abstract — does not necessarily matter towards the elicitation of a PMN response.

However, the above explanation does not necessarily take into account semantic and lexical aspects of speech processing that might be activated regardless of the semantic or semantics-free nature of a specific stimulus. In a context where a word is presented as baseline stimulus, semantic mapping and lexical activation processes would have already likely been activated. When listening to a word in a participant's native language, it is impossible to do so while ignoring the semantic aspect of that word. When a participant hears a word and removes a phoneme from that word, semantic mapping happens and word form and meaning are retrieved, regardless of how that happens. When a second stimulus is presented, such as in experiments by Newman et al. 2009 or Kujala et al. 2004, that is supposed to match the original stimulus minus the first phoneme, which in turn is also a real word at times, it would make sense that some form of lexical retrieval is also activated to check whether the two stimuli match. By removing the first phoneme from the word *clap*, which is lexically activated, we obtain the word *lap*, which is also lexically activated. Then, when comparing the resulting word with the incoming stimulus (i.e. either *lap* or a mismatching (nonce) word), it makes sense that semantic mapping is activated again to check for similarities between stimuli and it only presents different outcomes depending on the semantic nature of the stimulus, just as in the previously discussed example of Connolly and Phillips (1994).

The theory that the PMN is not simply a response to pre-lexical phonological and acoustic information is partially supported by the results of experiments one and two of the current research project. In experimental paradigms that never introduce any need for semantic retrieval and mapping, since participants were instructed from the start that all stimuli were free of meaning and they did not represent existing words in their language, no PMN was component was consistently elicited in all experiments. Furthermore, if the PMN was simply a marker of acoustic information merging with pre-lexical information, the PMN could have also been elicited by the stimuli in experiment three. Non linguistic, sine-wave tones in experiment three were carriers of acoustic information and pre-lexical "form" information which could be compared to, in a linguistic context, phonological information (pitch, duration, transitional probabilities, etc). However, no PMN-like response was elicited. For this very reason, evidence suggests that the original theory of Connolly and Phillips (1994) was most likely closer to the real nature of the PMN than its later reiterations (e.g. Newman et al. 2003). The PMN appears to be a response to phonological mapping but only in a context where semantic mapping and lexical retrieval are also present.

The amount of influence that lexical mapping has on the elicitation of the PMN component in unknown. Past research studies started using components such as the PMN to investigate the phonological processing abilities of patients in clinical settings, especially comparing that to their semantic processing abilities in the form of the N400 (e.g. Robson et al. 2017). However, it is likely that the PMN does not represent phonological processing of pre-lexical phonological information alone but that its observation is directly influenced and mediated by the presence and extent of semantic mapping in spoken-word recognition. In that case, measuring the elicitation of the PMN would not necessarily equate to measuring the extent of phonological processing abilities. Conversely, studying the extent to which a PMN component is observed as a response to phonological and semantic mismatch in subjects with confirmed impaired semantic processing abilities could help determine the influence semantic and lexical activation has on the observation of the PMN component in spoken-word recognition.

The findings of experiments one to three regarding the observation — or lack thereof — of a clear-cut PMN component are more in line with the original theory by Connolly and Phillips (1994), than they are with more recent interpretations of the PMN, who place the PMN as part of a lexical selection stage of spoken-word recognition. The evidence collected in the current research study cannot help determine to what exact level of processing of lexically-influenced spoken-word recognition the PMN belongs to (if any), especially because no PMN was elicited across the board. However, the findings suggest that the lack of lexical activation and semantic mapping in spoken-word perception, regardless of the amount of phonological, phonemic and acoustic mismatch in the perception of incoming stimuli, did not elicit a PMN in any of the forms (e.g. different topographical distribution) that have been previously reported in the literature. The observation of other components such as a late positivity linked to the P600 and stimulus categorisation and responses to the presentation of oddball stimuli suggest that participants were aware of mismatching, manipulated stimuli and that the lack of elicitation of a PMN was most likely not connected to limitations caused by the design of the experimental paradigm.

7.2.2 Other traces of phonological processing

Assuming that the PMN reflects word-form aspects of spoken-word recognition in contexts where semantic processing is also present, the question remains of whether phonemic and phonological mapping in semantics-free perception can be traced to the observation of any specific ERP response. In relation to this matter, two theories can be advanced that suggest that **i)** phonological mismatch in semantics-free stimuli emerges as part of neighbouring pre-N400 components such as the MMN and N1 or that **ii)** speech-like stimuli are perceived as speech, hence requiring some level of specifically phonological processing, only once semantic mapping is involved. The next sections of this chapter display both theories, discussing the findings of experiments two and experiment three in relation to the first theory and speculating on the implications theory number two would have on decade-long debates over information flow order and architectures of grammar.

7.2.3 Phonological processing and pre-N400 components

An ERP component such as the MMN, often linked to the presentation of mismatching stimuli in a sequence of control, matching stimuli (Näätänen et al. 1993; Näätänen and Alho 1995), was also linked to mismatch caused by a more abstract level of processing — specifically phonological processing (e.g. Pulvermüller 2001) — as opposed to being exclusively sensitive to more lower-level, acoustic and duration differences among stimuli. For this specific reason, an argument could be made that, if no PMN component is observed in a semantics-free context where phonological mismatch is present, a neighbouring component such as the MMN might be sensitive to some level of phonological processing instead.

If we were only to consider the findings of one experiment at a time, such as experiment two for example, we would not be able to assess whether an observed pre-N400 effect, such as an N1 or MMN-like component, was in response to phonological mismatch specifically. It could also be the case that its elicitation was primarily directed to the acoustic properties of incoming stimuli, such as changes in frequency, for instance. In experiment two alone, stimuli were both auditory and linguistically activated. Any early response to mismatch might have been prompted by physical (acoustic) differences or abstract (linguistic) features. For this reason experiment three, designed in tandem with experiment two, was aimed at presenting participants with auditory acoustic mismatch but no phonological mismatch by employing auditory but non-linguistic stimuli. Responses that, theoretically, were shared in both experiments could be attributed to stimulus categorisation and acoustic mismatch effects while responses that were specific to experiment two should be primarily linked to phonologically-specific aspects of stimulus categorisation. Metaphorically *subtracting* responses to acoustic non-phonological mismatch from responses to acoustic and phonological mismatch should highlight responses that are specific to phonological mismatch alone.

No patterns were observed between the two experiments that could directly be linked to phonologically-specific ERP components in experiment two. A comparable P600 response was observed among the findings of both experiment two and experiment three, suggesting that the response is not directly linked to linguistic aspects of the stimuli but rather to some common feature presents in either the stimuli or the methodology of both experiments. Namely, the observation of the P600 component was linked directly to violations in sequence presentations. While other responses were observed among the findings of experiment two that were not observed in experiment three, the lack of power of experiment three could be behind the lack of effects between the two conditions of experiment three. Most of the pre-N400 components observed in experiments one and two were characterised by very small amplitudes. The fact that the only component clearly recognised in experiment three was the P600, found to be the biggest response among all in earlier experiments too, could support the theory that power in experiment three was insufficient to discover more marginal effects.

An effect that is visible for the mismatch condition of experiment two and that it is not observed in experiment three (or one) at all, is a negative-going trend at 100 ms post stimulus onset in frontal and prefrontal scalp sites. This effect was originally linked to an N1-like trend in the discussion of *Chapter 5*. The effect was reported as significant among the findings of a LMM fitted to mean amplitude measurements for pre-frontal, anterio- and frontal scalp sites. Although some evidence of phonological influence on N1 elicitation has been proposed, this was primarily observed in clinical studies on dyslexia (e.g. Breznitz 2003; Fosker and Thierry 2005). All in all, there is not strong evidence to link this early effect to phonemic or phonological processing directly, considering that experiment one implemented the same stimuli and no comparable effect was observed.

The N1 component is elicited with no required focussed attention to stimulus presentation (Näätänen and Picton, 1987). The findings of experiment one and three display no statistically significant N1 (75-125 ms) component and the findings of experiment two appear to display a significant N1-like effect. If it was the case that the this N1-like response was present among the findings of experiment two but not among those of experiment three specifically because stimuli of experiment two are phonologically activated, the same effect should have been observed as a response to the presentation of mismatching stimuli in experiment one. For this reason, it is likely that the more prominent response in experiment two is due to a higher degree of focussed attention. However, the reason why the effect is not present among the results of experiment three could be attributed to lack of power and a higher noise threshold in the data. In experiment one, it is very much likely that the same effect would have been discovered perhaps with a higher number of trials aimed at balancing lower signal-to-noise ratio.

On the other hand, an effect that is most prominent in experiment one and less focussed — but still significant with the implementation of LMM — in experiment two, is a negative-going deflection between 150 and 200 ms post stimulus onset for the mismatch condition which has been linked, because of latency and topographical distribution, to an MMN-like effect. However, interestingly, the effect is not as prominent in experiment two when comparing estimates in both experiments, which should not be the case for the MMN. The MMN is modulated by attention to stimulus presentation but not reduced the more attention is present (Näätänen and Alho, 1995). The MMN has been linked to phonological processing in the past (Pulvermüller, 2001). However, the low power of experiment three, paired with a very small effect in experiment two, makes it almost impossible to determine whether the small difference between the two experiments was caused by a phonological processing effect embedded in the responses of experiments one and two or whether — given proper power — results from experiment three would have also displayed a similar component, confirming its response primarily to acoustic mismatch rather than to a more abstract, phonological or linguistic sensitivity.

The findings of experiments one to three are inconclusive in determining whether phonological and phonemic mismatch, while absent in the form of the PMN, appear through the elicitation of neighbouring ERP components. The experiments succeed in drawing a baseline for future work on this specific topic. Future experiments might be able to use paradigms that are specifically known to elicit components such as MMN, N1 and neighbouring pre-N400 responses. With the addition of phonological mapping in stimuli, studies will be able to investigate the extent to which these extensively studied components can be sensitive to chances in phonemic and phonological mapping, mismatch and, in more general terms, processing.

7.2.4 The role of lexical activation

In the absence of a clear-cut PMN component and the surfacing of phonological mapping responses in other pre-N400 ERP components, it could be suggested that responses to the perception of speech sounds and non-speech sounds are very similar and the biggest differences appear only once semantic context is directly introduced. For example, it should not be considered a coincidence that, while the N400 component has been extensively linked to semantic mapping specifically in language processing and that the P600 has primarily been observed as a response to syntactic — although not only directly linked to syntactic mapping — pre-N400 responses have often been found to be linked to other aspects of processing and mismatch in perception but, only secondarily, they have been found to be sensitive to linguistic specific aspects of speech and language. Components like N1 and MMN, for instance, have been found to respond differently to the presentation of simple tones and speech sounds across a variety of experiments (e.g. Näätänen and Picton 1987; Näätänen et al. 1993; Näätänen and Alho 1995; Pulvermüller 2001). Many of the differences can also be explained by the use of different methodologies, different levels of complexity of the stimuli and so on, not necessarily to the introduction of specifically linguistic information.

If we were to speculate further, that speech sounds are processed as language only once semantic information has been mapped — similarly to how the PMN appears to be elicited only if semantic processing also takes place — this would suggest that semantic mapping or, some level of semantic check, should be carried out and observed early in the perception chain if components such as the PMN, activated between 250 and 300 ms post stimulus onset, are determined by its presence or absence. This specific behaviour would, simultaneously, challenge architectures of grammar that claim grammatical modules act independently and, at the same time, provide a claim towards interactional models of information flow in speech perception. These implications, together with limitations of this theory and suggestions for future research aiming at disambiguating responses to phonological mapping in speech and speech-like perception are explored in more detail later in the current and following chapters.

7.3 Semantics required

The main theory suggested by the interpretation of current findings and previous research in the current chapter is that the elicitation of the PMN, albeit mirroring processes of phonological mapping, requires lexical information and activation to take place. More specifically, the PMN is seen as a marker of phonological mapping — whether it represents phonological mapping as a whole or specific processes such as mismatch is impossible to determine at this point — as part of a broader, non-particularly well-defined semantic activation layer. Other, more speculative theories and suggestions for future research on how the PMN was not observed in experiments one and two because it might not be necessarily mapped to phonological processing or because it might not exist as one clear-cut component are discussed in the final section of this chapter. Here, the aim is to present what implications a semantic-driven phonological processing PMN would have on modelling speech perception and grammatical architectures.

The theory proposed in this work is closest to the explanation for the elicitation of the PMN proposed originally by Connolly and Phillips (1994), without the contextualisation of the PMN as part of the *lexical selection* stage the Cohort model (Marslen-Wilson, 1984), rather than mirroring more modern theories such as the one previously discussed by Newman and Connolly (2003). The main implications arising from a non-observation of a PMN component in semantics-free contexts are that either an early semantic check determines whether phonological and semantic mapping in speech perception need to take place, which would suggest the existence of an interactional model of speech perception. Another concurring option could suggest that semantics-free and semantically activated stimuli are perceived and processed differently, more specifically as non-language and language respectively.

The two theories are not even necessarily exclusive of one another, as an earlier *semantic* check could determine whether the incoming stimuli are to be treated and processed as language or whether they should be processed as, for instance, auditory and acoustic non-speech stimuli. In *Chapter 2*, I introduced some of the most well known models of speech perception and spoken-word recognition and the role of major debates, such as information flow order, as part of the difficult task that is modelling speech perception. The theory that the PMN is a phonologically-sensitive component that is semantically activated suggests that models such as the TRACE model (McClelland and Elman, 1986), which suggests an interactional net of cognitive and linguistic processes, might be more fitting in explaining how word recognition takes place, with semantic processing prompting phonological mapping (through the PMN). This phenomenon, eventually, resolves in either semantic match or mismatch, represented in ERP research by the elicitation of the N400 component. A theory such as this one also suggests that an earlier semantic check could be reflected in pre-PMN ERP responses to speech or non speech stimulus presentation. Such a claim is, however, currently not backed up by any experimental evidence in this research project or existing literature on the PMN. However, evidence that links the elicitation of pre-PMN components such as the MMN and N1 components to linguistic mapping, for example, exists and it was summarised in preceding chapters.

It could also be hypothesised that, depending on contextual, extra-linguistic information or experimental task demands in the case of previous research, incoming stimuli are processed as either speech or non-speech by applying an offline selection process before the presentation of incoming stimuli. For instance, Tabullo et al. (2013) presented participants with sentences in their native language (i.e. Castilian Spanish) and in a semantics-free miniaturised artificial language (AL) researchers had taught subjects prior to the experiment. Mismatch during perception, both in the participants' native language and in the miniaturised AL caused the elicitation of an N400 component, often linked with semantic mapping and mismatch. Because participants were exposed to both real and *fake* (i.e. non semantically activated) speech, the default option for participants was that of treating all incoming information as (real) speech and to perform semantic mapping and lexical selection processes. These processes, in turn, resulted in mismatch when lexically non-activated stimuli were also used. Tabullo et al. (2013) suggest that the elicitation of the N400 is, however, linked to unfulfilled expectations regardless of lexical activation in the processing of artificial languages (Tabullo et al., 2013). The possibility that "language"-related components such as the N400 are observed in contexts where no semantic mapping or lexical retrieval is found could also be attributed to the idea that linking ERP components to specific grammatical modules is becoming much more of an outdated view of what the role of ERP responses really represents in the activation of cognitive processing in the brain.

The first of the two theories suggests that an interactional model of speech perception is behind the online interpretation of the semantic status of incoming stimuli in order to determine whether and what type of phonological mapping should be carried out as part of lexical retrieval. However, the hypothesis that contextual information can bias subjects in treating incoming stimuli as either language or not does not necessarily require an interactional model of spoken-word recognition. If it is in fact decided beforehand whether incoming stimuli are to be processed as linguistic information or simply auditory and acoustic information then grammatical modules and layers of speech perception and processing could act independently of one another and function in a feed-forward fashion as suggested by modules such as the Cohort model of word recognition (Marslen-Wilson, 1984) and others listed in *Chapter 2*. At this point, because of the lack of existing research and resources to determine why the PMN is not observed and, upon discovering that, what type of implications would that have on grammatical architecture modelling, the theories presented in the following paragraphs are mainly characterised by speculation and theoretical assumptions rather than being grounded in existing research.

7.4 ERP and grammatical modules

In scientific research in general as well as in electro-physiological research, the sum total knowledge gathered on specific roles of ERP components is usually linked to the interpretations made of experimental findings and existing literature on the topic. However, by definition, when more evidence is introduced and findings and theories are re-interpreted, the role of ERP components and their function can often change with the addition of more and more compelling evidence. In this specific research field, the above has been particularly true for almost all of the ERP components mentioned so far in this work. In ERP research, the change in interpretation is especially very noticeable because the base amount of knowledge that has so far been collected is relatively limited. For this reason, linked directly to the young nature of the field and ever-growing technical and computational advances, only little extra evidence is often necessary to tip the scales when it comes to discussing and interpreting the nature and function of ERP components in processes such as language comprehension among others.

More specifically, there has been a particular trend in the past decades where interpretations tend to widen — rather than to narrow down to a more specific function — the role and function of existing ERP components and each component eventually covers more and more aspects of cognition from what originally suggested, sometimes dramatically changing the interpretation we have of such responses. For instance, although the P600 component had been in the past labelled primarily as a syntactic response (Hagoort et al., 1993), new evidence was introduced that has associated it with processes, mapping and mismatch of information that is further and further away from syntactic mapping in language. These contexts include musical (Patel, 1998) and arithmetic (Núñez-Peña and Honrubia-Serrano, 2004) processing. Furthermore, as it is in the case of this specific research project and other similar experiments (e.g. Tabullo et al. 2013), the P600 has been linked to sequence violations, even in syntax-free or linguistics-free contexts. Generally, with the implementation of more and more experimental designs, ERP components are often found as a response to new types of stimuli and experimental paradigms, similarly to some of the findings of experiments one to three of this research project.

Similarly to the P600, but in an almost opposite fashion, the MMN component (Näätänen et al. 1993; Näätänen and Alho 1995), which was primarily thought to be a non-linguistically sensitive component and primarily a response to direct physical change in stimulus characteristics (e.g. frequency and duration chances), has been linked with direct sensitivity to phonological (Pulvermüller, 2001), syntactic (Shtyrov et al., 2003) and other linguistic information. The general trend, not only true for the aforementioned P600 and MMN components, but also found in other examples such as N400 and P3, that have been experimentally observed as responses to more and more types of mismatch and mapping than the original roles they had been given of semantic and odball component respectively. In a growing body of literature, the N400 has extensively been observed in instrumental musical processing, where no canonical semantic activation — as found in language — was present (e.g. Koelsch et al. 2004; Daltrozzo and Schön 2009; Slevc and Patel 2011). The common denominator for all changes in interpretation when it comes to the role of ERP components in cognition is that their role and function is often expanded from representing one specific type of function and, for language components, from representing one single grammatical module, as it was the case specifically for the P600 and the N400, originally strictly linked to processing syntactic and semantic information respectively.

Changes in interpretation of the role of components such as N400 and P600 have been carried out over the course of decades with the realisation of hundreds of studies and experiments that investigated their role as a response to different types of stimuli and experimental paradigms. When it comes to the PMN, the original interpretation of the component in the few experiments carried out to investigate its function has already undergone contrasting explanations (e.g. Connolly and Phillips 1994; Newman et al. 2003), suggesting that its function and role — and existence, as discussed later on — is still up for interpretation. Considering the general trend that moves the function of existing ERP components away from being interpreted as a direct representation of grammatical modules in language comprehension, it is not impossible that this is the case for a component such as the PMN, which has been observed as a response to phonological processing in some instances but not other. This would indicate that, perhaps, its response is not to phonological mapping directly but to some other contextual — linguistic or non linguistic — stimulus information that partially contains phonological mapping processes during semantic activation. Future research should aim at further dissociating the observation of the PMN from phonological and semantic processing, although the task is certainly not made easy considering that the research interest on the component has been particularly scarce. This makes it so that the gap in the literature is, as of today, probably considerably bigger than the existing knowledge we have regarding the role of the PMN. Although there is a chance that the PMN component is not directly linked to phonological (or semantic and linguistics) processing as part of language comprehension but that it mirrors a wider set of cognitive processes that partly contain semantic and phonological mapping, it is still unclear what the link is between the PMN, language and stimulus categorisation and mapping.

7.4.1 PMN: MMN and N400

The final implication regarding the non-observation of the PMN — and the most speculative of all the theories presented so far — is that the PMN was not elicited as a response to phonological processing not because the PMN does not represent phonological mapping but rather because the PMN, as a stand-alone clear-cut component, does not exist. The possibility, also briefly addressed in a recent meta analysis of existing PMN research (Lewendon et al., 2020), is primarily based both on inconsistencies among previous research studies, especially regarding the topographical distribution of the PMN, as well as on striking similarities between the PMN and neighbouring ERP components such as the MMN and N400.

Regarding the first observation, one of the ways a significant effect is interpreted as a result of a specific component over another, besides the interpretation and choice of experimental paradigm and component and peak latency, is topographical distribution. For example, the P600 component we have extensively addressed in previous chapters is considered a primarily central and parietal component (e.g. Hagoort et al. 1993; Hagoort et al. 1999; Hagoort and Brown 2000). While some evidence is present that places the P600 (or similar responses) in frontal regions of the scalp (e.g. Ledwidge 2017), the literature — and the results of the experiments of this research project — are overwhelmingly in favour of the distribution of the P600 component being primarily parietal. The fact that topographical distribution is an important factor when determining the nature of an observed response during experimentation is because most components, not just the P600, have a very consistent topographical distribution. Another example is the MMN component, usually observed as a frontocentral response (e.g. Näätänen et al. 1993; Näätänen and Alho 1995).

In the case of the PMN, Lewendon et al. (2020) proposed that, among the eleven studies on the elicitation of the PMN and PMN-like responses to phonological processing, no more than two papers agreed with each other on the topographical distribution of the component, with research observing a PMN-like response in frontal (Connolly et al., 1990), central and parietal (Connolly and Phillips 1994; Hagoort and Brown 2000) electrodes, as well as a flat distribution across the midline (Van Petten et al., 1999) and the scalp (Van Den Brink et al., 2001). With literature like the one on the P600, the component has been observed for the majority of cases as a parietal responses and, only in a limited handful of cases, distributed in other regions. This makes it so that, researchers interpreting new results might have an easier time in identifying the component as a P600-like response based on component latency and topographical distribution. However, because the PMN has been observed across a variety of locations across the scalp, any negative-going component whose latency is in the 250 to 300 ms range, regardless of topographical distribution, could be interpreted as a PMN response. On top of this, research that investigated the PMN was, in most cases, already predicting a PMN-like response based on previous studies. This bias could have prompted researchers to interpret the effect they already assumed was going to be a PMN component as the PMN component, despite the fact that topographical distribution was inconsistent with previous findings.

Unfortunately, this is a limitation of ERP research and, especially of exploratory ERP research. When very little information is present on the topic that is being investigated, a higher degree of interpretation and arbitrariness is required when interpreting the results of the experiment. The results of experiments one to three did not manage to elicit the PMN among all components observed. For this reason, the findings are not necessarily informative when it comes to determining whether the PMN is a real response or whether its observation in previous research was linked to the elicitation of multiple PMN-like components that were mistakenly interpreted as one single response. The PMN, in fact, shows similarities with neighbouring components such as the MMN and N400, as discussed at length in previous chapters. the similarities are not only topographical and temporal but also functional, with the MMN being linked to phonological processing (e.g. Pulvermüller 2001) and the PMN and N400 being often elicited in similar experimental paradigms.

Chapter 8

Conclusion

8.1 Future research & closing remarks

The aim of the current research was to test the elicitation of the PMN component across different experimental contexts involving semantics-free passive and active speech and speech-like perception. The non elicitation of the alleged phonological response is in itself further evidence of the behaviour of the PMN and the level of phonological mapping in spoken nonceword recognition. The proposed explanation for why no PMN was observed is that the PMN necessarily requires semantic activation for its elicitation, and therefore, it should not be treated as an exclusively phonological / pre-lexical ERP response.

The implications of a semantically-activated phonological mapping component have clear repercussions on architecture of grammar and speech perception models, which should then also account for module interaction. However, uncertainty surrounding the role of ERP components and their direct link to grammatical modules is at the base of other plausible theories for the missed elicitation of the PMN. Some examples include the idea that the PMN might not directly reflect phonological mapping or other grammatical modules or that the component might not be a clear-cut stand-alone response in itself, but rather a response linked to neighbouring components such as MMN and N400, as suggested in a recent meta analysis of the existing literature by Lewendon et al. (2020).

Experiments one to three of this research project all provided access to insights in the realm of speech, speech-like and auditory perception, with a particular focus on trying to distinguish phonological and auditory processing. The main strength of this set of experiments is that all of the experimental designs were very similarly structured. This allowed for direct and indirect comparisons in terms of what ERP responses were observed across all experiments and all conditions. Ideally, future research would build upon this, by proposing similar methodologies to those of experiments one to three, by either adding or subtracting information from the stimuli employed in the experimental design. For instance, in order to focus specifically on the differences between phonological and auditory perception and processing, no semantic activation was present throughout all experiments. This decision provided a direct way to determine whether the observation of the PMN was linked primarily to phonological mismatch, rather than to semantically-informed phonological mismatch. This paradigm also allowed for a direct investigation into the nuances of both the MMN and PMN components. However, while the difference between the two aforementioned components are certainly undeniable on paper, the PMN was found to be sensitive to semantic information as well as to phonological mapping (Connolly and Phillips, 1994). Separating the MMN and PMN responses, while at the same time introducing semantic processing, could help provide more evidence on the relationship between the PMN and later, more abstract components such as the N400 without influences of earlier responses such as the MMN.

An experiment that included auditory perception of lexically-activated non-speech stimuli could help provide a clearer picture of how auditory information is processed and how semantic mapping is carried out when no speech sounds are present. This is one example of how, by tweaking the methodologies employed in this work, one could be able to gain further insights into the role of MMN, PMN and N400 in the scenario of phonological and semantic processing in speech perception. ERP data are inherently noisy and unreliable, both because of their high unpredictability but also because of how little information we have on specific ERP components and their function across several realms of perception and processing. Building on already existing paradigms and slowly tweaking some of the factors is the best way to slowly and steadily build an array of contexts where ERP components are elicited, hopefully amounting to a precise atlas of ERP. The atlas will serve as the basis of experiments that, in the future, will tackle questions of a very important nature in linguistic theory, such as information processing in speech perception, architectures of grammar, and language acquisition theories.

The nature of the PMN component is, decades after its first discovery, still uncertain and future research should — although far from being a straightforward, trivial task — aim at systematically tackling all of the different theories highlighted in this and previous work, dissociating the component from semantic, phonological and acoustic processes. At the same time, the focus should be on studying variation in the component's topographical distribution, as of today one of the biggest inconsistencies when it comes to the sum total of the knowledge gathered on the PMN. At the same time, studies aiming to dissociate the PMN from other earlier responses such as the MMN should be carried out to rule out the hypothesis that the PMN is not but a later MMN response to phoneme or frequency changes in stimulus presentation.

Unfortunately, the little agreement on topographical distribution when it comes to the PMN, together with differing theories on the component's role in spoken-word recognition make it, simultaneously, harder to identify the component when faced with a similar negativity and easier to mistake any other negativity in similar latency ranges as the PMN. Although the findings of the current research project are not definitive in pinning down the role of the PMN component, they provide further evidence of the component's range of responses by not eliciting the PMN in relevant experimental environments. Especially when very little background knowledge is available and a number of assumptions need to made with the interpretation of the results, testing very specific contexts and experimental paradigms and applying very few methodological changes between them to maintain comparability is possibly the safest way to tackle exploratory research with under represented ERP components such as the PMN. There are limitations to exploratory ERP research in language comprehension, such as extensive gaps in the literature and conflicting theoretical frameworks. However, in-depth knowledge of ERP components will allow for a better understanding of mental processes involved in the perception and comprehension of language. These discoveries, in turn, would prompt important advances in linguistics and cognitive sciences, with exciting applications in clinical and experimental work as well as in providing evidence to support existing and new hypotheses in decade-long theoretical debates with regards to architectures of grammar and spoken-word recognition modelling.

Appendix A

Sequence violations and stimulus expectations

Considering the methodological choices and compromises that had to be made to design three experiments that offered comparable points of view and complementary methodologies, some responses to stimulus categorisation, presentation rates and sequence violation were expected, albeit not desired. However, one response that was not expected or, to a small extent, only expected in one of the three experiments was consistently observed across all three experiments of this research project. The effect in question is a parietal P600-like response for the mismatching experimental conditions in experiments one, two and three. Despite the P600 response not being critical to the interpretation of the absence of the PMN, its analysis offers a valuable perspective on how components that were primarily thought to be linked specifically to one type of processing — syntactic in the case of the P600 (Hagoort et al., 1993) — can now consistently be elicited with differing, non syntactic stimuli.

P600

The P600 is a positive-going event related potential component observed as a positive polarity between 500 and 700 ms over centroparietal scalp sites. The P600 is often associated with grammatical and syntactic processing of language (Hagoort et al. 1993; Hagoort et al. 1999). It is often elicited both through listening and reading tasks (Hagoort and Brown, 2000) and its amplitude is maximal at around 600 ms following the presentation of the target stimulus. The P600 component is not only elicited through grammatical and rule violations in language processing. In fact, it has also been observed as a response to the processing of arithmetic tasks (Núñez-Peña and Honrubia-Serrano, 2004) and music (Patel et al., 1998). While the majority of research on the P600 places the component as primarily posterior and parietal, later positive-going polarities thought to be related to syntactic processing and context resolution, similar roles that have been taken on by the P600 in existing research (e.g. Hagoort et al. 1999), have been found in frontal (Ledwidge, 2017), frontal and temporal (Friederici, 2006) and central (Silva-Pereyra and Carreiras, 2007) locations. Considering the P600 has been observed in contexts other than those characterised by syntactic violations in language, like those of rule violation in musical scales and arithmetic tasks, determining that the P600 is primarily connected to syntactic mapping would be detrimental to investigating the complete nature of the P600 component. It is likely that the component is linked to a broader category of cognitive processes which, in the realm of language processing, manifests as (but is not limited to) syntactic mapping.

In music, for instance, the P600 has been found as a response to an out of tune note (Zendel and Alexander, 2020) (arguably a very dissimilar context than that of syntactic mapping in language processing), mismatch in tonal endings (Schön and Besson, 2005), perception of inappropriate harmonies and tonal structures (Lagrois et al., 2018). The P600 has also been observed as a correlate of the P3 as a response to the presentation of oddball target stimuli (Sassenhagen and Fiebach, 2019) in language processing, again moving the focus of the P600 component away from strictly syntactic to a wider overset of processes. As if the role of the P600 was not complex enough, specifically considering it is found a response to many different levels of mapping and processing in language, music and other contexts, some studies have found the P600 to also be a response to morpho-syntactic violations (Hagoort et al., 1993), phrase structure violations (Friederici et al., 1999), long-distance dependencies (e.g. Fiebach et al. 2002), mismatching mathematical operations (Martín-Loeches et al., 2006) and violation in geometric shape sequences (Besson and Macar, 1987), among many other examples.

None of the experiments of this research project make use of any syntactic context. For this reason, it may seem that discussing the P600 as a relevant component to this work is outside the scope of the study. However, despite not including any syntactic context, late positive deflections that match both the topographical and temporal distribution of the P600 have been observed throughout all experiments of this research project. Specifically in experiments two and three (*Chapters 5* and *6*), participants were instructed to create stimulus sequences from the stimuli presented to them. When expectations were then flouted, both in the presentation of speech and non-speech stimuli, a P600-like effect was observed. These findings further help advance the idea that the sequences created by participants amounted to, to some extent, syntactic or rule-bound sequences just as musical scales or arithmetic functions are.

P600-like late positivity

An effect that was observed consistently across all three experiments for the mismatch condition is a late P600-like positivity maximal in parietal and parieto-occipital scalp sites between 500 and 700 ms post stimulus onset, peaking at around the 600 ms mark in all three experiments. The P600 component, with which this positive-going deflection is most likely associated, presents similar characteristics to the observed effect in regards to component latency and topographical distribution (Hagoort et al., 1993). As described in more detail above, the P600 component is often associated to the processing of syntactic anomalies (Gouvea et al., 2010) and ambiguities (Frisch et al., 2002) in language processing. However, it has also been recognised as a response to rule violation in musical processing (Patel et al., 1998) and arithmetic tasks (Núñez-Peña and Honrubia-Serrano, 2004). More recently, the P600 component has been revisited and its role in language processing has been questioned, with studies suggesting the P600 is also sensitive to semantic anomalies (Van Herten et al., 2005) and research observing P600-like effects in the perception of semantics-free artificial languages (Tabullo et al., 2013).

Because of its extensively recognised role as a a response to rule violation in musical processing (Patel et al. 1998; Patel 1998; Koelsch et al. 2005; Koelsch 2009), a P600 responses was expected among the effects of experiment three of this research project. In experiment three participants were presented with sequences of sine-wave tones varying in pitch which could be interpreted as musical sequences. However, in experiments one and two, no syntactic mapping or violations were present among the set of nonce-word stimuli used across both experiments. However, more and more experiments have been eliciting P600 in contexts where neither syntactic ambiguity nor syntactic mismatch were present and, at the same time, more and more experiments, like some of the ones mentioned above, have been contextualising the P600 as a response to overall rule and sequence violation, regardless of the context being syntax in language, music or arithmetic functions.

A study on the processing of semantics-free artificial languages (Tabullo et al., 2013) observed a P600-like response to the perception of mismatching sequences that were similar across native language and artificial language processing. The authors discuss that, because the P600 has been observed as a response to sequence violation in an extremely wide array of linguistic and extra linguistic contexts, its observation was expected in the presentation of mismatching artificial language sequences. More specifically, the authors suggest that the P600-like effect they observed — and an N400-like deflection that was also elicited in their experiments regardless of the fact that all stimuli were not semantically activated — could be interpreted as a response correlated to the cost of *"unfulfilled predictions about incoming stimuli"* (Tabullo et al., 2013). This contextualisation of the observed P600 response and the similarities between Tabullo et al's (2013) study and experiments one and two of this research project provide further evidence towards this idea of a more generalised function of the P600 — as opposed to a more syntactic-specific role in language processing — which is observed as a very consistent response, both in terms of latency and topographical distribution, in all three experiments of this research project.

In particular, in experiment one, a P600-like response was observed whenever the second item of each of the semantics-free stimulus pairs was manipulated, hence creating a syllable sequence violation. Experiment one is the experiment whose methodology is the closest to that of Tabullo et al's (2013), considering a set of four nonce-word stimuli were taught to participants as a miniature artificial language, where syntactic rules were summarised by nonce-word transitional probabilities inside each one of the two pairs. In experiment one, furthermore, the P600 component was also observed in contexts of passive listening, which suggests that the component can be elicited as a response to sequence violation regardless of the participant's focussed attention. This is not new evidence considering that participants as young as two-year old children have elicited P600-like responses as a correlate of syntactic processing during passive tasks (Oberecker et al., 2005). However, the findings of experiment one strengthen the already existing theory that the P600 component is not only a response to sequence violations but one that is independent of participant's focussed attention to stimulus presentation.

In experiments two and three, the P600 was observed as a response to task-directed sequence violations in the perception of nonce words and sine-wave tones alike. All in all, while the more generalised contextualisation of the P600 as a non-linguistic response to sequence, rule violation and unfulfilled predictions of incoming stimuli is not necessarily as informative when it comes to exploring more focal topics for this research project such as phonological mapping in speech perception, it provides a general, consistent perspective of how mismatching stimuli elicit similar responses across three different experimental paradigms. Once again this interpretation displays how any one ERP component can hardly only be linked to one specific level of response and processing and that, similar but not too similar types of mismatch and sequence violations across a variety of contexts and stimulus types can, on the other hand, elicit the same one, very consistent ERP response across all three experiments. This interpretation of ERP behaviour is discussed in more detail when presenting one of the other possible explanations behind the non-elicitation of a PMN component, advanced in the following chapter.

Bibliography

Akima, H., Gebhardt, A., Petzold, T., Maechler, M., et al. (2016). Package 'akima'.

Arbel, Y., Spencer, K. M., and Donchin, E. (2011). The N400 and the P300 are not all that independent. *Psychophysiology*, 48(6):861–875.

Astheimer, L. B. and Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, 49(12):3512–3516.

Bates, D., Sarkar, D., Bates, M. D., and Matrix, L. (2007). The lme4 package. *R package version*, 2(1):74.

Behroozmand, R., Korzyukov, O., and Larson, C. R. (2011). Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. *Clinical neurophysiology*, 122(12):2408–2417.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.

Besson, M. and Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, 24(1):14–25.

Braine, M. D., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., and Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of memory and language*, 29(5):591–610.

Breznitz, Z. (2003). Speed of phonological and orthographic processing as factors in dyslexia: Electrophysiological evidence. *Genetic, Social, and General Psychology Monographs*, 129(2):183.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive science*, 41:1318–1352.

Brouwer, H. and Hoeks, J. C. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in human neuroscience*, 7:758.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., and Michie, P. T. (1998). Decrement of the n1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology*, 31(1):51–68.

Causeur, D., Sheu, C.-F., and Causeur, M. D. (2018). *Package 'ERP'*.

Christiansen, M. H. (2001). *Using artificial language learning to study language evolution: Exploring the emergence of word order universals.* na.

Connine, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32(2):193–210.

Connolly, J. F., D'Arcy, R. C., Kujala, A., Alho, K., et al. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, 12(2):237–243.

Connolly, J. F. and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of cognitive neuroscience*, 6(3):256–266.

Connolly, J. F., Phillips, N. A., Stewart, S. H., and Brake, W. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and language*, 43(1):1–18.

Connolly, J. F., Stewart, S., and Phillips, N. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and language*, 39(2):302–318.

Corretge, R. (2012). Praat vocal toolkit: A praat plugin with automated scripts for voice processing. *Software package (http://www. praatvocaltoolkit. com/index. html)*.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1):87–114.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., and Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human brain mapping*, 9(4):183–191.

Daltrozzo, J. and Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of cognitive neuroscience*, 21(10):1882–1892.

D'Arcy, R. C., Connolly, J. F., and Crocker, S. F. (2000). Latency shifts in the n2b component track phonological deviations in spoken words. *Clinical Neurophysiology*, 111(1):40–44.

D'Arcy, R. C., Connolly, J. F., Service, E., Hawco, C. S., and Houlihan, M. E. (2004). Separating phonological and semantic processing in auditory sentence processing: A high-resolution event-related brain potential study. *Human brain mapping*, 22(1):40–51.

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

Desroches, A. S., Newman, R. L., and Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, 21(10):1893–1906.

Diaz, M. T. and Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain research*, 1146:85–100.

Dyke, F. B., Leiker, A. M., Grand, K. F., Godwin, M. M., Thompson, A. G., Rietschel, J. C., McDonald, C. G., and Miller, M. W. (2015). The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity. *International Journal of Psychophysiology*, 95(1):56–62.

Erlbeck, H., Kübler, A., Kotchoubey, B., and Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Frontiers in human neuroscience*, 8:654.

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of german wh-questions. *Journal of Memory and Language*, 47(2):250–272.

Finley, S. and Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of memory and language*, 61(3):423–437.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Folia, V., Uddén, J., De Vries, M., Forkstam, C., and Petersson, K. M. (2010). Artificial language learning in adults and children. *Language learning*, 60:188–220.

Fosker, T. and Thierry, G. (2005). Phonological oddballs in the focus of attention elicit a normal p3b in dyslexic adults. *Cognitive brain research*, 24(3):467–475.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., et al. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing*.

Friederici, A. D. (2006). The neural basis of sentence processing: Inferior frontal and temporal contributions. *Broca's region*, pages 196–217.

Friederici, A. D. and Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *Neuroimage*, 20:S8–S17.

Friederici, A. D., Steinhauer, K., and Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & cognition*, 27(3):438–453.

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, 99(1):529–534.

Frisch, S., Schlesewsky, M., Saddy, D., and Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, 85(3):B83–B92.

Froyen, D., Van Atteveldt, N., Bonte, M., and Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience letters*, 430(1):23–28.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance*, 6(1):110.

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

Getz, L. M. and Toscano, J. C. (2021). The time-course of speech perception revealed by temporally-sensitive neural measures. *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(2):e1541.

Glass, A. V., Gazzaniga, M. S., and Premack, D. (1973). Artificial language training in global aphasics. *Neuropsychologia*, 11(1):95–103.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological review*, 105(2):251.

Gómez, R. L. and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, 4(5):178–186.

Gott, P. S., Rabinowicz, A. L., and DeGiorgio, C. M. (1991). P300 auditory event-related potentials in nontraumatic coma: association with Glasgow Coma Score and awakening. *Archives of Neurology*, 48(12):1267–1270.

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes*, 25(2):149–188.

Gow Jr, D. W. and Olson, B. B. (2015). Lexical mediation of phonotactic frequency effects on spoken word recognition: A Granger causality analysis of MRI-constrained MEG/EEG data. *Journal of memory and language*, 82:41–55.

Grau, C., Fuentemilla, L., and Marco-Pallarés, J. (2007). Functional neural dynamics underlying auditory event-related N1 and N1 suppression response. *Neuroimage*, 36(3):522–531.

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review. *Psychophysiology*, 48(12):1711–1725.

Guillem, F., N'KAOUA, B., Rougier, A., and Claverie, B. (1995). Intracranial topography of event-related potentials (N400/P600) elicited during a continuous recognition memory task. *Psychophysiology*, 32(4):382–392.

Gunter, T. C., Stowe, L. A., and Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34(6):660–676.

Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes*, 8(4):439–483.

Hagoort, P. and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38(11):1531–1549.

Hagoort, P., Brown, C. M., and Osterhout, L. (1999). The neurocognition of syntactic processing. *The neurocognition of language*, pages 273–316.

Halekoh, U., Højsgaard, S., Højsgaard, M. S., and Matrix, I. (2017). *Package 'pbkrtest'*.

Hickok, G. and Poeppel, D. (2016). Neural basis of speech perception. *Neurobiology of language*, pages 299–310.

Hintzman, D. L. (1986). " schema abstraction" in a multiple-trace memory model. *Psychological review*, 93(4):411.

Invitto, S., Piraino, G., Mignozzi, A., Capone, S., Montagna, G., Siciliano, P. A., Mazzatenta, A., Rocco, G., De Feudis, I., Trotta, G. F., et al. (2018). Smell and meaning: an OERP study. In *Multidisciplinary Approaches to Neural Computing*, pages 289–300. Springer.

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, 15(2):159–201.

Kanske, P., Plitschka, J., and Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, 49(11):3121–3129.

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, 46(1):179–190.

Koelsch, S., Gunter, T. C., SchroÈger, E., Tervaniemi, M., Sammler, D., and Friederici, A. D. (2001). Differentiating ERAN and MMN: an ERP study. *NeuroReport*, 12(7):1385–1389.

Koelsch, S., Gunter, T. C., Wittfoth, M., and Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of cognitive neuroscience*, 17(10):1565–1577.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature neuroscience*, 7(3):302–307.

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., and Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with meg. *Cognitive brain research*, 21(1):106–113.

Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62:621–647.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.

Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947):161.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H., et al. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software*, 82(13):1–26.

Lagrois, M.-É., Peretz, I., and Zendel, B. R. (2018). Neurophysiological and behavioral differences between older and younger adults when processing violations of tonal structure in music. *Frontiers in neuroscience*, 12:54.

Ledwidge, P. S. (2017). *N400, P600, and late sustained frontal positivity event-related brain potentials reflect distinct psycholinguistic processes during the comprehension of contextually ambiguous narrative discourses: Implications for the neural organization of the language processing system.* PhD thesis, Department of Psychology, University of Nebraska Lincoln.

Leek, J. T. and Storey, J. D. (2008). A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences*, 105(48):18718–18723.

Lenth, R., Singmann, H., Love, J., and Buerkner, P. (2019). emmeans: estimated marginal means, aka least-squares means. R package.

Lewendon, J., Mortimore, L., and Egan, C. (2020). The phonological mapping negativity: History, inconsistency, and future direction. *Frontiers in Psychology*, 11:1967.

Lopez-Calderon, J. and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, 8:213.

Luck, S. J. (2012). *The Oxford handbook of event-related potential components (p. 329–360)*, chapter Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. Oxford university press.

Luck, S. J. and Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54(1):146–157.

Luck, S. J. and Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components.* Oxford university press.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of christmash past. *Cognitive Science*, 27(2):285–298.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. In *Attention and performance: Control of language processes*, pages 125–150. Erlbaum.

Martín-Loeches, M., Casado, P., Gonzalo, R., De Heras, L., and Fernández-Frías, C. (2006). Brain potentials to mathematical syntax problems. *Psychophysiology*, 43(6):579–591.

McClelland, J. L. and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1):1–86.

Näätänen, R. (1991). Mismatch negativity outside strong attentional focus: A commentary on Woldorff et al.(1991). *Psychophysiology*, 28(4):478–484.

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

Näätänen, R. and Alho, K. (1995). Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience*, 80(1-4):317–337.

Näätänen, R., Paavilainen, P., Titinen, H., Jiang, D., and Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, 30(5):436–450.

Näätänen, R. and Picton, T. (1987). The n1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

Newman, R., Connolly, J., Service, E., and McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, 40(4):640–647.

Newman, R. L. and Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological psychology*, 80(1):114–121.

Nigam, A., Hoffman, J. E., and Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of cognitive neuroscience*, 4(1):15–22.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3):189–234.

Núñez-Peña, M. I. and Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, 18(2):130–141.

Nyman, G., Alho, K., Laurinen, P., Paavilainen, P., Radil, T., Reinikainen, K., Sams, M., and

Näätänen, R. (1990). Mismatch negativity (MMN) for sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 77(6):436–444.

Oades, R. D. and Dittmann-Balcar, A. (1995). Mismatch negativity (MMN) is altered by directing attention. *Neuroreport*, 6(8):1187–1190.

Oberecker, R., Friedrich, M., and Friederici, A. D. (2005). Neural correlates of syntactic processing in two-year-olds. *Journal of cognitive neuroscience*, 17(10):1667–1678.

Patel, A. D. (1998). Syntactic processing in language and music: different cognitive operations, similar neural resources? *Music perception*, 16(1):27–42.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of cognitive neuroscience*, 10(6):717–733.

Pazo-Alvarez, P., Cadaveira, F., and Amenedo, E. (2003). Mmn in the visual modality: a review. *Biological psychology*, 63(3):199–236.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12(6):1038–1055.

Plourde, G. and Boylan, J. F. (1991). The long-latency auditory evoked potential as a measure of the level of consciousness during sufentanil anesthesia. *Journal of cardiothoracic and vascular anesthesia*, 5(6):577–583.

Polich, J. and McIsaac, H. K. (1994). Comparison of auditory P300 habituation from active and passive conditions. *International Journal of Psychophysiology*, 17(1):25–34.

Proverbio, A. M. and Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience letters*, 459(3):142–146.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in cognitive sciences*, 5(12):517–524.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabovsky, M. and McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, 132(1):68–89.

Robson, H., Pilkington, E., Evans, L., DeLuca, V., and Keidel, J. L. (2017). Phonological and semantic processing during comprehension in Wernicke's aphasia: An N400 and Phonological Mapping Negativity Study. *Neuropsychologia*, 100:144–154.

RStudio Team (2020). *RStudio: Integrated Development Environment for R.* RStudio, PBC., Boston, MA.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125(1):28.

Samuel, A. G. and Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48(2):416–434.

Sassenhagen, J. and Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, 200:425–436.

Scharenborg, O. and Boves, L. (2010). Computational modelling of spoken-word recognition processes: Design choices and evaluation. *Pragmatics & Cognition*, 18(1):136–164.

Schön, D. and Besson, M. (2005). Visually induced auditory expectancy in music reading: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, 17(4):694–705.

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3):341–349.

Scott, S. K. (2017). The neurobiology of speech perception. In *Twenty-first century psycholinguistics: Four cornerstones*, pages 141–156. Routledge.

Sheu, C.-F., Perthame, E., Lee, Y.-S., Causeur, D., et al. (2016). Accounting for time dependence in large-scale multiple testing of event-related potential data. *The Annals of Applied Statistics*, 10(1):219–245.

Shtyrov, Y., Pulvermüller, F., Näätänen, R., and Ilmoniemi, R. J. (2003). Grammar processing outside the focus of attention: an MEG study. *Journal of Cognitive Neuroscience*, 15(8):1195–1206.

Silva-Pereyra, J. F. and Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain research*, 1185:201–211.

Skrandies, W. and Reuther, N. (2008). Match and mismatch of taste, odor, and color is reflected by electrical activity in the human brain. *Journal of Psychophysiology*, 22(4):175.

Slevc, L. R. and Patel, A. D. (2011). Meaning in music and language: three key differences. *Physics of Life Reviews*, 8(2):110–111.

Stevens, C. J., Keller, P. E., and Tyler, M. D. (2013). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music*, 41(1):59–74.

Sun, Y., Zhang, N. R., Owen, A. B., et al. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *The Annals of Applied Statistics*, 6(4):1664–1688.

Systems, N. (2004). *Presentation [Computer Software]*. Albany, CA.

Tabullo, Á., Sevilla, Y., Segura, E., Zanutto, S., and Wainselboim, A. (2013). An ERP study of structural anomalies in native and semantic free artificial grammar: evidence for shared processing mechanisms. *Brain Research*, 1527:149–160.

Van Den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus n400 effects. *Journal of cognitive neuroscience*, 13(7):967–985.

Van Herten, M., Kolk, H. H., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2):241–255.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., and Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2):394.

Verbaten, M., Roelofs, J., Sjouw, W., and Slangen, J. L. (1986). Habituation of early and late visual ERP components and the orienting reaction: The effect of stimulus information. *International Journal of Psychophysiology*, 3(4):287–298.

Verhees, M. W., Chwilla, D. J., Tromp, J., and Vissers, C. T. (2015). Contributions of emotional state and attention to the processing of syntactic agreement errors: evidence from p600. *Frontiers in psychology*, 6:388.

Wang, A., Mouraux, A., Liang, M., Iannetti, G., and Lauwereyns, J. (2008). The enhancement of the N1 wave elicited by sensory stimuli presented at very short interstimulus. *PloS one*, 3(12).

Weber, A. and Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3):387–401.

Weber, C., Hahne, A., Friedrich, M., and Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research*, 18(2):149–161.

Weber, K. and Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology*, 45(6):920–925.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, 90(18):8722–8726.

Woldorff, M. G., Hackley, S. A., and Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28(1):30–42.

Woolrich, M. W., Beckmann, C. F., Nichols, T. E., and Smith, S. M. (2009). Statistical analysis of fmri data. In *fMRI techniques and protocols*, pages 179–236. Springer.

Wronka, E., Kaiser, J., and Coenen, A. M. (2008). The auditory P3 from passive and active three-stimulus oddball paradigm. *Acta Neurobiol Exp*, 68:362–372.

Zendel, B. R. and Alexander, E. J. (2020). Autodidacticism and music: Do self-taught musicians exhibit the same auditory processing advantages as formally trained musicians? *Frontiers in Neuroscience*, 14.