The MMN by another name? Exploring the autonomy of the Phonological Mapping (Mismatch) Negativity

Keywords: Event-related potentials, PMN, MMN, language, speech processing, neuroscience of language

Authors: Jen Lewendon[1], James Britton[1] & Stephen Politzer-Ahles[1]

1. Hong Kong Polytechnic University

Abstract

The Phonological Mapping Negativity (PMN) is an event-related potential component thought to index pre-lexical phonological processing. The response has long been considered distinct from the temporally-proximate Mismatch Negativity (MMN) - a distinction that primarily rests on the assumption that the PMN, unlike the MMN, cannot be elicited in inattentive contexts, thus implying differing underlying auditory-cortex mechanisms. Despite this, no study to date has established whether elicitation of an inattentive PMN response is possible. Here, we tested this assumption in two experiments during which participants heard phonological mismatches whilst engaging in a distractor task (experiment 1) or watching a film (experiment 2). Our results showed no consistent evidence for an inattentive PMN. Though attention may indeed serve to distinguish the two components, our results highlight consistent discrepancies in the temporal, topographical, and functional characteristics of the PMN that undermine efforts to establish its significance in the electrophysiological timeline of speech processing.

Introduction

Event-related potential (ERP) components are typically identified through a combination of their polarity, scalp distribution, latency, and sensitivity to task manipulations (Woodman, 2010). Substantial efforts to determine the specific manipulations that a range of ERP responses are (and are not) sensitive to underlie the use of such components as dependent measures in psychological research. This process of carefully characterizing ERP effects in terms of their functional sensitivity remains a crucial prerequisite to their use as markers of specific aspects of information processing (Donchin, 1984). The current study focuses on the Phonological Mapping Negativity (PMN), an ERP component established in the early 1990s, and since utilised as a dependent measure across a range of language research fields. Although there is a general consensus that the PMN represents an autonomous and discrete component, its spatiotemporal characteristics and functional sensitivities have much in common with the earlier Mismatch Negativity (MMN) and later N400 components. Such similarities have recently prompted debate about whether the PMN is really separate from these other components (Lewendon, Mortimore & Egan, 2020). In particular, whilst a vast amount of work has been dedicated to investigating the independence of the PMN from the later N400, there is little research available to support the widely-accepted distinction drawn between the PMN and the earlier MMN. Below, we describe what is known so far about the similarities and differences between the PMN and MMN, before outlining the current paper's contribution to this debate.

The PMN is said to index phonological processing at the pre-lexical level, occurring maximally around 300 ms post stimulus presentation across fronto-central electrode sites[1] (Connolly et al., 2001; Connolly & Phillips, 1994). The component is typically elicited in paradigms that generate phonological expectancy using words (Newman & Connolly, 2009) or sentences varying in cloze probability (Desroches, Newman, & Joanisse, 2009; Diaz & Swaab, 2007; Connolly, Service, D'Arcy, Kujala & Alho., 2001; Connolly & Phillips, 1994). For example, in a phoneme deletion task, Newman et al.

---

[1] See below for various reports of central and parietal PMN effects.

(2003) instructed participants to omit the initial phonemes from words (*clap* without the /k/), and subsequently played them with either a correct (*lap*) or incorrect (*cap*, *ap*, *nose*) target. When the expectation of specific phonological input (e.g., *lap*) was violated by an incorrect target (*cap*, *ap*, or *nose*), the PMN increased in amplitude relative to the expected correct condition, whilst lexicality did not appear to affect the response.

The increase in PMN amplitude to items that violate phonological predictions, independently of lexico-semantic factors, is thought to demonstrate its specific sensitivity to phonological manipulations. This selective sensitivity was a major focus of early research into the effect (Connolly, Phillips, Stewart, & Brake, 1992; Connolly & Phillips, 1994; Praamstra, Meyer, & Levelt, 1994; Hagoort, & Brown, 2000; van den Brink, Brown, & Hagoort, 2001; D'Arcy, Connolly, Service, Hawco, & Houlihan, 2004). Such work primarily sought to distinguish the PMN from the N400, which – albeit perhaps best known for its sensitivity to semantic manipulations – is widely acknowledged to be modulated by a range of phonological factors including phonological relatedness (Praamstra & Stegeman, 1993; Bölte and Coenen, 2002), and phonological prediction mismatch (Newman & Connolly, 2009). Notable overlaps in the functional sensitivity of the two components, alongside similarities in topography and latency, prompted considerable investigation into the degree to which the PMN might prove similarly sensitive to semantic factors. Nonetheless, despite substantial efforts, whether the two effects do indeed represent functionally distinct components remains the subject of ongoing debate (see Lewendon, Mortimore & Egan, 2020 for a review).

In the electrophysiological timeline of auditory language processing, an earlier response, typically elicited via the oddball paradigm, is the Mismatch Negativity (MMN). First discovered by Näätänen, Gaillard and Mäntysalo in 1978, the MMN is a frontocentral component peaking between 150–250 ms post stimulus onset, and is known to respond to violations of rules established by sequences of stimuli (Näätänen, 1992). In oddball paradigms, ‘deviant’, infrequent items such as tones, phonemes or syllables presented at random between ‘standard’, frequent stimuli elicit the response. Whilst overlaps (or proposed lack thereof) in the characteristics and functional sensitivity of the N400 and PMN have attracted considerable attention, the relationship between Mismatch Negativity and the PMN has been investigated to a far lesser extent. In some ways this appears intuitive, given the clear differences in standard paradigms and stimuli used to elicit the MMN and PMN. Nonetheless, when considered in light of similarities in their typical characteristics and functional sensitivity, certain striking similarities become apparent.

i) Scalp distribution

The MMN has a frontocentral topography when measured with a mastoid reference (Garrido et al., 2009; Näätänen et al., 2007). The generation of a MMN response is thought to result from a bilateral supratemporal process at the auditory cortex (producing a supratemporal MMN), and a frontal right-hemispheric process (producing a frontal MMN) (Paavilainen, 2013; Näätänen et al., 1978, Giard et al., 1990, Baldeweg et al., 1999, Rinne et al., 2000; Näätänen et al., 2007). The PMN is typically reported to be a frontocentral component (Newman et al. 2009; Connolly et al., 2001; D'Arcy et al., 2004; Lee, Harkrider & Hedrick, 2012), although a number of studies reveal PMN effects additionally extending towards temporal (Newman et al., 2003), central (Connolly et al., 1990), and parietal sites (D’Arcy et al., 2000), or with midline distribution (Connolly et al., 1992; Connolly & Phillips, 1994). In a recent review, Lewendon et al. (2020) proposed that whilst central and parietal PMN effects might plausibly be considered related, or even the result of early N400 activity, frontocentral PMN effects necessitated further consideration in regard to whether or not they are independent from the MMN.

Limited research has been conducted into the neurobiological processes underlying the PMN response, although two source-localisation studies suggest that the PMN may be generated in the anterior parts of the temporal cortex in phonological prediction tasks (Kujala et al., 2004) or from two bilateral sources in the inferior frontal regions, with left activation asymmetry in sentence-matching tasks (D’arcy et a., 2004). Despite the clearly overlapping scalp topographies of the MMN and frontal PMN reports, no study thus far has compared source loci of the PMN to that of the MMN elicited by a phoneme change (Näätänen & Alho, 1995).

ii) Latency

Perhaps then latency may serve to distinguish the two components? The MMN, which generally peaks between 150-250 ms post onset of the deviance (Kujala et al., 2007), occurs somewhat earlier than the PMN, which is typically maximal between 200-300 ms post stimulus onset[2]. Nonetheless, the latency of the MMN is notably variable, and it can occur as late as 300 ms if the stimuli or the rule that distinguishes standards and deviants is particularly complex or if the distinction between standards and deviants is subtle (Näätänen et al., 1989, 1997, 2007). Such variability in latency is interesting when considered in light of a recent neurobiological theory of language-related negativities outlined by Bornkessel-Schlesewesky and Schlesewesky (2019). In a move away from the traditional linguistic or cognitive functional interpretations of language-related ERP negativities, the authors outline a precision-weighted predictive coding account in which variable latency and topography of such components reflect stimulus complexity, and longer temporal receptive windows, as opposed to qualitatively different underlying mechanisms. Considering the PMN and MMN in light of this account in their review, Lewendon et al. (2020) propose that the two components could plausibly represent different points within a spectrum of electrophysiological responses to auditory prediction violation. That is, the classic MMN may represent an earlier, more rudimentary reaction generated from less complex information processing, with the later PMN a more sophisticated response to prediction violation generated by inputs or predictions of greater complexity.

iii) Sensitivity

The classic auditory MMN is one elicited by deviant sounds presented within a string of standard sounds, whilst the PMN responds to word and sentence-level phonological predictions. While these two methods of elicitation are superficially different, both ultimately rely on making the listener generate predictions about upcoming auditory forms. In an oddball paradigm these predictions result

---

[2] It should be noted that earlier research reports PMN windows beginning as early as 150 ms post stimulus onset (Van den Brink, Brown, & Hagoort, 2001; Connolly, Stewart & Phillips, 1990; Connolly, Phillips, Stewart & Brake, 1992), but that later research has typically selected later time windows for analysing the PMN.

from habituation to the standard stimuli, whilst in PMN paradigms the prediction is generated via task instructions (e.g., phoneme deletion/addition) or from the sentence context. Although the means of generating predictions differ between these paradigms, both approaches ultimately produce MMN and PMN effects that result from a mismatch between the prediction and the auditory input.

In light of the clear parallels between the sensitivities of the two components, arguments to support the distinction between the PMN and MMN might then fall back on the notion of higher-level top-down controlled prediction processes underlying PMN, as compared to lower-level automatic process for the MMN. However, evidence that the MMN is sensitive to phonology/phonotactics (Bonte et al. 2005; Weber et al., 2004; Eulitz & Lahiri, 2004) and even voice familiarity (Beauchemin et al., 2006), alongside recent reports of a ‘late MMN’ in response to auditory rule extraction (Zachau, et al., 2005), demonstrates its sensitivity to a range of higher-order manipulations, serving only to further blur distinctions between the functional sensitivities ascribed to the PMN and MMN.

While the PMN and MMN seem similar in terms of the characteristics described above, there remains one key characteristic that may differ between them: the effect of attention. The MMN is considered pre-attentive, in that it can be elicited by stimuli that participants are not attending to (Näätänen, Paavilainen, Rinned & Alho, 2007). The PMN, on the other hand, is thought to require participants' attention (Näätänen et al., 2007). Such a distinction has formed the basis of arguments that differing auditory-cortex mechanisms may underlie the two components. Despite this, to our knowledge no study to date has tested whether elicitation of an inattentive PMN response is possible. Whilst it remains plausible that attention may distinguish the two effects, if we are to claim that the PMN and MMN are distinct components on this basis, then there needs to be positive evidence showing that elicitation of the PMN does indeed require attention (i.e., not just that tasks involving attention to violations do elicit the PMN – as extant studies have shown – but also that tasks without attention to violation do not elicit PMN).

The present paper thus aims to investigate this putative feature of the PMN used to support its distinction from the MMN, i.e., the necessity of participant attention in order to elicit the response. In experiment

1, we tested whether an inattentive PMN response could be elicited via an auditory prediction task in which participants' attention was directed to either musical or linguistic stimuli. In experiment 2, we investigated whether an inattentive PMN response could be elicited in a more typical MMN-like paradigm in which participants attended to a silent film, whilst passively listening to phonological input.

1. Experiment 1

In experiment 1, we sought to determine whether an inattentive PMN effect could occur whilst participants were actively engaged in distraction task. To do so, we presented participants with speech and musical note stimuli simultaneously with the instruction to pay attention to just one of them. For example, in one condition participants were told that they would hear words like *banana*, and that they would also hear some musical notes at the same time but should ignore them; in another condition, participants were told that they would hear a series of musical notes, and that they would also hear a word at the same time but should ignore it. In each of these conditions, the actual stimulus remained the same (i.e. /bəˈ/[C] /nɑː/[B] /nə/[A], where A/B/C represent the musical notes played concurrently with the syllables), and all that differed was the instructions about what participants should attend to. Participants were asked to mentally delete either the first syllable or the first note (corresponding to the stimuli-type to which they attended). Subsequently, participants heard a target item which either matched or mismatched their pre-generated expectations. For example, if participants heard /bəˈ/[C] /nɑː/ [B] /nə/ [A] when instructed to pay attention to the words, then they should mentally delete the first syllable /bəˈ/ and thus expect to hear /ʼnɑː.nə/ next; in the next portion of the trial, if they did indeed hear /ʼnɑː.nə/ they should press the "match" key, whereas if they hear /rɑː.nə/ they should press the "mismatch" key. Crucially, this allowed us to expose participants to phonological violations under attentive conditions (e.g., participants hearing /rɑː/ [B] /nə/[A] when they were expecting *nana*) or inattentive conditions (e.g., participants hearing /rɑː/ [B] /nə/[A] when they were instructed to focus on notes and thus were expecting [B][A], but nevertheless had been exposed to /bəˈnɑː.nə/). We hypothesised that if the PMN occurs inattentively, phoneme prediction violations for unattended phoneme-mismatch stimuli should elicit a PMN response with the same characteristics (latency/topography) as the attentive phoneme-mismatch effect.

(Figure 1 here)

For example, a target like ‘nana’ (/ˈnɑː.nə/) would elicit a more negative ERP if it followed a phonologically mismatching prime like piranha (/pɪˈrɑː.nə/) than if it follows a phonologically matching prime like banana (/bəˈnɑː.nə/), and this would occur regardless of whether participants were instructed to make their match-mismatch judgment on the basis of the notes or the phonemes (see Figure 1). Alternatively, if the nature of phonological prediction required to elicit a PMN necessitates participant attention, a mismatch effect should occur for the attended condition only.

We predicted that, if the PMN occurs inattentively, the inattentive mismatch effect should result in significantly more negative mean amplitudes than for the inattentive match condition, across similar electrodes as compared to those for attentive mismatch effect within the 150 - 380 ms time window (i.e., the earliest and latest reported PMN onsets and offsets, see van den Brink, Brown, & Hagoort, 2001; Connolly, Stewart & Phillips, 1990; Connolly, Phillips, Stewart & Brake, 1992; Jachmann, Drenhaus, Staudte, & Crocker, 2019, Jones, Kuipers, Thierry 2016). Alternatively, if generation of the PMN necessitates attention, we would expect no significant difference between inattentive match vs inattentive mismatch conditions.

1.1. Materials and Methods

1.1.1. Participants

Fifty-four native English speakers took part in the experiment and gave written informed consent prior to participating (application no. HSEARS20210812003). Nine participants were excluded prior to data analysis – one due to disclosure of a congenital condition subsequent to data acquisition, one due to poor ICA decomposition, one due to a technical issue during EEG recording, and six due to low trial counts after artefact rejection and exclusion of incorrect trials (< 50%). Forty-five participants were retained in the final analysis (28 female, mean age = 38.4; *SD* = 9.4) with normal or corrected-to-normal vision, no learning disabilities, and self-reported normal hearing. All spoke English from birth at home as their native language. Eight participants were raised in bilingual households but all considered their proficiency in English to be equal or superior to that of their other language. Forty-two participants were right-handed, whilst three participants were left-handed[3]. Participants had on average 2.9 years musical experience (*SD* = 4.3), with 16 participants having no musical experience beyond school lessons, only 9 participants playing an instrument casually as an adult. For full participant demographics, see Supplementary File 1.

1.1.2. Stimuli

Word primes were 40 trisyllabic words with 2nd syllable stress (e.g., *fantastic*) and a mean Subtlex Zipf frequency of 3.56 (*SD* = 0.84). Targets were pseudowords that either corresponded to the prime without the first syllable (e.g., *tastic*) or with the incorrect initial phoneme (e.g., *castic*). Tune primes were tunes consisting of three-note melodies with an equal number of major and minor note patterns. Tunes followed fairly simple, common melodic patterns to ensure the major/minor key was clearly established and facilitate ease of memorisation. Stimuli were paired such that the correct target for a given word/tune prime formed the incorrect target for another word/tune prime, thus enabling full rotation of target stimuli across all conditions. Example stimuli are given in Table 1.

(Table 1 here)

Prime-target pairings were as follows (Example prime: /æm/(*A4*)/ˈbɪʃ/(*B4*)/əs/(*C5*)): (i) correct note & correct phoneme (e.g., /ˈbɪʃ/(*B4*)/əs/(*C5*)); (ii) correct note & incorrect phoneme (e.g., /ˈlɪʃ/(*B4*)/əs/(*C5*));

---

[3] Removal of non-right-handed participants did not change the trend of effects reported.

(iii) incorrect note & correct phoneme (e.g., /ˈbɪʃ/(*E5*)/əs/(*C5*)); (iv) incorrect note & incorrect phoneme (e.g., /ˈlɪʃ/(*E5*)/əs/(*C5*)). Pairings (iii) and (iv) formed a separate experiment, and were included to both form the tune task (i.e. to create an inattentive phoneme mismatch condition) and to ensure that there were an equal number of match and mismatch targets. Participants were instructed to attend either to the word or the tune, whilst ignoring other input. Thus, we used to the above pairings to form the following conditions as laid out in Table 2.

(Table 2 here)

Word primes were recorded by a native British female speaker in a sound attenuated room using an Audio-technica AT2035 cardioid condenser microphone at 44100 Hz sampling frequency. Following recording, the audio files first underwent noise reduction (freq. 12 dB, sensitivity 6, frequency smoothing bands 3) and were amplified by 11.292 dB in Audacity. Prime stimuli intensity was averaged at 70 dB (mean power in air: $9.99999357 \times 10^{-6}$ Watt/m²), and were subsequently segmented at zero-crossings to create three individual recordings for each syllable in Praat. Tune primes were created using Sibelius and saved at 44100 Hz sampling frequency. The recording was amplified (11.292 dB) and segmented to create individual audio files for each note that were 85% of the shortest matching syllable segment duration in any match/mismatch pair. For example, taking the given targets ***bi****tious* and ***li****cious* from Table 1 and assuming the segments *bi* and *li* were 100 ms and 120 ms respectively, both tones B4 and E5 for these words would be segmented to 85ms duration. Finally, fade-out was applied in Audacity® (Version 3.1.3, 1999-2021 Audacity Team) to first and last 50ms of each individual note recording, and the mean intensity for each note was matched to the corresponding segment intensity (e.g., *am* & A4 = 60dB, *bi* & B4 = 70dB, *tious* & C5 = 80dB). Tune and word sound files were then combined, and 10 ms ramp up/down was applied to the start and end of all prime and target stimuli.

1.1.3. Procedure

The experiment was performed using Presentation® software (Version 20.0 Build: 09.04.17, Neurobehavioral Systems, Inc., Berkeley, CA) and consisted of two sections, one in which participants were instructed to focus on the word portion of the stimuli and ignore the tune, and another in which they were instructed to focus on the tune portion of the stimuli and ignore the words, with task order counterbalanced across participants. Audio stimuli were presented via sound-cancelling ear inserts (Etymotic, Inc.), and participants were told to listen carefully to the stimuli and take away the first syllable/note of the prime in their head. They were then asked to hold the anticipated initial sound of the target item in their mind until target onset.

Both sections of the experiment consisted of a training session of 10 stimuli, in which participants practised either syllable or note deletion, followed by 8 experiment blocks of 20 items each. In any given trial, participants first saw a fixation cross (200 ms), which stayed on the screen during the presentation of an auditory prime. Next, an inter-stimulus interval randomly jittered between 1900 and 2100 ms separated the offset of the auditory prime and the onset of the auditory target. ISI duration was selected based on prior research using a similar paradigm and a 2 second ISI to allow time for generation of predictions (Newman et al., 2003; Newman & Connolly, 2009). Immediately following the offset of the audio target (average duration 830 ms, SD = 120 ms), a response screen consisting of the words 'match' and 'mismatch' on opposite sides of the screen was presented. Match/mismatch side was randomised throughout the experiment to avoid response preparation during the target period. Participants were asked to press a button to indicate whether the target matched or mismatched their expectations. The participant's button press immediately triggered the start of the subsequent trial. A schematic of the experimental procedure can be seen in Figure 2.

(Figure 2 here)

In each block, only one prime and corresponding melody from each of the prime pairs (e.g., *ambitious, delicious*) was heard, paired with one target item. Blocks were organised such that all tunes within a single block were either in a major or minor key for ease of melody prediction, and every block contained and equal number of conditions. The order of the blocks and the order of the items within a block were randomised. For full stimuli rotations and audio files, see Supplementary File 2. To ensure minimal audio latency jitter, the experiment was performed using the Presentation mixer Exclusive mode on a Dell Precision T1700, Windows 7 Professional (x64) with a Sound Blaster Z (Creative Technology) sound card, and a Cedrus StimTracker to send trigger events.

1.1.4. Data acquisition & pre-processing

EEG data were recorded and digitized at the sampling rate of 1000 Hz using a SynAmpsRT amplifier (Neuroscan, Charlotte, NC, U.S.) with 64 Ag/AgCl electrodes and an online reference located between Cz and CPz. Prior to recording, a cap was fitted to secure the EEG electrodes in place on the scalp at specific locations according to the extended international 10–20 system. Four bipolar facial electrodes were placed lateral to the outer canthi of each eye and in the inferior and superior areas of the left orbit provided recordings of the horizontal and vertical electrooculograms (EOG). Electrode impedances were reduced to < 5 kΩ and a 400Hz (gain 10) online low-pass filter was applied during recording. Offline, data were first high-pass filtered using 0.1 Hz (half-amp -6dB) IIR Butterworth Zero Phase shift filter, with a 24 db/oct roll-off and DC offset removed, and subsequently cleaned via visual inspection to remove pauses and breaks during the experiment from the EEG recording. Following this, data was re-referenced offline to the average of the left and right mastoids. Ocular correction was conducted using Independent Component Analysis (ICA)[4] following visual inspection of the data using the RUNICA algorithm, resulting in an average of 2.07 components removed per participant (range: 1-3). Data were then segmented into epochs ranging from −200 to 800 ms time-locked to the onset of the auditory target, and baseline corrected relative to 200 ms pre-stimulus activity. Automatic artifact rejection was conducted to remove extreme values (±100 µV) and data was low-pass filtered 30 Hz (24 dB/oct) IIR Butterworth Zero Phase shift filter. 1.02% of trials were removed during artifact rejection, and trials in which the participant responded incorrectly were excluded, leaving an average of 36.77/40

---

[4] NB. For two participants, ICA weights and scaling were taken from 1. An aggressively cleaned dataset, or 2. Aggressively high-pass filtered data (1.0Hz cut-off) to improve ICA decomposition.

(SD=3.64) trails in the attentive match condition, 31.22/40 (SD=4.01) in the attentive mismatch condition, 38.44/40 (SD=3.48) in the inattentive match condition, and 37.44/40 (SD=3.70) in the inattentive mismatch condition. All data pre-processing was conducted in MATLAB, aided by EEGLAB (Delorme & Makeig, 2004) functions, and the full pre-processing scripts can be found online.

1.1.5. Data analysis

Variation in reported PMN characteristics in literature using phoneme deletion paradigms (Newman & Connolly 2009; Newman, Connolly, Service & Mcivor, 2003) rendered selection of *a priori* topography and latency predictions difficult. Instead, we took the approach of conducting spatiotemporal cluster-based analyses (Maris & Oostenveld, 2007) for the contrasts between attentive mismatch & match and inattentive match & mismatch implemented in the FieldTrip toolbox. Given the potential influence of task, comparison of waveforms from the attentive match (in which participants' focus was directed towards the phonological task) and inattentive mismatch (where attention was directed towards the tone task) conditions was deemed too likely to reflect task-demand differences. At such, predictions for 'inattentive mismatch', as stated above, pertain only to inattentive mismatch vs. inattentive match comparisons. For each of the two comparisons (i.e., attentive mismatch vs attentive match , and inattentive match vs inattentive mismatch), paired t-tests across the 45 participants were conducted at every channel and every sample from 150 to 380 ms. Clusters were formed of spatiotemporally adjacent datapoints with significant tests (with one-tailed alpha <.05), and neighbours of at least 2 other channels that met the p-value criterion at the same time point. We quantified each cluster by summing the *t*-statistics within the cluster, and the highest such sum was used as the test statistic for the ensuing permutation test. The significance of this test statistic was non-parametrically evaluated by permuting the data 5000 times, with the clustering and measurement method as outlined above repeated for each permutation. The proportion of permutation test statistics larger than the observed cluster test statistic was the *p*-value for that comparison.

1.2. **Results**

Although tune mismatch conditions were included in the experimental paradigm, all analyses were conducted on trials with correct tunes, such that attentive and inattentive phonological match and mismatch were always heard alongside tune matches.

1.2.1. Behavioural Results

Overall accuracy in accepted trials was 94% in the attentive match condition, 75% in the attentive mismatch condition, 99% in the inattentive match condition, and 83% in the inattentive mismatch condition.

1.2.2. ERP Results

The spatiotemporal cluster-based analysis for the attentive mismatch effect revealed significantly more negative ERPs in response to phonologically mismatching targets as compared to matching targets (p =.044). As shown in the raster plot, this increase in negativity was driven by a cluster occurring about 288-380 ms across central, parietal and occipital electrodes (Figure 3).

(Figure 3 here)

(Figure 4 here)

A second cluster test on inattentive match vs inattentive mismatch failed to find more negative ERPs for mismatch than match (p=.19).

(Figure 5 here)

Finally, following visual inspection of the data we conducted exploratory analyses on a later (300 - 550 ms) time window for both attentive and inattentive mismatch. Spatiotemporal cluster-based analyses revealed a significant effect of attentive mismatch (p = .029), but no reliable effect of inattentive mismatch. Further details of our exploratory analyses can be found in Appendix 1.

1.3. Experiment 1: Discussion

In experiment 1, we tested whether a PMN effect could be elicited in an inattentive paradigm. Participants mentally deleted the first syllable or note of a prime item, and subsequently heard a target item that either matched or mismatched their pre-generated expectations. We directed participants to attend to either the word or the tune whilst ignoring other input, so as to create two otherwise identical conditions in which a phonological mismatch occurred attentively or inattentively. As expected, we found an attentive mismatch effect within a typical PMN window (i.e., 150-380 ms) across centroparietal electrodes, but the inattentive mismatch comparison failed to reach significance. Although in most conditions behavioural data showed accuracy to be high, accuracy in the attentive mismatch condition was notably lower than the other three conditions (75%), which is most likely due to difficulty identifying mismatching targets and an overall response bias towards match in cases of uncertainty.

The results of experiment 1 thus strongly support the notion of attention being a pre-requisite to PMN elicitation. Nonetheless, it remains possible that a number of factors intrinsic to the design of the present study may have result in a weaker (and thus less robustly detected) response. First, given the design of the experiment, task demands differed markedly from both typical inattentive prediction-error paradigms (i.e., MMN oddball experiments), and prior PMN experiments. Participants were actively engaged in a primary distraction task, in which they were required to make decisions regarding non-linguistic input. In a typical MMN paradigm on the other hand, participants are more commonly asked to attend passively to silent input (i.e., watch a silent film or read a book, as in Pulvermüller & Shtyrov, 2003; Pulvermüller et al, 2001; Korpilahti et al., 2001), whilst hearing the standard/deviant stimuli. In PMN experiments, participants are typically focussed solely upon the phonological input, forming predictions on the basis of sentence context or a given instruction. A further distinction between our methodology and that of typical PMN/MMN studies is the simultaneous presentation of auditory distraction stimuli. That is, not only are participants engaged in an active task, but the active task overlaps in modality with the critical manipulation, thus further decreasing the valence of the phonological mismatch as participants listen to more than one set of auditory input. It seems reasonable to think that the secondary focus on a demanding task required of participants could result in increased information processing complexity, and consequently decrease any mismatch response.

To further explore the myriad potential interpretations for experiment 1 we ran a second study. In experiment 2, participants engaged in an inattentive paradigm more similar to a typical MMN procedure, in which they watched a silent film whilst instructed to ignore auditory input. Stimuli consisted of 3-phoneme strings (e.g., /kʌg/), which were followed either by matching targets with the initial phoneme deleted (e.g., /ʌg/), or mismatching targets for which the initial vowel of the rhyme segment (e.g., /ɑːg/) differed. Through experiment 2 we sought to address the limitations of experiment 1 that may have contributed to a reduction in inattentive PMN effects. Firstly, the use of a silent film as distractor input enabled us to redirect participant attention, whilst removing potential complications resulting from the distractor task of experiment 1. The modification of the paradigm in this sense thus brough the design closer in line with that of typical inattentive MMN setups, removing both the necessity of a primary active distractor task and simultaneous presentation of an auditory distraction stimuli. Secondly, it is important to note both the low accuracy and the weak, centro-parietal, seemingly delayed - based on visual inspection of the waveforms and an exploratory spatiotemporal cluster analysis of the 300-550 ms time window - attentive PMN effects of experiment 1, although see the General Discussion for an in-depth discussion of the characteristics of PMN effects reported in the present paper. This pattern of results, in combination with the relatively high dataset rejection rate (13%, as opposed comparable studies with rejection rates around 5-10%), lends itself to the conclusion that the task at hand (both tone and syllable mismatch detection) may have been too difficult. Such a conclusion would support the attribution of our attentive PMN characteristics to task-complexity driven differences, in alignment with the Bornkessel-Shlesewesky (2019) predictive coding model that proposes that latency and topography differences may result from the nature and complexity of information being processed. To sum, in addition to addressing the aforementioned limitations, experiment 2 offered a substantially simplified paradigm through which to investigate inattentive phoneme deletion, being more similar in design to both typical MMN and PMN experiments than experiment 1.

Given the differences between classic PMN and MMN paradigms, it may be useful to outline the aspects drawn from each to form the design of experiment 2. As stated, classic MMN paradigms typically elicit responses to mismatches through repeated exposure to higher-frequency standards and infrequent deviants, whilst participants attend to distractor stimuli – occasionally involving a task such as playing a game or monitoring for targets and pressing a button, but most typically this "task" is to passively watch a film. As such, experiment 2 follows a typical MMN design insofar as the lack of active task or task instructions pertaining to the stimuli - instead using a silent film as distractor stimuli, and the absence of an attentive mismatch comparison. Importantly, given that MMN deviants are usually signalled through frequency within the experiment context (i.e., being a sound/phoneme/tone that occurs less often than standards) as opposed to the violation of any explicitly stated pattern or rule, we ran two blocks where mismatches accounted for either 25% or 50% of the total stimuli. The inclusion of the equal and low frequency mismatch conditions allowed us to determine whether elicitation of an inattentive PMN hinged purely on participants' ability to unconsciously detect violation of a rule (and thus present in both conditions), or - as in MMN paradigms - the infrequent occurrence of such a violation (thus present in only the low-frequency condition).

As previously stated, the means of generating predictions differ between PMN and MMN paradigms, and whilst both approaches ultimately produce effects that result from a mismatch between the prediction and the auditory input, elements of experiment 2 differ from that of a classic MMN experiment, instead adopting methodology closer to that of a PMN design. Whereas standard stimuli in an MMN experiment are usually formed of repeated or familiar events, the 'standards' in experiment 2 were formed by targets that correctly followed a phoneme deletion rule. That is, as opposed to deviants being inserted into a string of standard stimuli, the stimuli for experiment 2 formed natural pairs of primes and subsequent targets that were either congruent or incongruent with the pattern of initial phoneme deletion. Whilst this consists of a notable deviation from the typical MMN design, the manipulation closely resembles that of classic PMN experiments (see Newman et al., 2003; Newman & Connolly, 2009). If, in the present study, we wish to interpret any observed effects as proof of an inattentive PMN effect, it is crucial to ensure that our manipulation remains as close as possible to that associated with typical attentive PMN elicitation, but with the absence of attention.

In line with the visual inspection of the waveforms and the exploratory cluster analyses of experiment 1, we proposed to run two cluster tests, one on the original time-window (i.e., 150-380 ms) and one on the later time window as indicated by the results of experiment 1 (i.e., 300-550 ms). Given the outstanding questions regarding the disassociation and co-occurrence of the PMN and N400, our rationale for the dual time-window analysis procedure stemmed from the likelihood that a single, larger time window (i.e., 150-500 ms) might reduce the chances of capturing an early PMN effect, should a later N400-like response occur as in experiment 1.

2. Experiment 2

2.1. Materials and Methods

2.1.1. Participants

Twenty-seven native English speakers took part in the experiment and gave written informed consent prior to participating (application no. HSEARS20210812003). One participant was excluded due to low trial counts after artefact rejection and exclusion of incorrect trials (< 50%). Twenty-six participants were retained in the final analysis (16 female, mean age = 43.5; *SD* = 8.75) with normal or corrected-to-normal vision, no learning disabilities, and self-reported normal hearing. All spoke English from birth at home as their native language. Five participants were raised in bilingual households but all considered their proficiency in English to be equal or superior to that of their other language. Twentyfour participants were right-handed, whilst one participant was left-handed and one ambidextrous[5]. For full participant demographics, see Supplementary File 1.

2.1.2. Stimuli

The stimuli consisted of 42 onset segments (e.g., /k/, /j/, /v/) and 50 rhyme segments (e.g., /ʌg/, /up/, /ɑːp/), and 42 combinations of these to create a selection of non-words (e.g., /kʌg/, /jup/, /vɑːp/). Within the experiment design, non-word items featured as the primes, whilst rhyme segments were the targets. Targets formed two conditions, match – wherein the target matched the prime non-word with the onset segment deleted (e.g., /kʌg/ - /ʌg/), or mismatch – wherein the target mismatched the prime non-word based on the initial vowel of the rhyme segment (e.g., /kʌg/ - /ɑːg/). Each auditory target featured both in the match and mismatch condition, such that primes were created in pairs that shared the final phoneme, but differed in their vowel (see Table 3 for sample stimuli). Each segment and non-word combination were recorded by a female native English speaker using Audacity®. Noise reduction and amplification by 10.971dB were applied, as well as pitching up and pitching down for the first 5ms and last 5ms across all of the stimuli. For the target stimuli, the sound files were clipped so that the voice onset was exactly aligned with the start of the file. Mismatch items either accounted for 50% or 25% of the overall target stimuli.

(Table 3 here)

2.1.3. Procedure

The experiment was performed using Presentation® software (Version 20.0 Build: 09.04.17), Neurobehavioral Systems, Inc., Berkeley, CA). Participants were presented non-word composed of the two segments (onset & rhyme), followed by a variable ISI of 500 ms (+ 0-100 ms) which separated the prime from the subsequent target. Audio stimuli were presented via sound-cancelling ear inserts

---

[5] Removal of non-right-handed participants did not change the trend or significance of effects reported.

(Etymotic, Inc.), and participants were instructed to ignore the stimuli and instead concentrate on television episode presented on a laptop with the sound off and closed captions enabled.

The experiment itself consisted of two parts, one in which mismatch items featured as 50% of the target stimuli, and one in which they occurred 25% of the time, with each part separated by at least one minute. These sections were rotated across participants such that half the participants heard either the 50% or 25% section first, and the order of items within each section randomised. As per experiment 1, to ensure minimal audio latency jitter, the experiment was performed using the Presentation mixer Exclusive mode on a Dell Precision T1700, Windows 7 Professional (x64) with a Sound Blaster Z (Creative Technology) sound card, and a Cedrus StimTracker to send trigger events.

2.1.4. Data acquisition & pre-processing

EEG data was recorded and digitized at the sampling rate of 250 Hz using a SynAmpsRT amplifier (Neuroscan, Charlotte, NC, U.S.). All other data pre-processing was conducted as per experiment 1. Following ICA decomposition, an average of 1.92 components were removed per participant (range: 1-3). Following artifact rejection, 5.33% of trials were removed, leaving an average of 39.73/42 (SD=3.87) trails in the match condition, 39.88/42 (SD=4.30) in the 50% mismatch condition, and 39.65/42 (SD=3.68) in the 25% mismatch condition., As for experiment 1, the full pre-processing scripts can be found online.

2.1.5. Data analysis

As in experiment 1, we conducted spatiotemporal cluster-based analyses (Maris & Oostenveld, 2007) for the contrasts between mismatch (25%/50%) & match (50%) implemented in the FieldTrip toolbox. Note that the contrast between mismatch and match (75%) conditions was not considered due to the imbalance in the number of trials that constituted each condition, and the potential for this to introduce different signal-to-noise ratios into the comparison[6]. For each of the comparisons (i.e., mismatch 25% vs match 50%, mismatch 50% vs match 50%), paired t-tests across the 26 participants were conducted at every channel and every sample from 1) 150 to 380 ms and 2) 300 to 550 ms. Clusters were formed of spatiotemporally adjacent datapoints with significant tests (with uncorrected one-tailed alpha <.05), and neighbours of at least 2 other channels that met the p-value criterion at the same time point. We selected the strongest cluster (based on the sum of t-statistics within the cluster), and used the highest of which as the test statistic. The significance of the test statistic was non-parametrically evaluated by permuting the data 5000 times, with the clustering method as outlined above repeated for each permutation. The proportion of permutation test statistics larger than the observed cluster test statistic was the *p*-value for that region.

2.2. Results

The spatiotemporal cluster analysis for the 25% mismatch effect failed to find more negative ERPs for match vs mismatch in either the 150-380 ms (p =.45) or the 300-550 ms time-windows (p =.44). A second cluster test for the 50% mismatch effect also failed to find more negative ERP for match vs mismatch in either the 150-380 ms (p =.36) or the 300-550 ms time-windows (p =.36).

(Figure 6 here)

2.3. Experiment 2: Discussion

In experiment 2 we tested whether a PMN component could be elicited through the use of a standard phoneme-deletion manipulation during a fully-inattentive setting. We found no significant effect of phoneme mismatch for either the equal frequency (50%) or low frequency (25%) phoneme mismatches. A considerable amount of baseline noise is observable in all conditions, which we attribute to variable activity resulting from the task at hand – namely, watching a film and reading subtitles, an interpretation supported by the notably posterior distribution of the baseline noise, corresponding to C1 activity in

---

[6] Exploratory cluster-based analysis nonetheless also revealed non-significant results (p = .7) for the 25% mismatch/75% match contrast.

area V1 (the early visual cortex) and across occipital/parietal electrodes (i.e., P1, N1). Despite this, the alignment of N1/P1 components across both conditions suggests that this baseline noise would not have influenced our null results. The lack of effects reported in experiment 2 suggest that a PMN response to phoneme deletion manipulation cannot be elicited inattentively. However, it is perhaps important to consider a limitation of experiment 2 pertaining to the lack of an attentive control condition (which would have necessitated a considerable shift in task-demands relative to the inattentive set-up reported here). In the realisation of experiment 2 in present paper, we explore only whether an *inattentive* PMN effect would be elicited in a paradigm that resembles that of an MMN experiment. However, experiment 2 in its present state cannot tell us whether the same paradigm would have elicited an attentive PMN response. Without evidence of an attentive PMN in experiment 2, it is not possible to conclusively attribute the lack of PMN effects to the manipulation of attention (an alternative explanation could be argued that the design of experiment 2 might not have elicited any PMN effects – attentive or inattentive). This limitation notwithstanding, a large body of literature demonstrates PMN responses to phoneme prediction violations elicited via phoneme deletion – such as that utilised in experiment 2. Furthermore, in its present form, it is difficult to determine whether elicitation of an attentive PMN would have been feasible given the potential co-occurrence of P3 activity for less frequent mismatches. Despite the similarities in the phoneme-mismatch paradigm adopted by experiment 2 with that or prior literature, clear design differences between our methodology and traditional PMN investigations means that although experiment 2 serves as a second example of an instance in which the PMN does not occur inattentively, we exercise caution in the interpretation of our results. This considered, the findings of experiment 2, and their implication for the field are perhaps best discussed in combination with those of experiment 1.

3. General discussion

In the present study we tested the longstanding distinction drawn between the PMN and MMN on the basis of attention (i.e., the necessity for participants to attend to stimuli in order to elicit a PMN response). In experiment 1 we asked participants to focus either on phonological input, or on distractor tone stimuli. In contrast to the significant attentive PMN effect, we failed to find a significant inattentive PMN response in the cluster test run on an a priori-defined 150-380 ms time window. In experiment 2 we again tested whether an inattentive PMN response could be elicited, but this time whilst participants were fully inattentive (watching a film with subtitles) and instructed to ignore auditory input consisting of correct/incorrect phoneme deletions. We failed to find a significant effect for phonological mismatches that occurred either at infrequently (25%) or 50% of the time. The results of both experiments thus strongly suggest that elicitation of the PMN effect requires participant attention, thus distinguishing it from the earlier, pre-attentive MMN component.

3.1. Outstanding questions

Despite our seemingly clear-cut findings, there are a number of aspects of the results reported in the present study that warrant further discussion. In the following section we discuss the characteristics of the attentive PMN effects of experiment 1, the role that varying degrees of attention might have on PMN elicitation, and outstanding limitations of the present research to guide future investigations.

The attentive PMN of experiment 1 revealed by our cluster analysis is consistent with the latency of a number of prior PMN reports (Praamstra, Meyer, & Levelt, 1994; Jones, Kuipers, Thierry, 2016; D’Arcy et al., 2004), though somewhat extended as compared to studies in which time windows are selected via visual waveform inspection. The effect is, additionally, markedly centro-parietal relative to the typical PMN characterisation as a frontal, or fronto-central effect. Despite consistent reference to its frontal topography, a great number of studies report centro-parietal (Ho, Boshra, Schmidtke, Oralovac, Moro, Service & Connolly, 2019; Dufour et al., 2013; van den Brink & Hagoort, 2004; Hagoort & Brown, 2000; Desroches, Newman, & Joanisse, 2009), parietal (D’Arcy et al., 2000), and occipital (Diaz & Swaab, 2007) PMN effects. In fact, in a pre-registration of a systematic review of PMN effects, Lewendon, Egan, and Politzer-Ahles (2021) found that the vast majority of PMN experiments reported effects that spanned frontal, central, and parietal electrodes.

A crucial question therefore arises as to the comparability of central, parietal, and occipital PMN responses to frontal and fronto-central effects. Further still, given the initial premise that the overlapping topographies and latencies of the PMN and MMN lend themselves to conclusions regarding shared underlying mechanisms, it is perhaps unsurprising that we found no inattentive equivalent to a temporally delayed, centro-parietal attentive PMN response that shares few characteristics with the MMN.

The attentive PMN of the present study, alongside a great number of other reports (Lewendon et al., 2021), instead shares a great deal in common with the N400. The N400 can, as previously noted, be modulated via a range of phonological manipulations (Praamstra & Stegeman, 1993; Bölte and Coenen, 2002; Newman & Connolly, 2009). This clear overlap between the functional sensitivity of the PMN and N400 became a main focus of early literature on the PMN, with the key to distinguishing the two components thought to rest predominantly upon the PMNs supposed insensitivity to lexicality – a debate that remains ongoing (Lewendon et al., 2020). The N400's sensitivity to phonological factors, on the other hand, goes largely unchallenged. Given our current results, it might therefore seem plausible to conclude that the attentive mismatch effect we report in experiment 1 represents phonologically-induced N400-related activity, as opposed to a PMN response. Such an interpretation does, however, seem incompatible with the paradigm employed in the experiment. The generation of phonological predictions through silent phoneme deletion is closely based on a classic PMN paradigms (Newman et al., 2003; Newman et al., 2009), whilst the use of pseudoword targets devoid of semantic meaning should reduce the possibility of contamination from semantic N400 effects, if the conventional PMN/N400 distinction is to be assumed valid (Newman et al., 2003; Newman et al., 2009). The conclusion that the phonological mismatch effect present in our data is not a PMN effect thus seems incompatible with the very definition of the PMN as a marker of phonological prediction violation. If we are to argue that phoneme prediction violation using non-word target stimuli can generate N400, but not PMN effects, the notion of a separate, pre-lexical phonological mismatch response that precedes the N400 becomes difficult to defend[7].

How then should we account for the characteristics of our attentive PMN effect? Whilst a full theoretical account for the discrepancies in PMN distribution (to which our data now contributes) is beyond the scope of this paper, a number of things strike us as potentially relevant to this question. The first returns us to the precision-weighted predictive coding account proposed by Bornkessel-Schlesewesky & Shlesewesy (2019), in which the differing characteristics of components are thought to reflect the nature and complexity of the information being processed, as opposed to qualitatively different underlying mechanisms[8]. Extending upon Bornkessel-Schlesewesky and Shlesewesy's account, which argues that the centro-parietal N400 represents a long-latency MMN (with qualitatively similar mechanism), we argue that the PMN may also fall within this functionally and neurobiologically related family of negative ERP components, with information complexity increasing the latency, and potentially influencing the topography of the effect. Such an account situates the PMN directly between the MMN and N400 in terms of a continuum of responses to increasing precision-weighted prediction complexity.

Such an interpretation – that the complexity/difficulty of the task in experiment 1 may have influenced the characteristics of the attentive mismatch effect – is corroborated by clear indications of task difficulty in the behavioural data (low accuracy in both the attentive (75%) and inattentive mismatch (83%) conditions. In combination with an attentive mismatch effect that only just reached significance (p =.044), it seems reasonable to assume that removing attention to the manipulation would weaken any phoneme mismatch response. To sum, prior to conclusively establishing whether attention is a necessary precursor for PMN elicitation, it will be necessary to i) determine the factors that influence the fronto-central realisation of phonological mismatch effects, and ii) establish a simplified design that

---

[7] Note that we are not here proposing that the PMN and N400 *do* represent independent components with differing underlying cognitive mechanisms. Whilst we strongly feel that this distinction requires more evidence, we here offer our interpretation of the data if the PMN should be presumed to be separate from the N400.

[8] See also the separate, yet also potentially compatible account of Rabovsky, Hansen and McClelland (2018) who implemented simulations of a sentence gestalt model. In their computational model of language comprehension, N400 amplitude modulations are simulated as the change induced by an incoming stimulus (i.e., an incoming word) in an implicit probabilistic representation of meaning. This notion of a stimulus representing a cue to meaning 'which automatically change(s) an activation pattern that implicitly represents estimates of conditional probabilities of all aspects of meaning' shares notable similarities with the Bornkessel-Schlesewesky & Schlesewsky account.

results in more clear-cut attentive effects. Perhaps only in a paradigm producing clearly significant fronto-central attentive response can an inattentive response be further examined.

Given the important role attributed to task difficulty and information complexity in explaining the results of experiment 1, it becomes necessary here to acknowledge the substantially reduced complexity of experiment 2 and yet, the similar lack of inattentive mismatch effect. A key point to consider is perhaps the notion of attention to stimulus as a gradient, as opposed to a binary factor. It is plausible that a PMN effect may be elicited in conditions of reduced attention, but not total absence of attention. In experiment 1 participants were instructed to attend to either the tune or word stimuli, and thus although directing their attention to a manipulation other than the phoneme match/mismatch during the tune blocks, would have been attending to stimuli presented concurrently. Both the MMN and PMN ultimately rely on making the listener generate predictions about upcoming auditory forms. As noted, in an oddball paradigm these predictions result from habituation to the standard stimuli, whilst in PMN paradigms the prediction is generated via task instructions or sentence context. Thus, although the PMN and MMN may still share underlying mechanisms – responding to a mismatch between the prediction and the auditory input, it could be the case that the complexity of PMN manipulations (i.e., phoneme deletion) as opposed to habituation necessitates a certain degree of attention. Such attention may be necessary to allow participants to ascertain the pattern, and subsequent violation of said pattern that gives rise to a phonological mismatch. Subsequent work may wish to further investigate both the effect of reduced attention on PMN generation with/without task demands. This could be achieved through a paradigm similar to experiment 2, with the adaptation of 1. showing a film without subtitles, therefore reducing the level of distraction, 2. informing participants of the manipulation prior to commencing the experiment, therefore increasing awareness of the manipulation; or 3. using a phonological manipulation of greater simplicity, therefore increasing the likelihood of producing phonological violations despite an inattentive setting.

We should note that the Bornkessel-Schlesewesky & Schlesewesky (2019) account discussed at length, the Rabovsky, Hansen and McClelland (2018) computational model and, particularly, the theorisation of the results of the present study, necessitate further evidence and testing. Determining the degree to which the PMN can be situated within the Bornkessel-Schlesewesky and Shlesewesky ‘family’ of predictive coding error effects, further investigations into the varying characteristics of PMN responses, and work to determining the degree of attention necessary to produce a phonologically-induced mismatch response are crucial to our understanding of the timeline of auditory ERP responses. More broadly, further research is necessary to explore the extent to which these three components share functional and neurobiological mechanisms, and to establish whether the PMN can be considered a discrete, autonomous effect. Until the independence of the PMN from the MMN and N400 can be firmly established, its significance in the electrophysiological timeline of auditory speech processing remains difficult to determine.

Acknowledgments

The authors would like to acknowledge the kind help of Mr. Albert Chau in supporting the running of the experiment.

Declaration of interest statement

The authors report there are no competing interests to declare.

Data availability statement

The data that support the findings of this study are openly available on OSF at https://osf.io/ak4dc/?view_only=f52ea1b785dd4596bda079e04ad45d68

References

Baldeweg, T., Richardson, A., Watkins, S., Foale, C., & Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology:*

Official Journal of the American Neurological Association and the Child Neurology Society, *45*(4), 495-503.

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., & Lassonde, M. (2006). Electrophysiological markers of voice familiarity. *European Journal of Neuroscience*, *23*(11), 3081-3086.

Bölte, J., & Coenen, E. (2002). Is phonological information mapped onto semantic information in a one-to-one manner?. *Brain and Language*, *81*(1-3), 384-397.

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, *116*(12), 2765-2774.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in psychology*, *10*, 298.

Cermolacce, M., Scannella, S., Faugère, M., Vion-Dury, J., & Besson, M. (2014). "All that glitters is not… alone". Congruity effects in highly and less predictable sentence contexts. *Neurophysiologie Clinique/Clinical Neurophysiology*, *44*(2), 189-201.

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

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (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., Service, E., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). COGNITIVE NEUROSCIENCE AND NEUROPSYCHOLOGY-Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, *12*(2), 237-244.

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

D'Arcy, R. C., Connolly, J. F., Service, E., Hawco, C. S., & 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.

D'Arcy, R. C., Connolly, J. F., Service, E., Hawco, C. S., & 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., & 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., & 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., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain research*, *1146*, 85-100.

Donchin, E., Heffley, E., Hillyard, S. A., Loveless, N., Maltzman, I., Öhman, A., ... & Siddle, D. (1984). Cognition and event-related potentials: II. The orienting reflex and P300. *Annals of the New York Academy of Sciences*.

Dufour, S., Brunellière, A., & Frauenfelder, U. H. (2013). Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognitive science*, *37*(3), 489-507.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of cognitive neuroscience*, *16*(4), 577-583.

Fields, E. C., & Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. Psychophysiology, 57(2), e13468.

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

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

Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology*, *27*(6), 627-640.

Hagoort, P., & 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.

Ho, A., Boshra, R., Schmidtke, D., Oralova, G., Moro, A. L., Service, E., & Connolly, J. F. (2019). Electrophysiological evidence for the integral nature of tone in Mandarin spoken word recognition. *Neuropsychologia*, *131*, 325-332.

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers' gaze on situated language comprehension: evidence from event-related potentials. *Brain and cognition*, *135*, 103571.

Jones, M. W., Kuipers, J. R., & Thierry, G. (2016). ERPs reveal the time-course of aberrant visual-phonological binding in developmental dyslexia. *Frontiers in human neuroscience*, *10*, 71.

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, *76*(3), 332-339.

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., & 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.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological psychology*, *74*(1), 1-19.

Lee, J. Y., Harkrider, A. W., & Hedrick, M. S. (2012). Electrophysiological and behavioral measures of phonological processing of auditory nonsense V–CV–VCV stimuli. *Neuropsychologia*, *50*(5), 666-673.

Lewendon, J., Mortimore, L., & Egan, C. (2020). The Phonological Mapping (Mismatch) Negativity: history, inconsistency, and future direction. *Frontiers in psychology*, 1967.

Lewendon, J., Egan, C., & Politzer-Ahles, S.J. (2021) The Phonological Mapping Negativity: A systematic review. *Open Science Framework*. https://osf.io/u35xk/?view_only=76c57b6198f5477b9a9fd5c780cb374b

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

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man?. *Neuroscience letters*, *107*(1-3), 347-352.

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

Näätänen R, Alho K. Mismatch negativity–the measure for central sound representation accuracy. *Audiol Neuro-Otol* (1997) 2(5):341–53. doi: 10.1159/000259255

Näätänen, R., Teder, W., Alho, K., & Lavikainen, J. (1992). Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica*, *42*(4), 313-329.

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

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

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica*, *42*(4), 313-329.

Newman, R. L., & 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.

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

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. *International journal of psychophysiology*, *88*(2), 109-123.

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, *1*(2), 73-86.

Praamstra, P., Meyer, A. S., & Levelt, W. J. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of cognitive Neuroscience*, *6*(3), 204-219.

Pulvermüller, F., & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage*, *20*(1), 159-172.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., ... & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, *14*(3), 607-616.

Rabovsky, M., Hansen, S.S. & McClelland, J.L. (2018) Modelling the N400 brain potential as change in a probabilistic representation of meaning. Nat Hum Behav 2, 693–705.

Rinne, T., Alho, K., Ilmoniemi, R. J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *Neuroimage*, *12*(1), 14-19.

Van Den Brink, D., Brown, C. M., & 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 Petten, C., Coulson, S., Rubin, S., Plante, E., & 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.

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

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031-2046.

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *Neuroreport*, *16*(18), 2015-2019.

Table 1: Example stimuli. Note that to ensure full rotation across conditions, both correct and incorrect target words were paired with correct and incorrect target tones.

Prime word	**English IPA**	**Target word (correct)**	**Target word (incorrect)**	**Prime tune**	**Target tune (correct)**	**Target tune (incorrect)**
am**bitious**	/æmˈbɪʃ.əs/	bitious	licious	A4-**B4-C5**	B4-C5	E5-C5
de**licious**	/dɪˈlɪʃ.əs/	licious	bitious	G5-**E5-C5**	E5-C5	B4-C5
in**fecting**	/ɪnˈfek.tɪŋ/	fecting	jecting	G#3-**C#4-F4**	C#4-F4	G#4-F4
in**jecting**	/ɪnˈdʒek.tɪŋ/	jecting	fecting	G4-**G#4-F4**	G#4-F4	C#4-F4

Table 2: Experimental conditions

Condition	Participant Attention	Match
Attentive Match	Word	+note +phoneme
Inattentive Match	Tune	+note +phoneme
Attentive mismatch	Word	+note -phoneme
Inattentive mismatch	Tune	+note -phoneme

NB. +/- indicate target match or mismatch with reference to the prime

Table 3: Example stimuli.

Prime non-word	English IPA	Target (correct)	Target (incorrect)
cug	/kʌg/	/ʌg/	/ɑːg/
laig	/lɑːg/	/ɑːg/	/ʌg/
varp	/vɑːp/	/ɑːp/	/ɛp/
hep	/hɛp/	/ɛp/	/ɑːp/

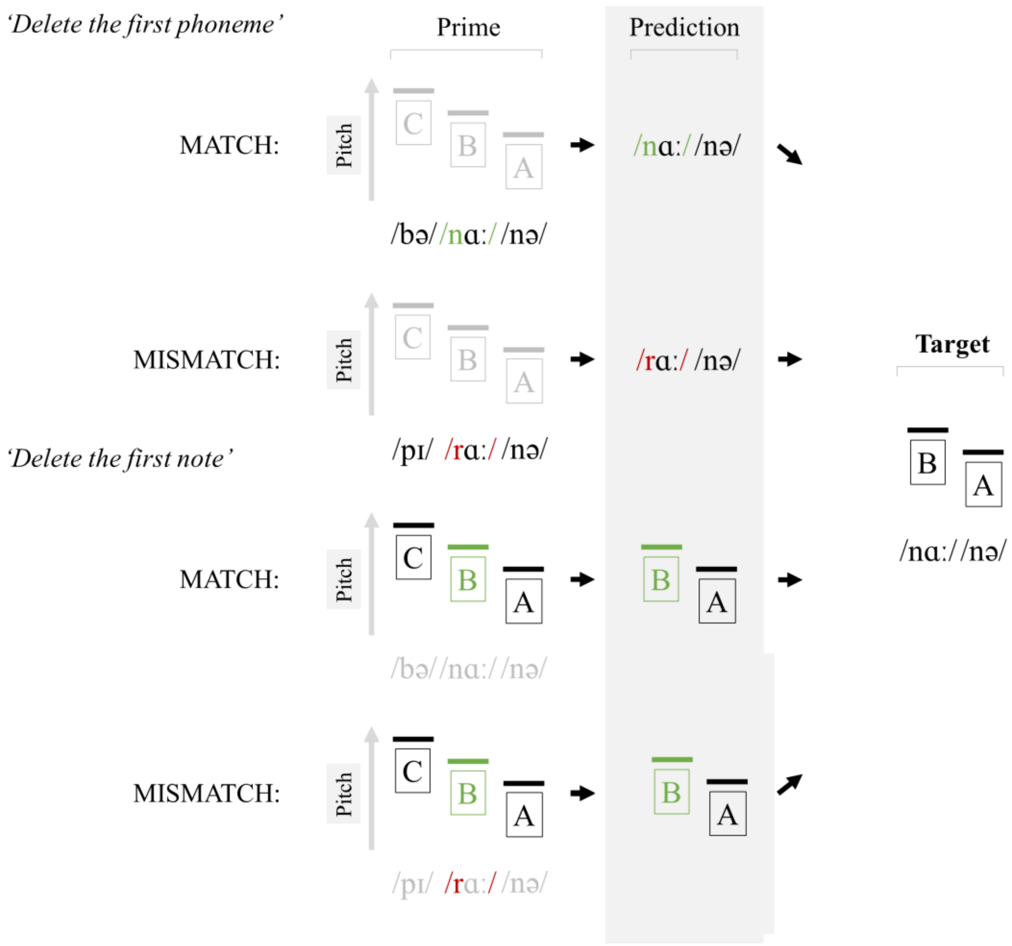

Figure 1: Primes which, when paired with the target, result in attentive or inattentive phoneme mismatch. NB. Greyed-out content for primes denotes unattended input.

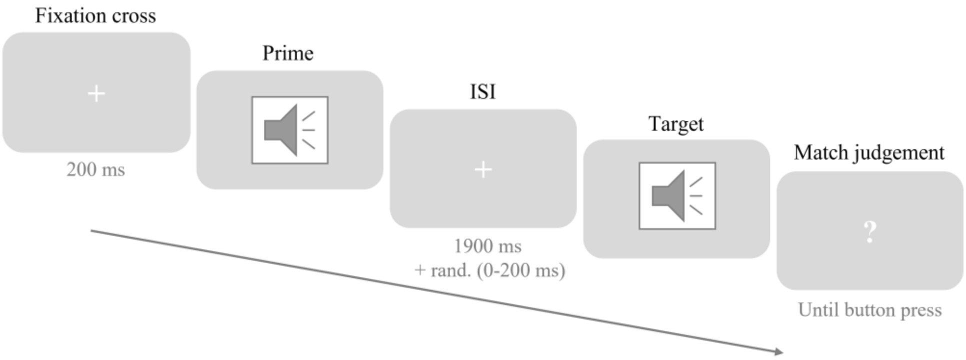

Figure 2: Experiment procedure

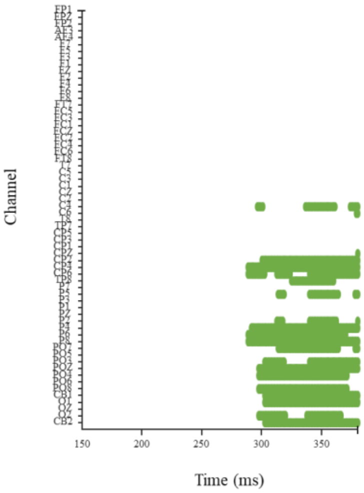

Figure 3: Raster plot for the cluster tests run on attentive match vs attentive mismatch.

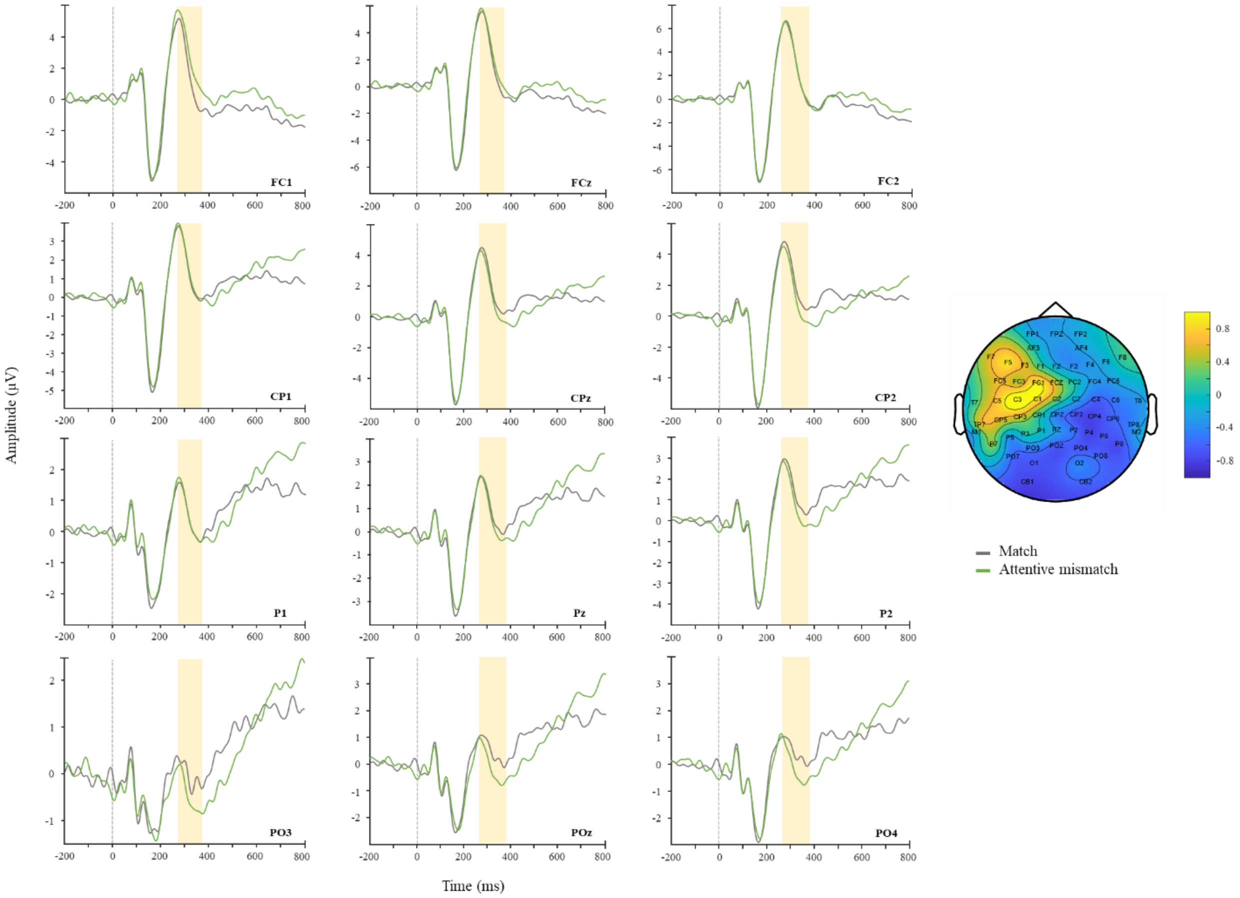

Figure 4: Topography and waveforms for the attentive match vs attentive mismatch comparison. Note that yellow windows indicate the cluster on which the test was based. Topography plot shows attentive mismatch minus attentive match between 288 – 380 ms.

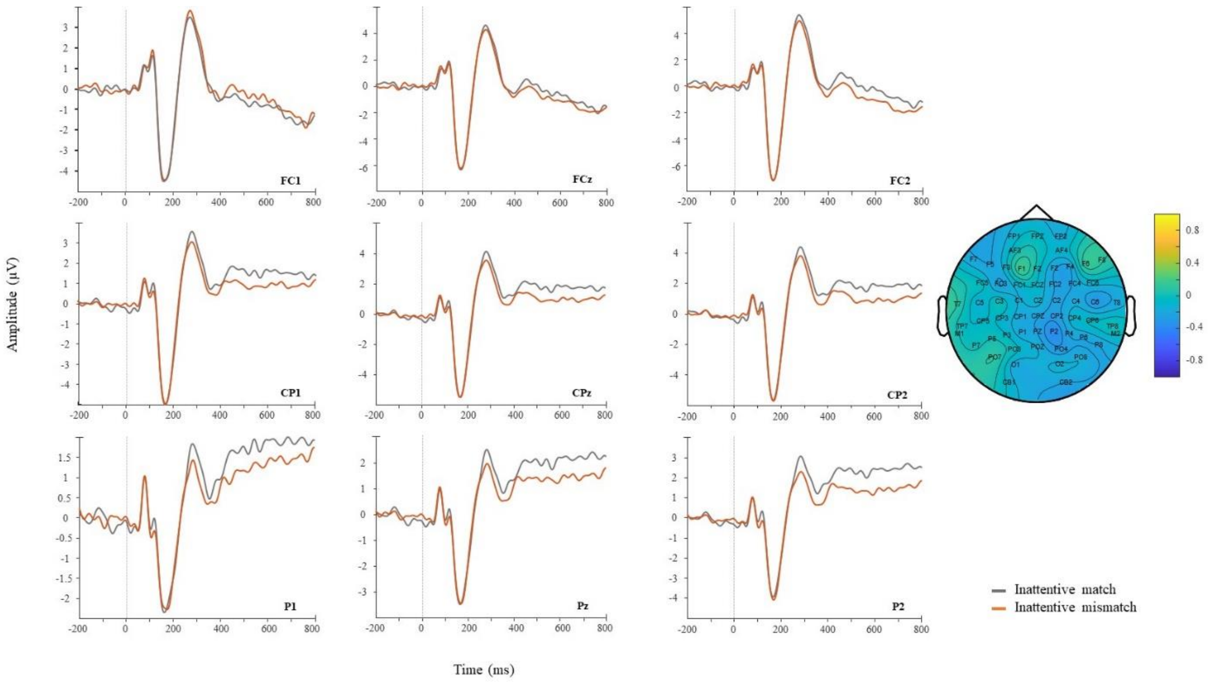

Figure 5: topography and waveforms for the inattentive match vs inattentive mismatch comparison. Topography plot shows inattentive mismatch minus inattentive match between 150 - 380ms.

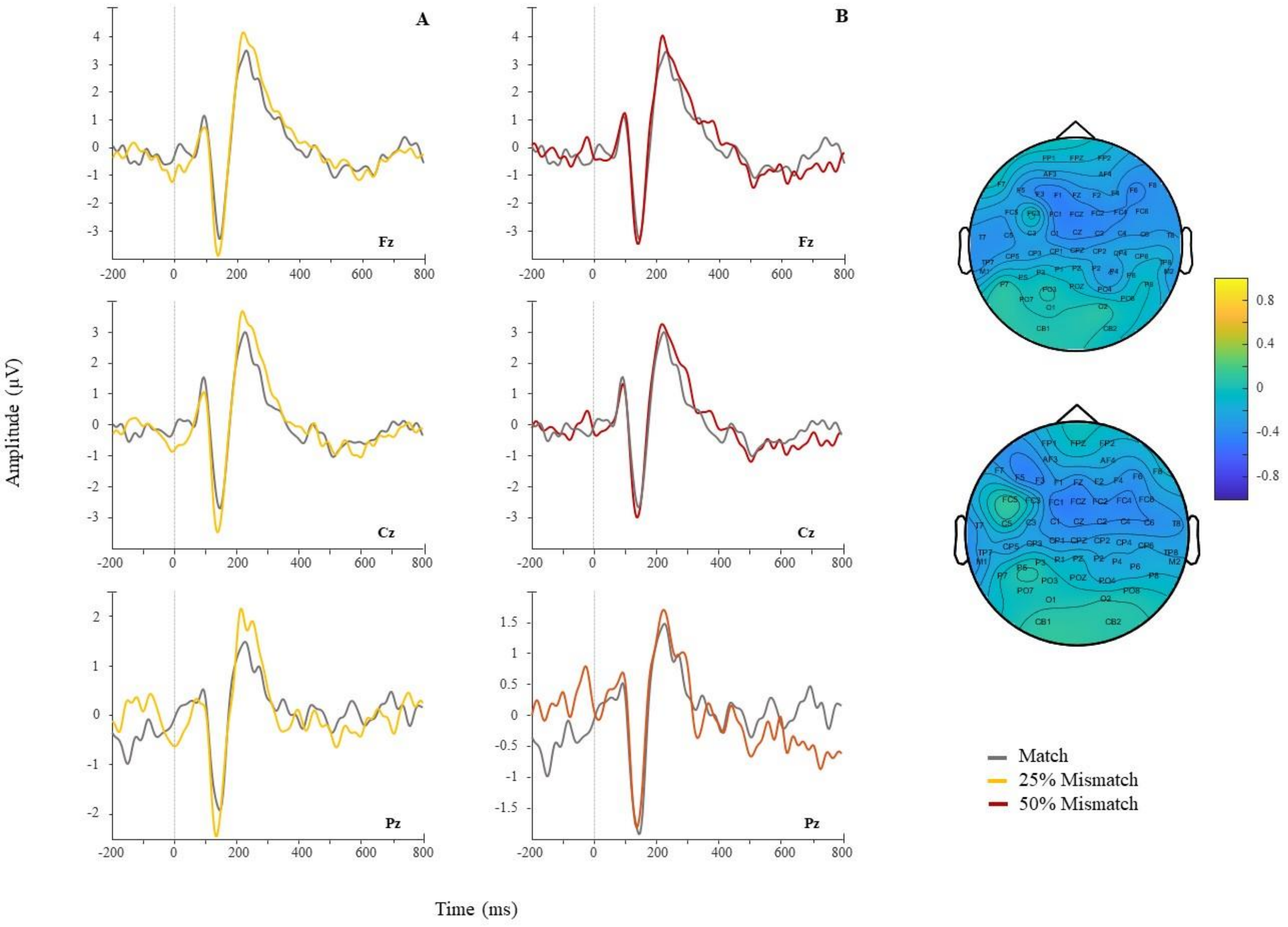

Figure 6: Topography and waveforms for the match vs mismatch (25%), and match vs mismatch (50%) comparisons. Topography plots show 25% mismatch vs match (top) and 50% mismatch vs match (bottom) between 150 – 380 ms.

Appendix 1

Despite the non-significant effect of inattentive mismatch within the PMN time window specified, visual inspection of the waveforms suggested a delayed difference between inattentive match vs inattentive mismatch conditions. To explore this, we conducted (i) a second spatiotemporal cluster-based analyses on an extended time window for the fully inattentive mismatch condition, and (ii) a more traditional ANOVA analysis. It is important to note that, given the exploratory nature of these tests, no strong conclusions were drawn on their basis regarding the hypotheses outlined in this paper.

3.1.1.1. Spatiotemporal Cluster-Based Analyses

We ran the exploratory spatiotemporal cluster-based analysis on a 300 – 550 ms time window. Following visual inspection of the waveforms we increased the cluster alpha to 0.1, as higher thresholds are typically better at detecting weak but sustained effects. Clusters were formed of spatiotemporally adjacent datapoints with significant tests (with uncorrected one-tailed alpha <.05), and neighbours of at least 2 other channels that met the p-value criterion at the same time point. As before, we used the 'maxsum' test statistic (i.e. summed the t-values within each cluster, and used the largest sum as the test statistic). The significance of the test statistic was non-parametrically evaluated by permuting the data 5000 times. The proportion of permutation test statistics larger than the observed cluster test statistic was the *p*-value for that region. For the attentive match/mismatch conditions, the cluster analysis revealed significantly more negative ERPs in response to phonologically mismatching targets as compared to matching targets (*p* =.014). As shown in the raster plot, this increase in negativity was driven by a cluster occurring between 300-518 ms predominantly across central, parietal, and occipital electrodes (Figure 7a). For the inattentive match/mismatch conditions, the cluster analysis revealed marginally significantly more negative ERPs in response to phonologically mismatching targets as compared to matching targets (p=.045) occurring between 422-550 ms across frontal, central and parietal electrodes (Figure 7b).

3.1.1.2. ANOVA

We ran an exploratory ANOVA on inattentive match vs inattentive mismatch conditions. Mean amplitude were calculated across the 430-550 ms time widow, on the average of frontocentral, central, and parietal electrodes (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2; P1, Pz, P2), as indicated by both the waveforms and the exploratory cluster analysis. No significant difference was found between match and mismatching trials (*p* = 0.253).