TOWARDS AN AUTOMATED MEASURE OF LEVELS OF COGNITIVE FUNCTION IN UNRESPONSIVE PATIENTS

TOWARDS AN AUTOMATED MEASURE OF LEVELS OF COGNITIVE FUNCTION IN UNRESPONSIVE PATIENTS

By

RICHARD L. MAH, B. Sc. (Hons)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

McMaster University

© Copyright by Richard L. Mah, September 10, 2018

McMaster University DOCTOR OF PHILOSOPHY (2018) Hamilton, Ontario (Cognitive Science of Language)

TITLE: Towards an automated measure of levels of cognitive function in unresponsive patients

AUTHOR: Richard L. Mah, B.Sc. (Hons) (University of Ottawa)

SUPERVISOR: Professor John F. Connolly

NUMBER OF PAGES: xxiv, 175

Abstract

This thesis aims to evaluate the clinical utility of several auditory paradigms designed to elicit various event-related scalp potentials (ERPs). Through four papers, this thesis (i) determines which paradigms best elicit the desired components in healthy controls, (ii) evaluates methods of confirming the presence of the MMN in clinical ERP data, and (iii) examines the use of spectral entropy, and specifically the use of wavelet signal decomposition to determine the periodicity of spectral entropy in order to target the use of these paradigms for diagnostic use.

Chapter 2 first sets out a framework for extended monitoring of patients in coma by selecting paradigms that performed well in healthy control populations. From an initial group of six paradigms designed to elicit the MMN, P300, and N400, two are selected that were able to elicit the desired ERPs from the healthy controls. This study is the first to examine these various paradigms within the same participants, as well as across two different age groups (younger and older adults).

Chapters 3 and 4 provide evidence that the MMN–a component previously thought to be stable over time–appears to fluctuate in detectability in patients in coma. In addition to the traditional visual inspection method of MMN detection, four other methods of verifying the presence of the MMN were evaluated: the

topographic consistency test, a serial t-test, a spatiotemporal cluster analysis, and Bayesian t-tests. In all four patients presented, the MMN appears to change in detectability over a period of approximately 24 hours. The spatiotemporal cluster analysis and Bayesian t-tests both proved to be suitable for use in confirming visual inspection judgments of the presence of the MMN in this set of patient data, and were able to overcome problems from external noise in the signal. These results suggest that patients should be tested multiple times to increase the likelihood of capturing a period where the MMN is detectable and reducing the chance of a false negative.

Finally, Chapter 5 examines the application of a spectral entropy signal analysis to the same patient data. Period of higher spectral entropy are indicative of a more complex EEG signal, which in turn has been thought to index conscious experience. This analysis was used to determine both if the patients had periods of higher spectral entropy, and if they did, what the periodicity of that fluctuation of spectral entropy would be. Previous work has shown that patients in a minimally conscious state (MCS) can show periods of around 70 minutes, which is similar to healthy, conscious individuals. Of the three patients whose data was appropriate to use in this analysis, one showed a periodicity of around 70 minutes, one did not show a signal with a strong main periodicity, and one had two main periodicities which was indicative of being contaminated by external noise. Even though the analysis method is extremely sensitive to external noise, it does show promise as a means of targeting the cognitive assessments, as these should be given when the patient is likely to be more conscious. This is especially important considering the evidence presented that the MMN fluctuates in comatose patients, so targeting the delivery of these tests can further reduce the false negative rate.

Overall, we have established which ERP paradigms have the best chance of eliciting the components of interest, which in turn can be used for coma prognostication. We have presented evidence suggesting that the MMN fluctuates in its detectability, which provides a caution to clinicians using this methodology to perform repeat testing to better capture the MMN. As well, we have suggested methods to further confirm the presence of the MMN in noisy patient data. Finally, we provide a method of using spectral entropy for determining periods during which these tests should be performed to maximize the likelihood of capturing the components of interest. Taken together, this work brings us closer to an automated measure of the levels of cognitive function in unresponsive patients.

Acknowledgements

I would first like to thank my supervisor, Dr. John Connolly, who has over the past seven years been both a mentor and advisor. I am grateful for you bringing me into your lab all those years ago and giving me this opportunity.

I give thanks to my committee members, Dr. Victor Kuperman and Dr. Elisabet Service. Even though we had all intended at the outset that you both would only be temporary committee members, I am very grateful that you stuck with me throughout this journey. I also give thanks to Dr. Cindy Hamelic and Dr. Alison Fox-Robichaud, who very generously sponsored my studies in the Hamilton General Hospital, and provided me with support throughout my time there. The patient work I have completed would not have been possible without their help, and for that, I am very grateful.

I owe more gratitude than can fit on this page to my lab members, past and present: Amabilis Harrison, Tsee Leng Choy, Angela Harrison, Magdalena Partyka, Rober Boshra, Kyle Ruiter, and other members of the Language, Memory, and Brain lab, both past and present. Without their help, I would not have been able to collect all of the data that went into the study in Chapter 2. It has been a true pleasure working with you all.

I also owe thanks to past and present members of the Department of Linguistics and Languages: Tiffany Deschamps, Daniel Schmidtke, Jitka Bartosova,

Cassandra Chapman, Constance Imbault, Diane Doran, Chelsea Whitwell, Laura Beaudin, and countless others. I cannot thank each of you enough for the many discussions and debates over the years, the moral support and your friendship. I would not have made it this far without all of you.

I also thank my family for supporting me from back home out west. Mom, Dad, Grandma, Kelly, Linden, and Shannon: even though I never really could explain what it was that I was doing, and my trips home never seemed to be long or frequent enough, thanks for constantly believing in me. Maybe now I can go get a real job!

I would be remiss if I did not also acknowledge the agencies who helped me perform my work through their financial support: McMaster University, and specifically the Department of Linguistics and Languages, both for supporting my graduate work and for supporting my travel to present that work to the scientific community; the Government of Ontario for supporting me through the Ontario Graduate Scholarship program; and the Centre for Advanced Research in Experimental and Applied Linguistics (ARiEAL) for supporting me with a research fellowship.

Finally, I must give thanks to the families of the patients who gave up time with their loved ones while in the Hamilton General Intensive Care Unit while I ran my studies, and to the patients themselves. All of this work would have been impossible without them and their contributions.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

ABI Acquired brain injury

ANOVA Analysis of variance

BAEP Brainstem auditory evoked potential

CNS Central nervous system

CV Coefficient of variability

CRS-R Coma Recovery Scale – Revised

CFN Common First Name

CT Computed tomography

dB Decibel

DOC Disorder(s) of consciousness

EEG Electroencephalography

EOG Electrooculogram

ED Emergency department

ERP Event-related potential

FN Familiar Novel

FOUR Full Outline of Unresponsiveness

GCS Glasgow Coma Scale

GFP Global field power

Hz Hertz

ICA Independent components analysis

ICU Intensive care unit

JZS-BF Jeffrey-Zellner-Siow Bayes factor

MRI Magentic resonance imaging

MEG Magnetoencephalography

µV Microvolt

ms Millisecond

MMN Mismatch negativity

NSOW Non-salient Other Word

ROI Region of interest

Ag/AgCl Silver/silver chloride

SSEP Somatosensory evoked potential

SPL Sound pressure level

SD Standard deviation

SON Subject's own name

TCT Topographic consistency test

UFN Unfamiliar Novel

UWS Unresponsive wakefulness syndrome

Declaration of Academic Achievements

Chapter 2

This chapter is a reprint of PLOS One as: **Mah RL** & Connolly JF (2018) A protocol for extended monitoring of levels of cognitive function in unresponsive patients. PLOS ONE 13(7): e0200793. https://doi.org/10.1371/journal.pone.0200793

RM and JC designed the stimulation battery. RM implemented the method, collected control data, and conducted the statistical analyses. RM was the primary writer of the manuscript and JC helped edit and provided part of the discussion.

Chapter 3

This chapter has been submitted to the British Medical Journal Case Reports as **Mah, R.L.**, Connolly J.F., Hamielec C., & Fox-Robichaud A.E. Advancing prognostication by the detection of fluctuating states of consciousness in coma as reflected by an electrophysiological response.

RLM was responsible for collecting and analyzing the data, preparing the figures, and writing and editing the manuscript. JFC conceived of the research

idea and designed the experiment, contributed to the discussion of findings, critically revised the manuscript, and approved the final manuscript. CH provided clinical support to collect the data, edited the manuscript, and approved the final manuscript. AF-R provided clinical support to collect the data, edited the manuscript, and approved the final manuscript.

Chapter 4

This chapter will be submitted to the Journal of Neuroscience as **Mah, R.L.** & Connolly J.F. Electrophysiological markers of variations in perceptual and cognitive processing in coma.

RM collected the patient data and conducted the statistical analyses. RM was the primary writer of the manuscript and JC helped edit.

Chapter 5

This chapter will be submitted to the Journal of Neurology as **Mah, R.L.** & Connolly J.F. Characterizing EEG ultradian rhythmicity differences in coma using spectral entropy.

RM collected the patient data and conducted the analyses. RM was the primary writer of the manuscript and JC helped edit.

1 Introduction

Despite its importance for care decisions, predicting the outcome of a coma remains a challenge in clinical practice. Current techniques usually involve a consensus diagnosis with a team of doctors looking at the results of multiple behavioural assessments, such as the Glasgow Coma Scale (GCS) (Teasdale & Jennett, 1974) or the Full Outline of UnResponsiveness (FOUR) (Wijdicks, Bamlet, Maramattom, Manno, & McClelland, 2005), reflexes, and neuroimaging tests (like CT or MRI). There is a need to make these diagnoses more objective rather than simply best educated guesses. Studies have shown (Schnakers et al., 2009) that there is a high chance of misdiagnosis in patients who emerge from coma and progress into other states of altered consciousness. Predicting how someone will progress while still in a coma is even more difficult.

Before discussing how one predicts the outcome of a coma, it is important first to briefly discuss what coma is and how it relates to consciousness. While the concept of consciousness is still a hotly debated topic both in the fields of neuroscience and philosophy, it is generally understood that in order to have a conscious experience, there must be an active integration of information from multiple sources. This, in turn, requires alertness and the mediation of this

information through executive control (c.f. the Information Integration theory of consciousness by Tononi, 2004).

The concept of consciousness becomes more contentious when one looks through the lens of philosophy versus the lens of neuroscience. And within each field, there still remains some controversy as to how exactly one defines consciousness. In a philosophical sense, some scholars argue that there is only a unitary concept that is *consciousness* and somehow most people are able to intuitively understand it, even if they are unable to define it (Antony, 2001). Other scholars (like Block, 1995), propose a more divided view, where there is a *phenomenal* (or P-consciousness) experience, where one experiences everything in the world through one's body, which creates a feeling of what it is to be in that situation–or a *quale*. To complement this, Block posits an *access* (or A-consciousness) resulting from the information in one's mind becoming available for self-report. Based on this dichotomy, our perceptions, introspections, and memories are all products of access consciousness. It seems intuitively easy to see how access consciousness works, but understanding phenomenal consciousness is much harder, giving rise to the *hard problem of consciousness* in philosophy (Chalmers, 1995).

Irrespective of the theory of consciousness one subscribes to, medicine has given a very concrete definition to what a coma is. This definition says that coma is the state of apparently absent or suspended consciousness, which can complicate a wide range of clinical conditions (Young, Ropper, & Bolton, 1998). It is also described as the general unarousability, apparent absence of sleep/wake cycles, and the inability for environmental interaction, and is often associated with severe, diffuse bihemispheric lesions and/or brain stem injury, but can also result from a disruption of the reticular activating system. (American Congress

of Rehabilitation Medicine, 1995)

As many of the tests used in coma prognostication either require a behavioural result from the patient, or a subjective interpretation from a clinician, it would seem preferable to have a more objective means of determining the patient's probable outcome. Event-related potentials (ERPs) are a convenient technique that allow a patient's brain to respond, even if the patient is unable to. In the next section, we will briefly introduce ERPs, the method of their recording and analysis, and how they can be used for the study of consciousness.

Event-related potentials

ERPs are measures of coordinated changes in the brain's electrical activity, as seen through an electroencephalogram (EEG), that are elicited by a physical stimulus, or an internal, psychological event (Picton, Lins, & Sherg, 1995). They reflect the synchronous firing of post-synaptic potentials in large groups of neurons in the cortex.

The ERP is not readily detectable in the continuously recorded EEG signal. The amplitude of an individual ERP time-locked to a stimulus event is quite small relative to the amplitude of all other recorded electrical activity. In order to extract the ERP signal, we must overcome the background noise that covers up our signal of interest. The easiest, and one of the most powerful, methods to do this is signal averaging (Dawson, 1954).

This method involves the collection of continuous EEG data, which is sampled regularly at a rate of between 500 to 1000 Hz. This continuous signal is then segmented up into epochs based on a time-locking event, such as a stimulus or behavioural response. Each epoch usually begins prior to the onset of the

stimulus to capture some of the random pre-stimulus activity, and then continues for some fixed time thereafter. For example, a signal may begin 100 ms prior to the stimulus onset and continue for 500 ms afterwards in order to capture an ERP that occurs at about 200 ms. Once all of the trials have been segmented, they are grouped by category of interest (such as all of one type of stimulus), aligned based on stimulus onset, and then the voltages for each time point are averaged for each channel recorded.

By averaging the signals from the epochs, one is able to increase the signal-to-noise ratio of a weak signal (*i.e.* the ERP we are looking for) and have it emerge from the noise from unrelated neural processes. As the ERP signal is related to the onset of the stimulus, the weak signals will sum together. By the same token, since the unrelated noise is not related to the onset of the stimulus, it should essentially sum to zero through destructive interference. As the number of trials increases, so too does the signal-to-noise ratio.

ERPs are usually classified based on their polarity (whether they are positive or negative as detected on the scalp), their latency with respect to the eliciting stimulus onset, their distribution across the scalp, and (sometimes) their sequential ordering. Very early components in the auditory modality are generated by auditory brainstem pathways, and are usually referred to as brainstem auditory evoked potentials (BAEPs). These occur in the first 10 ms following a stimulus and are useful in verifying the integrity and function of the brainstem. Later occurring potentials reflect higher order processes, and are those that are generated by the cortex. These can peak from 50 ms to around 800 ms following a stimulus and are affected by cognitive and psychological processes.

Included in the category of evoked potentials are those generated in response to somatosensory input, or somatosensory evoked potentials (SSEPs). These

potentials are often used to assess the function and integrity of the spinal cord. They involve the stimulation of a peripheral nerve (like the tibial, median, or ulnar nerve). These potentials usually occur at about 20 ms following the stimulus, but are quite small in amplitude.

The notion of using ERPs as an objective means of assessment of patients in altered states of consciousness is not a new one. Some of the first prognostic studies involved the use of the so called P300 response in nontraumatic comas; correlating its presence with favourable outcomes (di Giorgio, Rabinowicz, & Gott, 1993; Gott, Rabinowicz, & DeGiorgio, 1991; Yingling, Hosobuchi, & Harrington, 1990).

The P300 has traditionally been elicited using variants of auditory oddball tasks, in which a train of repeating stimuli is interrupted by a different oddball stimulus. The participants are told to respond mentally or physically to each oddball target stimulus and ignore all other stimuli. As with most ERP components, the P300 is characterized by both amplitude and latency. Amplitude is measured from the largest positive peak of the waveform within a time window (traditionally 250-500 ms) to the mean prestimulus baseline voltage. The latency is measured from stimulus onset to the maximal point of the same positive peak. For a more thorough discussion of the P300 and its neural generators, see Polich (2007).

In the cases where the P300 was used for coma prognostication, the presence of the P300 in the patient was correlated with their survival and eventual emergence. For instance, Gott et al. (1991) using a frequent tone and an infrequent oddball tone found that 30% (6/20 patients) that were tested had a detectable P300, and that 83% (5/6) of those with the P300 awoke. They determined that the presence of the P300 was associated significantly with awakening, but the

absence of the P300 did not preclude it. di Giorgio et al. (1993) found that the mean GCS score of patients with a P300 was significantly higher than that of those without.

An often-cited shortcoming of using the P300 for predicting emergence from coma is its reliance on attention to elicit a strong response. The proposed solution was the use of another response–the mismatch negativity (MMN). The MMN can be strongly elicited irrespective of the subject's attention to the stimulus.

Much like the P300, the MMN is an ERP component that is elicited to deviant tones in auditory oddball paradigms. In the auditory modality, as is the case of the work presented in the following chapters, the MMN can be elicited in response to a deviance in pitch, intensity, or duration (Näätänen, 1992). The difference between the P300 and the MMN that helps solve the attention problem with the P300 is that the MMN can be elicited with or without the participant attending to the stimulus. In fact, the MMN has been elicited in various states of consciousness, such as normal awareness, sleep (Sallinen, Kaartinen, & Lyytinen, 1994; Sculthorpe, Ouellet, & Campbell, 2009), and minimally conscious and vegetative states (Kotchoubey et al., 2005).

Based on his original model of attention and automaticity in auditory processing Näätänen (1990) proposed that there are two passive routes of processing that are able to interrupt the central executive. This interruption is thought to provide the passively analyzed information to the perceptual and cognitive systems for further analysis. The first route detects changes in the transient energy of a stimulus and is associated with the N1 ERP component. The other route detects deviations from a series of similar stimuli and is associated with the MMN.

In contrast to the P300, the MMN is reported as the subtraction of the re-

sponse to the frequent "standard" tones from the infrequent "deviant" tones—in "difference" waves. It appears as a negative peak in the frontocentral and central electrodes at about 150 to 250 ms. (Näätänen, Paavilainen, Rinne, & Alho, 2007) The magnitude of the response can vary with the parameters of the experiment, like the time between stimuli, the type of deviant used, or just the intensity of the deviant compared to the standards.

Currently, the paradigm with the best predictive performance for outcomes in comatose patients is presented in Fischer et al. (1999). Using an auditory oddball paradigm to elicit the MMN, they were able to detect the MMN in 33/128 patients and the N100 in 84/128 patients. After three months, 95 patients regained consciousness with most of them having moderate to severe disability. Among these 95 patients, 30 were from the group who had shown the MMN and 70 were from the group who had shown a N100.

These previous investigations only give positive or negative (*i.e.* the patient will live or die) predictions. What would be more desirable is to know the specific state of the patient when they emerge from their coma. Will they emerge into a vegetative state, will they be minimally conscious, or will they fully awaken? What will their language ability be when they emerge–will they be able to speak fluently, or will they require extensive language rehabilitation? Not only are the answers to these questions important for clinicians for creation of an appropriate treatment plan, but they help families prepare for the eventual outcome of their loved one.

A major concern is whether the patient will be able to use language for communication if they emerge from their coma. One way of tackling the question of language ability upon emergence is the use of language-related ERPs. These include the N400 response, which is an ERP that is sensitive to contextually-

based expectancies, and more specifically, displays the greatest amplitudes in response to violations to semantic expectancies. This component requires the integration of both pre-existing knowledge and newly parsed information. It is a late-occurring response associated broadly with comprehension of speech, text, and other stimuli possessing “meaning" (Kutas & Federmeier, 2011).

Semantic expectations can be set up through the use of word pairs which may or may not have underlying semantic relationships, as in Holcomb and Neville (1990). For example, the words *doctor* and *nurse* have a semantic relationship (*i.e.* they both work together in a hospital and are both medical professionals), whereas the words *doctor* and *bread* do not have a semantic relationship. The expectations can also be created by using a stronger semantic context by using sentences with incongruous, unexpected, or infrequent endings, like in Connolly and Phillips (1994). In this case, a sentence begins normally like *“I take my coffee with cream and . . .”*, but can either end in a highly expected word (*sugar*), abnormally (*feet*), with the word having a first syllable that is the same as the highly expected word but then ending differently (*shoes*), or with a lower cloze frequency but still semantically appropriate word (*sweetener*). In both of these situations, the semantically incongruous endings should elicit the N400. The component is a negative-going waveform, which is usually detected between 250 to 500 ms, and peaks at around 400 ms. It is usually maximally found in centro-parietal electrode sites.

There have been a few investigations into whether the N400 is able to be elicited from patients with disorders of consciousness and whether its presence has any ability to predict the return of consciousness to a patient (Kotchoubey et al., 2005; Rämä et al., 2010; Schoenle & Witzke, 2004). For example, Schoenle and Witzke (2004) studied 120 patients with severe brain damage who ranged

from "not in vegetative state" to those who were in a vegetative state. They found those who were in a near-vegetative state showed significantly more intact semantic capacity than those who were diagnosed as vegetative. Rämä et al. (2010) found, in a group of 13 comatose patients, those with intact temporal lobes to be able to produce N400s in response to semantically unrelated word pairs in Finnish, when compared to patients with temporal lobe damage.

As noted by Steppacher et al. (2013), the utility of the N400 as a binary prognostic indicator remains somewhat unclear due to the small sample sizes of past studies, although it should not be discounted as a predictor. What is more unclear is whether the N400 is a good index of language ability for patients upon emergence, and whether it is a worthwhile tool for use during their recovery.

This dissertation uses ERPs with the aim to address a number of unresolved issues that emerged while developing and testing a stimulation battery for the prediction of coma outcomes. These questions are:

1. What paradigms elicit the strongest responses in the absence of explicit instruction to attend?

2. Is the MMN a reliable measure of emergence out of coma? Or, more specifically: is it necessary to perform repeated measurements to ensure a reliable result?

3. How do the underlying rhythms of consciousness of comatose patients compare to those with other disorders of consciousness?

These questions will be addressed through the development of a stimulation battery that was subsequently used to elicit neural responses from comatose patients in the Hamilton General Hospital Intensive Care Unit. The remainder

of this chapter serves as brief overview of how these questions will be examined through these data.

1. What paradigms elicit the strongest responses in the absence of explicit instruction to attend?

As noted earlier, both the MMN and P300 have been shown to be predictive of positive outcomes from coma, however there have been definite shortcomings identified with the P300 in relation to attending to the stimulus. There are a number of possible paradigms that could be used to elicit the MMN, the P300, and the N400. The present study selected a collection of them that are expected to elicit the strongest responses when participants are not instructed to pay attention to the stimuli.

In Chapter 2, five paradigms are described that were chosen for evaluation: an auditory oddball paradigm (like that used in Fischer, Dailler, & Morlet, 2008) that includes novel stimuli (as used in Holeckova, Fischer, Giard, Delpuech, & Morlet, 2006) to elicit both the MMN and the P300, a pattern violation mismatch (as in Sculthorpe & Campbell, 2011) to elicit the MMN, the Subject's Own Name paradigm (adapted from Holeckova et al., 2006) to elicit the P300, semantic violation sentences (from Connolly & Phillips, 1994) and word-word priming (from Holcomb & Neville, 1990), both to elicit the N400.

These paradigms were tested on two groups of adults: younger undergraduate students, and older adults from the surrounding community. These groups were chosen because the vast majority of cases of comas that are seen in the Hamilton General ICU are young adults with traumatic brain injuries, and older adults with complications arising from stroke.

In all cases, the participants were instructed to ignore the sounds that were being played through the headphones and instead watch a video. This was

meant to simulate the worst case scenario for a comatose patient unable to control their attention or follow instructions.

Based on the data presented in Chapter 2, two paradigms were selected as the best performing while also eliciting all of the desired components of interest: the auditory oddball with novel stimuli to elicit the MMN and P300, and the semantic violation sentences to elicit the N400. In the case of the auditory oddball paradigm, robust MMN and P300 responses were seen in both age groups. This one paradigm performed significantly better than the pattern violation mismatch and Subject's Own Name paradigms, and only required half the stimulation time. For all of the paradigms tested, there was a significant effect of age, where older adults had reduced component amplitudes compared to the younger adults. Reduced responses present a challenge for the use of this method as a clinical tool, as the age of the patient is not something that we have control over. We can, however, try to improve the strength of the response in other ways to try and counteract this effect.

Both paradigms that gave rise to the most robust ERPs were the ones with the strongest contexts built by the stimuli. The auditory oddball paradigm had a strong contrast between the tones used to elicit the MMN and the two novel stimuli. The two novel stimuli were also very different from each other—one being the Subject's Own Name, which has been shown to elicit a strong recognition response, and the other being a dog bark, which is very different and abrupt, and should elicit a strong reorientation response. The semantic violation sentences built up a much stronger semantic bias when whole sentences were presented, whereas the word-word priming paradigm was unable to construct such a strong context.

2. Is the MMN a reliable measure of emergence from coma?

This question is specifically motivated by the underlying assumption in Fischer (1999)–that if the MMN is present in a patient, it should be there no matter when the patient is tested. Ideally, a clinical test should give the same result no matter when the test is done.

Chapters 3 and 4 examine this assumption through the analysis of patient data collected over the course of 24 hours and using various statistical methods. As is the case in Fischer et al. (1999), visual inspection of the waveforms by electrophysiologists is done. In addition to this, four different statistical methods were used to detect the MMN within each stimulation block.

The first method was a one-sided serial t-test (Marchand, D'Arcy, & Connolly, 2002), which looked for intervals over the averaged waveforms where the response to the deviant tone was significantly more negative than the response to the standard tone. This was done by computing the point-by-point t-scores for overlapping 20 ms windows, and significant intervals during the 120 – 240 ms window were considered to contain the MMN.

The second method was a topographic consistency test (Koenig & Melie-García, 2010), which searched for intervals containing electrical activity that was both event-related and spatially consistent. This was done by computing the Global Field Power (GFP) for the known average ERP signal and comparing it to a surrogate null GFP distribution generated by a random shuffling of electrode labels. Again, significant intervals within the 120 – 240 ms window were considered to contain the MMN.

The third method was a spatiotemporal clustering analysis (Maris & Oostenveld, 2007), which computed one-tailed dependent sample t-tests for individual MMN trials and electrode locations, where the largest cluster of significant intervals and electrodes were retained. These were then compared to a null

distribution generated by a random shuffling of trials and electrode labels.

The final method was to compute the Jeffrey-Zellner-Siow Bayes factor (JZS-BF) to test the strength of the evidence of a significant difference between the standard and deviant responses. (Rouder, Speckman, Sun, Morey, & Iverson, 2009)

From these patients, it appears that the serial t-test and topographic consistency tests were not very good at reliably detecting the MMN, at least when compared to manual visual detection. On the other hand, the spatiotemporal clustering method and the Bayesian t-tests were both able to confirm the presence of the MMN after visual detection.

Throughout each patient's testing session, there was an apparent waxing and waning of the MMN. Whether this is actually due to the MMN not being consistently present remains unclear. However, it does suggest the need to test patients multiple times, or until a positive result is seen.

3. How do the underlying rhythms of consciousness of comatose patients compare to those with other disorders of consciousness?

Thus far, a method for eliciting robust ERPs of interest has been selected and validated, and some evidence has been presented which suggests that the MMN may fluctuate in comatose patients over time. The fluctuations of the MMN also give us reason to believe that there may be other fluctuations in the patient's condition. For example, a recent investigation into spectral entropy in patients who were vegetative and minimally conscious by Piarulli et al. (2016) showed rhythmic fluctuations in the minimally conscious individual's spectral entropy over a period of four hours. This fluctuation was, on average, in 70 minute periods.

There has been some evidence of comatose patients recalling events that took

place while they were in their comas, like in the study done by Thonnard et al. (2013). If these are truly memories of patient experiences during their comas, and not an artifact of the brain trying to integrate disordered information while also recovering from a traumatic injury, then these individuals would have had to have had some sort of conscious experience during those times.

As we are able to use entropy as a means of quantifying the complexity of a signal, spectral entropy allows the estimation of how uniform a signal's power spectral distribution is. Linking spectral entropy to the level of consciousness of an individual is relatively straightforward. Low spectral entropy is indicative of low levels (or the absence) of consciousness, whereas high spectral entropy is associated with a more complex conscious experience. By continuously characterizing the EEG signal of a patient, it becomes possible to identify periods where the signal is more complex.

A consequence of identifying periods during which comatose patients are more likely to be more conscious or aware is that the timing of assessments can be better chosen. As was the case in Chapters 3 and 4, there were periods during which the MMN was not detectable in patients, but determining if there was a specific periodicity was difficult due to the randomized timing of the stimulus block delivery.

In order to evaluate the effectiveness and viability of applying this technique to patients in a critical care environment, three candidate patients were subjected to the spectral entropy analysis method, all of whom were comatose throughout the four hours of testing.

One of the patients showed very minimal spectral entropy activity throughout the entire period, with no dominant peak. Another patient had two dominant patterns of activity, one with a 35 minute period, and one with a 114 minute

period. This patient seemed to display artifactual activity that may have influenced the wavelet analysis. The third patient was the most interesting case–the wavelet analysis was most consistent with the MCS group, with a single dominant period of 86 minutes, but the EEG features were more consistent with the VS group. This patient was the same patient who was the subject of the case study in Chapter 3. He eventually emerged from his coma and has returned to a normal life.

It appeared that the method of extracting the spectral entropy of the signal and then subjecting those values to a wavelet analysis was very susceptible to external noise. On the other hand, when the data was stable and clean of external noise, the methodology was able to provide some interesting results as were seen in the third patient. As is noted in Chapters 5, if this methodology is to be used, care must be taken to prepare the testing environment to remove any extraneous noise. However, if the patient and environment are adequately prepared, this analysis method does appear to be well suited for application in a continuous testing situation to determine the best time to administer further tests to evaluate the patient's state of consciousness.

In summary, this thesis presents a set of paradigms that were selected to best elicit ERP components that are predictive of awakening from coma. The measures were validated on a sample of healthy control participants. We showed that the MMN varies in detectability over a period of 24 hours in a small sample of comatose patients. Finally, the rhythmicity of the spectral entropy of the same small sample of comatose patients was investigated and compared to that of patients with other disorders of consciousness.

References

American Congress of Rehabilitation Medicine. (1995). Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Archives of Physical Medicine and Rehabilitation*, *76*(2), 205–209.

Antony, M. (2001). Is consciousness ambiguous? *Journal of Consciousness Studies*, *8*(2), 19-44.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

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.

Dawson, G. D. (1954). A summation technique for the detection of small evoked potentials. *Electroencephalography and clinical neurophysiology*, *6*, 65–84.

di Giorgio, C., Rabinowicz, A., & Gott, P. (1993). Predictive value of P300 event-related potentials compared with EEG and somatosensory evoked potentials in non-traumatic coma. *Acta Neurologica Scandinavica*, *87*(5), 423–427.

Fischer, C., Dailler, F., & Morlet, D. (2008). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical Neurophysiology*, *119*(10), 2224–30.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*, *110*(9), 1601–1610.

Gott, P., Rabinowicz, A., & DeGiorgio, C. (1991). P300 auditory event-related potentials in nontraumatic coma: Association with glasgow coma score and awakening. *Archives of Neurology*, *48*(12), 1267-1270.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312.

Holeckova, I., Fischer, C., Giard, M.-H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. *Brain Research*, *1082*(1), 142–152.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, *23*(3), 233–242.

Kotchoubey, B., Lang, S., Mezger, G., Schmalohr, D., Schneck, M., Semmler, A., ... Birbaumer, N. (2005). Information processing in severe disorders of consciousness: vegetative state and minimally conscious state. *Clinical Neurophysiology*, *116*(10), 2441–2453.

Kutas, M., & 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.

Marchand, Y., D'Arcy, R. C., & Connolly, J. F. (2002, nov). Linking neurophysiological and neuropsychological measures for aphasia assessment. *Clinical Neurophysiology*, *113*(11), 1715–1722.

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

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–233.

Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

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.

Piarulli, A., Bergamasco, M., Thibaut, A., Cologan, V., Gosseries, O., & Laureys, S. (2016, Sep 01). Eeg ultradian rhythmicity differences in disorders of consciousness during wakefulness. *Journal of Neurology*, *263*(9), 1746–1760.

Picton, T., Lins, O., & Sherg, M. (1995). The recording and analysis of event-related potentials. In F. Boller, R. Johnson, & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 10). Elsevier.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Rämä, P., Relander-Syrjänen, K., Öhman, J., Laakso, A., Näätänen, R., & Kujala, T. (2010). Semantic processing in comatose patients with intact temporal lobes as reflected by the n400 event-related potential. *Neuroscience letters*, *474*(2), 88–92.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225-237.

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1994). Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of K-complex? *Electroencephalography and Clinical Neurophysiology*, *91*(2), 140 - 148.

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., ... Laureys, S. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*, *9*(1), 1.

Schoenle, P. W., & Witzke, W. (2004). How vegetative is the vegetative state? preserved semantic processing in vs patients–evidence from n 400 event-related potentials. *NeuroRehabilitation*, *19*(4), 329–334.

Sculthorpe, L. D., & Campbell, K. B. (2011). Evidence that the mismatch negativity to pattern violations does not vary with deviant probability. *Clinical Neurophysiology*, *122*(11), 2236–2245.

Sculthorpe, L. D., Ouellet, D. R., & Campbell, K. B. (2009). MMN elicitation during natural sleep to violations of an auditory pattern. *Brain Research*, *1290*(Supplement C), 52 - 62.

Steppacher, I., Eickhoff, S., Jordanov, T., Kaps, M., Witzke, W., & Kissler, J. (2013). N400 predicts recovery from disorders of consciousness. *Annals of neurology*, *73*(5), 594–602.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and imparied conciousness. A practical scale. *Lancet*, *2*(7872), 81–84.

Thonnard, M., Charland-Verville, V., Brédart, S., Dehon, H., Ledoux, D., Laureys, S., & Vanhaudenhuyse, A. (2013). Characteristics of Near-Death Experiences Memories as Compared to Real and Imagined Events Memories. *PLoS ONE*, *8*(3), 1–5.

Tononi, G. (2004). An information integration theory of consciousness. *BMC*

Neuroscience, *5*(1), 42.

Wijdicks, E. F., Bamlet, W. R., Maramattom, B. V., Manno, E. M., & McClelland, R. L. (2005). Validation of a new coma scale: The FOUR score. *Annals of Neurology*, *58*(4), 585–593.

Yingling, C. D., Hosobuchi, Y., & Harrington, M. (1990). P300 as a predictor of recovery from coma. *The Lancet*, *336*(8719), 873.

Young, G., Ropper, A., & Bolton, C. (1998). *Coma and impaired consciousness: A clinical perspective*. McGraw-Hill, Health Professions Division.

2 A protocol for extended monitoring of levels of cognitive function in unresponsive patients

This chapter is reproduced from PLOS One as: Mah RL & Connolly JF (2018) A protocol for extended monitoring of levels of cognitive function in unresponsive patients. PLOS ONE 13(7): e0200793. https://doi.org/10.1371/journal.pone.0200793

Abstract

Generally, prognostication of coma outcome currently combines behavioral, reflex, and possibly neuroimaging tests that are interpreted by an attending physician. Electroencephalography, particularly, event-related brain potentials (ERP) have received attention due to evidence demonstrating the positive predictive value of certain ERP including the mismatch negativity (MMN) and the P3a, for coma emergence. We describe a set of ERP paradigms designed to require and reflect increasing levels of cognitive processing with the added objective of determining the influence of each paradigm's context strength on its ability to elicit ERPs. These paradigms were then used without explicit instructions to participants to attend to the stimuli to determine which paradigms possessed sufficient context "strength" to elicit ERPs in the absence of active participation on the part of the subject; a circumstance often encountered in brain injury patients. These paradigms were then validated on two groups of adults–younger and older, and the difference due to active participation was validated on another group of younger adults. Results show that paradigms with stronger stimulus context features performed better than those with weaker contexts, and that older adults generally had significantly attenuated and delayed responses compared to younger adults. Based on these findings, it is recommended the use of the auditory oddball paradigm that includes novel stimuli to elicit the mismatch negativity and P300, and semantic violation sentences to elicit the N400. These findings also reinforce the procedure of instructing

participants about the requirements of a protocol–regardless of the patient's diagnosis or apparent state–in order to help those who are able to attend to show the most robust responses possible.

2.1 Introduction

The use of event related potentials (ERPs) for the assessment of patients in altered states of consciousness has been a topic of research for the past few decades. Some of the first prognostic studies involved the use of the P300 in non-traumatic comas, where the presence of the P300 was correlated with positive outcomes (di Giorgio, Rabinowicz, & Gott, 1993; Gott, Rabinowicz, & DeGiorgio, 1991; Yingling, Hosobuchi, & Harrington, 1990).

Some researchers have taken issue with the use of the P300, which is heavily influenced by the participant attending to the stimuli. The proposed solution was to use an ERP component, which could be evoked irrespective of attention, like the mismatch negativity (MMN). One of the first to do this was Kane et al., who presented tones in an auditory oddball sequence (Kane, Curry, Butler, & Cummins, 1993). Since the MMN has been strongly elicited irrespective of attention, they believed it to be a more reliable measure for patients in altered states of consciousness (Näätänen, 1995; Näätänen, Gaillard, & Mäntysalo, 1978). This study showed that in each case the MMN was detected, the patient would soon regain consciousness. They noted that the "MMN response is the earliest available indicator of awakening from coma" while acknowledging that the MMN "does not provide prognostic information about functional outcome, [but] it may help to define objectively the duration of coma" (Kane et al., 1993).

Fischer et al. built upon this work by expanding the use of the MMN in comatose patients (Fischer et al., 1999). These studies, however, utilized the traditional method of ERP analysis that is time consuming, requires specialized tools, and visual inspection by a trained electrophysiologist. Fischer and colleagues used stimuli which were delivered in short blocks, and only those blocks

that were visually identified as containing the N100 and P200 were further scrutinized for the MMN (Fischer et al., 1999). Depending on the quality of the recorded data, this could lead to large quantities of discarded data, which ultimately makes the technique harder for clinicians to use. Despite these methods running the risk of data loss and the requirement of expert examination, they remain the gold standard for much of the clinical research literature (Gabriel et al., 2016; Naccache et al., 2016).

More recent investigations by Tzovara et al. have tried to reduce the quantity of data needed to generate meaningful information regarding the presence or absence of the MMN (Tzovara et al., 2013). Using an automated classification technique to quantify the neural response of each individual instead of an expert visual inspection, they were able to take a whole data set and blindly classify it, thereby increasing the amount of useful data. Furthermore, even though the model was able to accurately classify the neural responses of non-survivors, it was the positive or negative progression over time that was the major predictor of outcomes.

A notable commonality of this literature is the frequent use of greatly different MMN elicitation protocols and analysis methods to confirm its presence. This difference in method may be a contributing factor to the wide range in test specificity seen across studies. The inconsistency in test specificities is one of the reasons preventing clinicians from bringing ERP tests into a health care setting (see Discussion).

The current clinical gold standard of outcome prognosis include the Glasgow Coma Scale (GCS) (Teasdale & Jennett, 1974), which relies heavily on the patient's ability to produce behavioral responses to external stimuli and commands. However, early evoked potentials, like somatosensory responses and brainstem

auditory potentials are also of high prognostic value (Fischer, Dailler, & Morlet, 2008; Fischer et al., 1999). Tests that utilize later cortical responses like the MMN have further improved the prognostic value of these tests while continuing the move away strictly behaviorally based ones (Gawryluk, D'Arcy, Connolly, & Weaver, 2010). Such improvement is necessary considering the misdiagnosis rates for the unresponsive wakefulness syndrome (UWS) of between 41% (Schnakers et al., 2009) to 43% (Andrews, Murphy, Munday, & Littlewood, 1996) when relying on traditional behaviorally-based consensus methods. More structured behavioral assessment can significantly improve the diagnosis of UWS (Schnakers et al., 2009) and there is every reason to believe that functional brain measures would further improve diagnostic accuracy (Gosseries, Di, Laureys, & Boly, 2014; Gosseries, Zasler, & Laureys, 2014).

Kane suggested that the use of attention-modulated ERP components (e.g., P300) was a hindrance to their prognostic power because the response is larger in attentive individuals and is even absent in some healthy controls (Kane et al., 1993). However, it is precisely the sensitivity of certain responses to changes in attention that increases the power of these methods to examine levels of function in disorders of consciousness (DOC) patients and with improved protocols the absence of such responses in controls can be reduced dramatically.

In addition to the P300 reflecting different levels of cognitive activity, it is also sensitive to more sophisticated cognitive process such as memory-based processing including recognition of one's own name. For example, Holeckova et al. used a classic auditory oddball paradigm comprised of standard and infrequently occurring deviant tone stimuli, but was also able to elicit the P300 in response to rare and more salient novel sounds, such as the Subject's Own Name (SON) (Holeckova, Fischer, Giard, Delpuech, & Morlet, 2006). The SON

test is of particular interest as it is known to capture attention in the absence of effort (Wood & Cowan, 1995) and elicits a robust P300 response (Berlad & Pratt, 1995; Folmer & Yingling, 1997; Müller & Kutas, 1996; Perrin, García-Larrea, Mauguière, & Bastuji, 1999).

Building on this, another component that is sensitive to contextually-based expectancies is the N400. This component requires the integration of both pre-existing knowledge and newly parsed information. It is a late-occurring response associated broadly with comprehension of speech, text, and other stimuli possessing "meaning" (Kutas & Federmeier, 2011).

Expectations can be set up through the use of word pairs that may or may not have underlying semantic relationships (Holcomb & Neville, 1990), or sentences with incongruous, unexpected or infrequent endings (Connolly & Phillips, 1994). With word pairs, the N400 occurs to the second word when it violates the semantic context created by the first word. A significantly stronger semantic context is created by sentences so that a word that fails to meet contextually-based expectations results in a large N400 in contrast to little or no response to a contextually appropriate word (Kutas & Federmeier, 2011). In these cases, the mismatching words must be interpreted within a particular context, the processing of which requires elaborated attention and memory. Within the context of DOC, the evaluation of receptive language functions provides an objective measure of cognitive processing and, by implication, the level of consciousness.

This study addresses these issues by examining the various stimulation paradigms often used in the literature and selecting those that are best able to elicit strong ERP components of interest. The strongest paradigms will be integrated into a framework for extended and repeated testing of patients in comas for the prediction of both coma emergence and functional outcome. The

framework's design allows for data to be collected over extended durations and at milestone points during a patient's recovery, giving a better understanding of a patient's recovery trajectory and to further examine the stability of these components in brain injured patients over time. This framework does not prescribe a specific analysis method to replace expert visual inspection but rather aims to generate data that is agnostic toward the method of analysis. The components elicited should be strong enough to be detected using both traditional methods as well as newer machine learning-based methods.

2.2 Methods

2.2.1 Participants

Two groups of participants were recruited based on age range. Twenty-six native English speaking undergraduate students (19 females) from McMaster University were recruited for the younger adult group. Thirteen native English speaking adults (6 females) from the Hamilton community were recruited for the older adult group. Participants were 18 to 22 and 66 to 76 years old (M = 19.8, SD = 1.44; M = 69.8, SD = 3.35) for the younger and older adult groups, respectively. All participants were dextral (M = 85.6; Edinburgh Handedness Inventory Laterally Quotient Range: 40–100; Oldfield (1971)), had no history of neurological diseases or disorders, and had normal or corrected-to-normal vision. Undergraduate students received course credit for their participation, and older adults received $20.

A third group of participants was recruited for the follow up study and included twelve native English speaking students (10 females) from McMaster

University. These participants were 19 to 25 (M = 21.0, SD = 1.78), were dextral (M = 93.7; Laterally Quotient Range: 40–100; Oldfield (1971)), had no history of neurological diseases or disorders, and had normal or corrected-to-normal vision.

2.2.2 Electrophysiological methods

The electroencephalogram (EEG) was recorded continuously (bandpass = 0.01–100 Hz and sampled at 512 Hz) using a 64 channel Biosemi ActiveTwo system (Biosemi, Amsterdam, The Netherlands) using a 10-20 elastic cap with Ag/AgCl electrodes. The electrooculogram (EOG) was recorded from electrodes placed above and at the outer canthus of the left eye. References were recorded bilaterally from the mastoids and at the nasion for offline referencing.

Data preprocessing was conducted using BrainVision Analyzer 2. All recordings were visually inspected and epochs containing artifacts (e.g., muscle, movement) removed. Individual subtask recordings were filtered offline with a bandpass of 0.1–30 Hz. Ocular artifacts were corrected using the *Ocular ICA* transformation provided by BrainVision Analyzer 2.

Trials were segmented into epochs depending on the component of interest. Different pre-stimulus intervals for each component of interest were chosen based on the analysis methods used in the original work. For segments containing the MMN: 100 ms pre-stimulus to 500 ms post-stimulus (as in Fischer et al. (2008)); the P300: 200 ms pre-stimulus to 1000 ms post-stimulus (as in Holeckova et al. (2006)); the N400: 100 ms pre-stimulus to 1000 ms post-stimulus (as in Connolly and Phillips (1994)). Segments for each subtask were baseline corrected together to remove mean pre-stimulus activity. A final ar-

tifact rejection was performed automatically, removing segments with voltage steps greater than 50μV, voltage differences greater than 200μV in 200 ms, and channels with low activity (<0.5μV). Segments were averaged together per condition, per participant for each subtask.

In each condition, peaks were automatically detected for each channel independently within the following epochs: N1: 110–190 ms, P2: 180–280 ms, MMN: 120–240 ms, P300: 270–450 ms, N400: 300–700 ms. For further analysis, the latency and the mean amplitude of a 50 ms epoch around each peak were determined for each condition and participant.

2.2.3 Assessment battery

A battery of tasks was developed to evaluate increasing levels of auditory, cognitive, and linguistic processing. As this battery will be used to assess the level of consciousness of comatose patients, participants were informed that the auditory stimuli were of no relevance to the study and were free to view a silent film. The working hypothesis is that comatose patients are incapable of processing environmental stimuli, so the instructions were intended to better approximate the possible variability in the mental state of patients. The third group of participants received additional instructions (see Behavioural manipulation section of Procedure). The total time required to administer these tests was approximately 90 minutes. A brief description of each task follows.

All stimulus items were normalized using WaveGain, which is a program that applies the ReplayGain standard to sound files. The ReplayGain standard is a normalization technique that is based on the perceptual loudness of sounds. Auditory stimulus delivery was calibrated to 89 dB SPL using a continuous 800

Hz tone.

2.2.3.1 Oddball mismatch

The SON protocol (Fischer et al., 2008; Holeckova et al., 2006) has been demonstrated to invoke a P3b component to the subject's own name. This protocol was combined with an auditory oddball task like that in Fischer et al. (2008) to which an additional novel sound was included in order to elicit the P3a (c.f. Friedman, Cycowicz, and Gaeta (2001)). These protocols were chosen due to their prior use with clinical populations including comatose patients and those with UWS. The modifications introduced in the present study were intended to capture additional but related information on the processing levels already captured by the original protocols.

Stimuli consisted of standard (80%) and deviant (14%) tones, a familiar novel sound (the SON) (3%), and an unfamiliar novel sound that carried no linguistic content (a dog bark) (3%). Tones were digitally generated sine waves of 800 Hz, with a standard tone duration of 75 ms and a deviant tone duration of 30 ms. The familiar novel was a digital recording of the subject's name spoken by a native speaker of Canadian English in a neutral voice. The unfamiliar novel was a digital recording of a dog barking. Stimuli were presented pseudorandomly (no deviant or novel stimulus was preceded by less than two standard tones) in one block of 2000 items with a stimulus onset asynchrony (SOA) for the tones being 610 ms and 1220 ms for the novels.

2.2.3.2 Pattern violation mismatch

A different MMN-eliciting task was included to determine whether the type of expectancy violation would affect the resulting MMN component. This task was

a global pattern violation task (adapted from Sculthorpe and Campbell (2011)) and used the same tone stimuli as in the Oddball Mismatch task.

Tones were presented in an alternating pattern (i.e., ABABAB with, for example, A being the longer tone and B the shorter) so that violations were produced when the alternating sequence was altered by the double repetition of one of the stimuli (e.g., ABAB**BB**AB). Stimuli were presented in one block of 2000 items with 8% of the items being first repetition deviants and 8% being second repetition deviants. The stimulus onset asynchrony was 610 ms.

2.2.3.3 Subject's own name (SON)

Similar to the oddball mismatch task above, this task (adapted from Holeckova et al. (2006)) used the subject's own name to elicit the P300. This task was included to compare the P300 characteristics to names presented within the oddball mismatch to those presented alongside other names and words.

The subject's first name is presented alongside five other Common First Names (CFN) (two of the same gender as the participant and three of the opposite), and a list of ten mono- or di-syllabic Non-Salient Other Words (NSOW). These non-salient other words were high in frequency and matched the length of the subject's own name. All words were digital recordings of a speaker of Canadian English reading the words in a neutral voice. Each of these items were presented 60 times for a total of 480 trials. The speaker did not represent a familiar voice or person to the participant.

2.2.3.4 Semantic violation sentences

This task was a replication without behavioral responses of the terminal-word semantic violation paradigm (as in Connolly and Phillips (1994)) and used 144

sentences of six to twelve words recorded using natural speech. All sentences were digital recordings of a speaker of Canadian English reading in a neutral voice. These sentences were divided into groups of 36 having either: a semantically congruent terminal word (Phoneme Match-Semantic Match), *He takes his coffee with milk and SUGAR*; a semantically incongruent terminal word (Phoneme Mismatch-Semantic Mismatch), *The pizza was too hot to SING*; a terminal word with low cloze probability (Phoneme Mismatch-Semantic Match), *The pigs wallowed in the PEN*; or a phonological foil having a terminal word that began with the same initial sound as the semantically congruent word (Phoneme Match-Semantic Mismatch), *The gambler had a streak of bad LUGGAGE*.

2.2.3.5 Word-word priming

The final element of the battery was a replication without behavioral responses of the word-word auditory priming task from Holcomb and Neville (1990). All words were recorded using natural speech by a speaker of Canadian English reading in a neutral voice. All word pairs had a valid word of English as a prime. The targets included valid words of English that were either semantically congruent or incongruent, English pseudowords, or noise that was generated by reversing a valid English word. A total of 160 pairs of words were presented with 1150 ms between words and three seconds between trials.

2.2.4 Procedure

The study was approved by the Hamilton Integrated Research Ethics Board, Hamilton, Ontario, Canada. All persons gave their written informed consent prior to their inclusion in the study, in accordance with the ethical standards

of the Declaration of Helsinki. Participants completed a brief demographic questionnaire to ensure they met all inclusion criteria. While seated in a dimly lit room, auditory stimuli were delivered via Etymotic ER-1 insert earphones and participants were instructed to watch a silent video and disregard the sounds occurring in the background. Between each subtest, participants were given a brief break.

2.2.4.1 Behavioral manipulation

The third group of participants were only given a subset of the paradigms in the passive condition, as described above, and then were given the same paradigms with instructions to press buttons depending on the paradigm.

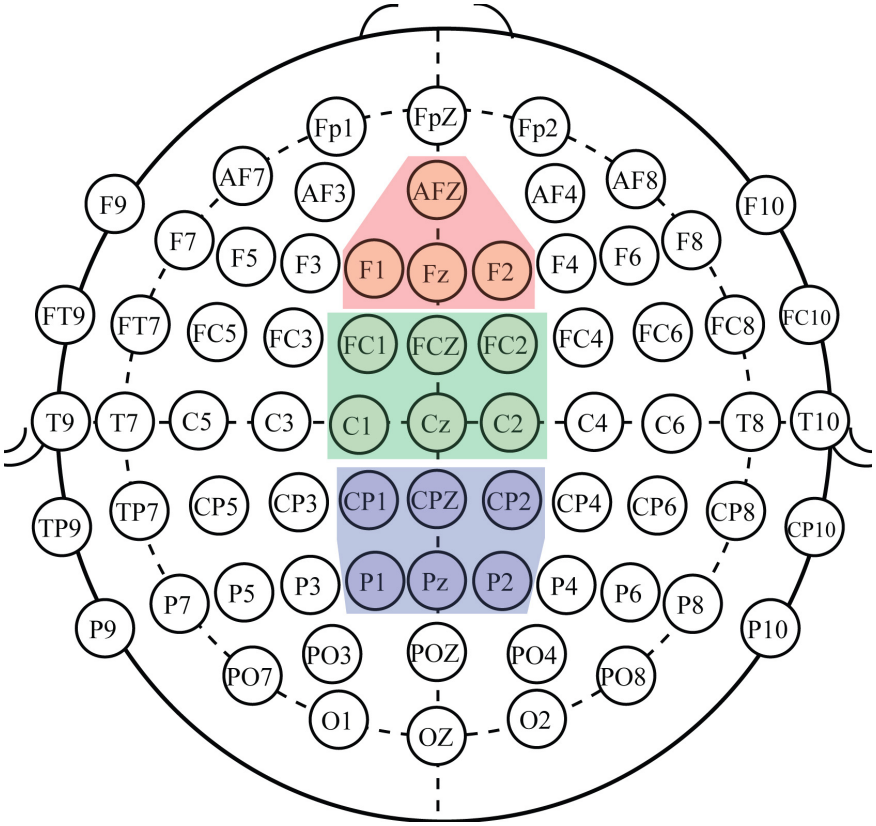
For the SON paradigm, the participants were instructed to press a button when they heard their own name. For the semantic violation sentences paradigm, they were instructed to press one button when the sentence was grammatical, and another if the sentence was ungrammatical. For the word-word priming paradigm, they were instructed to press one button if the target word was a valid English word, and another if the target word was not a valid word of English.

2.2.5 Statistical analyses

The ERP peak amplitude values were extracted by computing the mean value of 50 ms windows centered on the detected ERP peak. ERP peak amplitude and latency data were analyzed using separate mixed-design analysis of variance (ANOVA) models. The ANOVAs were conducted as omnibus tests with Greenhouse-Geisser corrections being applied to the degrees of freedom. All corrected probabilities are reported. Analyses were conducted separately for

each detected peak's amplitude and latency in each subtask individually using *Group* (younger and older adult) as a between-subjects factor, and *Condition* (levels dependent on protocol) and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as within-subjects factors. The ROIs were defined per the NEMO ROI plan (NEMO Consortium, 2012), and as shown in Fig 2.1.

FIGURE 2.1 – Graphical representation electrode members of the Regions of Interest (ROIs). Mid Frontal (red, 4 electrodes): F1, F2, Fz, AFz. Mid Central (green, 6 electrodes): C1, C2, Cz, FC1, FC2, FCz. Mid Parietal (blue, 6 electrodes): P1, P2, Pz, CP1, CP2, CPz.



To determine if the MMN was present for the pattern violation mismatch paradigm, the mean MMN peak amplitude for each combination of group, ROI, and deviant was compared for significant differences from zero using a one-tailed t-test (Näätänen, Pakarinen, Rinne, & Takegata, 2004).

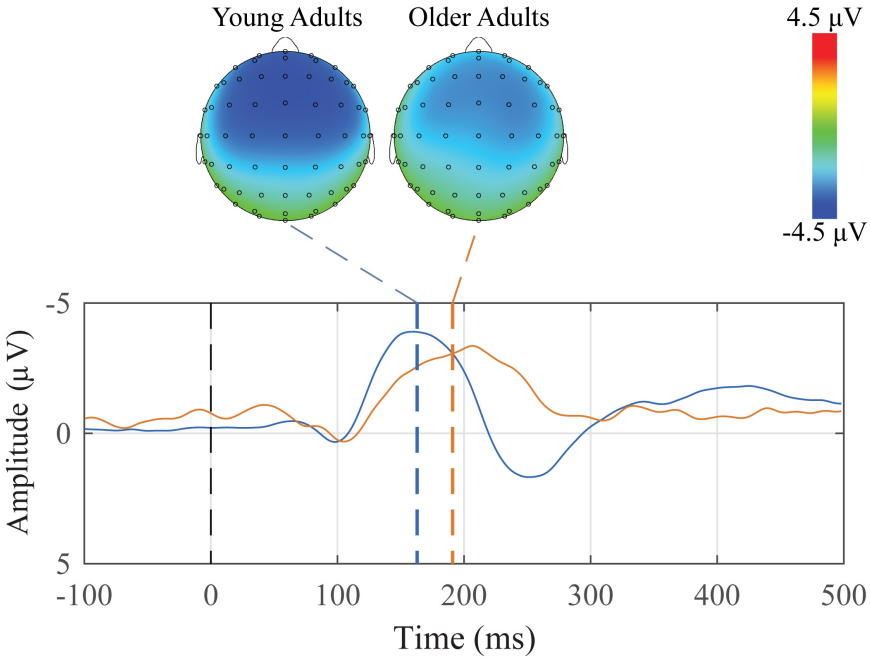
2.3 Results

2.3.1 Oddball mismatch

2.3.1.1 MMN

Grand average subtraction waveforms for both groups with peak topographic maps are presented in Fig 2.2. For younger adults, 21 of 26 participants showed the oddball mismatch, and for older adults, 12 of 13 participants showed the oddball mismatch.

FIGURE 2.2 – Grand average difference waveforms in the Mid Central ROI of the oddball mismatch MMN and corresponding peak topographic maps. The mean difference response for the younger adult (blue) and older adult (orange) groups are plotted. Dashed colored lines indicate the mean group latency from individually scored MMN peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



It is readily apparent that the MMN occurs earlier in the younger adults

than in the older adults. It is also clear that waveform morphology differed between the two groups with younger adults exhibiting a small post-MMN positivity (P200) but no such response being seen in older adults. Finally, the MMN topographical maps show a stronger and more focused scalp distribution for younger adults than for older adults.

Descriptive statistics for the amplitude and latency for both groups and the three ROIs are presented in Table 2.1. Mixed-design ANOVAs were conducted separately for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as a within-subject factor. Since the MMN is most readily detected by using subtractions between the deviant and standard conditions, there is no *Condition* factor in the ANOVA.

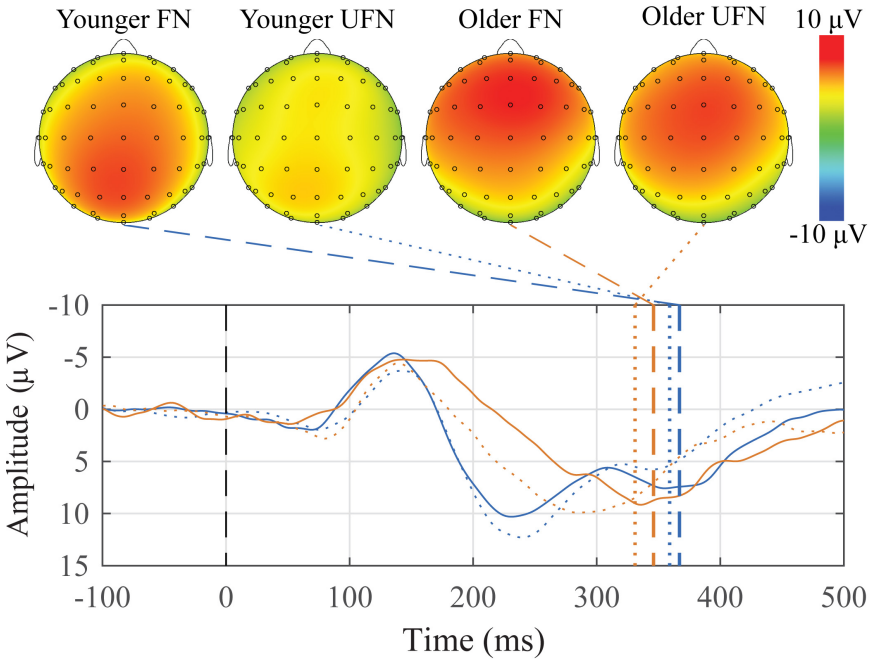
There was only a main effect of ROI for amplitude (F(2,74) = 44.001, $p <$ 0.001), with both the Mid Central and Mid Frontal ROIs being significantly more negative than the Mid Parietal ROI (all $p's <$ 0.001). There were both main effects of group (F(1,37) = 25.694, $p <$ 0.001) and ROI (F(2,74) = 13.443, $p <$ 0.001) for latency. The peaks for younger adults were significantly earlier than the older adults, and both the Mid Central and Mid Frontal ROIs occurring significantly earlier than the Mid Parietal ROI (all $p's <$ 0.02).

2.3.1.2 P300 Response to Novels

Grand average waveforms with peak topographic maps for both groups are presented in Fig 2.3. For younger adults, 25 of 26 participants had both a P300 response to the FN and UFN, and for older adults, 13 of 13 participants had both a P300 response to the FN and UFN.

Descriptive statistics for the amplitude and latency for both groups, the three

FIGURE 2.3 – Grand average waveforms at Mid Central ROI to the familiar and unfamiliar novels and corresponding peak topographic maps within the oddball mismatch. The mean responses to the familiar novel (FN) for younger adults (blue) and older adults (green), and the unfamiliar novel (UFN) for younger adults (orange) and older adults (red) groups are plotted. Dashed colored lines indicate the mean group latency from individually scored P300 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



	Younger Adults		Older Adults	
ROI	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	-3.23 (0.20)	158.5 (1.5)	-2.88 (0.16)	186.0 (2.9)
Mid Central	-3.10 (0.14)	163.3 (1.4)	-3.00 (0.10)	191.2 (2.3)
Mid Parietal	-2.37 (0.14)	167.9 (1.7)	-2.19 (0.11)	196.8 (2.2)

TABLE 2.1 – Means and standard errors of the mean of MMN peak amplitudes and latencies for the oddball mismatch.

ROIs, and both conditions are presented in Table 2.2. Both groups exhibited large P300s to both novel sounds, with the familiar novel (FN) generating the larger amplitude on average.

To first determine if there was a significant difference between the two P3 generating conditions and the standard baseline condition, and if there were any group differences, separate mixed-design ANOVAs were conducted for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor, and *Condition* (Standard, Familiar novel, Unfamiliar novel) as a within-subjects factor. There was only a main effect of condition for amplitude (F(2,74) = 55.390, $p < 0.001$), and both a main effect of group (F(1,37) = 7.736, $p < 0.001$) and condition (F(2,74) = 29.031, $p < 0.001$) for latency.

Post-hoc tests showed that all three conditions differed significantly from each other in terms of amplitude (all $p's < 0.001$), but only the FN and UFN conditions differed compared to the standard condition in terms of latency (all $p's < 0.001$).

To better understand the difference between the two P3 generating conditions and their scalp distribution, another pair of mixed-design ANOVAs were conducted for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor, and *Condition* (Familiar novel and Unfamiliar novel) and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as within-subjects factors.

For amplitude, there were main effects of both condition (F(1,37) = 5.020, $p = 0.031$) and ROI (F(2,74) = 3.758, $p = 0.048$), a Group x ROI interaction (F(2,74) = 11.055, $p < 0.001$), and a Group x Condition x ROI interaction (F(2,74) = 10.507, $p < 0.001$).

Post hoc tests showed that overall, the FN was significantly more positive than the UFN, the Mid Central ROI was significantly more positive than the

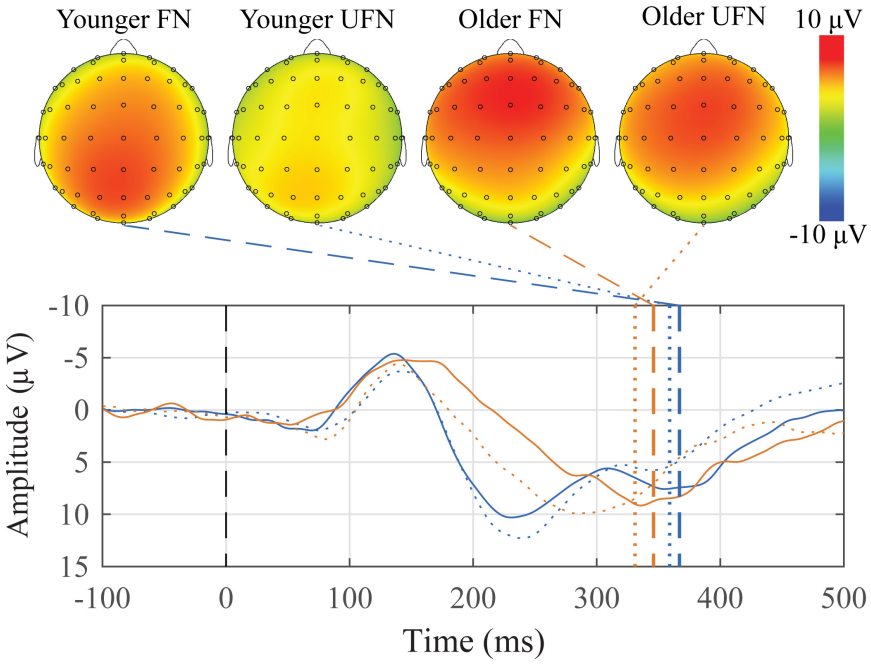
		Younger Adults		Older Adults	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	FN	6.58 (0.43)	364.4 (4.0)	9.49 (0.60)	353.3 (6.2)
	UFN	5.87 (0.51)	362.4 (4.2)	7.71 (0.75)	336.6 (7.1)
Mid Central	FN	8.26 (0.38)	366.6 (3.8)	9.47 (0.56)	345.9 (4.1)
	UFN	6.49 (0.44)	359.2 (4.6)	8.72 (0.64)	330.5 (5.0)
Mid Parietal	FN	9.18 (0.37)	375.7 (4.0)	7.53 (0.56)	377.0 (5.7)
	UFN	6.92 (0.37)	372.2 (5.0)	7.04 (0.61)	365.9 (9.7)

TABLE 2.2 – Means and standard errors of the mean of P300 peak amplitudes and latencies in response to novel stimuli in the oddball mismatch.

Mid Frontal ROI, but the Mid Parietal ROI was not significantly different than the other ROIs. The three-way interaction is shown graphically in Fig 2.4.

For latency, there was only a main effect of ROI (F(2,74) = 11.882, $p < 0.001$), where the P300 scored in the Mid Parietal ROI peaked significantly later than those in the other two ROIs (all $p's < 0.001$).

FIGURE 2.4 – The amplitude Group x Condition x ROI interaction to the P300 response within the oddball mismatch. The mean values of each combination of group, condition, and ROI are plotted. Younger adults are represented with red circles, and older adults with blue triangles. Error bars represent ± 0.5 SEM.

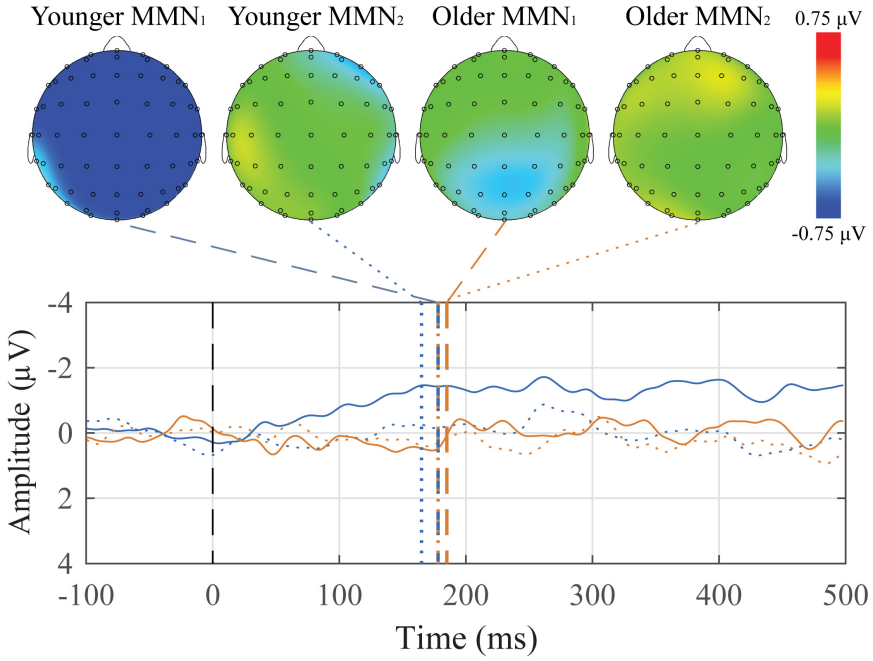


2.3.2 Pattern violation mismatch

Grand average waveforms and corresponding peak topographic maps for both groups are presented in Fig 2.5. For younger adults, 7 of 25 participants (one participant did not complete this paradigm) showed a MMN to the pattern violation mismatch. For older adults, 4 of 12 participants (one participant did

not complete this paradigm) showed a MMN to the pattern violation mismatch.

FIGURE 2.5 – Grand average difference waveforms in the Mid Frontal ROI of the pattern violation mismatch MMN to the first and second deviants and corresponding peak topographic maps. The mean difference response to the first and second deviants for the younger adult and older adult groups are plotted. Young adult first deviant (blue), younger adult second deviant (orange), older adult first deviant (green), and older adult second deviant (red). Dashed colored lines indicate the mean group latency from individually scored MMN peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



Two difference waves were generated: the difference between the first deviant and the standard (MMN_1), and the difference between the repetition of the deviant and the standard (MMN_2).

A summary of the mean amplitudes and latencies of these peaks for each group is presented in Table 2.3.

The results from the MMN-detection t-tests are given in Table 2.4. P-values have been adjusted using a per-group Bonferroni correction.

Since only the response to the first deviant in the mid central and mid

		Younger Adults		Older Adults	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	MMN_1	-1.69 (0.15)	177.5 (2.6)	-0.17 (0.11)	185.4 (3.4)
	MMN_2	-0.86 (0.14)	165.3 (2.7)	-0.05 (0.24)	177.5 (4.4)
Mid Central	MMN_1	-1.54 (0.01)	174.1 (2.2)	-0.40 (0.07)	181.1 (3.1)
	MMN_2	-0.71 (0.12)	169.7 (2.2)	-0.20 (0.17)	173.1 (3.8)
Mid Parietal	MMN_1	-1.37 (0.11)	173.2 (2.3)	-0.62 (0.09)	180.0 (3.0)
	MMN_2	-0.45 (0.12)	175.1 (2.8)	-0.33 (0.12)	172.3 (3.8)

TABLE 2.3 – Means and standard errors of the mean of MMN peak amplitudes and latencies for pattern violation mismatches.

Group	ROI	Condition	t-value	Degrees of Freedom	Adjusted p-value
Younger Adults	Mid Frontal	MMN_1	-10	100	<0.001
		MMN_2	-6	100	<0.001
	Mid Central	MMN_1	-10	100	<0.001
		MMN_2	-6	100	<0.001
	Mid Parietal	MMN_1	-10	100	<0.001
		MMN_2	-3	100	0.003
Older Adults	Mid Frontal	MMN_1	-2	50	0.370
		MMN_2	-0.4	50	1.000
	Mid Central	MMN_1	-5	70	<0.001
		MMN_2	-1	70	0.754
	Mid Parietal	MMN_1	-7	70	<0.001
		MMN_2	-3	70	0.0277

TABLE 2.4 – Means and standard errors of the mean of MMN peak amplitudes and latencies for pattern violation mismatches.

parietal ROIs was found to be significantly different from zero, the mixed-design ANOVAs were separately conducted for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor, and *ROI* (Mid Central and Mid Parietal) as a within-subjects factor, and only included peaks from the first deviant.

A significant main effect of Group (F(1,35) = 5.740, $p = 0.022$) and a Group x ROI interaction (F(1,35) = 6.920, $p = 0.008$) were found for peak amplitude. The younger adults had more negative peaks across both regions, with the mid central region being most negative. The older adults, however, had a more negative response in the mid parietal region than compared to the mid central. There were no significant differences with regard to the latencies.

2.3.3 Subject's own name (SON)

The grand average waveforms and peak topographic maps for SON, Common First Names, and Non-salient Other Words for younger and older adults are presented in Fig 2.6. For younger adults, 10 of 26 participants had a P300 response to their own name, and for older adults, 6 of 13 participants had a P300 response to their own name.

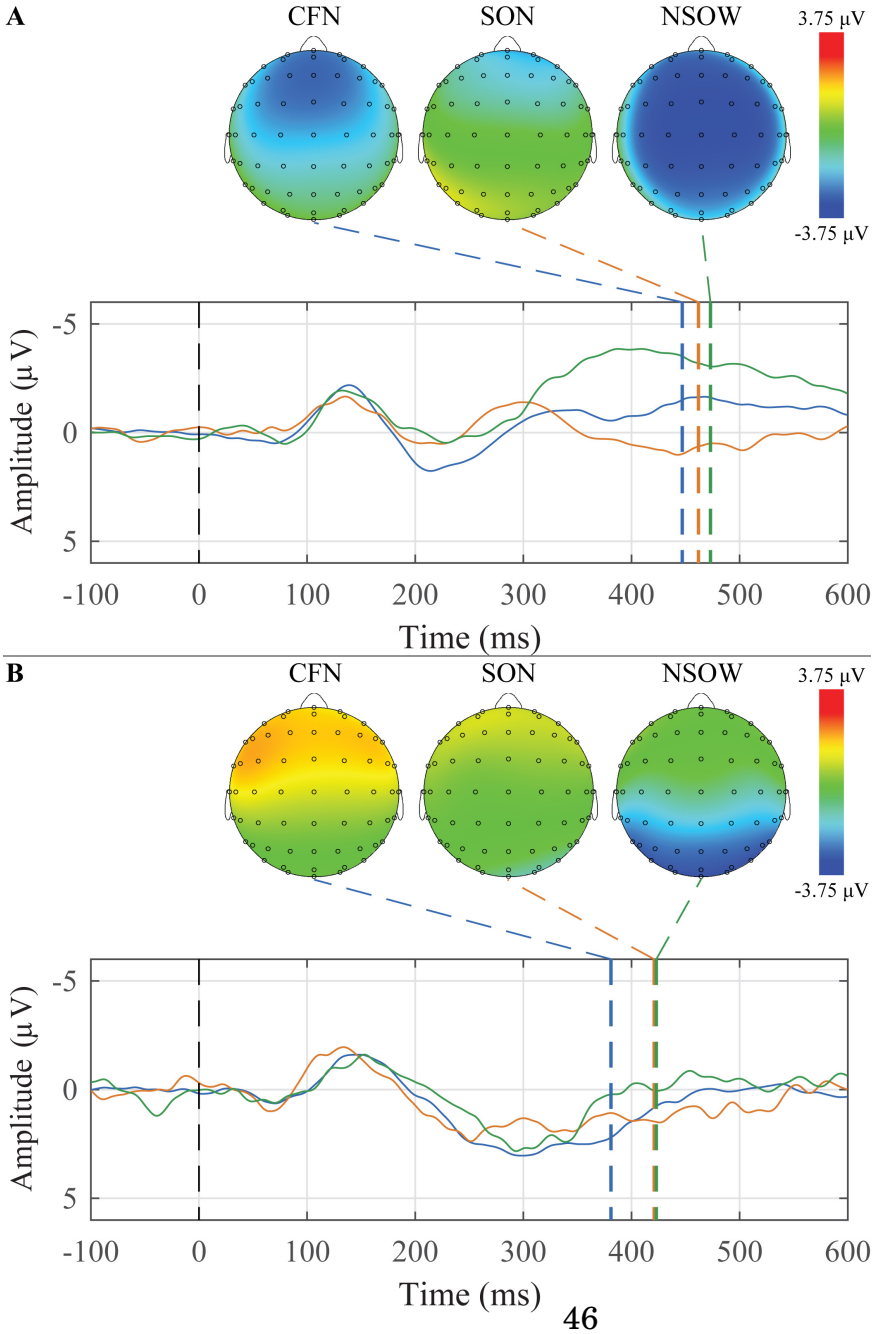
Descriptive statistics for the amplitude and latency for both groups, the three ROIs, and both conditions are presented in Table 2.5.

Mixed-design ANOVAs were conducted separately for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor, and *Condition* (Subject's Own Name, Non-salient Other Words, and Common First Names) and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as within-subjects factors.

For peak amplitude, there were significant main effects of group (F(1,37)

FIGURE 2.6 – Grand average waveforms to a list of Common First Names, the Subject's Own Name, and a list of Non-salient Other Words and their corresponding peak topographic maps within the Subject's Own Name paradigm.

A: The younger adult group's average responses in the Mid Parietal ROI to the Common First Names (blue), Subject's Own Name (orange), and the list of Non-salient Other Words (green) are plotted. B: The older adult group's average responses in the Mid Frontal ROI to the Common First Names (blue), Subject's Own Name (orange), and the list of Non-salient Other Words (green) are plotted.



		Younger Adults		Older Adults	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	SON	-0.10 (0.33)	444.8 (7.3)	2.44 (0.34)	420.8 (9.4)
	NSOW	-1.58 (0.37)	451.4 (7.0)	1.90 (0.48)	422.8 (11.8)
	CFN	-1.01 (0.18)	414.1 (7.8)	2.45 (0.33)	381.2 (7.6)
Mid Central	SON	0.54 (0.28)	433.2 (5.4)	1.69 (0.27)	428.1 (8.2)
	NSOW	-1.63 (0.28)	450.6 (6.0)	1.53 (0.30)	412.0 (9.9)
	CFN	-0.68 (0.14)	420.8 (6.2)	2.08 (0.23)	387.3 (7.5)
Mid Parietal	SON	1.35 (0.23)	461.6 (5.5)	0.85 (0.29)	482.5 (8.5)
	NSOW	-1.07 (0.25)	472.8 (6.5)	-0.29 (0.23)	428.3 (10.9)
	CFN	-0.20 (0.12)	446.8 (6.4)	0.89 (0.14)	432.7 (10.6)

TABLE 2.5 – Means and standard errors of the mean of P300 peak amplitudes and latencies for the Subjects Own Name paradigm.

= 11.05, $p = 0.002$) and condition (F(2,74) = 4.13, $p = 0.025$). There were also significant Group x ROI (F(2,74) = 31.17, $p < 0.001$) and Condition x ROI (F(4,148) = 3.99, $p = 0.015$) interactions. Generally, the older adults had peaks that were more positive than the younger adults. The Own Name condition was the most positive, followed by other names, and then other words. The interaction effects appear to be driven by the mid parietal region, where the difference between the two groups is smaller.

For peak latency, there was only a significant main effect of ROI (F(2,74) = 18.529, $p < 0.001$) with the mid parietal peaks occurring later than the other two regions.

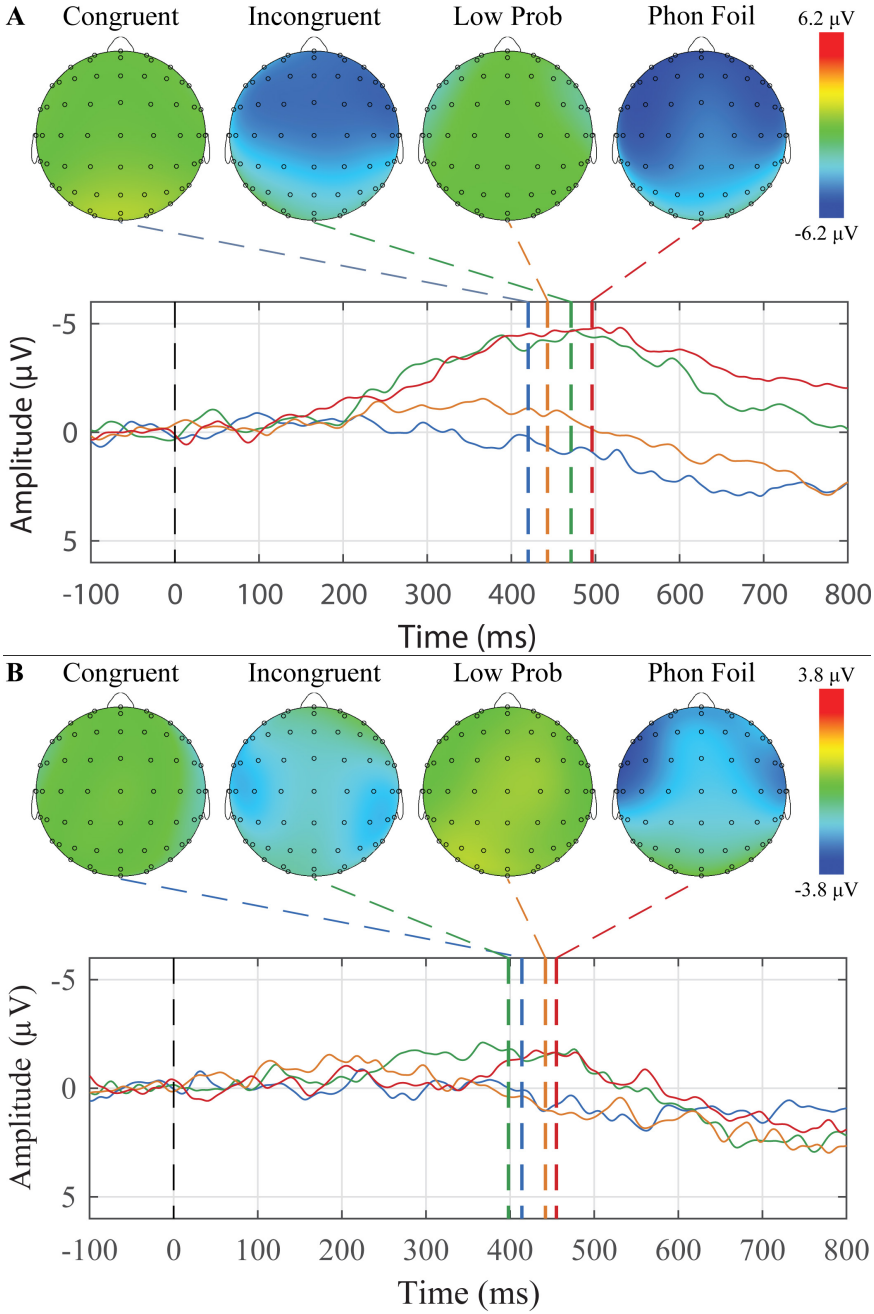
2.3.4 Semantic violation sentences

The grand average waveforms and peak topographic maps for the semantic violation sentences for younger and older adults are presented in Fig 2.7. N400 peaks were scored for each condition, group, and ROI. The mean amplitudes and latencies are presented in Table 2.6. For younger adults, 19 of 26 participants had a N400 response to incongruent terminal words, and for older adults, 5 of 13 participants had a N400 response to incongruent terminal words.

Mixed-design ANOVAs were separately conducted for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor, and *Condition* (Congruent, Incongruent, Phonological Foil, and Low Probability) and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as within-subjects factors. For amplitude, there were significant main effects of Group (F(1,37) = 11.783, $p = 0.001$), Condition (F(3,111) = 12.364, $p < 0.001$), and ROI (F(2,74) = 9.297, $p = 0.003$). Young adult peaks were overall significantly more negative than older

FIGURE 2.7 – Grand average waveforms to Congruent, Incongruent, Low Probability, and Phonological Foil terminal words and their corresponding peak topographic maps within the semantic violation sentences paradigm.

A: The Youth group's average responses to the list of Congruent (blue), Low Probability (orange), Incongruent (green) and Phonological Foil (red) terminal words are plotted. B: The Elderly group's average responses to the list of Congruent (blue), Low Probability (orange), Incongruent (green) and Phonological Foil (red) terminal words are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Colored bands indicate 50 ms windows used to extract mean amplitudes for use in ANOVA. Scalp topography maps show voltage distributions at mean group peak latencies.



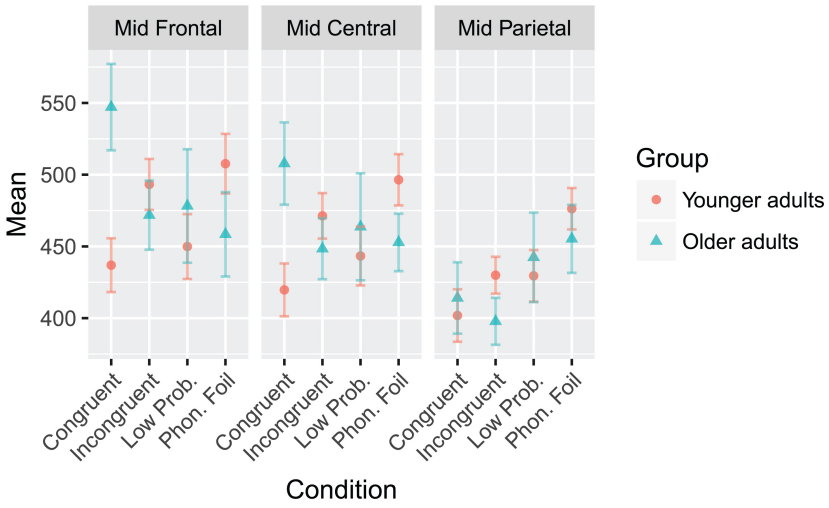
		Younger Adults		Older Adults	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	Incongruent	-5.22 (0.29)	493.2 (9.8)	-2.56 (0.36)	471.8 (14.3)
	Low Probability	-2.94 (0.35)	449.9 (12.1)	-1.68 (0.33)	478.2 (19.7)
	Phonological Foil	-5.66 (0.34)	507.7 (10.7)	-2.35 (0.38)	458.4 (15.6)
	Congruent	-1.48 (0.36)	436.9 (10.0)	-1.58 (0.36)	547.1 (17.3)
Mid Central	Incongruent	-5.34 (0.22)	471.3 (7.5)	-2.69 (0.30)	448.4 (10.5)
	Low Probability	-2.77 (0.25)	443.3 (8.8)	-1.61 (0.03)	463.7 (15.3)
	Phonological Foil	-5.47 (0.23)	496.4 (7.8)	-2.64 (0.28)	452.8 (10.1)
	Congruent	-1.59 (0.25)	419.7 (8.6)	-1.12 (0.32)	507.8 (15.1)
Mid Parietal	Incongruent	-4.28 (0.20)	429.9 (6.4)	-2.87 (0.34)	397.8 (8.1)
	Low Probability	-2.31 (0.22)	429.5 (8.1)	-1.38 (0.16)	442.4 (14.5)
	Phonological Foil	-5.15 (0.22)	476.3 (6.4)	-2.07 (0.22)	455.3 (10.1)
	Congruent	-0.91 (0.20)	401.8 (8.5)	-0.80 (0.33)	414.0 (11.3)

TABLE 2.6 – Means and standard errors of the mean of N400 peak amplitudes and latencies for semantic violation sentences.

adults ($p < 0.001$). Post hoc comparisons showed that Incongruent and Phonological Foil sentences were both significantly more negative than Low Probability sentences ($p's < 0.001$), which in turn were significantly more negative than Congruent ($p = 0.007$). There was no significant difference between Incongruent and Phonological Foil sentences. The mid parietal ROI was significantly had peaks that were significantly more negative than both the mid frontal and mid central ROIs, but those two ROIs were not significantly different from each other ($p's < 0.001$).

For latency, there was a significant main effect of ROI (F(2,74) = 32.457, $p < 0.001$), as well as Group x Condition (F(3,111) = 3.521, $p = 0.026$), Condition x ROI (F(6,222) = 3.990, $p = 0.003$), and Group x Condition x ROI (F(6,222) = 2.898, $p = 0.019$) interactions. Overall, peaks in the mid parietal region occurred earlier than both of the other regions ($p's < 0.001$), The three-way interaction is shown graphically in Fig 2.8.

FIGURE 2.8 – The latency Group x Condition x ROI interaction to the N400 response for the semantic violation sentences paradigm. The mean values of each combination of group, condition, and ROI are plotted. Younger adults are represented with red circles, and older adults with blue triangles. Error bars represent ±0.5 SEM.



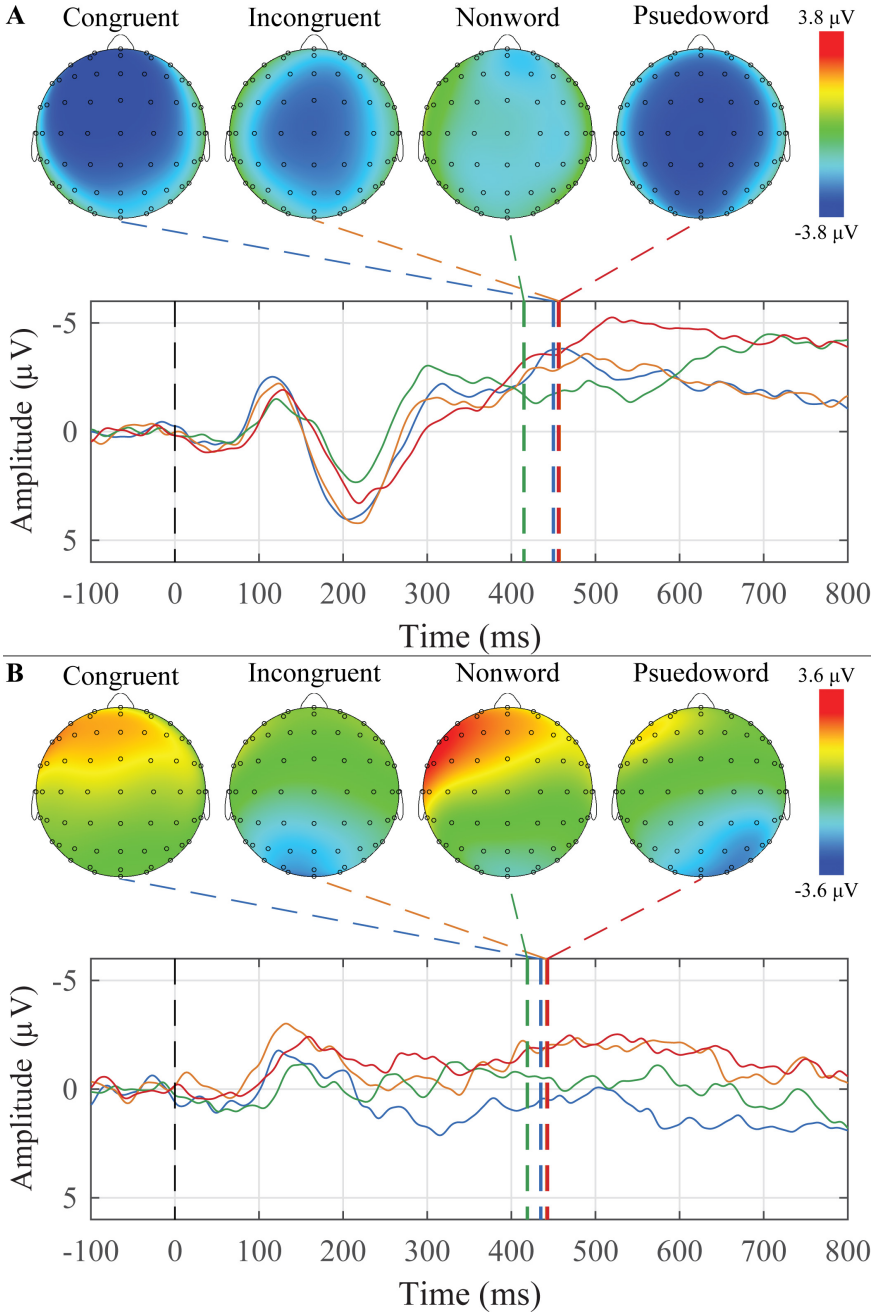
2.3.5 Word-word priming

The grand average waveforms and peak topographic maps for the word-word priming paradigm for younger and older adults are presented in Fig 2.9. N400 peaks were scored for each condition and each group. The mean amplitudes and latencies are presented in Table 2.7. For younger adults, 7 of 25 participants (one participant did not complete this paradigm) had a N400 response to incongruent target words, and for older adults, 4 of 11 participants (two participants did not complete this paradigm) had a N400 response to incongruent target words.

Mixed-design ANOVAs were conducted separately for amplitude and latency with *Group* (younger and older adults) as a between-subjects factor, and *Condition* (Semantically Congruent, Semantically Incongruent, Non-word, and Pseudoword) and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as within-subjects factors.

FIGURE 2.9 – Grand average waveforms to Congruent, Incongruent, Nonword, and Pseudoword target words and their corresponding peak topographic maps within the word-word priming paradigm.

A: The younger adult group's average responses in the Mid Central ROI to the list of Congruent (blue), Incongruent (orange), Nonword (green) and Pseudoword (red) target words are plotted. B: The older adult group's average responses in the Mid Parietal ROI to the list of Congruent (blue), Incongruent (orange), Nonword (green) and Pseudoword (red) target words are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



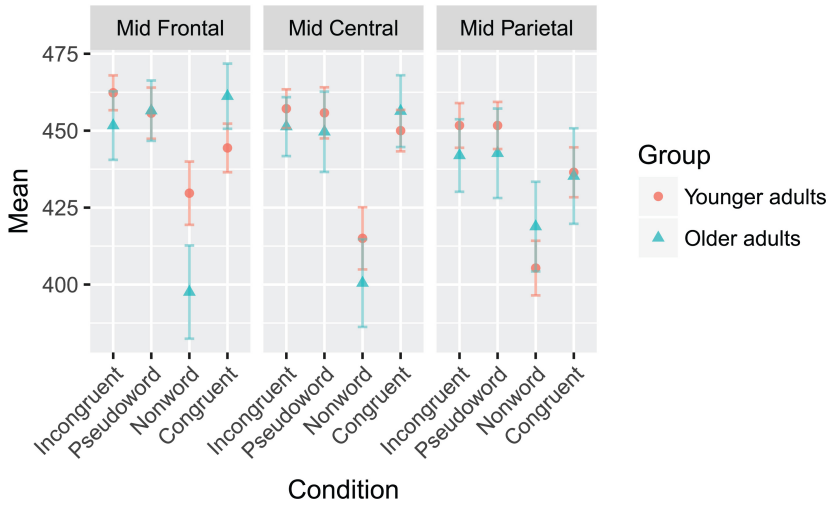
		Younger Adults		Older Adults	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	Incongruent	-3.64 (0.41)	462.3 (3.0)	-0.58 (0.48)	451.7 (5.5)
	Pseudoword	-4.36 (0.42)	455.7 (4.3)	0.83 (0.39)	456.5 (5.8)
	Nonword	-3.40 (0.44)	429.7 (5.5)	0.56 (0.57)	397.6 (7.6)
	Congruent	-3.86 (0.38)	444.4 (4.3)	1.30 (0.57)	461.2 (5.6)
Mid Central	Incongruent	-4.06 (0.35)	457.1 (2.7)	-1.49 (0.40)	451.3 (4.7)
	Pseudoword	-4.70 (0.29)	455.8 (3.6)	-0.74 (0.33)	449.6 (5.6)
	Nonword	-3.14 (0.35)	415.0 (4.4)	0.50 (0.34)	400.5 (6.2)
	Congruent	-3.72 (0.32)	450.0 (3.2)	0.95 (0.47)	456.4 (5.2)
Mid Parietal	Incongruent	-3.58 (0.30)	451.7 (3.3)	-2.73 (0.42)	441.9 (5.5)
	Pseudoword	-4.20 (0.21)	451.7 (3.5)	-2.62 (0.28)	442.7 (5.9)
	Nonword	-3.26 (0.30)	405.4 (3.7)	-0.47 (0.24)	418.8 (6.3)
	Congruent	-3.29 (0.28)	436.5 (3.9)	-0.40 (0.39)	435.3 (6.8)

TABLE 2.7 – Means and standard errors of the mean of N400 peak amplitudes and latencies for word-word priming.

For amplitude, there were significant main effects of group (F(1,34) = 10.424, $p = 0.003$) and ROI (F(2,68) = 8.499, $p = 0.003$), and a Group x ROI interaction (F(2,68) = 23.155, $p < 0.001$). Younger adults peaks were generally more negative than the older adults, and the mid parietal region was significantly more negative than the mid frontal region overall. Older adults appear to have a more focused negativity towards the mid parietal region, whereas the younger adults have a large negativity that is present across all three regions.

There were main effects of condition (F(3,102) = 8.488, $p < 0.001$) and ROI (F(2,68) = 6.015, $p = 0.009$) in terms of peak latency, as well as a significant Group x Condition x ROI interaction (F(6,204) = 3.765, $p = 0.009$). All non-word targets had significantly earlier peaks compared to all other conditions (all p's $<$ 0.001), and peaks in the mid parietal region occurred significantly earlier than those in the other two regions (all p's < 0.001). The three-way interaction is shown graphically in Fig 2.10.

FIGURE 2.10 – The latency Group x Condition x ROI interaction to the N400 response for the word-word priming paradigm. The mean values of each combination of group, condition, and ROI are plotted. Younger adults are represented with red circles, and older adults with blue triangles. Error bars represent ±0.5 SEM.



2.3.6 Behavioral manipulation

The analysis for the three paradigms used in the behavioral manipulation was the same as described above.

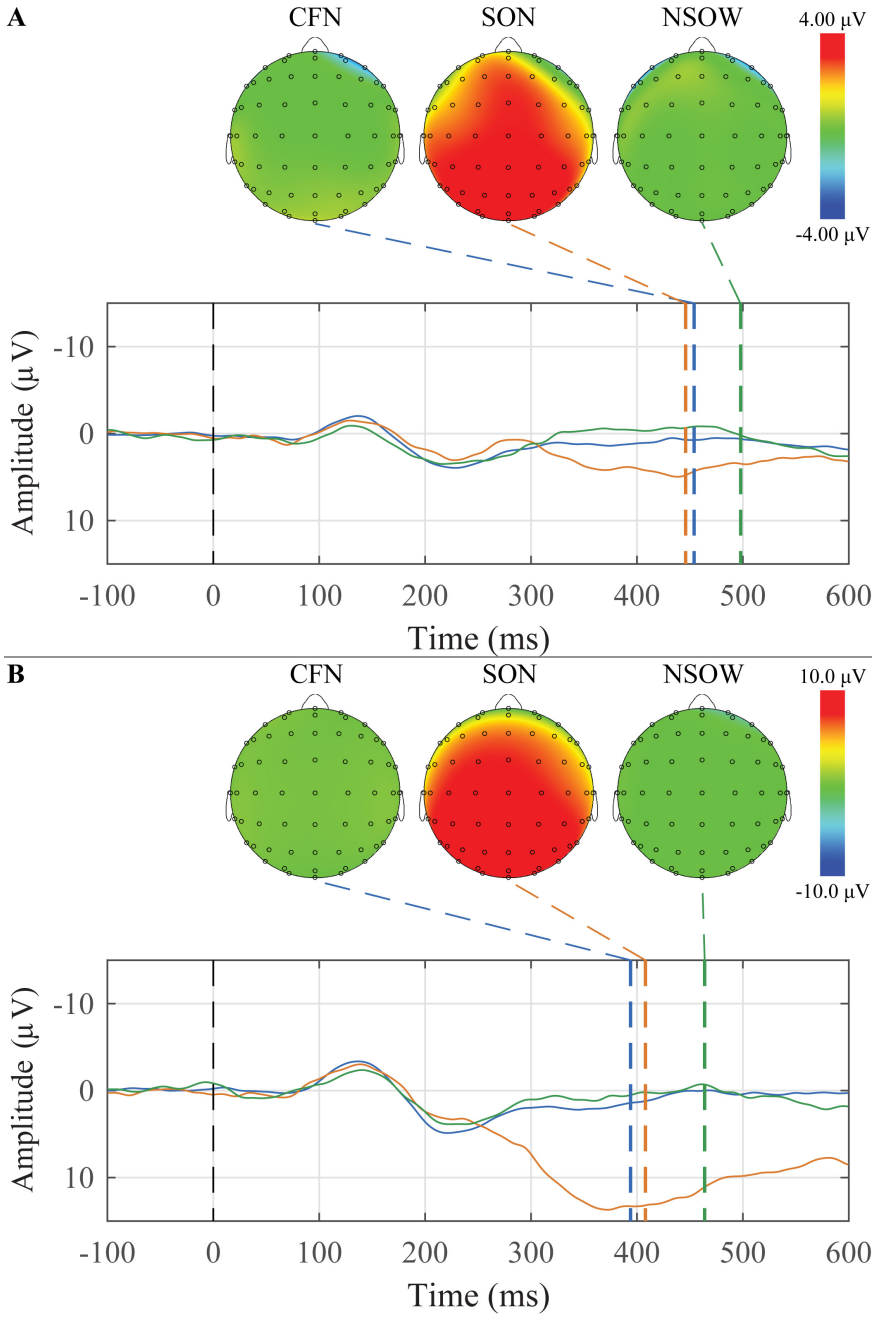
2.3.6.1 SON

The grand average waveforms and peak topographic maps for SON, Common First Names, and Non-salient Other Words for the Passive and Active tasks are presented in Fig 2.11. Descriptive statistics for the amplitude and latency for both groups, the three ROIs, and both conditions are presented in Table 2.8. In the passive condition, 9 of 13 participants had a P300 response to their own name, whereas in the active condition, 13 of 13 participants had a P300 response to their own name.

Repeated measures ANOVAs were conducted separately for amplitude and la-

FIGURE 2.11 – Grand average waveforms at Cz to a list of Common First Names, the Subject's Own Name, and a list of Non-salient Other Words and their corresponding peak topographic maps within the Subject's Own Name paradigm with the behavioral manipulation.

A: The average responses to the Common First Names (blue), Subject's Own Name (orange), and the list of Non-salient Other Words (green) in the Passive condition are plotted. B: The average responses to the Common First Names (blue), Subject's Own Name (orange), and the list of Non-salient Other Words (green) in the Active condition are plotted. Dashed colored lines indicate the mean group latency from individually scored P300 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



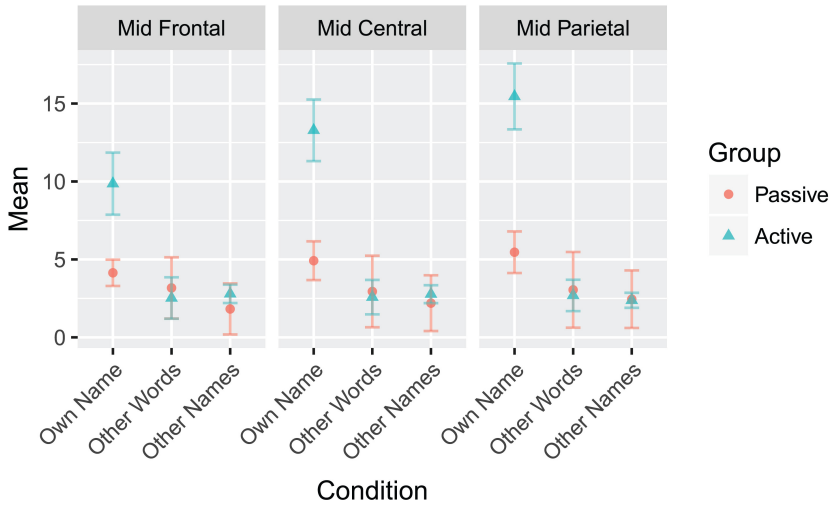
		Active		Passive	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	SON	9.86 (1.00)	385.1 (5.7)	4.14 (0.44)	418.1 (9.8)
	NSOW	2.52 (0.65)	411.5 (10.8)	3.17 (0.96)	452.3 (13.1)
	CFN	2.79 (0.29)	385.7 (9.6)	1.82 (0.80)	454.3 (14.5)
Mid Central	SON	13.28 (0.81)	400.5 (7.2)	4.92 (0.50)	425.6 (7.6)
	NSOW	2.58 (0.45)	444.3 (10.7)	2.94 (0.91)	477.3 (10.5)
	CFN	2.77 (0.23)	387.8 (7.2)	2.2 (0.70)	449.5 (11.1)
Mid Parietal	SON	15.46 (0.84)	408.1 (7.2)	5.46 (0.53)	446.2 (7.7)
	NSOW	2.69 (0.41)	464.4 (11.4)	3.05 (0.96)	498.3 (10.8)
	CFN	2.38 (0.2)	394.0 (6.9)	2.45 (0.73)	454.0 (9.8)

TABLE 2.8 – Means and standard errors of the mean of P300 peak amplitudes and latencies for the active and passive versions of the Subjects Own Name paradigm.

tency with *Behavioral Condition* (Passive and Active), *Task Condition* (Subject's Own Name, Non-salient Other Words, and Common First Names), and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as factors. For amplitude, there were significant main effects of task condition (F(2,44) = 20.370, $p < 0.001$) and ROI (F(2,44) = 5.760, $p = 0.019$), as well as Behavioural Condition x Task Condition (F(2,44) = 8.450, $p = 0.004$), Task Condition x ROI (F(4,88) = 11.210, $p < 0.001$), and Behavioural Condition x Task Condition x ROI (F(4,88) = 5.530, $p = 0.009$) interactions. Peaks to the subject's own name condition were significantly more positive than the other two conditions across behavioural condition and regions. Peaks were also, on average, significantly more positive in the mid parietal region than in the mid frontal region, but no significant difference was found between the mid central region and the other regions. The Behaviour x Task Condition interaction is attributable to the larger P300 amplitude to the SON in the Active compared to the Passive condition; an effect not found in the other two task conditions. The three-way interaction is shown graphically in Fig 2.12.

Peak latency had only significant main effects of behavioural condition (F(1,22) = 5.465, $p = 0.029$), task condition (F(2,44) = 4.190, $p = 0.030$), and ROI (F(2,44) = 4.190, $p = 0.008$). Overall, the P3 peaked earlier in the active condition compared to the passive condition. The P3 also peaked earlier to the subject's own name and to other names than to other words, and was on average later in the mid parietal region than the other two regions.

FIGURE 2.12 – The amplitude Behavioural Condition x Task Condition x ROI interaction to the P3 response for the Subject's Own Name paradigm. The mean values of each combination of behavioural condition, task condition, and ROI are plotted. The passive task condition is represented with red circles, and the active task condition with blue triangles. Error bars represent ±0.5 SEM.



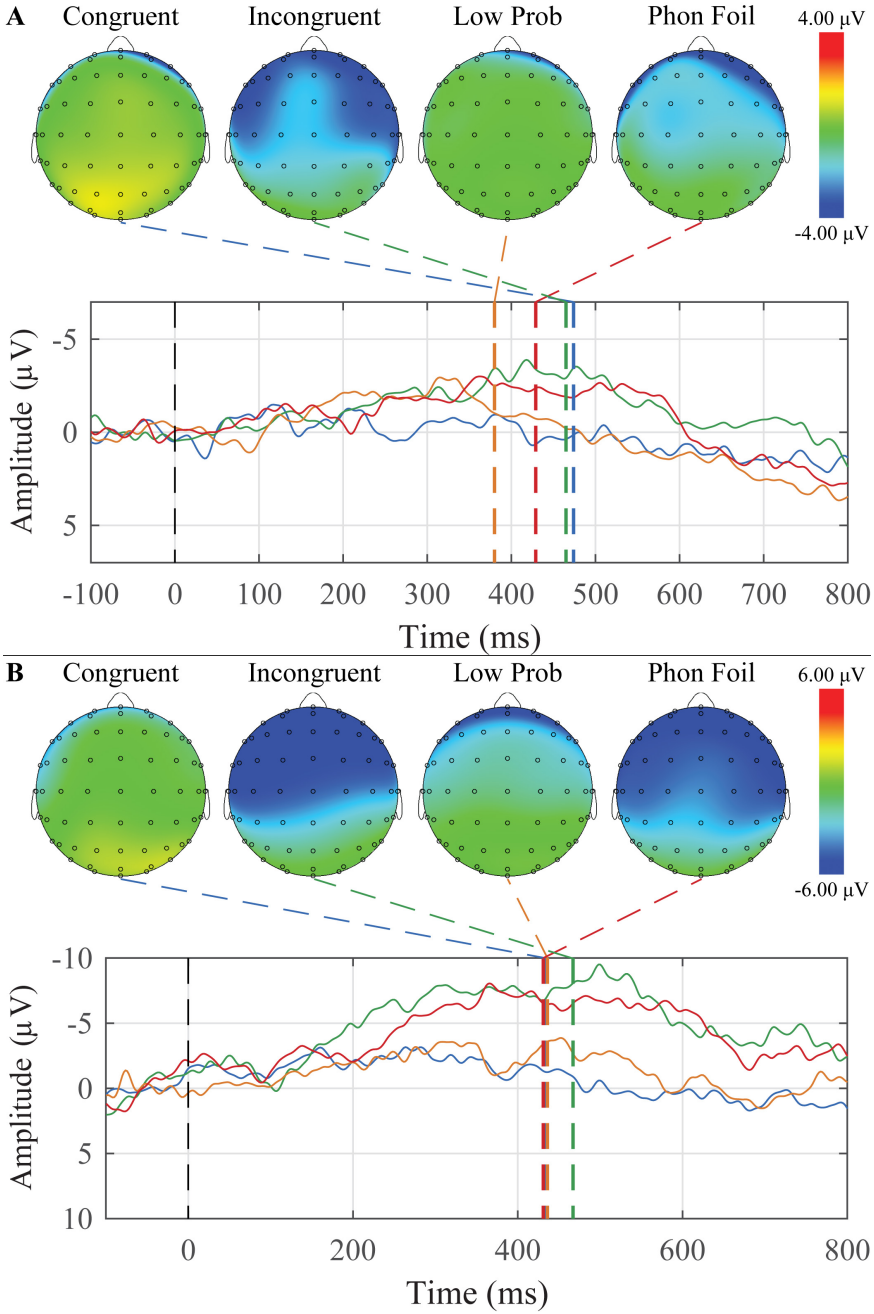
2.3.6.2 Semantic violation sentences

The grand average waveforms and peak topographic maps for the semantic violation sentences for the Passive and Active tasks are presented in Fig 2.13. N400 peaks were scored for each task condition and each behavioral condition. The mean amplitudes and latencies are presented in Table 2.9. In the passive condition, 4 of 13 participants had a N400 response to incongruent terminal words, whereas in the active condition, 9 of 13 participants had a N400 response to incongruent terminal words.

Repeated measures ANOVAs were conducted separately for amplitude and latency with *Behavioral Condition* (Passive and Active), *Task Condition* (Congruent, Incongruent, Phonological Foil, and Low Probability), and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as factors. For amplitude, there were

FIGURE 2.13 – Grand average waveforms at Cz to Congruent, Incongruent, Low Probability, and Phonological Foil terminal words and their corresponding peak topographic maps within the semantic violation sentences paradigm with the behavioral manipulation.

A: The average responses to the list of Congruent (blue), Low Probability (orange), Incongruent (green) and Phonological Foil (red) terminal words in the Passive condition are plotted. B: The average responses to the list of Congruent (blue), Low Probability (orange), Incongruent (green) and Phonological Foil (red) terminal words in the Active condition are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



		Active		Passive	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	Incongruent	-10.29 (0.84)	467.3 (13.4)	-5.26 (1.03)	465.2 (17.8)
	Low Probability	-4.98 (0.94)	435.8 (14.4)	-3.34 (0.65)	379.6 (13.7)
	Phonological Foil	-10.16 (1.10)	431.4 (13.8)	-4.42 (0.60)	429.1 (14.6)
	Congruent	-5.03 (0.87)	432.6 (18.1)	-3.17 (0.59)	473.8 (19.0)
Mid Central	Incongruent	-7.83 (0.45)	427.1 (11.7)	-4.54 (0.80)	453.1 (12.8)
	Low Probability	-3.16 (0.54)	406.1 (11.1)	-2.54 (0.53)	391.1 (11.4)
	Phonological Foil	-7.58 (0.61)	424.4 (8.5)	-4.24 (0.51)	424.2 (11.3)
	Congruent	-3.11 (0.43)	381.9 (11.1)	-2.21 (0.43)	467.3 (15.9)
Mid Parietal	Incongruent	-4.65 (0.31)	407.1 (10.8)	-3.44 (0.81)	417.5 (11.4)
	Low Probability	-1.30 (0.47)	396.1 (11.6)	-1.81 (0.56)	401.3 (10.0)
	Phonological Foil	-4.85 (0.53)	419.3 (8.1)	-3.13 (0.50)	421.8 (9.9)
	Congruent	-1.76 (0.41)	356.5 (7.9)	-1.31 (0.50)	442.7 (14.9)

TABLE 2.9 – Means and standard errors of the mean of N400 peak amplitudes and latencies for the active and passive versions of semantic violation sentences.

significant main effects of Task Condition (F(3,66) = 4.099, $p = 0.023$), and ROI (F(2,44) = 25.535, $p < 0.001$), and a significant Behavioural Condition x ROI interaction (F(2,44) = 5.559, $p = 0.023$). Overall, the response to Incongruent endings was the most negative, although there was no significant difference between Incongruent and Phonological Foil endings. All three regions were significantly different from each other, with the mid frontal being the most negative, and the mid parietal being the least negative. The interaction appears to be driven by a reduction in the difference between the active and passive behavioural condition in the mid parietal region, compared to the other regions.

For latency, there was only a significant main effect of ROI (F(2,44) = 12.036, $p < 0.001$) with mid parietal peaks occurring significantly earlier than those in the mid frontal region.

2.3.6.3 Word-word priming

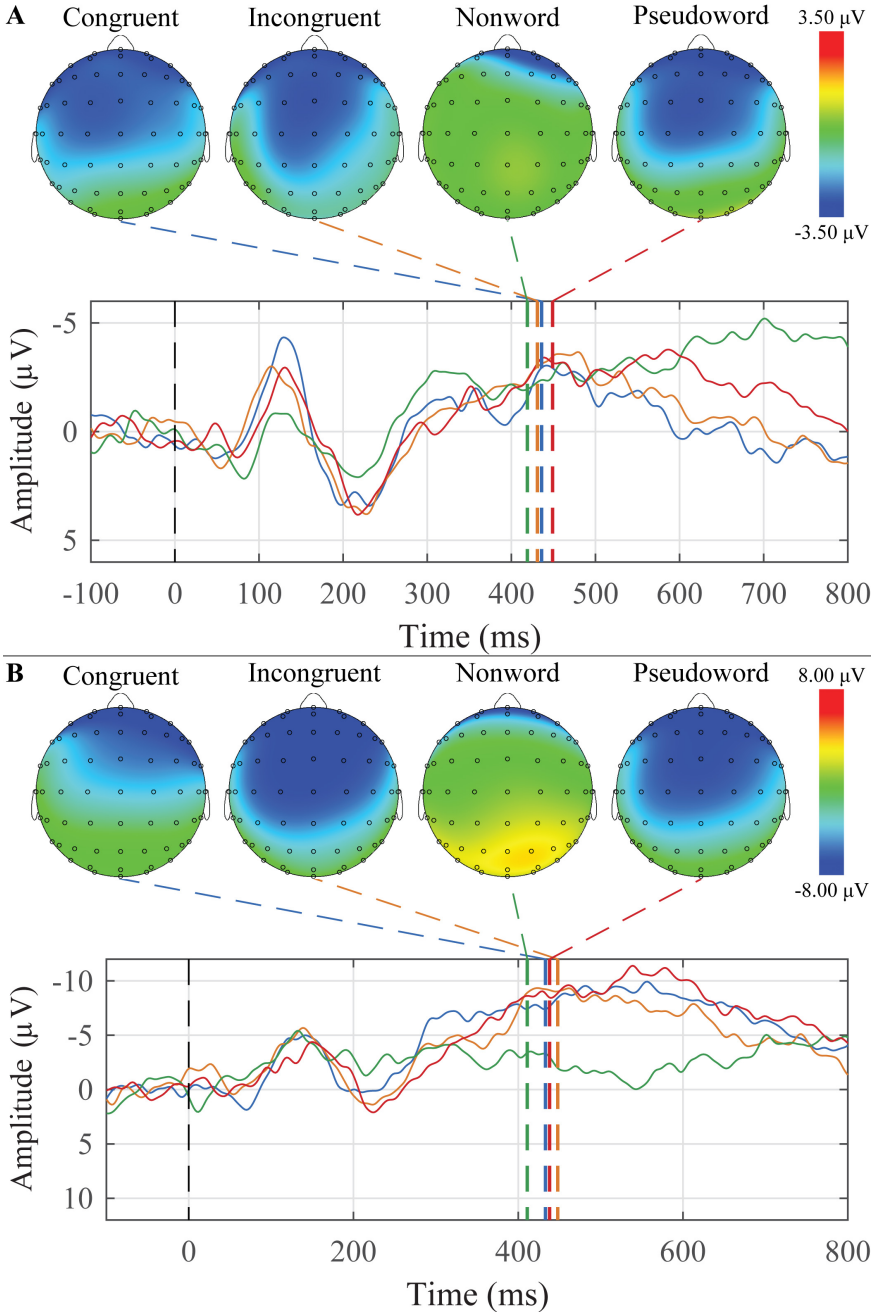
The grand average waveforms and peak topographic maps for the word-word priming paradigm for the Passive and Active conditions are presented in Fig 2.14. N400 peaks were scored for each condition and each group. The mean amplitudes and latencies are presented in Table 2.10. In the passive condition, 2 of 13 participants had a N400 response to incongruent target words, whereas in the active condition, 5 of 13 participants had a N400 response to incongruent terminal words.

Repeated measures ANOVAs were conducted separately for amplitude and latency with *Behavioral Condition* (Passive and Active), *Task Condition* (Semantically Congruent, Semantically Incongruent, Non-word, and Pseudoword), and *ROI* (Mid Frontal, Mid Central, and Mid Parietal) as factors.

For peak amplitude, there was a significant main effect of Task Condition

FIGURE 2.14 – Grand average waveforms at Cz to Congruent, Incongruent, Nonword, and Pseudoword target words and their corresponding peak topographic maps within the word-word priming paradigm with the behavioral manipulation.

A: The average responses to the list of Congruent (blue), Incongruent (orange), Nonword (green) and Pseudoword (red) target words in the Passive condition are plotted. B: The average responses to the list of Congruent (blue), Incongruent (orange), Nonword (green) and Pseudoword (red) target words in the Active condition are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.



		Active		Passive	
ROI	Condition	Amplitude (SEM)	Latency (SEM)	Amplitude (SEM)	Latency (SEM)
Mid Frontal	Incongruent	-9.94 (0.65)	448.0 (5.4)	-3.97 (0.73)	430.8 (6.3)
	Pseudoword	-10.29 (1.18)	438.3 (6.7)	-4.06 (0.88)	448.7 (5.4)
	Nonword	-5.10 (1.31)	410.6 (6.6)	-4.22 (0.83)	418.5 (8.0)
	Congruent	-10.46 (1.35)	433.4 (7.0)	-3.80 (0.84)	436.4 (6.6)
Mid Central	Incongruent	-8.84 (0.49)	437.3 (4.0)	-3.77 (0.51)	431.6 (4.8)
	Pseudoword	-8.46 (0.92)	423.3 (5.2)	-3.55 (0.59)	429.1 (4.6)
	Nonword	-1.28 (0.83)	389.3 (4.2)	-2.67 (0.76)	402.1 (6.4)
	Congruent	-7.95 (0.94)	421.5 (6.2)	-3.54 (0.58)	432.1 (6.6)
Mid Parietal	Incongruent	-5.29 (0.46)	434.3 (4.1)	-2.89 (0.39)	442.0 (5.0)
	Pseudoword	-4.71 (0.83)	423.7 (4.2)	-1.80 (0.63)	418.5 (4.6)
	Nonword	1.62 (0.60)	373.0 (3.5)	-1.76 (0.72)	393.1 (6.2)
	Congruent	-3.49 (0.73)	406.6 (6.3)	-2.25 (0.54)	427.5 (6.7)

TABLE 2.10 – Means and standard errors of the mean of N400 peak amplitudes and latencies for the active and passive versions of word-word priming.

(F(3,66) = 4.032, $p = 0.015$) with the responses to the Congruent, Incongruent, and Pseudoword targets all being significantly more negative than the response to Non-word targets (all p's < 0.001), and a significant main effect of ROI (F(2,44) = 26.671, $p < 0.001$), with the mid frontal and mid central regions being significantly more negative than the mid parietal. Additionally, there was a significant Behavioural Condition x Task Condition (F(3,66) = 3.054, $p = 0.042$) interaction where all target words except for the Non-word targets had a more negative response when actively responded to, and a significant Behavioural Condition x ROI (F(2,44) = 7.430, $p = 0.010$) interaction where the active condition became less negative in more posterior sites, but the passive condition was the same throughout.

For peak latency, there were only significant main effects of Task Condition (F(3,66) = 4.841, $p = 0.005$), with Non-word targets occurring significantly earlier than the responses to all other targets (all p's < 0.001), and ROI (F(2,44) = 13.364 $p < 0.001$) with mid parietal peaks occurring earlier than those in the mid frontal region.

2.4 Discussion

We recorded ERPs from 26 younger and 13 older healthy adults in five paradigms eliciting the MMN, P300, and N400 components. As one of the goals of this study was to evaluate and select paradigms that were capable of strongly eliciting the ERP components of interest, we will examine each grouping in turn. The number of participants who exhibited an ERP component in each paradigm is given in Table 2.11.

Passive-only	**Younger**	**Older**
Oddball mismatch (MMN)	21/26	12/13
Pattern violation mismatch	7/25	4/12
Oddball mismatch (P300 to novels)	25/26	13/13
SON	10/26	6/13
Semantic violation sentences	19/26	5/13
Word-word priming	7/25	4/11
Behavioural manipulation	**Passive**	**Active**
SON	9/13	13/13
Semantic violation sentences	4/13	9/13
Word-word priming	2/13	5/13

TABLE 2.11 – Counts of participants exhibiting the desired ERP response in each paradigm.

Two paradigms were used to elicit the MMN, differing primarily in the type of violation. As illustrated in Figs 2.2 and 2.5, although both paradigms produced negativities, those generated using the classical oddball paradigm (Fig 2.2) were larger (in some cases 4 μV), and more clearly defined than those seen in the pattern violation paradigm (Fig 2.5).

The oddball MMN had significant latency differences between the age groups, with older adult participants having later peak latencies compared to the younger adult group, and latency differences between conditions. The pattern violation paradigm did not generate strong MMNs, however there was a difference in amplitudes between groups, with younger participants having amplitudes that were almost four times larger than the older participants. This age-related attenuation effect is consistent with other studies which also found reduced MMN amplitudes at relatively short inter-stimulus intervals in older adult participants. (Cooper, Todd, McGill, & Michie, 2006; Pekkonen et al., 1996; Schroeder, Ritter, & Vaughan, 1995; Woods, 1992) Since the strongest and most reliable MMN was generated with the auditory oddball paradigm, that would be one of the paradigms included in the suggested battery.

Two paradigms were used to elicit the P300: novel sounds within an auditory oddball paradigm (Fig 2.3), and a subject's own name (SON) paradigm that included other names and words (Fig 2.6). The difference in the size of the waveforms between the two paradigms is quite noticeable. The N1 and P2 components in the oddball paradigm are almost 300% larger than those in the SON paradigm. Unexpectedly, the P300 to the subject's own name, which should be quite sizable, is late and very small in the SON paradigm employed here; but appears in the correct time window and is larger when embedded within tones. Within the oddball paradigm, there is a significant difference in amplitude between the familiar and unfamiliar novel conditions and the standard tone. There was also a significant difference between groups and between conditions. The P300 in the SON paradigm was expected to be largest to the subject's own name, however that is not the case for the older participants (Fig 2.6). These participants exhibited a positivity to the list of other common names, but not their own. Younger participants showed a later and sustained positivity to their own name compared to the other common names; even this response is, however, relatively small. Both groups exhibit an N400-like component to the list of other words. Overall, the younger participants displayed larger waveforms, which is seen as a significant main effect of group. These data would appear to support the previous claims that the P300 may not always be a reliable enough measure to be used in a clinical setting. (Connolly & D'Arcy, 2000; Connolly, D'Arcy, Newman, & Kemps, 2000; Picton, 1992) As the P300 was most reliably elicited to the novel stimuli in the auditory oddball paradigm without requiring attending to the sounds, we would again recommend the use of this paradigm in the suggested battery. This provides the added benefit of eliciting two components with only one stimulation paradigm.

The N400 was elicited using two paradigms of increasing semantic context: sentences with a terminal-word manipulation (Fig 2.7) and word-word priming (Fig 2.9). In both paradigms, no explicit instruction was given to participants to attend to any semantic relations between the words. In the case of the word-word priming, there was again an effect of age with the younger participants having larger N400's than the older. Across groups, there was a significant difference between the non-word targets and all other target words, with the non-words having earlier peak latencies.

The semantic violation sentences showed significant age effects, with younger participants having larger N400's than older. They also showed significant condition effects, with incongruent and phonological foil endings resulting in larger N400 amplitudes than the low probability and congruent conditions. The responses to the incongruent endings in the semantic violation sentences were generally more negative than the responses to the incongruent targets in the word-word priming. Considering the results up to this point, the semantic violation sentences appear to have a better ability to elicit the N400 without explicit instruction to attend to the sentences.

To better understand the effect of attention on these paradigms, we recorded ERPs from 13 other young healthy adults while they first passively experienced the stimuli, and then actively responded to what they were hearing.

Overall, the P300 responses to the SON and the N400 responses to semantically incongruent sentence endings and target words were much larger in the active task condition than in the passive task condition. This is in line with the amplitude differences reported in (Polich & McIsaac, 1994) where actively responded to oddball stimuli elicited more positive P300 responses than those that were passively listened to. This gives reason to always provide instruction to attend

to the stimuli irrespective of the participant's ability.

The use of ERPs to assess the clinical state of an individual is not without its complexities. ERPs are in some ways ideally suited to examine the cognitive consequences of brain injury because different ERP components are so strongly related to specific cognitive functions (Duncan et al., 2009).

EEG and ERPs in clinical contexts have many advantages: 1. The ability of most people to tolerate the less intimidating environment that characterize other brain recording systems (e.g., MRI); 2. The close relationship between particular ERP components and particular sensory, perceptual and cognitive processes – a feature shared only with the more expensive magnetoencephalography (MEG) methodology; 3. The lowest costs of any neuroimaging/recording method; and, 4. The exquisite sensitivity of ERP measures to many of the most common manifestations of CNS pathology, particularly acquired brain injury (ABI)–that is, generalized response latency delays, reduced response amplitudes and most notably domain-specific changes in latency/amplitude that reveal compromised functional integrity in attention, memory and language (see Harrison and Connolly (2013)).

Over the years, however, there has been the belief that ERPs, and in particular the oddball P300, are not sufficiently stable to serve as clinical tools. (Connolly & D'Arcy, 2000; Connolly et al., 2000; Picton, 1992) This view has often been conflated to include later-occurring responses that are related to higher level cognitive functions. While the MMN has long demonstrated its stability and relevance as a clinical tool, there remains some skepticism regarding the oddball P300 (Polich & Herbst, 2000). However, when evaluated against gold standard medical assays, the P300 recorded using the oddball paradigm fares well.

Using the coefficient of variation (CV) (Howell, 2012) and normative data from

various standard biomedical test norms (Statland, 1987), a comparison was made of these tests' CVs with CVs derived from literature-based data for P300 amplitude and latency (Polich & Herbst, 2000). Lower CVs imply a leptokurtic distribution rather than a platykurtic distribution making lower CVs preferable because any observed atypical response can more reliably be attributed to true abnormality rather than an uncontrolled source of variance. P300 amplitudes' CV values were comparable to assays for triglycerides used for assessing heart health – these values being amongst the highest in the collection of standard clinical assays evaluated. In contrast, CVs for P300 latencies were comparable to the lowest CVs recorded for standard clinical assays such as those for hemoglobin and potassium, some tests for thyrotropin, and considerably lower than assays for cholesterol and glucose (Polich & Herbst, 2000).

As impressive as these findings are, however, there remains a question about the test-retest reliability of the oddball P300 in individual subjects; and without reliability at the individual subject level there will be limited adoption of this protocol in clinical settings. Although a strong case is made for the clinical utility of the P300, Polich and Herbst acknowledge that protocols enabling improved sensitivity and discriminability are needed before P300 is adopted more widely for use in clinical settings (Polich & Herbst, 2000).

Further criticisms of the utility of the P300 in clinical settings include Picton (1992), who suggested that it reflected little beyond the fact that an individual was capable of responding differentially to frequently and infrequently occurring stimuli. He also identified the lack of relevance of the oddball P300 for a patient because it measured such a relatively inconsequential activity. A further limitation of the oddball P300 is its non-specificity; the response is frequently found to be delayed in latency and/or smaller in amplitude in a wide range

of CNS pathologies but specific to none. These characteristics are not in and of themselves "fatal flaws" for the P300 if the paradigm can be constructed to target a specific function in a highly reliable manner.

Bekinschtein et al. (2009) used a local/global paradigm and found in several experiments that a late positivity (the P3b) was obtained reliably if and only if participants were consciously aware of the global regularity pattern and violations of that pattern. That is, an unengaged participant did not exhibit a P3b–a finding that is compatible with earlier criticisms of the traditional P300 paradigms. In fact, Bekinschtein et al. included a "mind wandering" condition in their work and demonstrated an absence of the late positivity indicative of a failure to recognize the global stimulus structure and its violations. At the same time, however, the MMN was observed reliably and in accord with their findings in MCS patients was interpreted as reflecting "conscious processing of local regularities".

Despite concerns surrounding the use of classic oddball paradigms as clinical tests, they retain clinical assessment potential. However, as is often the case with many clinically useful assessment tests, classic oddball tests are best employed in conjunction with other tests that address more specific cognitive processes. Also, the entire test context should not be ignored any more than it would be ignored in a more traditional neuropsychological assessment context. That is, the choice of test, test sequence, and the initial difficulty level should be chosen on the basis of patient performance, and to the extent possible, clinical judgment.

Connolly and colleagues have proposed a complementary set of criticisms and suggestions that address the relevance of the testing paradigms to the individual patient and to the pathology being targeted; an approach that increases the

reliability of the P300 at the individual subject level (Connolly & D'Arcy, 2000; Connolly et al., 2000). Building on earlier arguments (Picton, 1992), they have noted that it is possible to address both the relevance to the patient issue and the reliability of the P300 by implementing protocols that are transparently relevant to the patient and ask more of the patient than differentiating stimuli that occur often from those that do not. For example, Connolly, Major, Allen, and D'Arcy (1999) adapted the vocabulary tests in the Wechsler Intelligence Scale for Children III (WISC-III) (Wechsler, 1991) and the Wechsler Adult Intelligence Scale-Revised as a Neuropsychological Instrument (WAIS-R-NI) (Kaplan, Fein, Morris, & Delis, 1991) in order to test receptive language skills and vocabulary knowledge in a group of younger adults.

The specificity of the P300 component as an assessment measure emphasizes the importance of task choice. In particular, specificity for the P300 was found to be 90.5% for the WISC-III, that indicated that 19/21 healthy participants showed the expected response statistically while specificity for the WAIS-R was 85.7% (18/21 participants). It is important to note that when these two psychometric tests were combined, the P300 specificity value was 90.5% (19/21 participants). These specificity values were obtained when participants' behavior (button presses to correct and incorrect choices) was taken into account by creating the averaged P300 with correctly identified deviant stimuli only. Connolly and colleagues recognized that relying on an ERP component whose specificity was possibly dependent on identifying a participant's behavioral accuracy and removing data obtained only when participants responded inaccurately might be of little value in trying to assess the cognitive ability of a patient incapable of executing any type of behavioral response. However, when behavioral responses were not taken into account and averaged P300 responses were obtained wit-

hout removing incorrectly identified trials (thus "simulating" a non-responsive participant), specificity values for the WISC-III actually increased to 95.2% (20/21 participants), but declined for the WAIS-R, 66.6% (14/21 participants). But when the two tests were combined for assessment purposes, specificity was 90.5% (19/21 participants)–precisely the same value obtained when behavior was accounted for. These findings demonstrate not merely the utility of this particular combination of tests in providing significant information about individuals' cognitive function but more to the point, they provide a demonstration that the use of ERPs generally and, in this case, P300 specifically, can be used with confidence in nonresponsive populations in the knowledge that one is obtaining accurate and valuable psychometric information about that individual's cognitive status.

Of course, context within a particular protocol is equally important. We have shown that the classic MMN oddball design of a frequently occurring series of "standard" tones interspersed with a less frequently occurring "deviant" tone results in larger MMN responses than an alternating pattern of tones (A,B) serving as the standard with a tone repetition (e.g., A,B,B,A,B) serving as the deviant. The P300 generated the strongest response when the subject's own name was put together with tones, which are acoustically very different, rather than with other words and names, which are very similar in nature. The strongest N400 response was generated when the semantic violation was put in a sentence context rather than in a less semantically rich word-word priming environment. (Lau, Almeida, Hines, & Poeppel, 2009; Zeelenberg, Pecher, Shiffrin, & Raaijmakers, 2003)

The role of context strength is critically important when evaluating language comprehension generally and even more so when using ERPs in clinical envi-

ronments. The choice of either a weaker word-word priming task instead of a stronger sentence design or the failure to ensure that sentences are structured for maximum contextual strength can result in healthy controls failing to show hypothesized N400 effects. This failure to establish a clear baseline control makes interpretation of patients' failures to show N400 responses impossible to interpret. (Cruse et al., 2014; Rohaut et al., 2015)

When comparing the two cohorts of younger participants in their performance on the two N400 paradigms, we see a decline in the number of responding participants in the smaller, second cohort. However, the pattern of the decline in performance with the reduction in strength of context still holds, both in the passive and active conditions in the second cohort. In both behavioural conditions, the number of responding participants to the word-word priming paradigm are nearly half of those who responded to the semantic violation sentences. While there are paradigms that are able to elicit the N400 with even stronger contexts, such as those used by van Berkum, Hagoort, and Brown (1999) that place the violation in the middle of the sentence or within a larger discourse, care must be taken to balance the cognitive processing requirements of those paradigms with their contextual strength. These paradigms may be too taxing on patients with traumatic brain injuries or diminished capacity, which may lead to false negatives or cases with too few good trials for averaging.

It is important to reiterate that the only instruction given to the participants in the first cohort was that they need not pay attention to the stimuli. As has been noted elsewhere (see Kutas and Federmeier (2011) for a review of N400 and attention, and Morlet, Ruby, André-Obadia, and Fischer (2017) for a discussion of the P300 and attention), not attending to the stimuli can significantly reduce the amplitude of an ERP response. As was seen in from the first cohort, the most

robust responses were generated when the stimuli were salient and contextually different enough to attract attention regardless of whether the stimuli were being attended to explicitly or not. An enhancement of this effect was seen when the participants were instructed to pay attention to the stimuli.

These results demonstrate the effects of attention on ERPs and by implication the processes they reflect. They also show that these effects are altered by the aging process. For example, N400 amplitudes to semantic violations in sentences were observed in both the younger and older adult participants; however, N400 amplitudes were reduced by 50% in the older adults. This same pattern of age-related attenuation is seen in several other paradigms, like the P300 response to the oddball names, the pattern violation MMN, and the word-word priming. Given that at least 75% of strokes in Canada are in people over the age of 65 (Public Health Agency of Canada, 2017), and that coma is often a consequence of stroke, it is very likely that this age-related ERP degradation may become a constant background feature that should be acknowledged and accounted for in future work. Similarly, providing instructions to patients regardless of diagnosis and apparent state of consciousness as well as using the most stable paradigms available from the literature are all procedurally essential for providing patients the opportunity to generate the most robust responses of which they may capable.

In summary, we recommend the use of an auditory oddball paradigm that includes novel stimuli to elicit the MMN and P300, and semantic violation sentences to elicit the N400. As we have demonstrated and as illustrated in Table 2.11, on average, almost 90% of all participants showed a MMN to the auditory oddball paradigm, and nearly 100% of all participants showed a P300 to the novel stimuli in this paradigm. While the proportion of participants responding to

the semantic violation sentences was markedly lower than the auditory oddball paradigm, the stronger context provided by the sentences improved the response rate compared to using a word-word priming paradigm. There was also an enhancing effect of paying attention to rather than ignoring the semantic violation sentences, with the number of participants showing a N400 doubling. This combination of paradigms appears to allow for a robust response in the absence of explicit attention, and are reinforced when the participant is instructed to pay attention. It is also important to note that when applying these tests to clinical environments, that the absence of a positive response should not be interpreted as a negative response. The purpose of these assessments, at least in their current state, should be to better inform clinicians and enable to start rehabilitation treatments earlier in cases of positive results.

2.5 Supporting information

FIGURE 2.S1 – Grand average difference waveforms of the oddball mismatch MMN and corresponding peak topographic maps. The mean difference response for the younger adult (blue) and older adult (orange) groups are plotted for the (A) Mid Frontal, (B) Mid Central, and (C) Mid Parietal ROIs. Dashed colored lines indicate the mean group latency from individually scored MMN peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

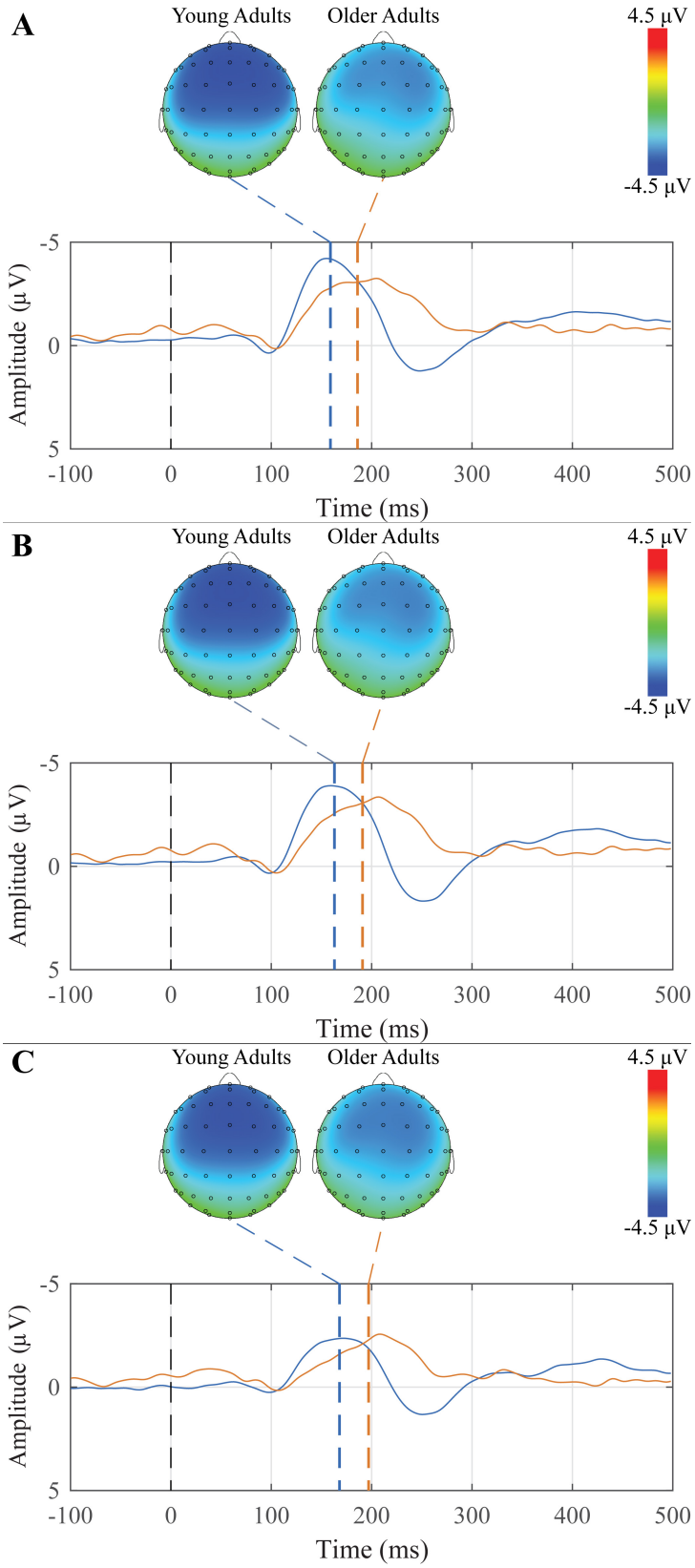

A
Young Adults
Older Adults
4.5 µV
-4.5 µV
Amplitude (µV)
-5
0
5
-100
0
100
200
300
400
500
Time (ms)
B
Young Adults
Older Adults
4.5 µV
-4.5 µV
Amplitude (µV)
-5
0
5
-100
0
100
200
300
400
500
Time (ms)
C
Young Adults
Older Adults
4.5 µV
-4.5 µV
Amplitude (µV)
-5
0
5
-100
0
100
200
300
400
500
Time (ms)

FIGURE 2.S2 – Grand average waveforms to the familiar and unfamiliar novels and corresponding peak topographic maps within the oddball mismatch. The mean responses to the familiar novel (FN) for younger adults (blue) and older adults (green), and the unfamiliar novel (UFN) for younger adults (orange) and older adults (red) groups are plotted for the (A) Mid Frontal, (B) Mid Central, and (C) Mid Parietal ROIs. Dashed colored lines indicate the mean group latency from individually scored P300 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

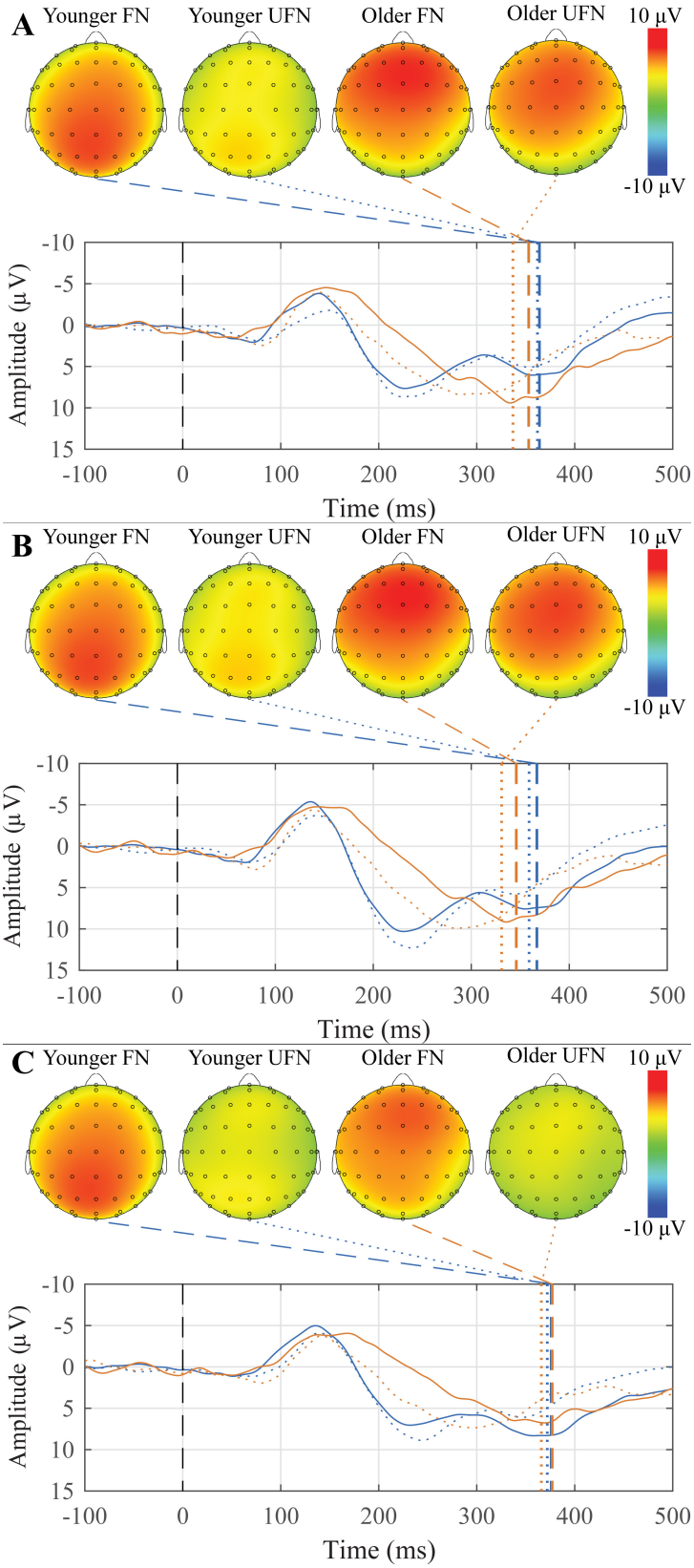

A
Younger FN
Younger UFN
Older FN
Older UFN
10 μV
-10 μV
Amplitude (μV)
-10
-5
0
5
10
15
-100
0
100
200
300
400
500
Time (ms)
B
Younger FN
Younger UFN
Older FN
Older UFN
10 μV
-10 μV
Amplitude (μV)
Time (ms)
C
Younger FN
Younger UFN
Older FN
Older UFN
10 μV
-10 μV
Amplitude (μV)
Time (ms)

FIGURE 2.S3 – Grand average difference waveforms of the pattern violation mismatch MMN to the first and second deviants and corresponding peak topographic maps. The mean difference response to the first and second deviants for the young adult and older adult groups are plotted for the (A) Mid Frontal, (B) Mid Central, and (C) Mid Parietal ROIs. Young adult first deviant (blue), young adult second deviant (orange), older adult first deviant (green), and older adult second deviant (red). Dashed colored lines indicate the mean group latency from individually scored MMN peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

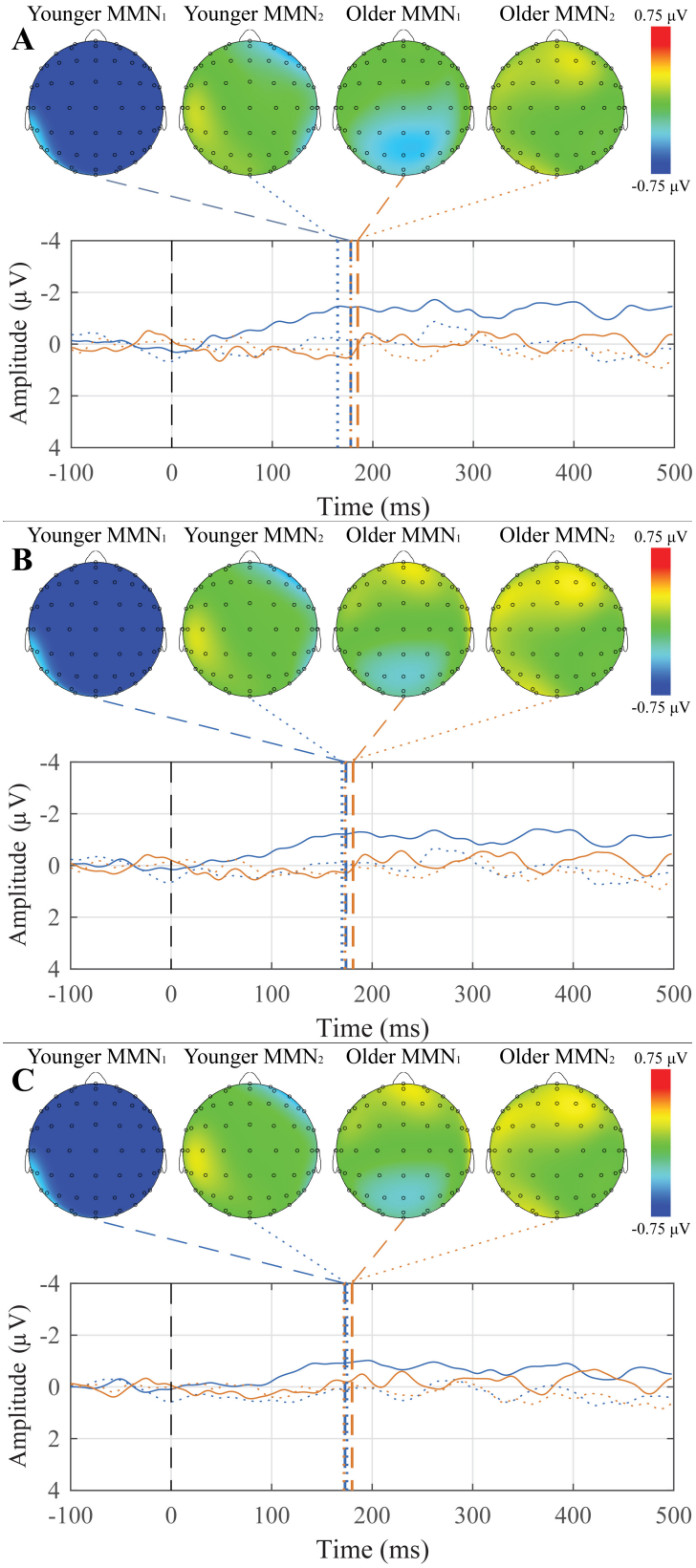

A
Younger MMN1
Younger MMN2
Older MMN1
Older MMN2
0.75 µV
-0.75 µV
Amplitude (µV)
-4
-2
0
2
4
-100
0
100
200
300
400
500
Time (ms)
B
Younger MMN1
Younger MMN2
Older MMN1
Older MMN2
0.75 µV
-0.75 µV
Amplitude (µV)
Time (ms)
C
Younger MMN1
Younger MMN2
Older MMN1
Older MMN2
0.75 µV
-0.75 µV
Amplitude (µV)
Time (ms)

FIGURE 2.S4 – Grand average waveforms at all ROIs to a list of Common First Names, the Subject's Own Name, and a list of Non-salient Other Words and their corresponding peak topographic maps within the Subject's Own Name paradigm. The younger adult group's average responses in the (A) Mid Frontal, (C) Mid Central, (E) Mid Parietal ROIs, and the older adult group's average responses in the (B) Mid Frontal, (D) Mid Central, (F) Mid Parietal ROIs to Common First Names (blue), Subject's Own Name (orange), and the list of Non-salient Other Words (green). Dashed colored lines indicate the mean group latency from individually scored P300 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

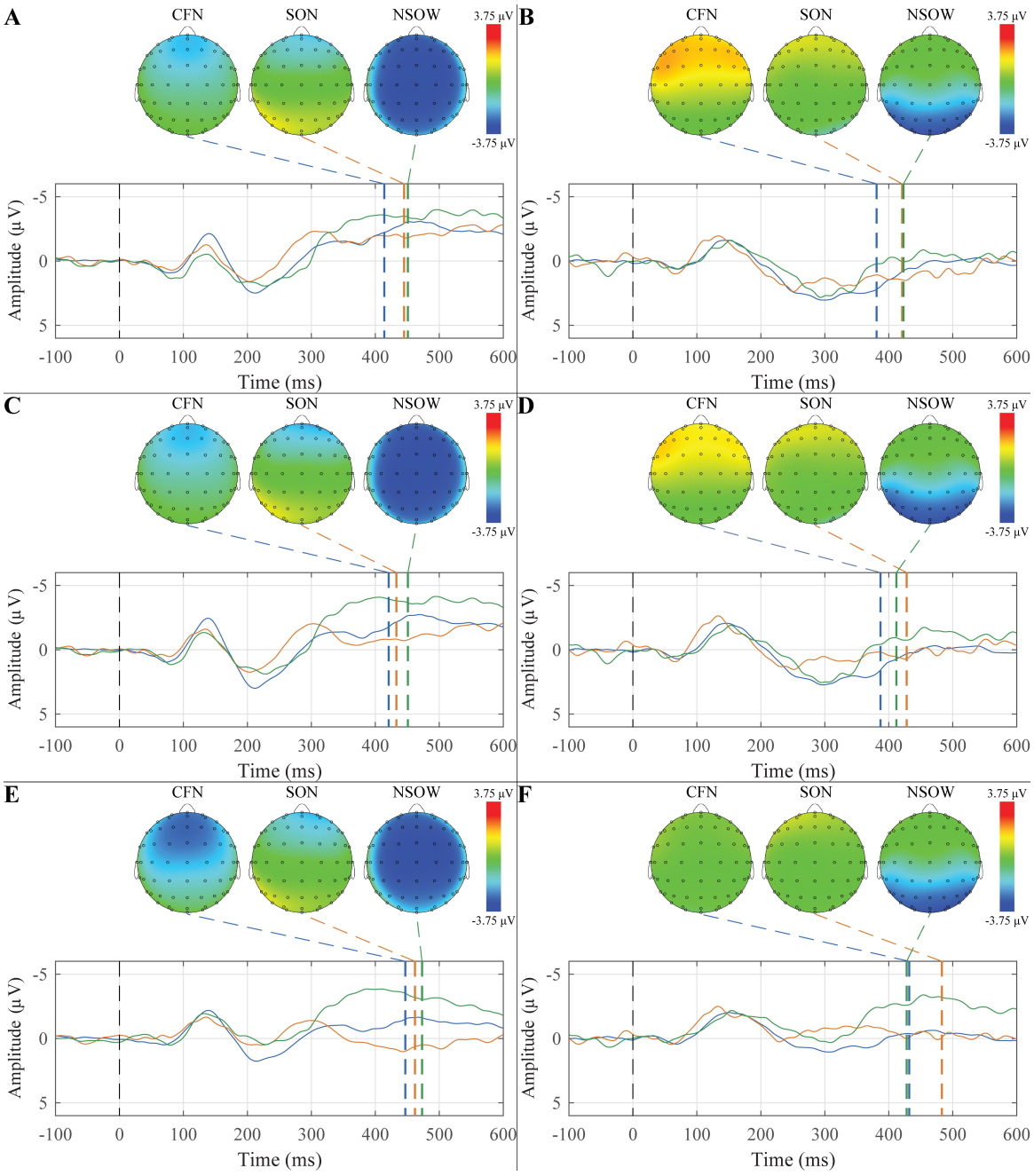

A
B
C
D
E
F
CFN
SON
NSOW
3.75 µV
-3.75 µV
Amplitude (µV)
-5
0
5
-100
0
100
200
300
400
500
600
Time (ms)

FIGURE 2.S5 – Grand average waveforms at all ROIs to Congruent, Incongruent, Low Probability, and Phonological Foil terminal words and their corresponding peak topographic maps within the semantic violation sentences paradigm. The younger adult group's average responses in the (A) Mid Frontal, (C) Mid Central, (E) Mid Parietal ROIs, and the older adult group's average responses in the (B) Mid Frontal, (D) Mid Central, (F) Mid Parietal ROIs to Congruent (blue), Low Probability (orange), Incongruent (green) and Phonological Foil (red) terminal words are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

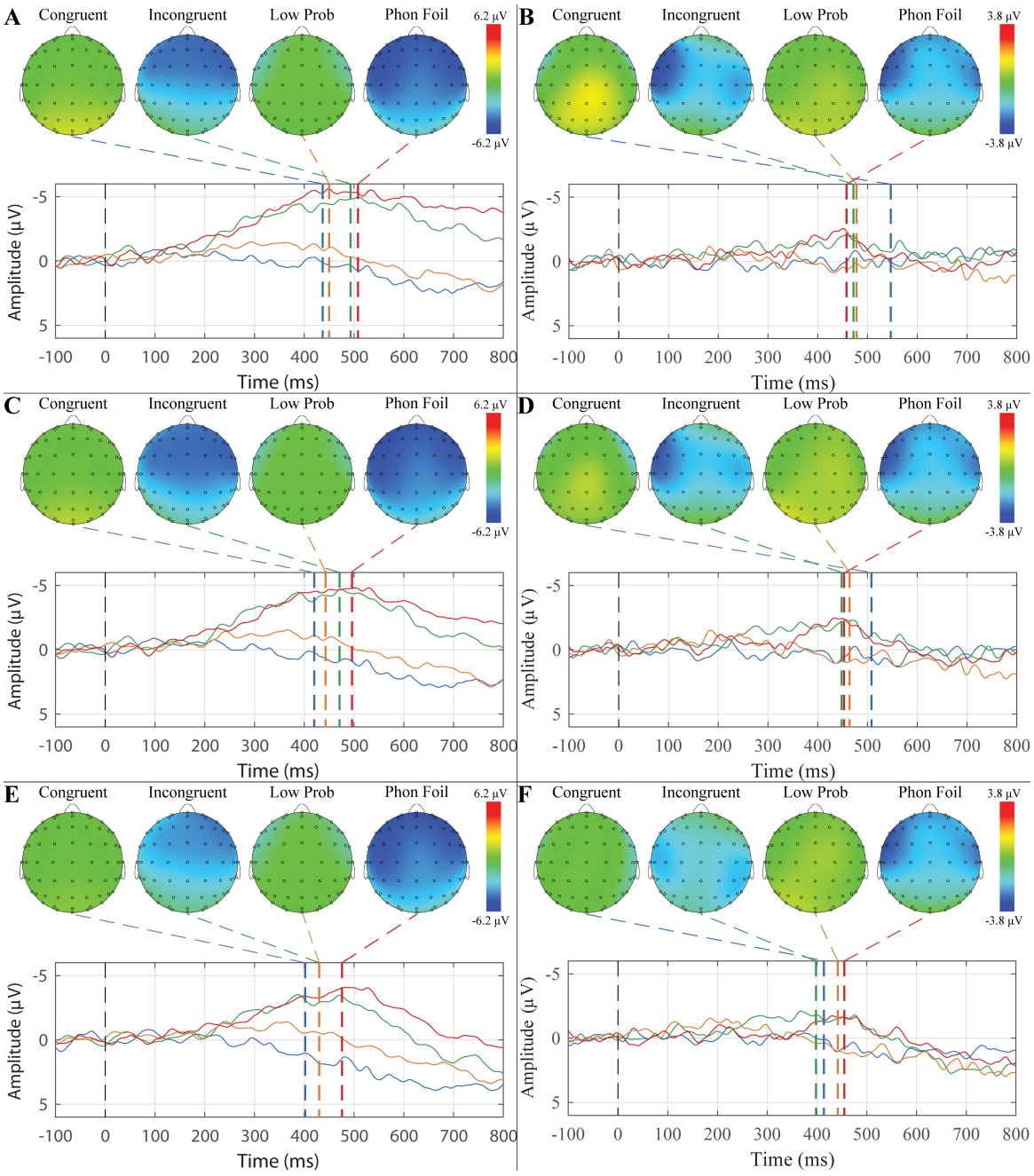

A
Congruent
Incongruent
Low Prob
Phon Foil
6.2 µV
-6.2 µV
Amplitude (µV)
Time (ms)
B
3.8 µV
-3.8 µV
C
D
E
F

FIGURE 2.S6 – Grand average waveforms at all ROIs to Congruent, Incongruent, Nonword, and Pseudoword target words and their corresponding peak topographic maps within the word-word priming paradigm. The younger adult group's average responses in the (A) Mid Frontal, (C) Mid Central, (E) Mid Parietal ROIs, and the older adult group's average responses in the (B) Mid Frontal, (D) Mid Central, (F) Mid Parietal ROIs to the list of Congruent (blue), Incongruent (orange), Nonword (green) and Pseudoword (red) target words are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

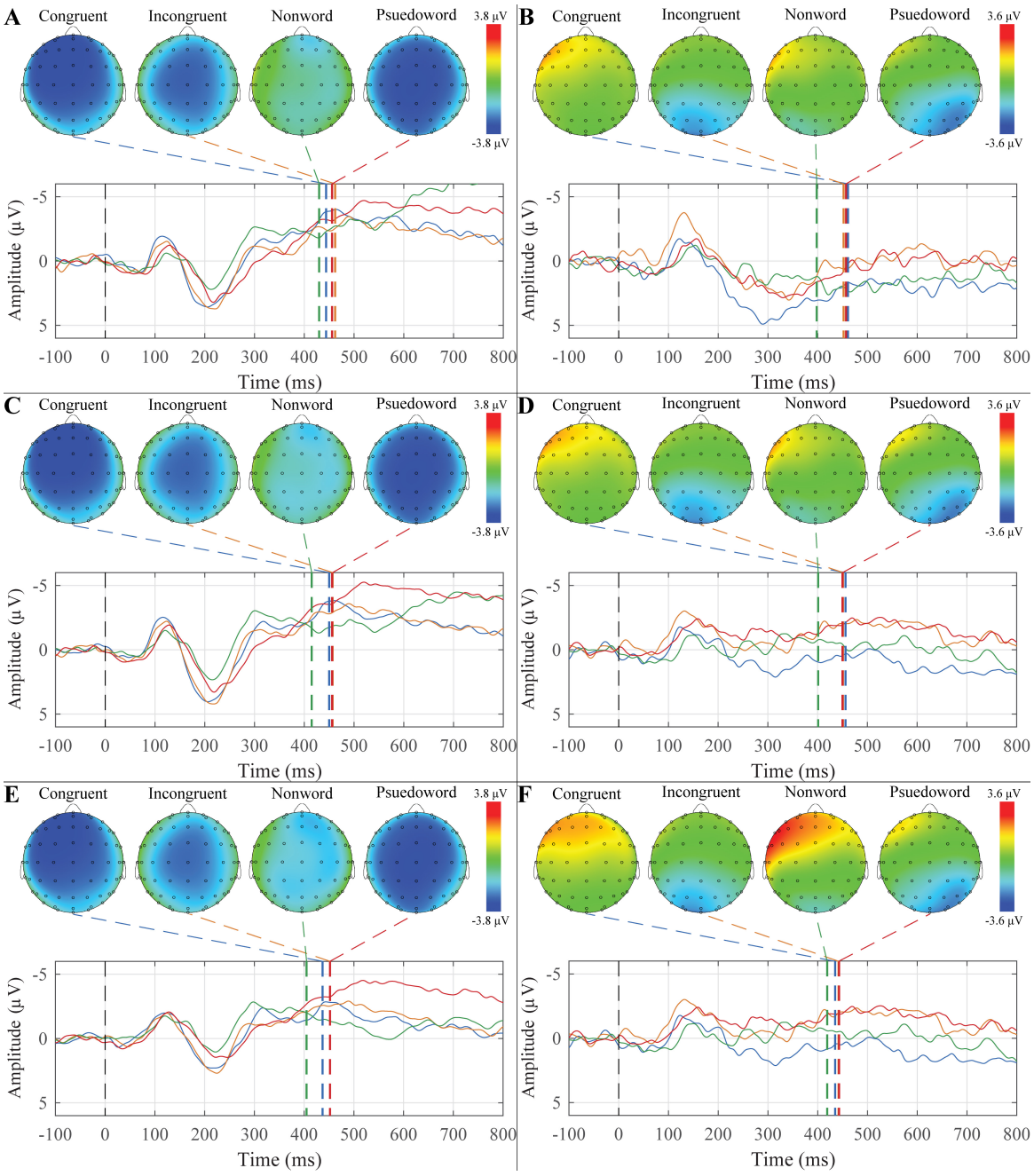

A
Congruent
Incongruent
Nonword
Psuedoword
3.8 µV
-3.8 µV
Amplitude (µV)
Time (ms)
B
Congruent
Incongruent
Nonword
Psuedoword
3.6 µV
-3.6 µV
Amplitude (µV)
Time (ms)
C
Congruent
Incongruent
Nonword
Psuedoword
3.8 µV
-3.8 µV
Amplitude (µV)
Time (ms)
D
Congruent
Incongruent
Nonword
Psuedoword
3.6 µV
-3.6 µV
Amplitude (µV)
Time (ms)
E
Congruent
Incongruent
Nonword
Psuedoword
3.8 µV
-3.8 µV
Amplitude (µV)
Time (ms)
F
Congruent
Incongruent
Nonword
Psuedoword
3.6 µV
-3.6 µV
Amplitude (µV)
Time (ms)

FIGURE 2.S7 – Grand average waveforms at all ROIs to a list of Common First Names, the Subject's Own Name, and a list of Non-salient Other Words and their corresponding peak topographic maps within the Subject's Own Name paradigm with the behavioral manipulation. The active condition average responses in the (A) Mid Frontal, (C) Mid Central, (E) Mid Parietal ROIs, and the passive condition average responses in the (B) Mid Frontal, (D) Mid Central, (F) Mid Parietal ROIs to the the Common First Names (blue), Subject's Own Name (orange), and the list of Non-salient Other Words (green) are plotted. Dashed colored lines indicate the mean group latency from individually scored P300 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

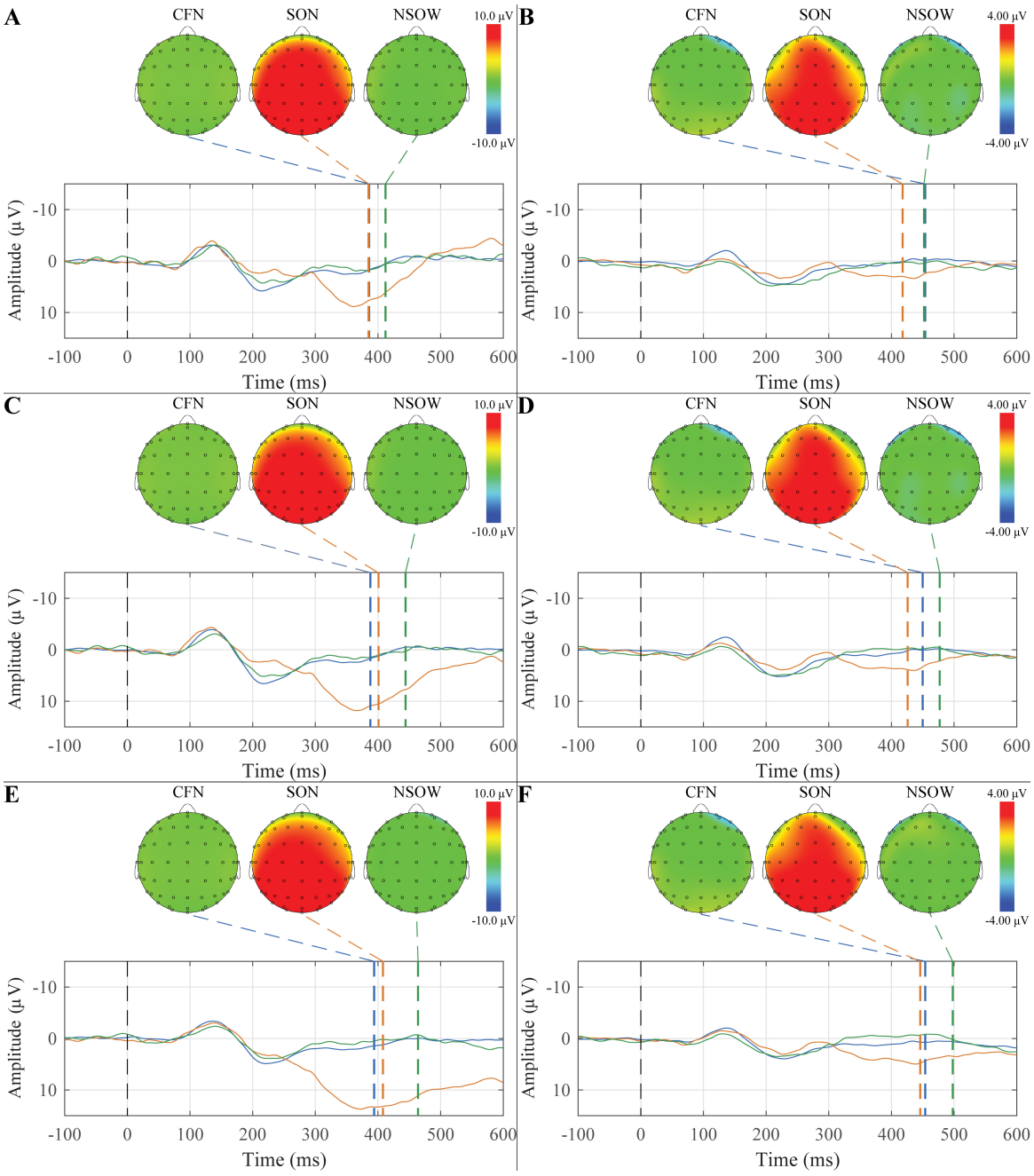

A
B
C
D
E
F
CFN
SON
NSOW
10.0 μV
-10.0 μV
4.00 μV
-4.00 μV
Amplitude (μV)
Time (ms)
-100
0
100
200
300
400
500
600
-10
0
10

FIGURE 2.S8 – Grand average waveforms at all ROIs to Congruent, Incongruent, Low Probability, and Phonological Foil terminal words and their corresponding peak topographic maps within the semantic violation sentences paradigm with the behavioral manipulation. The active condition average responses in the (A) Mid Frontal, (C) Mid Central, (E) Mid Parietal ROIs, and the passive condition average responses in the (B) Mid Frontal, (D) Mid Central, (F) Mid Parietal ROIs to Congruent (blue), Low Probability (orange), Incongruent (green) and Phonological Foil (red) terminal words are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.

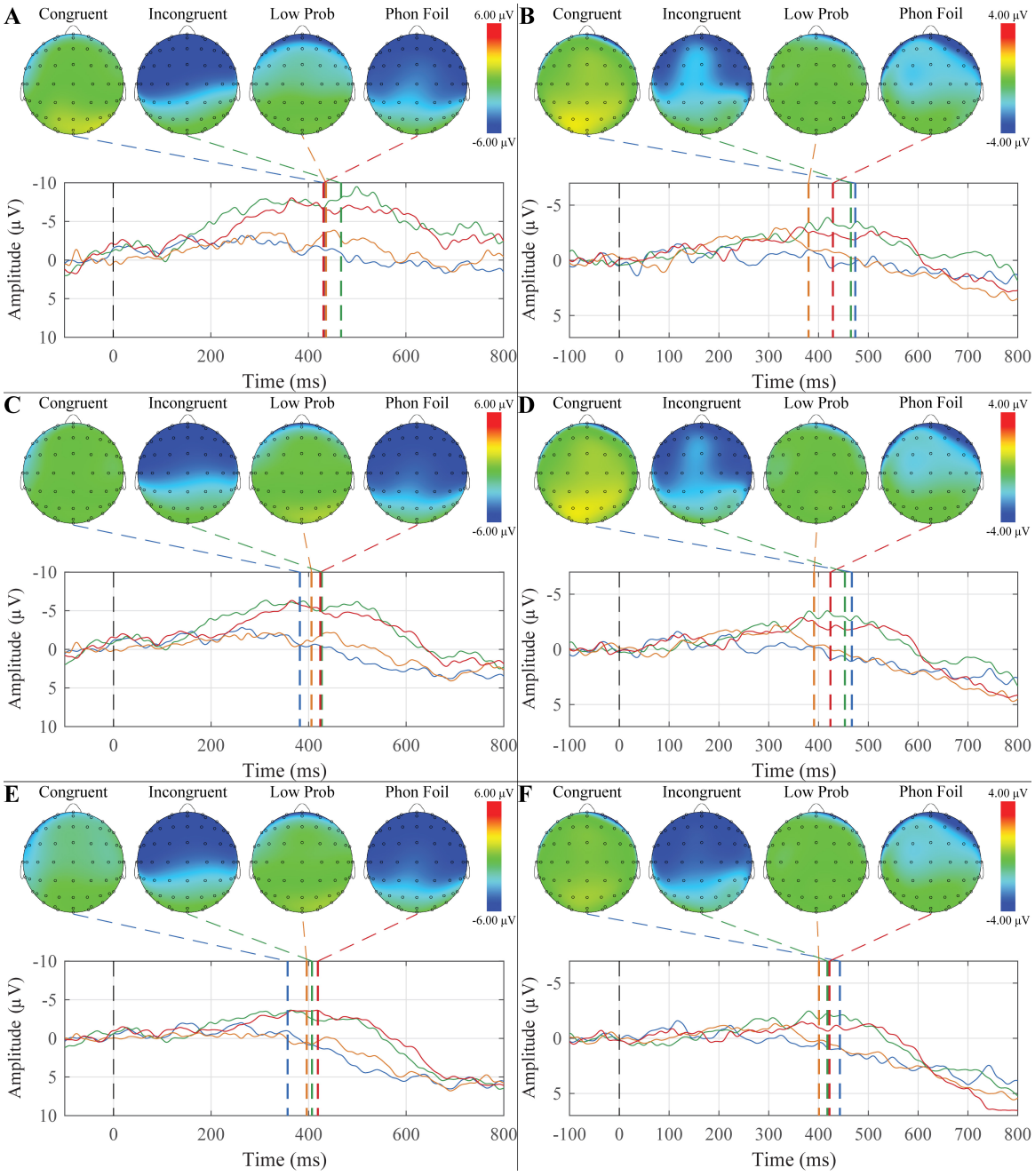
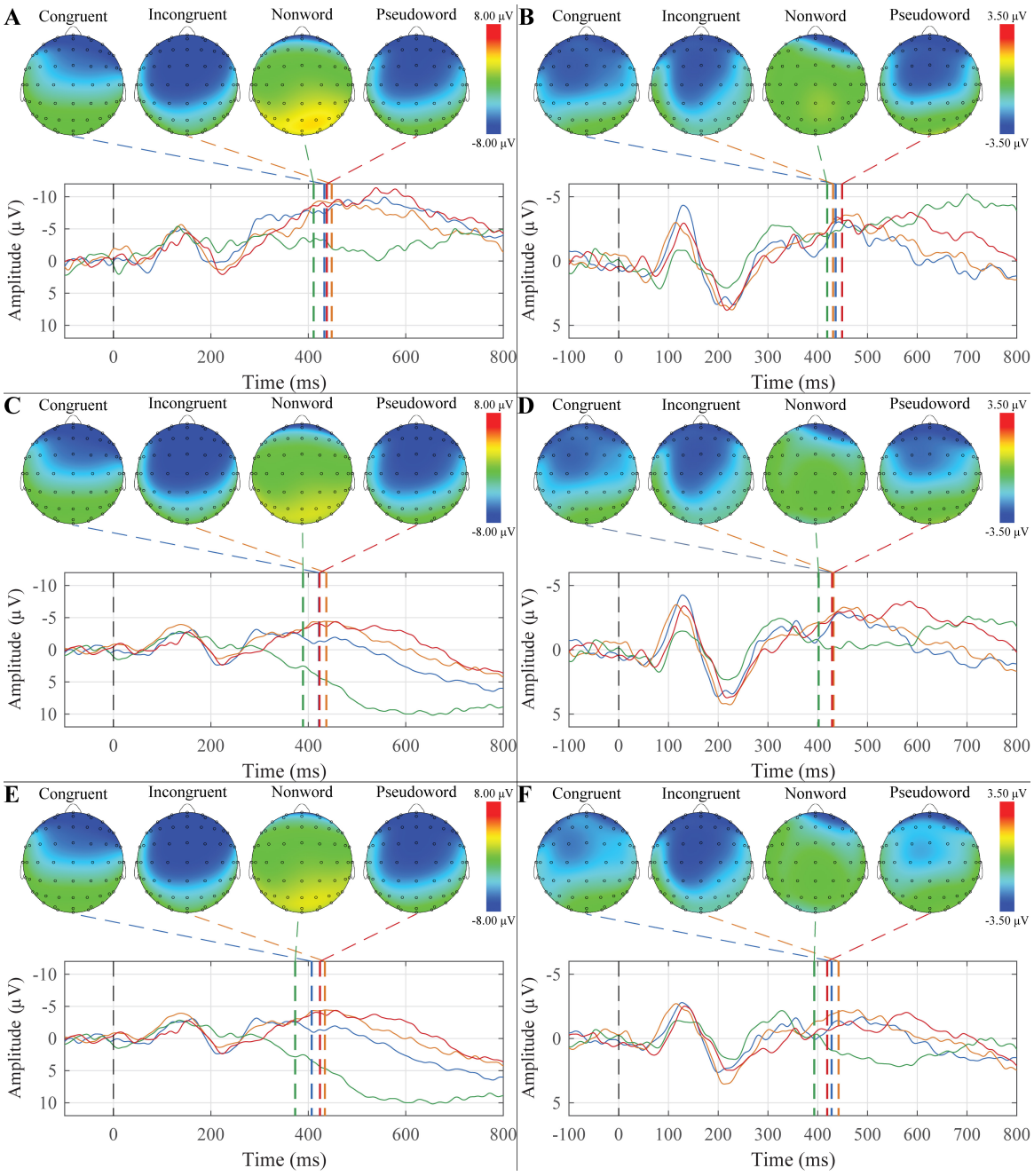

A
Congruent
Incongruent
Low Prob
Phon Foil
6.00 μV
-6.00 μV
Amplitude (μV)
Time (ms)
B
4.00 μV
-4.00 μV
C
D
E
F

FIGURE 2.S9 – Grand average waveforms at all ROIs to Congruent, Incongruent, Nonword, and Pseudoword target words and their corresponding peak topographic maps within the word-word priming paradigm with the behavioral manipulation. The active condition average responses in the (A) Mid Frontal, (C) Mid Central, (E) Mid Parietal ROIs, and the passive condition average responses in the (B) Mid Frontal, (D) Mid Central, (F) Mid Parietal ROIs to the list of Congruent (blue), Incongruent (orange), Nonword (green) and Pseudoword (red) target words are plotted. Dashed colored lines indicate the mean group latency from individually scored N400 peak latencies. Scalp topography maps show voltage distributions at mean group peak latencies.


A
Congruent
Incongruent
Nonword
Pseudoword
8.00 μV
-8.00 μV
Amplitude (μV)
Time (ms)
B
3.50 μV
-3.50 μV
C
D
E
F

References

Andrews, K., Murphy, L., Munday, R., & Littlewood, C. (1996). Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. *British Medical Journal*, *313*(7048), 13–16.

Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009, feb). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1672–7.

Berlad, I., & Pratt, H. (1995). P300 in response to the subject's own name. *Electroencephalography and Clinical Neurophysiology*, *96*(5), 472–474.

Connolly, J. F., & D'Arcy, R. C. (2000). Innovations in neuropsychological assessment using event-related brain potentials. *International Journal of Psychophysiology*, *37*(1), 31–47.

Connolly, J. F., D'Arcy, R. C., Newman, R. L., & Kemps, R. (2000). The application of cognitive event-related brain potentials (ERPs) in language-impaired individuals: review and case studies. *International Journal of Psychophysiology*, *38*(1), 55–70.

Connolly, J. F., Major, A., Allen, S., & D'Arcy, R. (1999). Performance on WISC-III and WAIS-R NI vocabulary subtests assessed with event-related brain potentials: an innovative method of assessment. *Journal of clinical and experimental neuropsychology*, *21*(4), 444–464.

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.

Cooper, R. J., Todd, J., McGill, K., & Michie, P. T. (2006). Auditory sensory memory and the aging brain: a mismatch negativity study. *Neurobiology of Aging*, *27*(5), 752–762.

Cruse, D., Beukema, S., Chennu, S., Malins, J. G., Owen, A. M., & McRae, K. (2014). The reliability of the N400 in single subjects: implications for patients with disorders of consciousness. *NeuroImage: Clinical*, *4*, 788–799.

di Giorgio, C., Rabinowicz, A., & Gott, P. (1993). Predictive value of P300 event-related potentials compared with EEG and somatosensory evoked

potentials in non-traumatic coma. *Acta Neurologica Scandinavica*, *87*(5), 423–427.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., ... Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908.

Fischer, C., Dailler, F., & Morlet, D. (2008). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical Neurophysiology*, *119*(10), 2224–30.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*, *110*(9), 1601–1610.

Folmer, R. L., & Yingling, C. D. (1997). Auditory p3 responses to name stimuli. *Brain and Language*, *56*(2), 306–311.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373.

Gabriel, D., Muzard, E., Henriques, J., Mignot, C., Pazart, L., André-Obadia, N., ... Moulin, T. (2016). Replicability and impact of statistics in the detection of neural responses of consciousness. *Brain*, *139*(6), e30.

Gawryluk, J. R., D'Arcy, R. C., Connolly, J. F., & Weaver, D. F. (2010). Improving the clinical assessment of consciousness with advances in electrophysiological and neuroimaging techniques. *BMC Neurology*, *10*(1), 1.

Gosseries, O., Di, H., Laureys, S., & Boly, M. (2014). Measuring consciousness in severely damaged brains. *Annual Review of Neuroscience*, *37*, 457–478.

Gosseries, O., Zasler, N. D., & Laureys, S. (2014). Recent advances in disorders of consciousness: focus on the diagnosis. *Brain injury*, *28*(9), 1141–1150.

Gott, P., Rabinowicz, A., & DeGiorgio, C. (1991). P300 auditory event-related potentials in nontraumatic coma: Association with glasgow coma score and awakening. *Archives of Neurology*, *48*(12), 1267-1270.

Harrison, A. H., & Connolly, J. F. (2013). inding a way in: a review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1403–1419.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming

in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312.

Holeckova, I., Fischer, C., Giard, M.-H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. *Brain Research*, *1082*(1), 142–152.

Howell, D. (2012). *Statistical methods for psychology*. Thomson Wadsworth.

Kane, N., Curry, S., Butler, S., & Cummins, B. (1993). Electrophysiological indicator of awakening from coma. *The Lancet*, *341*(8846), 688 -.

Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). The WAIS-R as a neuropsychological instrument [Computer software manual]. New York, NY.

Kutas, M., & 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.

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and language*, *111*(3), 161–172.

Müller, H. M., & Kutas, M. (1996). What's in a name? electrophysiological differences between spoken nouns, proper names and one's own name. *NeuroReport*, *8*(1), 221–225.

Morlet, D., Ruby, P., André-Obadia, N., & Fischer, C. (2017). The auditory oddball paradigm revised to improve bedside detection of consciousness in behaviorally unresponsive patients. *Psychophysiology*, *54*(11), 1644–1662.

Naccache, L., Sitt, J., King, J.-R., Rohaut, B., Faugeras, F., Chennu, S., ... Dehaene, S. (2016). Reply: Replicability and impact of statistics in the detection of neural responses of consciousness. *Brain*, *139*(6), e31.

NEMO Consortium. (2012). *NEMO Technical Report TR-2012-008: Scalp Regions of Interest (ROI)* (Tech. Rep.). Atlanta, GA: Neuro ElectroMagnetic Ontologies.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and hearing*, *16*(1), 6–18.

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

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch ne-

gativity (mmn): towards the optimal paradigm. *Clinical Neurophysiology*, *115*(1), 140–144.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Pekkonen, E., Rinne, T., Reinikainen, K., Kujala, T., Alho, K., & Naatanen, R. (1996). Aging effects on auditory processing: an event-related potential study. *Experimental aging research*, *22*(2), 171–184.

Perrin, F., García-Larrea, L., Mauguière, F., & Bastuji, H. (1999). A differential brain response to the subject's own name persists during sleep. *Clinical Neurophysiology*, *110*(12), 2153–2164.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456–456.

Polich, J., & Herbst, K. L. (2000). P300 as a clinical assay: rationale, evaluation, and findings. *International Journal of Psychophysiology*, *38*(1), 3–19.

Polich, J., & McIsaac, H. K. (1994). Comparison of auditory p300 habituation from active and passive conditions. *International Journal of Psychophysiology*, *17*(1), 25 - 34.

Public Health Agency of Canada. (2017). *Canadian chronic disease surveillance system*.

Rohaut, B., Faugeras, F., Chausson, N., King, J.-R., El Karoui, I., Cohen, L., & Naccache, L. (2015). Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia*, *66*, 279–292.

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., ... Laureys, S. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*, *9*(1), 1.

Schroeder, M., Ritter, W., & Vaughan, H. (1995). The mismatch negativity to novel stimuli reflects cognitive decline. *Annals of the New York Academy of Sciences*, *769*(1), 399–401.

Sculthorpe, L. D., & Campbell, K. B. (2011). Evidence that the mismatch negativity to pattern violations does not vary with deviant probability. *Clinical Neurophysiology*, *122*(11), 2236–2245.

Statland, B. (1987). *Clinical decision levels for lab tests*. Medical Economics Books.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and imparied conciousness. A practical scale. *Lancet*, *2*(7872), 81–84.

Tzovara, A., Rossetti, A. O., Spierer, L., Grivel, J., Murray, M. M., Oddo, M., & De Lucia, M. (2013). Progression of auditory discrimination based on neural decoding predicts awakening from coma. *Brain*, *136*(1), 81–89.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671.

Wechsler, D. (1991). The Wechsler intelligence scale for children–third edition. [Computer software manual]. San Antonio, TX.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255.

Woods, D. L. (1992). Auditory selective attention in middle-aged and elderly subjects: an event-related brain potential study. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *84*(5), 456–468.

Yingling, C. D., Hosobuchi, Y., & Harrington, M. (1990). P300 as a predictor of recovery from coma. *The Lancet*, *336*(8719), 873.

Zeelenberg, R., Pecher, D., Shiffrin, R. M., & Raaijmakers, J. G. (2003). Semantic context effects and priming in word association. *Psychonomic Bulletin & Review*, *10*(3), 653–660.

3 Advancing prognostication by the detection of fluctuating states of consciousness in coma as reflected by an electrophysiological response

This chapter has been submitted to the British Medical Journal Case Reports as Richard L. Mah, John F. Connolly, Cindy Hamielec, and Alison E. Fox-Robichaud. Advancing prognostication by the detection of fluctuating states of consciousness in coma as reflected by an electrophysiological response.

Abstract

Mismatch negativity (MMN) is an event-related brain potential related to early attentional processing of auditory information that requires a state of consciousness, although not necessarily awareness. This study demonstrates for the first time a fluctuating pattern of the MMN in a waxing/waning manner across a 24+ hour period in a coma patient (GCS subscores: E1, V1t, M2). An irregular "ultradian" rhythm was confirmed repeatedly across the 24+ hour period using several assessment procedures. This finding provides a procedure for obtaining a prognostic marker for emergence that is more sensitive than previous electrophysiological procedures, while replicating their fundamental discoveries. In the current example, the patient emerged and ultimately resumed more or less normal daily life; a finding that may be related to the nature of the waxing/waning cycle of the MMN.

3.1 Background

Patients who exhibit general unresponsiveness, the apparent absence of sleep/wake cycles, and a lack of environmental interaction are considered to be comatose. This condition is often associated with severe, diffuse bi-hemispheric lesions and/or brain stem injury, but can also result from a disruption of the reticular activating system. (American Congress of Rehabilitation Medicine, 1995)

Current diagnosis of coma requires the use various behavioural tests like the Glasgow Coma Scale (GCS) (Teasdale & Jennett, 1974) or the Full Outline of Unresponsiveness (FOUR) (Wijdicks, Bamlet, Maramattom, Manno, & McClelland, 2005), along with neuroimaging techniques like MRI or CT to evaluate the extent of the patient's anatomical injury. Prognostication in these cases requires the ongoing review of these tests, but remains a subjective evaluation of observable behavioural features and is dependent on the individual clinician performing the assessment. There is no evidence to suggest the accuracy of behaviourally-based prognostication of coma outcome is any better than similar evaluative methods with unresponsive wakefulness syndrome (UWS). (Andrews, Murphy, Munday, & Littlewood, 1996; Candelieri, Cortese, Dolce, Riganello, & Sannita, 2011; Childs, Mercer, & Childs, 1993; Schnakers et al., 2009)

The presence of certain event-related cortical potentials (ERPs), and in particular the mismatch negativity (MMN) and P300, have been shown to be highly predictive of positive outcomes from coma. In fact, presence of the MMN has been noted in some cases to be 100% predictive of awakening from coma. (Fischer, Dailler, & Morlet, 2008; Fischer et al., 1999)

While there appears to be great potential in this technique as a new objective

prognostic test for clinicians, we present evidence to suggest that care must be taken when using these methods to predict coma outcome. Specifically, patient monitoring over longer time periods appears to increase the chance of capturing a period during which the MMN can be detected. (Kane, Curry, Butler, & Cummins, 1993)

We present data from a patient who was monitored for a 24-hour period during which he received constant auditory stimulation to elicit ERPs. By using statistically-intensive methods of detecting the MMN, we not only found that the MMN appears and disappears visually, but the reliability of that detected MMN increases and decreases over time.

3.2 Case Presentation

The patient, a 29-year old male, was involved in a motorcycle-automobile accident. He sustained significant injuries and was assessed with a Glasgow Coma Scale (GCS) score of 3 as recorded at the scene of the accident. Prior to leaving the ED, he was intubated and ventilated.

Neurological examination occurred 14 days after admission and provided a poor prognosis, due primarily to diffuse axonal injury (DAI) and hypoxic ischemia. Computed tomography scans conducted at hospital admission had shown DAI, subarachnoid haemorrhage and subdural haematoma. A clinical EEG using 21 scalp electrodes was conducted 21 days after admission and noted generalized slowing and signs of a developing alpha coma. Additional support for a poor prognosis due to DAI was small amplitude EEG and failure to respond to afferent stimulation. Further details of this patient's case presentation and investigations as well as additional insight into his condition can be found in a

different experimental investigation of his state of consciousness. (Blain-Moraes et al., 2016) The patient was included in this study 27 days after admission when he was not receiving barbiturates or sedatives. The patient's diagnosis of coma had not changed and his GCS score was 4.

EEG recordings began at 19:30 and continued until 23:00 the following day. Informed consent for inclusion in the study was given by the patient's next of kin. The patient demonstrated no significant behavioural changes throughout the study. Auditory stimuli (Mah & Connolly, 2018) were presented consisting of 100 ms duration tones occurring 80% of the time (standard tone) and less frequent (deviant tone) 30 ms duration tones (14%) used to elicit the mismatch negativity (MMN). EEG was recorded with a 0.1-100 Hz bandpass and sampled at 512 Hz (offline filter was 3-30 Hz) (Fischer, Morlet, & Giard, 2000) from 32 channels (according to the 10-20 system) using using an active recording system. Other stimuli unrelated to the MMN–the patient's name (3%) and unexpected (3%) environmental sounds (e.g., dog bark) were presented also. The continuous EEG was segmented into epochs of 600 ms: 100 ms pre- and 500 ms post-stimulus. Epochs were averaged together per condition for each auditory stimulation block and several methods of analysis were used to test for the MMN in each block. First, each average was visually inspected by two electrophysiologists for the presence or absence of the MMN. Second, a one-tailed serial t-test (Marchand, D'Arcy, & Connolly, 2002) was conducted to find intervals where the MMN to the deviant condition was significantly more negative than to the standard condition. Third, a topographic consistency test (TCT) (Koenig & Melie-García, 2010) was used to locate intervals containing significantly consistent event-related scalp distributions, reflecting the event-related engagement of a constant set of brain regions. Fourth, a spatiotemporal clustering analysis was used to find groups of

neighbouring electrodes with activity that was significantly different between conditions, but consistent amongst those electrodes. (Maris & Oostenveld, 2007) Finally, the Jeffrey-Zellner-Siow Bayes factor (JZS-BF) was calculated for the difference between the standard and deviant averages for each block and was used to test the strength of the evidence for each observed effect size. (Rouder, Speckman, Sun, Morey, & Iverson, 2009) The Cauchy distribution with width of 0.707 was used as the prior distribution. (Jeffreys, 1961) The interpretation of the JZS-BF was based on the following ranges: between 1/3 and 3 - anecdotal, 3 to 10 - substantial, 10 to 100 - strong. This study was approved by the Hamilton Integrated Research Ethics Board (HIREB).

3.3 Investigations

Computed tomography: Subarachnoid haemorrhage and subdural haematoma plus evidence of diffuse axonal injury. (Blain-Moraes et al., 2016) Conventional EEG: Generalized slowing accompanied by signs of alpha coma with support signs of diffuse axonal injury.

3.4 Treatment

The patient was not medicated at the time of testing.

3.5 Outcome and Follow-up

In total, 14 blocks of data were acquired over the course of 29 hours. Table 3.1 provides a summary of the significance testing results of each block.

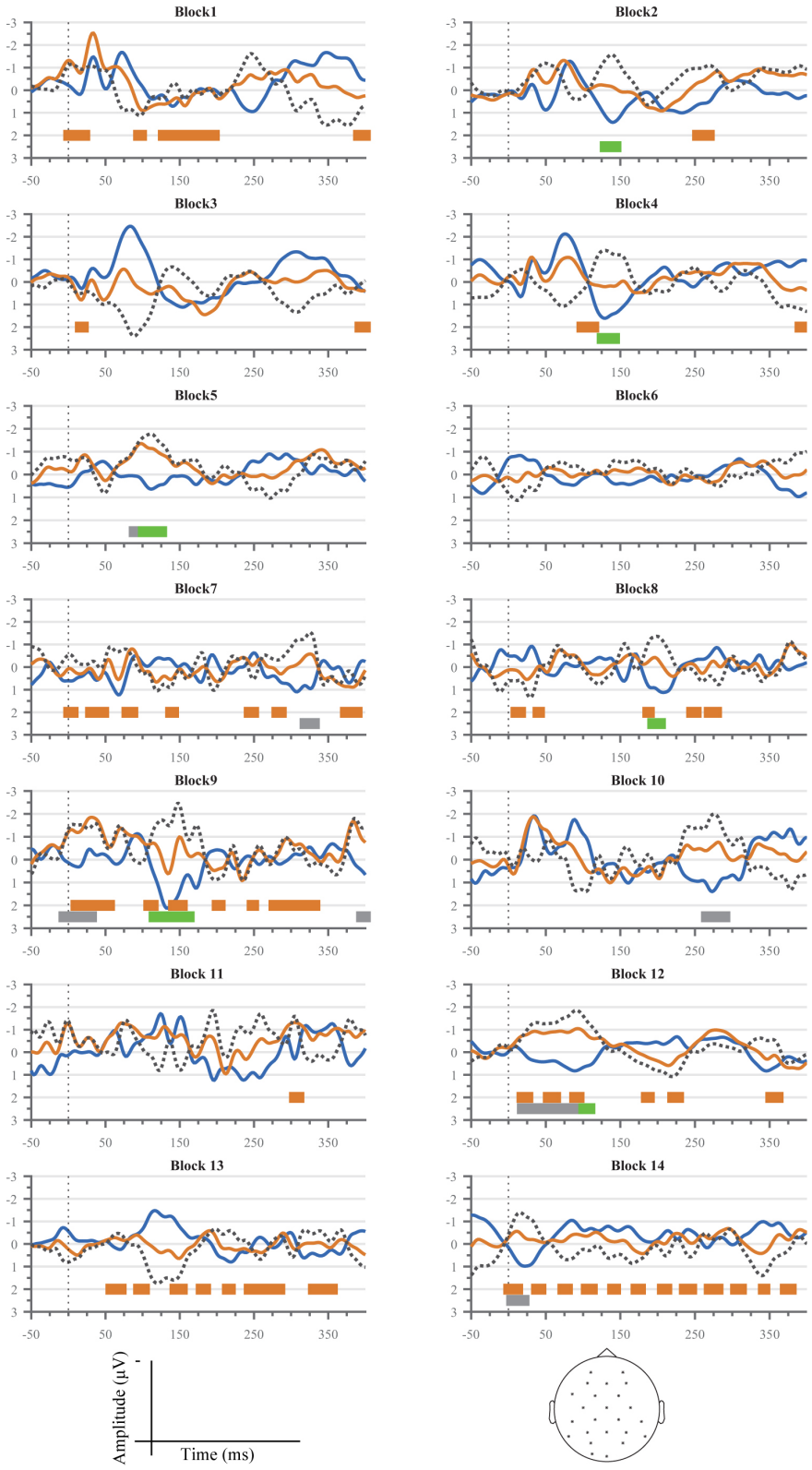
TABLE 3.1 – Summary of the results of visual inspection, serial t-test, topographic consistency test, cluster-based spatiotemporal analysis, and Bayesian t-tests for each MMN stimulation block.

Block	Start time*	Vis. Insp.	Ser. t-test	TCT	Cluster	Bayes
1	01:27:56	-	-	+	-	-
2	02:40:32	+	+	+	-	-
3	06:13:36	-	-	-	-	-
4	07:41:41	+	+	+	-	-
5	09:50:43	+	+	-	-	-
6	11:02:25	-	-	-	-	-
7	13:28:54	-	-	-	-	-
8	15:21:54	-	-	-	-	-
9	16:31:58	+	+	+	+	+
10	18:54:56	-	-	-	-	-
11	21:16:13	-	-	-	-	-
12	00:58:39	-	-	-	-	-
13	26:54:08	-	-	-	-	-
14	28:44:10	-	-	-	-	-

*Denotes start time of block since beginning of data collection.

The average waveforms for each block are presented in Fig 3.1, showing the standard, deviant, and difference waves at the Cz electrode. Significant intervals outside of the MMN time window from the serial t-test are denoted by a grey bar below the waveform, and those intervals within the time window are denoted by a green bar. Intervals with consistent topographic similarity as computed by the topographic consistency test are denoted by an orange bar. The largest significant cluster of electrodes within the MMN time window from Block 9 is presented at the bottom of Fig 3.1.

FIGURE 3.1 – MMN waveforms for each stimulation block with significant intervals from serial t-test and TCT. Waveforms for the standard (blue), deviant (orange), and their subtraction (dotted) at the electrode Cz. Significant intervals outside of the MMN time window from the serial t-test are denoted by a grey bar below the waveform, and those intervals within the time window are denoted by a green bar. Intervals with consistent topographic similarity as computed by the topographic consistency test are denoted by an orange bar. Locations of the electrodes forming the largest significant cluster within the MMN time window from Block 9 are plotted on the head map.


Block1
Block2
Block3
Block4
Block5
Block6
Block7
Block8
Block9
Block 10
Block 11
Block 12
Block 13
Block 14
-3 -2 -1 0 1 2 3
-50 50 150 250 350
Amplitude (μV)
Time (ms)

The patient began to follow commands and regained consciousness approximately 49 days post-trauma. He was subsequently transferred from the ICU to a Slow-to-Recover Rehabilitation Inpatient unit where he remained for 15 days. Following this, he was alert and oriented, able to walk independently and only required a walker for longer distances. He was discharged to his home where he was able to care for himself independently.

3.6 Discussion

Over the course of approximately 30 hours, 14 blocks of EEG data were recorded during which the MMN protocol was presented. Of these blocks, visual inspection determined that four of them (blocks 2, 4, 5, and 9) contained the MMN. Using the one-tailed serial t-test, these same four blocks also showed significant intervals within the MMN time. The topographic consistency test showed significantly coordinated activity within the MMN time window for four blocks (blocks 1, 2, 4, and 9). The spatiotemporal cluster analysis and the Bayesian t-tests were significant in block 9 only.

This study provides evidence that the MMN–a stable and reliably observed component in healthy controls–can vary in its detectability over time in coma. While it appears that visual inspection by a trained electrophysiologist remains the gold standard method for detecting the MMN, (Fischer et al., 2008, 1999; Gabriel et al., 2016; Naccache et al., 2016), we show here that other statistical measures may be helpful in confirming the presence of the MMN.

The positive results in all tests from block 9 show the value of using multiple detection methods, especially if an electrophysiologist is not available to assess the waveforms. By using more objective measures, the confidence in the detection

and identification process is increased while also making the test easier to use for healthcare professionals.

In consideration of the "cycling" presence/absence of the MMN over a 29-hour period, it is apparent that an "absence" of the MMN during a single test occasion should not be interpreted as a definitive absence of the MMN in a patient. Numerous studies have shown that presence of the MMN in coma is correlated highly with emergence. Over 90% of coma patients exhibiting a MMN ultimately emerge from coma—a high positive predictive value–while no MMN response was observed in over 90% of patients judged to be non-awake–a reflection of the high specificity of the MMN. However, the low sensitivity of the MMN is reflected in the finding that only about 30% of patients emerging from coma exhibited a MMN. (Fischer et al., 1999; Morlet & Fischer, 2014) The present results suggest strongly that the low sensitivity of the MMN may be attributable to the traditional one-occasion testing procedure failing to detect the MMN due to this newly discovered "ultradian" fluctuation or rhythm of the MMN in coma. The clinical testing implications of this case study is that MMN testing should be repeated several times over the course of several hours to increase the likelihood of detecting the response and increasing the sensitivity of the MMN in this context. The clinical assessment implications of this case are that a yet-to-be determined number of coma patients transition into a conscious state as reflected by the MMN (Dykstra, Cariani, & Gutschalk, 2017; Tavakoli, Varma, & Campbell, 2017) and that while this state is not necessarily accompanied by conscious awareness, it does indicate a state not seen during anaesthesia-induced unconsciousness. (Blain-Moraes et al., 2016)

3.7 Learning Points

- The mismatch negativity cycles in its presence/absence over a 24-hour period in this comatose patient.
- The presence of the mismatch negativity has been shown to be highly predictive of positive outcomes from coma, however testing for it at a single point in time increases the risk of a false negative.
- Using complementary statistical analysis methods alongside visual inspection of waveforms can help confirm the presence of the mismatch negativity in comatose patients.
- Inclusion of this novel prognostic procedure may be beneficial to the case management of patients who are comatose.

References

American Congress of Rehabilitation Medicine. (1995). Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Archives of Physical Medicine and Rehabilitation*, *76*(2), 205 - 209.

Andrews, K., Murphy, L., Munday, R., & Littlewood, C. (1996, jul). Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. *BMJ*, *313*(7048), 13 LP – 16.

Blain-Moraes, S., Boshra, R., Ma, H. K., Mah, R., Ruiter, K., Avidan, M., ... Mashour, G. A. (2016). Normal Brain Response to Propofol in Advance of Recovery from Unresponsive Wakefulness Syndrome. *Frontiers in Human Neuroscience*, *10*(June), 1–6.

Candelieri, A., Cortese, M. D., Dolce, G., Riganello, F., & Sannita, W. G. (2011, jul). Visual Pursuit: Within-Day Variability in the Severe Disorder of Consciousness. *Journal of Neurotrauma*, *28*(10), 2013–2017.

Childs, N. L., Mercer, W. N., & Childs, H. W. (1993, Aug). Accuracy of diagnosis of persistent vegetative state. *Neurology*, *43*(8), 1465 LP – 1465.

Dykstra, A. R., Cariani, P. A., & Gutschalk, A. (2017). A roadmap for the study of conscious audition and its neural basis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714). doi: 10.1098/rstb.2016.0103

Fischer, C., Dailler, F., & Morlet, D. (2008). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical Neurophysiology*, *119*, 2224–30.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*, *110*(9), 1601–1610.

Fischer, C., Morlet, D., & Giard, M. (2000). Mismatch negativity and N100 in comatose patients. *Audiology & neuro-otology*, *5*(3-4), 192–7. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/10859413` doi: 13880

Gabriel, D., Muzard, E., Henriques, J., Mignot, C., Pazart, L., André-Obadia, N., ... Moulin, T. (2016). Replicability and impact of statistics in the detection of neural responses of consciousness. *Brain*, *139*(6), e30.

Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). Oxford: Oxford

University Press.

Kane, N. M., Curry, S. H., Butler, S. R., & Cummins, B. H. (1993, mar). Electrophysiological indicator of awakening from coma. *Lancet*, *341*(8846), 688.

Koenig, T., & Melie-García, L. (2010, Sep 01). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, *23*(3), 233–242.

Mah, R. L., & Connolly, J. F. (2018, jul). A framework for the extended monitoring of levels of cognitive function in unresponsive patients. *PLOS ONE*, *13*(7), e0200793. doi: 10.1371/journal.pone.0200793

Marchand, Y., D'Arcy, R. C., & Connolly, J. F. (2002). Linking neurophysiological and neuropsychological measures for aphasia assessment. *Clinical Neurophysiology*, *113*(11), 1715–1722.

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

Morlet, D., & Fischer, C. (2014). MMN and novelty P3 in coma and other altered states of consciousness: A review. *Brain Topography*, *27*(4), 467–479.

Naccache, L., Sitt, J., King, J.-R., Rohaut, B., Faugeras, F., Chennu, S., ... Dehaene, S. (2016). Reply: Replicability and impact of statistics in the detection of neural responses of consciousness. *Brain*, *139*(6), e31.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225–237. doi: 10.3758/PBR.16.2.225

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., ... Laureys, S. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: Clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*, *9*, 1–5.

Tavakoli, P., Varma, S., & Campbell, K. (2017). Highly relevant stimuli may passively elicit processes associated with consciousness during the sleep onset period. *Consciousness and Cognition*, *58*(May 2017), 60–74.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: A practical scale. *The Lancet*, *304*(7872), 81 - 84.

Wijdicks, E. F. M., Bamlet, W. R., Maramattom, B. V., Manno, E. M., & McClel-

land, R. L. (2005). Validation of a new coma scale: The four score. *Annals of Neurology*, *58*(4), 585–593.

4 Electrophysiological markers of variations in perceptual and cognitive processing in coma

This chapter is in preparation for submission to the Journal of Neuroscience as Richard L. Mah and John F. Connolly. Electrophysiological markers of variations in perceptual and cognitive processing in coma.

Abstract

Mismatch negativity (MMN) is an event-related brain potential related to early attentional processing of auditory information, whose generation has been shown to require a state of consciousness, although not necessarily awareness. This study expands on the case presented in Mah and Connolly (2017) by presenting three cases of patients in coma who were monitored for at least 16 hours. Four methods of confirming the presence of the MMN were tested on these three patients, and all patients showed fluctuations in the detectability of the MMN. In the end, both the spatiotemporal clustering and Bayesian t-test analysis methods were the best at confirming the presence of a visually detected MMN. Additional patient testing is necessary to determine the cause of the fluctuations in the detectability of the MMN, however it does give cause for repeat testing of patients who may not show a MMN due to the time of day during which they were tested.

4.1 Introduction

Coma is a condition which is characterized by patients who exhibit general unarousability, apparently absent sleep/wake cycles, and the inability to interact with their environment. This condition is often associated with severe, diffuse bihemispheric lesions and/or brain stem injury, but can also result from a disruption of the reticular activating system. (American Congress of Rehabilitation Medicine, 1995)

The current standard of diagnosis of coma involves the administration of various behavioural tests such as the Glasgow Coma Scale (GCS) (Teasdale & Jennett, 1974) or the Full Outline of UnResponsiveness (FOUR) (Wijdicks, Bamlet, Maramattom, Manno, & McClelland, 2005). In addition to these tests, clinicians will often use neuroimaging techniques such as MRI or CT to fully evaluate the extent of the patient's injuries. Prognostication in these cases requires the ongoing review of these tests, but is highly subjective and dependent on the individual clinician performing the assessment.

In more recent years, there has been a strong push to move away from these highly subjective tests and towards more objective measures. The use of papillary reflexes, early cortical somatosensory evoked potentials (SSEPs), as well as continuous EEG measurements to find isoelectric or burst suppression patterns have been of varying usefulness (Robinson, Micklesen, Tirschwell, & Lew, 2003; Zandbergen, de Haan, Stoutenbeek, Koelman, & Hijdra, 1998). Also of interest have been certain event-related cortical potentials (ERPs), which occur much later than the SSEPs (100 to 200 ms). Specifically of interest are the mismatch negativity (MMN) and P300, both of which have been shown to be highly predictive of outcomes from coma (Fischer, Dailler, & Morlet, 2008; Fischer et al.,

1999). It has been noted that in some cases, presence of the MMN was 100% predictive of awakening from coma.

While it is encouraging that these methods have been producing such strong positive predictions and the potential that these methods have to become new clinical tests, it is also necessary to acknowledge some pitfalls to this methodology. Specifically, we present evidence to suggest that care must be taken when using these methods to predict outcomes from comas, and that longer term monitoring of a patient appears to increase the chance of capturing a period where the MMN can be detected.

In the work of Fischer et al. (1999) and those that followed, the MMN is always found via visual detection by a skilled electrophysiologist. This method does not lend itself easily to use in a clinical setting as a new diagnostic test. To make it more appealing to clinicians, this visual inspection step would need to be removed or at least the reliance on it reduced.

This paper aims to expand on the single patient case presented in Mah and Connolly (2017) with data from three additional comatose patients who were monitored for at least sixteen hours, during which they received constant auditory stimulation to elicit ERPs. We also examined the performance and ability of multiple different statistical methods to detect the MMN and further confirm its presence in the waveforms.

4.2 Materials and Methods

The sample consisted of three patients (two males and one female), aged between 21 and 69 years. All were in the Intensive Care Unit at Hamilton General Hospital and were comatose with a Glasgow Coma Scale score of less than 8 at

the time of recording. All patients had undergone at least one brain CT or MRI. The recordings began between 10 and 22 days post-trauma. A summary of their individual demographics is presented in Table 4.1.

TABLE 4.1 – Patient demographic information

Patient	Sex	Age	Time post injury	Blocks recorded
1	F	69	22 days	6
2	M	56	10 days	9
3	M	21	13 days	11

At the time of the study, no patient's medications included any sedatives or barbiturates. In all cases, the patients demonstrated no significant behavioural changes throughout the study.

A substitute decision maker provided informed consent for the study, which was approved by the Hamilton Integrated Research Ethics Board.

4.2.1 Experimental Design

An auditory stimulation series was presented to each patient (Mah & Connolly, 2018) and only the Oddball Mismatch blocks were included in the analysis. These blocks were designed to elicit three ERP waveforms (N100, Mismatch Negativity (MMN), P300) and consisted of an auditory oddball series of four sounds: standard tones (80%), deviant tones (14%), familiar novel (FN; 3%), and unfamiliar novel sounds (UFN; 3%).

The tones were digitally generated sine waves of 800 Hz, with a standard tone duration of 75 ms and a deviant tone duration of 30 ms. The familiar novel was a digital recording of the subject's name spoken by a native speaker of Canadian English in a neutral voice. The unfamiliar novel was a digital recording of a dog barking. Stimuli were presented pseudorandomly (no deviant or novel stimulus

was preceded by less than two standard tones) in one block of 2000 items with a stimulus onset asynchrony (SOA) for the tones being 610 ms and 1220 ms for the novels.

4.2.2 EEG recording and pre-processing

The electroencephalogram (EEG) was recorded continuously at the patient's bedside in the intensive care unit (bandpass = 0.01–100 Hz and sampled at 512 Hz) using either an **8 or 32** channel Biosemi ActiveTwo system (Biosemi, Amsterdam, The Netherlands) with a 10-20 elastic cap holding Ag/AgCl electrodes. The electrooculogram (EOG) was recorded from electrodes placed above and at the outer canthus of the left eye. References were recorded bilaterally from the mastoids and at the nasion for offline referencing.

Data preprocessing was conducted using BrainVision Analyzer 2. All recordings were visually inspected and epochs containing artifacts (e.g., muscle, movement) removed. Recordings were filtered offline with a bandpass of 0.1–30 Hz. Ocular artifacts were corrected using the *Ocular ICA* transformation provided by BrainVision Analyzer 2. The continuous EEG was segmented into epochs of 600 ms total: 100 ms pre-stimulus and 500 post-stimulus.

A final artifact rejection was performed automatically, removing segments with voltage steps greater than 50µV, voltage differences greater than 200µV in 200 ms, and channels with low activity ($<$0.5µV).

These epochs were averaged together per condition for each block, and for each block, several methods of analysis were used to test for the MMN.

4.2.3 Statistical Analysis

In total, five methods of detecting the mismatch negativity were used.

First, each average was visually inspected by two electrophysiologists for the presence or absence of the MMN.

Second, a one-sided serial t-test (Marchand, D'Arcy, & Connolly, 2002) was conducted to find intervals where the deviant condition was significantly more negative than the standard condition. This method computes the point-by-point t-scores for overlapping windows of 20 ms length. Blocks with statistical significance within the MMN time window of interest (120–240 ms) were considered to contain the MMN.

Third, a topographic consistency test (Koenig & Melie-García, 2010) was used to locate intervals containing significantly consistent event-related scalp distributions, reflecting the event-related engagement of a constant set of brain regions. To do this, global field power (GFP) of the grand average ERP is calculated at each time point along the waveform. In order to determine whether the GFP is being generated by a consistent set of neural sources, the electrode labels are randomly shuffled, averaged, and the resulting GFP recorded. This is repeated 1000 times to form a null distribution, which is then compared to the true GFP. This results in a series of p-values for each time point. Blocks with significant scalp distributions within the MMN time window of interest (120–240 ms) were considered to contain the MMN.

Fourth, a spatiotemporal clustering analysis was used to find groups of neighbouring electrodes that were significantly different between conditions, but consistent between electrodes. This analysis was completed using the spatiotemporal permutation-based cluster analysis implemented in the Fieldtrip toolbox

(Oostenveld, Fries, Maris, & Schoffelen, 2011). For each block, one-tailed dependent samples t-tests (deviant $<$ standard) were used to compare the standard and deviant conditions of the MMN. Spatiotemporally adjacent t-values with p-values $<$0.05 were then clustered based on their spatiotemporal proximity. T-values within each cluster were summed and the largest cluster retained. To correct for multiple comparisons, 1000 permutations were computed and then compared to the known data (Maris & Oostenveld, 2007). Blocks with significant negative clusters within the MMN time window of interest (120–240 ms) were considered to contain the MMN.

Finally, to complement the t-tests, the Jeffrey-Zellner-Siow Bayes factor (JZS-BF) was calculated for the difference between the peak mean amplitudes at electrode Cz of the standard and deviant waveforms for each block and was used to test the strength of the evidence. (Rouder, Speckman, Sun, Morey, & Iverson, 2009) This computation was done using the *BayesFactor* package (Morey & Rouder, 2015) for R (R Core Team, 2016). The Cauchy distribution with width of 0.707 was used as the prior distribution. The interpretation of the JZS-BF was: between 1/3 and 3 - anecdotal, 3 to 10 - substantial, 10 to 100 - strong. (Jeffreys, 1961)

4.3 Results

Each patient was monitored for a period of at least 16 hours, over which at least 6 blocks meant to elicit the MMN were recorded. A summary of the results for each patient are given in Tables 4.2-4.4.

The average waveforms for each patient and each block are presented Figures 4.1–4.3 and show the standard, deviant, and subtraction at the electrode

TABLE 4.2 – Summary of results for Patient 1.
+ denotes a positive result, - a negative result. For Bayes column, + denotes anecdotal evidence, ++ substantial evidence, and +++ strong evidence.

Block	Time since beginning of study (HH:MM:SS)	Vis. Insp.	Ser. t-test	TCT	Cluster	Bayes
1	00:10:42	-	+	-	-	-
2	02:15:34	-	+	-	-	-
3	10:25:17	+	+	+	+	+
4	12:11:31	-	+	-	-	-
5	13:44:39	+	+	-	-	-
6	16:05:20	-	+	+	-	-

TABLE 4.3 – Summary of results for Patient 2.
+ denotes a positive result, - a negative result. For Bayes column, + denotes anecdotal evidence, ++ substantial evidence, and +++ strong evidence.

Block	Time since beginning of study (HH:MM:SS)	Vis. Insp.	Ser. t-test	TCT	Cluster	Bayes
1	01:12:54	+	+	+	-	+
2	02:17:15	+	+	+	-	++
3	03:39:35	+	+	+	+	+++
4	06:07:32	+	+	+	-	-
5	07:44:12	+	+	-	-	+
6	10:18:03	-	+	+	-	-
7	12:07:51	-	+	-	-	-
8	14:18:19	-	-	-	-	-
9	16:08:07	-	+	+	-	++

Cz. Significant intervals outside of the MMN time window from the serial t-test are denoted by a grey bar below the waveform, and those intervals within the time window are denoted by a green bar. Intervals with consistent topographic similarity as computed by the topographic consistency test are denoted by an orange bar.

FIGURE 4.1 – MMN waveforms for each stimulation block with significant intervals from serial t-test and TCT for Patient 1. Waveforms for the standard (blue), deviant (orange), and their subtraction (dotted) at the electrode Cz. Significant intervals outside of the MMN time window from the serial t-test are denoted by a grey bar below the waveform, and those intervals within the time window are denoted by a green bar. Intervals with consistent topographic similarity as computed by the topographic consistency test are denoted by an orange bar.

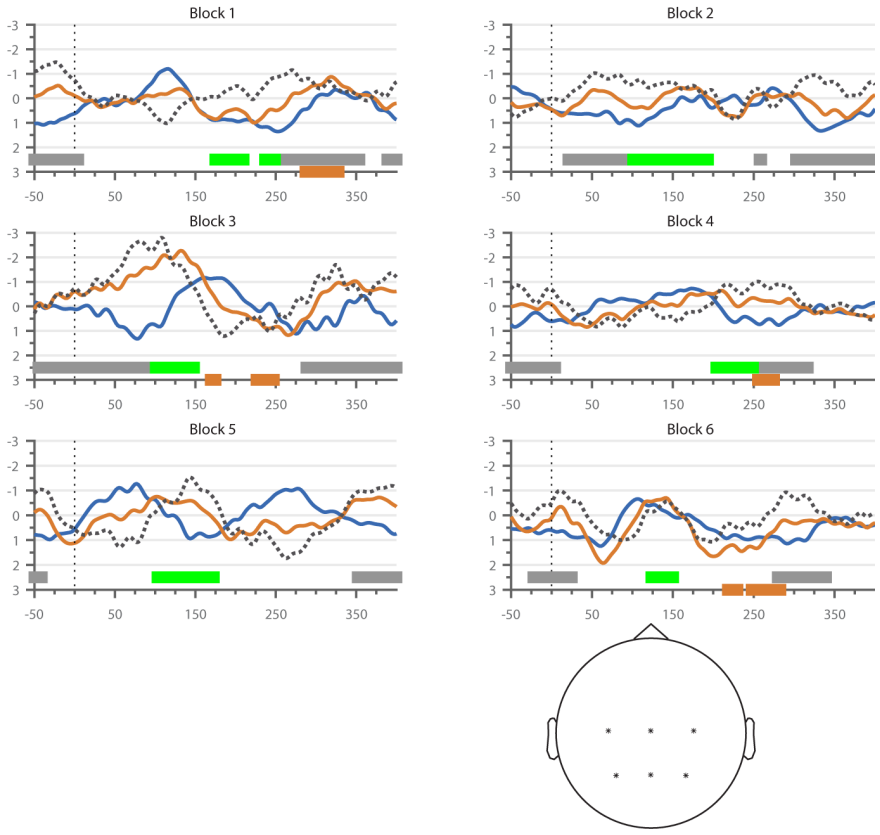


FIGURE 4.2 – MMN waveforms for each stimulation block with significant intervals from serial t-test and TCT for Patient 2. Waveforms for the standard (blue), deviant (orange), and their subtraction (dotted) at the electrode Cz. Significant intervals outside of the MMN time window from the serial t-test are denoted by a grey bar below the waveform, and those intervals within the time window are denoted by a green bar. Intervals with consistent topographic similarity as computed by the topographic consistency test are denoted by an orange bar.

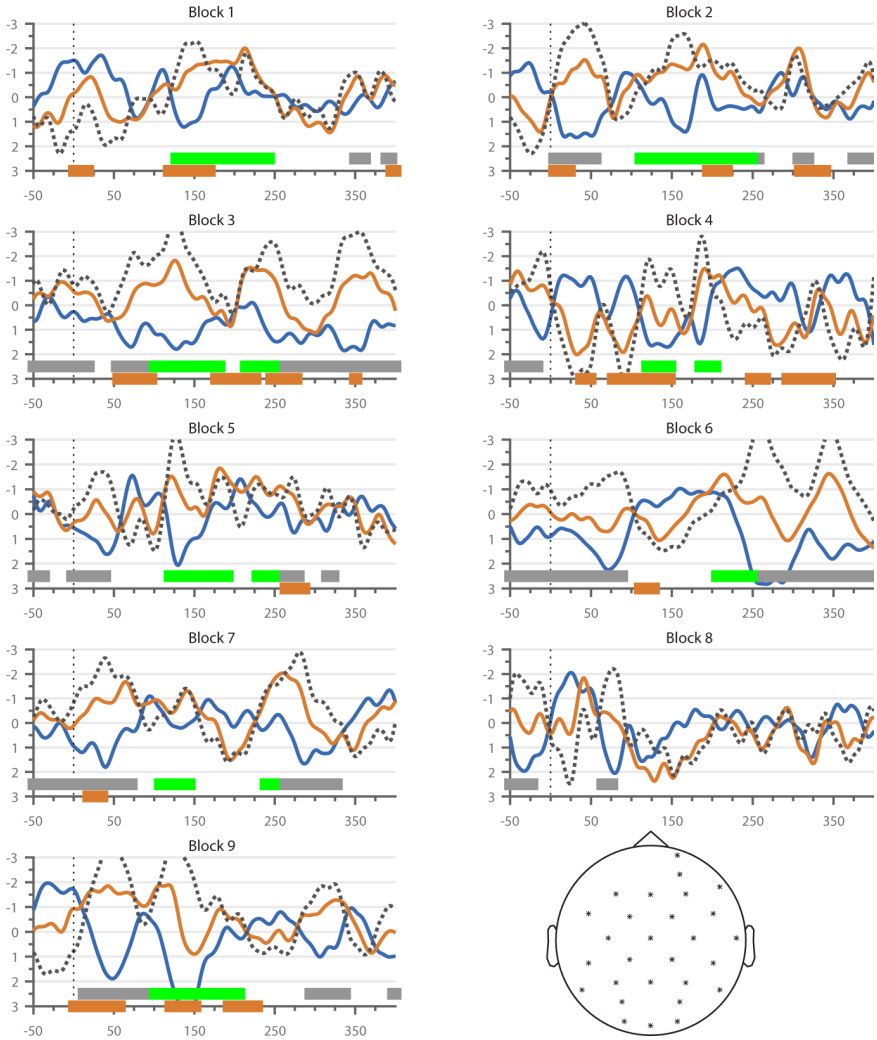


FIGURE 4.3 – MMN waveforms for each stimulation block with significant intervals from serial t-test and TCT for Patient 3. Waveforms for the standard (blue), deviant (orange), and their subtraction (dotted) at the electrode Cz. Significant intervals outside of the MMN time window from the serial t-test are denoted by a grey bar below the waveform, and those intervals within the time window are denoted by a green bar. Intervals with consistent topographic similarity as computed by the topographic consistency test are denoted by an orange bar.

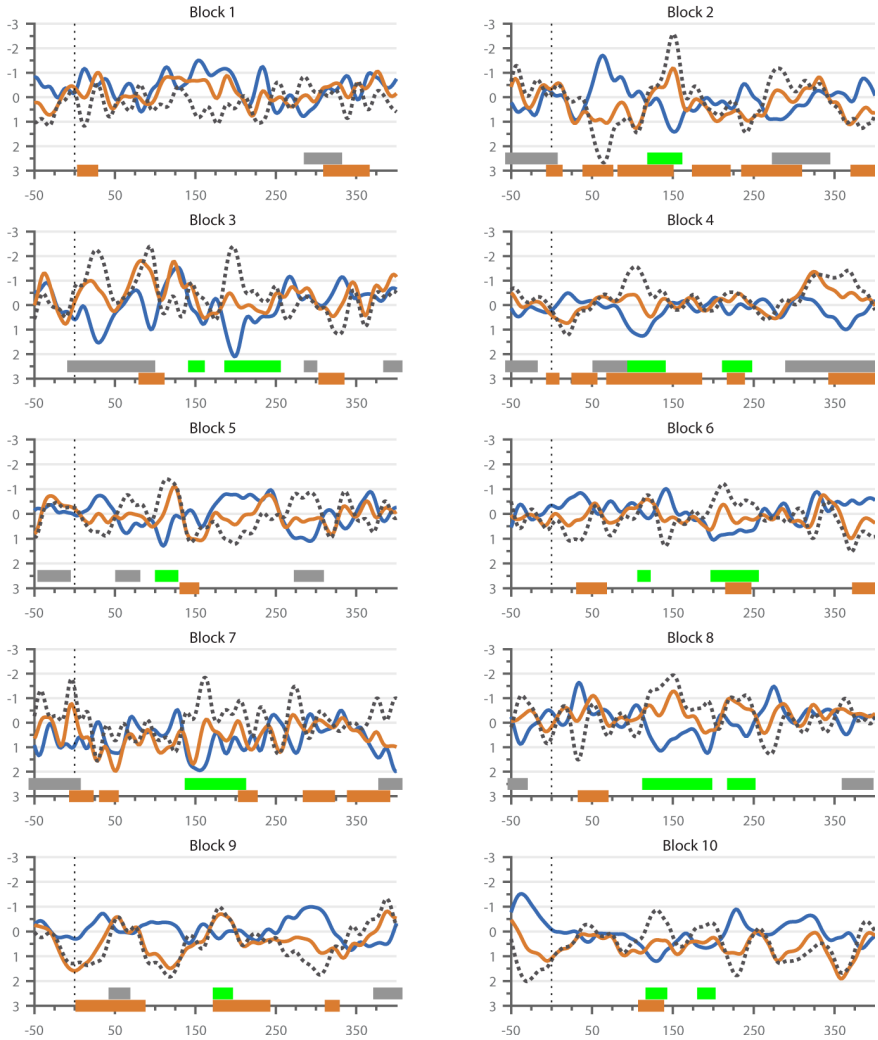


TABLE 4.4 – Summary of results for Patient 3.
+ denotes a positive result, - a negative result. For Bayes column, + denotes anecdotal evidence, ++ substantial evidence, and +++ strong evidence.

Block	Time since beginning of study (HH:MM:SS)	Vis. Insp.	Ser. t-test	TCT	Cluster	Bayes
1	00:42:30	-	-	-	-	-
2	04:58:39	+	+	+	-	-
3	05:56:59	+	+	+	-	-
4	08:09:15	+	+	+	-	-
5	10:00:47	-	+	+	-	-
6	11:00:50	-	+	-	-	-
7	13:21:21	-	+	-	-	-
8	16:18:51	+	+	+	-	+
9	18:50:00	-	+	+	-	-
10	20:36:31	-	+	+	-	-

4.3.1 Patient outcomes

Patient 1 passed away in the ICU 23 days post-injury, after the withdrawal of ventilation at the request of the family.

Patient 2 passed away in the step-down unit 16 days post-injury, after the withdrawal of ventilation six days prior at the request on the family.

Patient 3 was discharged to the Acquired Brain Injury unit 30 days post-injury to complete their slow-to-recover program, and was subsequently discharged to his home three months post-injury.

4.4 Discussion

There were three major aims to this study: first, to expand on the single case presented in (Mah & Connolly, 2017); second, to better determine how statistical methods of increasing intensity might aid in the detection of the MMN; and

third, to further examine the inconsistent nature of the MMN.

Three patients were monitored over the course of at least 16 hours and at least 6 blocks of MMN-eliciting data were collected. For patient 1, two blocks were considered to have the MMN after visual inspection, the serial t-test showed that all blocks had significant intervals within the MMN window of interest, two blocks had spatially-correlated activity according to the TCT, and only one cluster had a significant cluster and Bayesian t-tests. Patient 2 had four MMN blocks after visual inspection, all but one significant after serial t-test, all but three after TCT, two blocks with significant clusters, and five blocks showing evidence with the Bayesian t-tests. Patient 3 had four MMN blocks after visual inspection, all but one significant after serial t-test, all but three after TCT, and only one block showing evidence with the Bayesian t-tests.

In all three cases, by using visual inspection, the MMN is very clearly evident in at least one block, but also indiscernible from noise in at least one block. At least in Fischer et al. (1999), visual inspection was their standard for detection of the MMN. Had these patients been tested in a similar fashion to those in Fischer et al. (1999) (*i.e.* only at one point in time), there would have been at least a 60% chance of missing the MMN and giving them a poorer prognosis.

It would appear that the criterion used in the serial t-test is a bit too lax to make any meaningful contribution to a confirmation of the MMN. Across the 15 collected blocks between the three patients, only two came back with no significant intervals within the MMN window. Even though the t-test is being conducted as a one-tailed test (standard greater than deviant), it is still showing significance to intervals that would appear to be artefactual noise. Perhaps narrowing the relevant time window and increasing the number of consecutive significant points would help reduce some of these false positive results.

The topographic consistency test was also somewhat of a poor performer. While there were times when it showed significantly spatially correlated activity along side other positive tests, there were times when it was one of the only significant tests (*i.e.* Patient 2, Block 6; Patient 3, Blocks 5, 9, and 10). This test also seems to suffer poor performance when using lower density electrode arrays. There appears to be a strong co-occurrence of visual confirmation of the MMN and a positive TCT result. In the case of Patient 1, only 8 electrodes were used, which may have contributed to a potential false negative (as seen in Block 5) and a potential false positive (as seen in Block 6). It may also be the case that the TCT does not perform well with very noisy clinical data. While all efforts were taken to reduce the amount of noise in the recordings, the ICU is a very electrically noisy environment, and there was still quite a bit of higher frequency noise.

The spatiotemporal cluster-based analysis appeared to be a good confirmatory test. In all instances where there was a positive cluster result, there would always be at least a positive visual inspection, serial t-test, and TCT result. It also would co-occur with some amount of evidence from the Bayesian t-tests. As is seen in Figures 4.1–4.3, the topographic distribution of these clusters is typical for the MMN–strongly clustered around the Cz electrode, at about 150 ms.

Finally, the Bayesian t-tests were also a good confirmatory test. In cases where the MMN was very evident in visual inspection, the test gave a result of "strong evidence". In cases where the MMN was less evident, but was still judged to be present, the test would give "anecdotal" to "substantial" evidence. The only case where the test gave a positive result when the visual inspection did not was in Patient 2, Block 9. In this case, it is likely that the waveforms contained some other non-related rhythmic artifact, especially in the standard condition.

To summarize, it appears that the serial t-test and the topographic consistency tests were not very good at providing reliable confirmations of the MMN, at least in this patient sample. The spatiotemporal clustering method and the Bayesian t-tests both appeared to give good confirmations of the MMN after visual inspection. It seems as though these data-driven methods are very susceptible to any sort of artifact left in the waveform, but these are often very difficult to avoid in a clinical intensive care setting.

Perhaps alternative data-driven methods that should be seriously considered are those which use machine learning, as in Tzovara et al. (2013) or Armanfard, Komeili, Reilly, Mah, and Connolly (2016). These methods both seek to improve the accuracy of the detection of the MMN while reducing the quantity of trials needed for detection.

Turning now to the apparent waxing and waning of the MMN, much like was seen in (Mah & Connolly, 2017), there appears to be a variation in the presence of the MMN in the recorded signal. The MMN is usually thought of as a stable component and able to be elicited in various states of consciousness (*i.e.* normal awareness, sleep (Sallinen, Kaartinen, & Lyytinen, 1994), minimally conscious or vegetative states (Kotchoubey et al., 2005), etc.), but this variability in detection suggests that that may not actually be the case. Perhaps the MMN is not generated during some periods for these comatose patients due to their injuries. It may also be the case that there is just too much noise from other neural sources (like increased slow wave activity, or other dysrhythmias or generalized suppression (Nuwer, Hovda, Schrader, & Vespa, 2005)) for the MMN to be detected. In either case, more patient data is necessary before choosing one alternative over the other.

References

American Congress of Rehabilitation Medicine. (1995). Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Archives of Physical Medicine and Rehabilitation*, *76*(2), 205–209.

Armanfard, N., Komeili, M., Reilly, J. P., Mah, R., & Connolly, J. F. (2016). Automatic and continuous assessment of ERPs for mismatch negativity detection. In *2016 38th annual international conference of the ieee engineering in medicine and biology society (embs)* (pp. 969–972). doi: 10.1109/EMBC.2016.7590863

Fischer, C., Dailler, F., & Morlet, D. (2008, oct). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, *119*(10), 2224–30.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical neurophysiology*, *110*(9), 1601–1610.

Jeffreys, H. (1961). *The theory of probability*. OUP Oxford.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, *23*(3), 233–242.

Kotchoubey, B., Lang, S., Mezger, G., Schmalohr, D., Schneck, M., Semmler, A., ... Birbaumer, N. (2005). Information processing in severe disorders of consciousness: Vegetative state and minimally conscious state. *Clinical Neurophysiology*, *116*(10), 2441 - 2453.

Mah, R. L., & Connolly, J. F. (2017). *Electrophysiological markers of variations in perceptual and cognitive processing in a comatose patient in intensive care.* (In preparation for submission to the British Medical Journal Case Reports)

Mah, R. L., & Connolly, J. F. (2018, jul). A framework for the extended monitoring of levels of cognitive function in unresponsive patients. *PLOS ONE*, *13*(7), e0200793. doi: 10.1371/journal.pone.0200793

Marchand, Y., D'Arcy, R. C., & Connolly, J. F. (2002, nov). Linking neurophysio-

logical and neuropsychological measures for aphasia assessment. *Clinical Neurophysiology*, *113*(11), 1715–1722.

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

Morey, R. D., & Rouder, J. N. (2015). Bayesfactor: Computation of bayes factors for common designs [Computer software manual]. (R package version 0.9.12-2)

Nuwer, M. R., Hovda, D. A., Schrader, L. M., & Vespa, P. M. (2005). Routine and quantitative eeg in mild traumatic brain injury. *Clinical Neurophysiology*, *116*(9), 2001 - 2025.

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*, 1.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Robinson, L. R., Micklesen, P. J., Tirschwell, D. L., & Lew, H. L. (2003). Predictive value of somatosensory evoked potentials for awakening from coma. *Critical care medicine*, *31*(3), 960–967.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225-237.

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1994). Is the appearance of mismatch negativity during stage 2 sleep related to the elicitation of k-complex? *Electroencephalography and Clinical Neurophysiology*, *91*(2), 140 - 148.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and imparied conciousness. A practical scale. *Lancet*, *2*(7872), 81–84.

Tzovara, A., Rossetti, A. O., Spierer, L., Grivel, J., Murray, M. M., Oddo, M., & De Lucia, M. (2013). Progression of auditory discrimination based on neural decoding predicts awakening from coma. *Brain*, *136*(1), 81–89.

Wijdicks, E. F., Bamlet, W. R., Maramattom, B. V., Manno, E. M., & McClelland, R. L. (2005). Validation of a new coma scale: The FOUR score. *Annals of Neurology*, *58*(4), 585–593.

Zandbergen, E. G., de Haan, R. J., Stoutenbeek, C. P., Koelman, J. H., & Hijdra, A. (1998). Systematic review of early prediction of poor outcome in anoxicischaemic coma. *The Lancet*, *352*(9143), 1808 - 1812.

5 Characterizing EEG ultradian rhythmicity differences in coma using spectral entropy

This chapter is in preparation for submission to the Journal of as Richard L. Mah and John F. Connolly. Characterizing EEG ultradian rhythmicity differences in coma using spectral entropy.

Abstract

Prior work has shown that patients with some disorders of consciousness such as those in a minimally conscious state (MCS) or vegetative state/unresponsive wakefulness syndrome (VS/UWS) can be differentiated with the use of spectral entropy, and specifically, the periodicity of thereof. To determine whether this methodology would be appropriate to use with patients who are comatose, we analyzed EEG recordings from three patients in coma using the wavelet analysis technique from Piarulli et al. (2016). Total and relative band powers (delta, theta, alpha, upper and lower beta bands) and spectral entropy were estimated (Fz, Cz, and Pz electrodes), as well as the spectral entropy. In terms of band relative power features, Patient 1 had a few features that were within the range of the MCS group, Patient 2 had several features that were within the range of the MCS group, and Patient 3 only had two features in one channel that were within the range of the VS/UWS group. Only one patient had spectral entropy fluctuations that were in the 70 min range. The others either did not have fluctuations, or had fluctuations that appeared to be contaminated with artifact. Overall, even though this methodology appears to be extremely sensitive to external noise, it does appear to hold potential for use as a way to target when cognitive assessments should be delivered to maximize the likelihood that the patient will be more aware.

5.1 Introduction

Determining the likely progression of a patient with severe injuries is central to the job of critical care doctors. In the case of severe traumatic brain injuries and patients who arrive either unconscious or in an altered state of consciousness, clinicians must determine whether a patient is likely to regain normal consciousness, if they will remain comatose, or progress into a vegetative state/unresponsive wakefulness syndrome (VS/UWS) (Laureys et al., 2010) or minimally conscious state (MCS) (Giacino et al., 2002). It is also important to better predict the trajectory of a patient's recovery to ensure the best use of medical resources. Depending on the complexity of care required, and the type of hospital the patient is in, the costs of a day of ICU care in Canada in 2013 were estimated to be between $3200 and $4200, which is approximately three times the cost of a normal ward bed. (Canadian Institute for Health Information, 2016)

While the current clinical standard of care to diagnose and categorize patients with disorders of consciousness involves the use of behavioural tests such as the Glasgow Coma Scale (GCS) (Teasdale & Jennett, 1974) and Full Outline of Unresponsiveness (FOUR) (Wijdicks, Bamlet, Maramattom, Manno, & McClelland, 2005) for comatose or emerging patients, and the Coma Recovery Scale-Revised (CRS-R) (Giacino, Kalmar, & Whyte, 2004) for VS/UWS/MCS patients. Scientists and clinicians have been working towards methods that reduce the reliance on these behavioural tests, and to move towards more objective measures of conscious states. This is especially important for those patients who may have fluctuating levels of consciousness leading to inconsistent ability to follow commands, particularily when trying to distinguish a patient between MCS minus

or plus. (Bruno, Vanhaudenhuyse, Thibaut, Moonen, & Laureys, 2011) Moreover, the misdiagnosis rates for patients who are VS/UWS has been found to be somewhere between 41% (Schnakers et al., 2009) to 43% (Andrews, Murphy, Munday, & Littlewood, 1996) when relying on traditional behaviorally-based consensus methods.

For comatose patients, work using the mismatch negativity (MMN) and P300 to predict favourable outcomes has shown high specificity and sensitivity. (Fischer, Dailler, & Morlet, 2008; Fischer et al., 1999) We have, however, noted that the MMN appears to go through periods where the it is more or less detectable. That is to say, depending on when the test is performed, the MMN may be missed simply because the patient is not the most ideal state to elicit a MMN.

The GCS has also been noted to have severe shortcomings, especially in patients who have been intubated to protect their airway. Because they are unable to properly vocalize while intubated, their verbal subscore may be artificially lowered. It has been further criticized by others for being used in situations where it was never intended for, and being highly variable depending on who is scoring it. (Green, 2011; Laureys, Bodart, & Gosseries, 2014)

Together, it appears that not only should these behavioural tests not be the sole basis of a diagnosis or prognosis, but it becomes more and more difficult to properly categorize a patient when the level of outward behaviour is next to zero. Piarulli, et al. (Piarulli et al., 2016) suggested that using spectral entropy would be a good correlate for the level of consciousness in VS/UWS/MCS patients, especially in those who had fluctuating levels of consciousness. They recommended targeting the administration of clinical assessments to periods when the entropy was the highest, and that the spectral entropy appeared to cycle at about 70 minute intervals in MCS patients.

Looking at the GCS, the only defining characteristic between a VS/UWS patient and a comatose one is spontaneous eye opening in the VS/UWS patient. This left the question of whether this method of measuring spectral entropy would categorize a comatose patient more closely as a VS/UWS or MCS patient, or perhaps as a third, distinct category.

While the application of EEG-based entropy measures on classification of patients in a comatose state has previously been investigated (Gosseries et al., 2011), this work has been focused on single point measurements of state entropy. In the present study, we apply the analysis methodology from Piarulli, et al. (Piarulli et al., 2016) to long-duration EEG recordings in patients who are in a comatose state. We compare the results from three comatose patients to each other and to the two clinical groups, MCS and VS/UWS, previously reported in three ways: the relative power in five frequency bands, the mean spectral entropy, and the fluctuations of spectral entropy over time.

5.2 Methods

5.2.1 Patients

Three comatose patients were included in the study. At the time of the study, all patients had a GCS score less than 6. The patients had a mean age of 35 $\pm$ 17.8 (standard deviation), and all were male and had traumatic etiologies. They were studied without sedative medication and all approximately 14 days post-injury. No patient had continuous epileptiform activity, suppression or burst-suppression patterns present in their EEG. All patients were assessed in the intensive care unit. GCS assessments were performed both by clinical

staff during the course of the patients' treatment, and were confirmed at the commencement of the EEG recording. The GCS has three subscales, which include verbal and motor responses from the patient as well as responses from the patient's eyes. Scoring is based on the presence or absence of responses to various stimuli, and range from a minimum score of 3 (no movement at all in response to painful stimulation) to a maximum score of 15 (full orientation to surroundings, ability to converse and follow commands). The verbal subscore is reduced if the patient has been intubated, however this should be interpreted in the context of the individual assessment. Clinical and demographic details of the patients are reported in Table 5.2.4. The study was approved by the Hamilton Integrated Research Ethics Board, and written informed consent was obtained from the patients' legal representatives.

5.2.2 EEG recordings

The electroencephalogram (EEG) was recorded continuously at the patient's bedside in the intensive care unit (bandpass = 0.01–100 Hz and sampled at 512 Hz) using a 32-channel Biosemi ActiveTwo system (Biosemi, Amsterdam, The Netherlands) with a 10-20 elastic cap holding Ag/AgCl electrodes. Consistent with previous work (Piarulli et al., 2016), to ensure the recorded signals would be of a high and stable quality, only the first four hours of the recordings were retained and analyzed.

5.2.3 Signal pre-processing

All analyses were performed in Matlab (Mathworks, Natic, MA, USA) with the preprocessing steps using functions from the EEGLAB toolbox (Delorme &

Patient	Age	Gender	Time since injury (days)	Etiology	GCS sub-scores	Time to awakening (days)	State upon awakening
1	55	M	11	Traumatic	E1 V1 M2	N/A*	N/A
2	21	M	13	Traumatic	E2 V1 M3	20	Following commands, not verbalizing
3	29	M	19	Traumatic	E1 V1t M2	49	Following commands, oriented

E eye, *V* verbal (*t* indicates patient was intubated), *M* motor

Patient was given palliative care and life-sustaining measures were discontinued

TABLE 5.1 – Demographics of patients included in study

Makeig, 2004). Only a subset of channels (F3, Fz, F4, C3, Cz, C4, T3, P3, Pz, P4, Oz) were retained for the analysis.

The signals were referenced to the nasion, and then bandpass filtered from 1 to 45 Hz using a zero-phase FIR filter. The filtered signal was then segmented into four second consecutive epochs. Epochs with a peak-to-peak voltage range greater than 100 µV were excluded from the analysis.

5.2.4 Signal analysis

The signals were analyzed consistent to previous work (Piarulli et al., 2016), and the specific steps used in the present study are briefly outlined below.[1]

5.2.4.1 Feature extraction

For each retained epoch, the total and relative power for five frequency bands were computed: delta (1–3.75 Hz, theta (4–7.75 Hz), alpha (8–11.75 Hz), low beta (12–17.75 Hz), and high beta (18–24.75 Hz). For each channel in each epoch, the power spectral density was estimated by computing the Fourier Transform of the signal convolved with a Hamming window, then squaring the magnitude of its output. Relative power was computed as the ratio between the total band power and the total power between 1–25 Hz.

The definition for spectral entropy from (Piarulli et al., 2016) is given in Eq. (5.1),

$$\text{se} = -\frac{\sum_{k=1}^{K}(P_{f_k}\log_2 P_{f_k})}{\log_2 K}, 1\text{Hz} \leq f_k \leq 25\text{Hz}, \tag{5.1}$$

where P_{f_k} is the normalized power spectral density at frequency f_k, and was also computed for each channel in each epoch.

[1]Thanks to A. Piarulli for his helpful elaborations to the method provided in the original publication.

The four second epochs were then grouped into five minute consecutive intervals. Within these intervals, the mean total and relative band powers for each channel were estimated by averaging the epochs together. The spectral entropy mean was also estimated, along with the standard deviation, and the coefficient of variation.

The mean values of each feature from each patient were then compared to the confidence intervals of the corresponding feature of both the MCS and VS/UWS groups from (Piarulli et al., 2016).

5.2.4.2 Wavelet analysis

In a similar fashion to the the feature extraction, for each patient, eight time series were extracted: spectral entropy for Fz, Cz, and Pz, and the log-transformed delta, theta, alpha, high and low beta total powers for Fz.

Each time series was wavelet transformed using Matlab functions from (Torrence & Compo, 1998)[2]. Each time series was zero padded to a length of the next power of two (in this case, 48 data points were padded to 64 data points) to minimize edge effects and to allow resolution of periods up to 120 minutes. A Morlet wavelet was used, with the number of wavelet cycles (ω_0) set to 4. The spacing between scales (δj) was set to 0.1, with the smallest wavelet scale being $2\delta j$. The full set of scales was computed in the transformation, although only 26 of them (from 20 to 114 minutes) were retained for further analysis.

After each time series had been transformed, the mean amplitude spectrum distribution was estimated by averaging the amplitude spectrum along the 240 min interval. For each patient, the period contributing the most power to

[2]Wavelet analysis functions are available for download at http://paos.colorado.edu/research/wavelets/

the spectral distribution of the Fz spectral entropy was determined, and that period was transformed back to the time domain by applying the inverse wavelet transformation.

The peak periodicities of the Fz power in each of the five frequency bands of interest were compared to that of the Fz spectral entropy. For each patient, the time series of the log-transformed Fz band powers were correlated to the spectral entropy time series producing a correlation coefficient. To assess the coefficient's significance, a permutation test was done where the data from each time course was randomized, then the correlation was done again. This procedure was repeated 1000 times to build a distribution of correlation coefficients, against which the true coefficient was compared. The proportion of coefficients disregarding sign that were greater than the true value to the total generated gave the estimated p-value.

5.3 Results

5.3.1 Band relative power

The mean relative band power and confidence intervals for each band were computed for each patient (Tables 5.2, 5.3, and 5.4) and then compared to the means and confidence intervals from (Piarulli et al., 2016).

For patient 1, low beta band in the Fz channel was within the MCS range. High beta was within the VS range, but (Piarulli et al., 2016) did not find a significant difference between the MCS and VS groups for this band. Patient 2 had mean band powers within the MCS range for the delta, theta, and the low and high beta bands. Patient 3 only had theta and alpha band powers within the VS

range. All other bands were outside of the reported ranges for both clinical groups.

Comparing the means and confidence intervals of the three patients for the relative band powers did not yield any significantly similar means. Even if the confidence intervals were expanded slightly to 90%, no patient mean is within the confidence interval of another.

5.3.2 Spectral entropy

A spectral entropy time course for the Fz, Cz, and Pz channels was computed using the mean values over five minute intervals. For both patients 1 and 2, their mean value of spectral entropy over the whole four hour period of testing for the Fz and Pz channels were within the range of the MCS clinical group. Additionally, patient 2 was within the MCS group's range for the Cz channel. None of the patients had mean standard deviation or coefficient of variation values within the MCS or VS group ranges. All patients had data that was more variable than both clinical groups; that is to say their mean standard deviation and coefficients of variation were higher than the MCS and VS groups. Patient 3 had a mean spectral entropy value below the VS group, which in turn was lower than the MCS group.

Comparing the means and confidence intervals of the three patients for the measures of spectral entropy, only the mean spectral entropy of patient 1 was within the confidence interval of another–patient 2. This was seen in Fz, Cz, and Pz, but only held in the direction of patient 1 to patient 2. The spectral entropy standard deviation for patient 3 was within the ranges of both patient 1 (in the Pz channel) and patient 2 (in the Cz channel). Patient 2 was also within

the range of patient 3's standard deviation in the Cz channel.

5.3.3 Wavelet decomposition of spectral entropy

The time courses of the spectral entropy and log-transformed band powers for each patient were submitted to a wavelet analysis to identify any dominant oscillatory components.

As seen in Figures 5.1 and 5.2, Patient 1 does not appear to have any dominant oscillatory components. The period corresponding to the best fit is 92 minutes, although the mean spectral entropy amplitude at this point is quite small at less than 0.02.

Patient 2 is shows a different pattern, as there appear to be two dominant periods, although their activity is somewhat limited to the middle of the four hour period. The first period is at 35 minutes, and has a MSE amplitude of about 0.05. The second period does not have a well defined peak, as the amplitude increases starting at 53 minutes and does not stop increasing at the end of the range (114 minutes).

Patient 3 appears to be more consistent with the MCS group from (Piarulli et al., 2016), as the dominant period is at 86 minutes, which is close the MCS group range of 53–80 minutes that was reported. Even the mean amplitude spectrum for this patient looks more consistent with the MCS group spectrum as reported.

The period corresponding to the peak for each patient was identified and those values were used to estimate the time course for the inverse wavelet transformation and are shown in Figure 5.3.

FIGURE 5.1 – Spectral entropy amplitude spectrum (sa) time course for the Fz channel for each patient. The x-axis identifies the time during the four hour recording, the y-axis identifies the period of the oscillation of the wavelet. Colors from white to dark red show increasing contributions to the spectral amplitude variations.

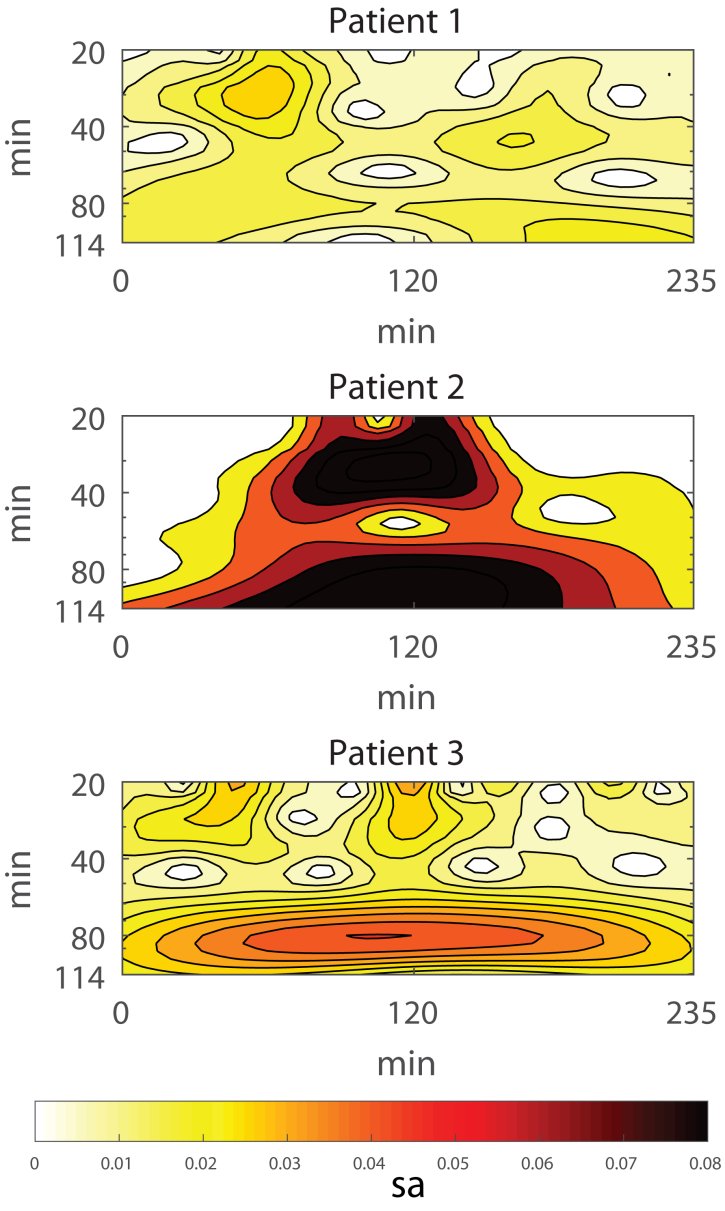


FIGURE 5.2 – Mean contributions of oscillations with periods from 20 minutes to 120 minutes to the Fz spectral entropy time variations (MSA) for each patient. The mean is enclosed within a 95% confidence interval.

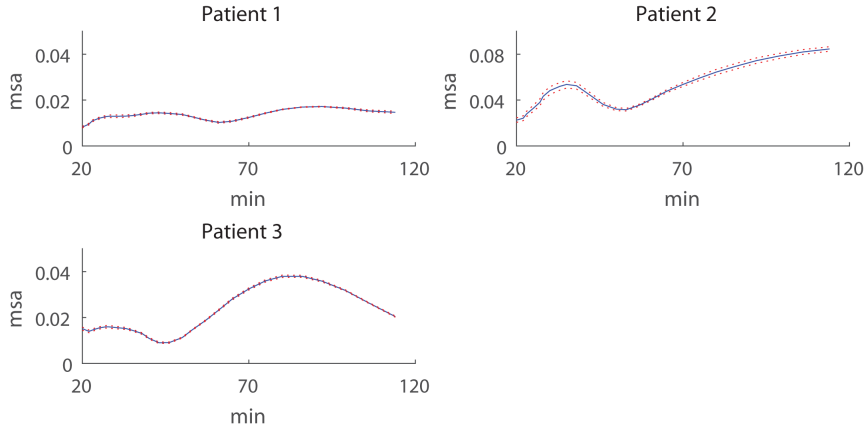
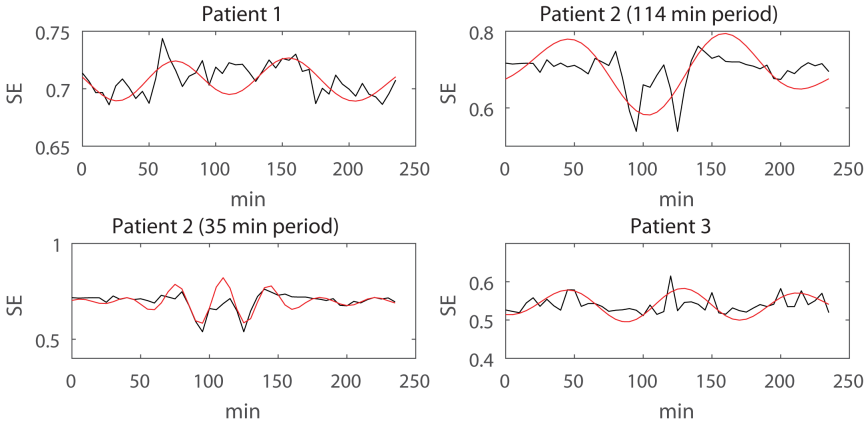


FIGURE 5.3 – The five-minute mean spectral entropy at Fz for each patient (black) with the inverse wavelet transformation of the main oscillatory period (red). Inverse wavelet transformations have been normalized to the scale of the original spectral entropy signal.

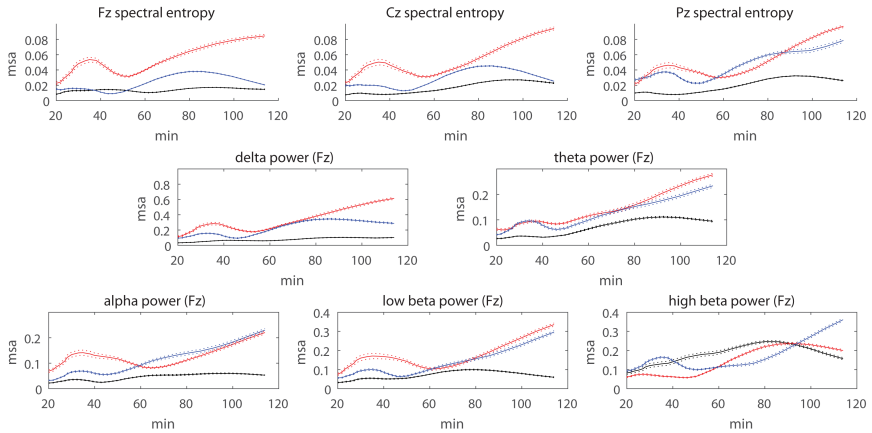


5.3.4 Correlations between band powers and spectral entropy

Again, consistent with (Piarulli et al., 2016), a within-subjects correlation between the spectral entropy time course and the EEG band power time courses were performed using total band powers (Table 5.6). The significance of each of the correlations was assessed against null distribution of permuted correlations. Additionally, the peak periods for each of the bands of interest were compared to the peak period of the spectral entropy for each patient (as in Figure 5.4 and Table 5.5).

With respect to periodicity, Patient 1 had the same peak periods across the delta and theta bands as that of spectral entropy at 92 minutes, and a close peak in the alpha band at 99 minutes. Patient 2 had peak periods of 114 minutes for spectral entropy and all bands except for the high beta band. Patient 3 only had a similar period between spectral entropy and the delta band at 86 minutes.

FIGURE 5.4 – Amplitude spectra for patient 1 (black lines), patient 2 (red lines), and patient 3 (blue lines) for spectral entropy at Fz, Cz, and Pz, and estimated band powers. All lines are enclosed by 95% confidence intervals (dotted lines).



Turning now to correlations between spectral entropy and band powers at the Fz channel (Table 5.6 of Supplementary Material), there were significant anti-correlations in all patients in the delta ($p < 0.001$) and theta bands ($p < 0.04$). There was a significant anti-correlation in patient 2 in the alpha band ($p < 0.001$). The low beta band was positively correlated to spectral entropy for patient 1 ($p = 0.047$), but negatively correlated for patient 2 ($p < 0.001$). Finally, there were significant positive correlations in the high beta band for both patients 1 and 2 ($p < 0.01$).

5.4 Discussion

We analyzed four hour EEG recordings in three comatose patients while they were in the intensive care unit. These patients were then compared to the two patient groups, minimally conscious (MCS) and vegetative state (VS), previously reported (Piarulli et al., 2016).

In general, when comparing the features generated from the spectral entropy and relative band powers across three channels to these clinical groups, patient 1 had a few features that were within the range of the MCS group, patient 2 had several features that were within the range of the MCS group, and patient 3 only had two features in one channel that were within the range of the VS group. The mean spectral entropy amplitude for patients 1 and 2 were slightly higher than both groups on the whole, whereas patient 3 was somewhat lower than both groups. In all cases, the coefficient of variation was higher for all patients than either reported patient group. Both Piarulli et al. (Piarulli et al., 2016), and Gosseries et al. (Gosseries et al., 2011) showed that higher EEG spectral entropy values were present in MCS patients than compared to VS patients.

That does not hold in this case, as patients 1 and 2 were both higher on average than both of the other groups. Whether this is due to the patients being in a more conscious state but being behaviourally unresponsive, or whether it was simply due to noise in the recording is unclear.

Since all patients included in this sample were in a comatose state, no EMG or EOG recordings were made since their eyes would be closed and they exhibited no spontaneous movement. Additionally, the CRS-R was not performed, as the test would not have been appropriate for their state. Since the CRS-R was not performed, the correlations between behaviour and spectral entropy could not be assessed. None of the patients exhibited voluntary movements nor spontaneous eye opening, and some have noted shortcomings with the GCS with regard to its sensitivity, especially in patients with multiple severe injuries (Grote, Böcker, Mutschler, Bouillon, & Lefering, 2011). It may be useful to correlate a different assessment, such as the Full Outline of Unresponsiveness (FOUR) (Wijdicks et al., 2005), which has been shown to have a strong correlation with the GCS but is more sensitive. The FOUR is also able to properly assess the level of consciousness of patients who are intubated, which is often the case of comatose patients.(Iyer et al., 2009; Khanal, Bhandari, Shrestha, Acharya, & Marhatta, 2016) The FOUR has also been shown to better diagnose a patient in a vegetative or minimally conscious state (Schnakers et al., 2006), although it is still beneficial to use the CRS-R in conjunction.

Relative band powers show higher power in the delta band than in the theta and alpha bands for all three patients. In the case of patient 3, almost the entire relative power is contained in the delta band. As has been reported previously (Piarulli et al., 2016; Sitt et al., 2014), decreasing power in the delta band is indicative of a change from VS to a return to full consciousness. These results

would suggest that, at least at the time of testing, these comatose patients would be more similar to patients in a vegetative state.

It appears that spectral entropy is extremely sensitive to external noise. While this in and of itself is not surprising, as any change in the amount or type of noise in the signal would directly change the spectral entropy, it is surprising to see the extent to which the noise can affect the rest of the analysis. Specifically, the analyses of patient 2 appears to be heavily influenced by changes in the external electrical noise from the ICU. This is clearly evident in Figure 5.3, where there is a second major period of 35 minutes that fits in with the two large drops in spectral entropy. In turn, these two drops in spectral entropy are also exactly anti-correlated with changes in relative band powers resulting in extremely significant correlations. Furthermore, from Figure 5.1, the spectrum for patient 2 looks nothing like any of the other spectra reported here or elsewhere (Piarulli et al., 2016). It would seem that the use of this analysis methodology requires a higher standard of signal quality and stability than one would need for other methodologies, and should be taken into account when preparing for data acquisition.

Of the three cases reported in the present study, patient 3 was the most interesting. The spectral entropy amplitude spectrum in Figure 5.1 visually appears to be more similar to the MCS group. Similarly, the mean spectral entropy amplitude has a well defined peak area of 86 minutes. However, the raw values of spectral entropy as well as the mean relative band powers were either the same or below that of the VS group, which in turn was below that of the MCS group.

Piarulli et al. (Piarulli et al., 2016) suggested that the increase in the CRS-R scores were followed by an increase in spectral entropy periodicity, which

brought them closer to those of healthy individuals. They also point out that two VS patients showed local periodicities that were more similar to the MCS group, suggesting local preservation of structures underneath those specific electrodes. This may also be the case for patient 3, however they point out some of the same shortcomings as we do, namely low sample sizes and the inability to generalize from so few individuals.

There are several differences between the present study and previous work (Piarulli et al., 2016). Some of these differences come from the circumstances under which the current data were collected, as these were not collected specifically for this purpose, but rather as part of another study. These data involved the patients receiving auditory stimulation for the duration of the recording, although the previous report did not specifically state whether the patients were stimulated or were simply in a resting state.

Other differences come from these patients being in a comatose state and being in a hospital environment where they were receiving complex care. As was mentioned before, they did not have their eyes open and were not moving for the duration of the recording, so no EMG or EOG recordings were made to compare with the EEG signals. The surrounding environment of the ICU is also not an ideal recording environment, both because of the constant care the patients are receiving and the higher than ideal electrical noise around them. Some of this can be reduced by recording data with the intent to perform this analysis and work to remove any potential sources of noise from the environment.

Since the sample of comatose patients was so small, it was not possible to make the same group-level comparisons as previous work with other clinical groups. Comparisons were made between individuals and the clinical groups, but these are not terribly informative in a group context. Inclusion of additional patient

cases would aid in the interpretability of these data and further validate the use of this analysis methodology with comatose patients.

Despite the shortcomings of this methodology and its extreme sensitivity to the external environment, we agree that the application of this methodology in an automated fashion would be beneficial for the other complementary EEG assessments, especially those that perform best when the patient is most aware. Much like we have reported in earlier work (Mah & Connolly, 2018; Mah, Connolly, Hamielec, & Fox-Robichaud, 2018), and has been noted by others (Giacino, Fins, Laureys, & Schiff, 2014), fluctuations in levels of consciousness can lead to increased false negative findings unless repeated measurements are made. By searching for times when the patient is likely to be more aware or at a higher level of consciousness and assessing them during these periods, the likelihood of false negatives can be reduced.

5.5 Supplemental Materials

TABLE 5.2 – Features for all patients at electrode Fz

Patient 1 (Fz)

	Mean	SD	Lower CI	Upper CI
Delta	0.5551	0.0250	0.5480	0.5622
Theta	0.2836	0.0163	0.2790	0.2882
Alpha	0.1069	0.0109	0.1038	0.1100
Low Beta	0.0393	0.0055	0.0377	0.0409
High Beta	0.0147	0.0071	0.0127	0.0167
SE Mean	0.7084	0.0132	0.7047	0.7121
SE SD	0.0566	0.0058	0.0550	0.0582
SE COV	0.0993	0.0097	0.0966	0.1020

Patient 2 (Fz)

	Mean	SD	Lower CI	Upper CI
Delta	0.5899	0.0715	0.5697	0.6101
Theta	0.2322	0.0429	0.2201	0.2443
Alpha	0.0823	0.0175	0.0773	0.0873
Low Beta	0.0644	0.0168	0.0596	0.0692
High Beta	0.0305	0.0114	0.0273	0.0337
SE Mean	0.6973	0.0442	0.6848	0.7098
SE SD	0.0845	0.0123	0.0810	0.0880
SE COV	0.1192	0.0219	0.1130	0.1254

Patient 3 (Fz)

	Mean	SD	Lower CI	Upper CI
Delta	0.8459	0.0339	0.8363	0.8555
Theta	0.1207	0.0239	0.1139	0.1275
Alpha	0.0199	0.0063	0.0181	0.0217
Low Beta	0.0086	0.0028	0.0078	0.0094
High Beta	0.0047	0.0020	0.0041	0.0053
SE Mean	0.5395	0.0212	0.5335	0.5455
SE SD	0.0713	0.0109	0.0682	0.0744
SE COV	0.1478	0.0135	0.1440	0.1516

TABLE 5.3 – Features for all patients at electrode Cz

Patient 1 (Cz)

	Mean	SD	Lower CI	Upper CI
Delta	0.5015	0.0297	0.4931	0.5099
Theta	0.2893	0.0158	0.2848	0.2938
Alpha	0.1388	0.0132	0.1351	0.1425
Low Beta	0.0536	0.0065	0.0518	0.0554
High Beta	0.0165	0.0053	0.0150	0.0180
SE Mean	0.7213	0.0151	0.7170	0.7256
SE SD	0.0741	0.0101	0.0712	0.0770
SE COV	0.0933	0.0097	0.0906	0.0960

Patient 2 (Cz)

	Mean	SD	Lower CI	Upper CI
Delta	0.5687	0.0709	0.5486	0.5888
Theta	0.2270	0.0406	0.2155	0.2385
Alpha	0.0925	0.0208	0.0866	0.0984
Low Beta	0.0780	0.0180	0.0729	0.0831
High Beta	0.0331	0.0113	0.0299	0.0363
SE Mean	0.7105	0.0453	0.6977	0.7233
SE SD	0.0802	0.0123	0.0767	0.0837
SE COV	0.1192	0.0219	0.1130	0.1254

Patient 3 (Cz)

	Mean	SD	Lower CI	Upper CI
Delta	0.8453	0.0429	0.8332	0.8574
Theta	0.1179	0.0322	0.1088	0.1270
Alpha	0.0243	0.0078	0.0221	0.0265
Low Beta	0.0084	0.0029	0.0076	0.0092
High Beta	0.0041	0.0016	0.0036	0.0046
SE Mean	0.5321	0.0250	0.5250	0.5392
SE SD	0.0787	0.0114	0.0755	0.0819
SE COV	0.1478	0.0135	0.1440	0.1516

TABLE 5.4 – Features for all patients at electrode Pz

Patient 1 (Pz)

	Mean	SD	Lower CI	Upper CI
Delta	0.4817	0.0335	0.4722	0.4912
Theta	0.3150	0.0183	0.3098	0.3202
Alpha	0.1454	0.0139	0.1415	0.1493
Low Beta	0.0426	0.0057	0.0410	0.0442
High Beta	0.0149	0.0042	0.0137	0.0161
SE Mean	0.7066	0.0171	0.7018	0.7114
SE SD	0.0781	0.0103	0.0752	0.0810
SE COV	0.0993	0.0097	0.0966	0.1020

Patient 2 (Pz)

	Mean	SD	Lower CI	Upper CI
Delta	0.5577	0.0767	0.5360	0.5794
Theta	0.2314	0.0444	0.2188	0.2440
Alpha	0.1298	0.0415	0.1181	0.1415
Low Beta	0.0585	0.0222	0.0522	0.0648
High Beta	0.0222	0.0076	0.0200	0.0244
SE Mean	0.6966	0.0434	0.6843	0.7089
SE SD	0.0734	0.0150	0.0692	0.0776
SE COV	0.1192	0.0219	0.1130	0.1254

Patient 3 (Pz)

	Mean	SD	Lower CI	Upper CI
Delta	0.8556	0.0402	0.8442	0.8670
Theta	0.1113	0.0304	0.1027	0.1199
Alpha	0.0213	0.0070	0.0193	0.0233
Low Beta	0.0077	0.0026	0.0070	0.0084
High Beta	0.0038	0.0013	0.0034	0.0042
SE Mean	0.5112	0.0414	0.4995	0.5229
SE SD	0.0809	0.0093	0.0783	0.0835
SE COV	0.1478	0.0135	0.1440	0.1516

TABLE 5.5 – Periods of peak wavelet intensity for each patient and frequency band at electrode Fz

Patient	SE	Delta	Theta	Alpha	Low Beta	High Beta
1	92	92	92	99	80	80
2	114	114	114	114	114	92
3	86	86	114	114	114	114

TABLE 5.6 – Correlation coefficients (r) and permutation test p-values for each patient and frequency band at electrode Fz

Patient	Delta		Theta		Alpha		Low Beta		High Beta	
	r	p-val	r	p-val	r	p-val	r	p-val	r	p-val
1	**-0.6907**	0.000	**-0.4105**	0.002	-0.1986	0.168	**0.284**	0.047	**0.3636**	0.008
2	**-0.9528**	0.000	**-0.7803**	0.000	**-0.7661**	0.000	**-0.6301**	0.000	**0.5303**	0.000
3	**-0.6265**	0.000	**-0.2935**	0.037	-0.1030	0.466	-0.1207	0.413	-0.0750	0.636

References

Andrews, K., Murphy, L., Munday, R., & Littlewood, C. (1996). Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. *Bmj*, *313*(7048), 13–16.

Bruno, M. A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., & Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: Recent advances in our understanding of disorders of consciousness. *Journal of Neurology*, *258*(7), 1373–1384.

Canadian Institute for Health Information. (2016). *Care in Canadian ICUs* (No. August). Ottawa, ON: CIHI.

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.

Fischer, C., Dailler, F., & Morlet, D. (2008, October). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical neurophysiology*, *119*(10), 2224–30. doi: 10.1016/j.clinph.2008.03.035

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical neurophysiology*, *110*(9), 1601–1610.

Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., & Katz, D. I. (2002). The minimally conscious state. *Neurology*, *58*(3), 349–353.

Giacino, J. T., Fins, J. J., Laureys, S., & Schiff, N. D. (2014). Disorders of consciousness after acquired brain injury: The state of the science. *Nature Reviews Neurology*, *10*(2), 99–114.

Giacino, J. T., Kalmar, K., & Whyte, J. (2004). The jfk coma recovery scale-revised: Measurement characteristics and diagnostic utility. *Archives of physical medicine and rehabilitation*, *85*(12), 2020–2029.

Gosseries, O., Schnakers, C., Ledoux, D., Vanhaudenhuyse, A., Bruno, M.-A., Demertzi, A., ... Laureys, S. (2011). Automated EEG entropy measurements in coma , vegetative state / unresponsive wakefulness. *Functional Neurology*, *26*(1), 25–30.

Green, S. M. (2011). Cheerio, laddie! Bidding farewell to the Glasgow Coma Scale. *Annals of Emergency Medicine*, *58*(5), 427–430.

Grote, S., Böcker, W., Mutschler, W., Bouillon, B., & Lefering, R. (2011). Diagnostic Value of the Glasgow Coma Scale for Traumatic Brain Injury in 18,002 Patients with Severe Multiple Injuries. *Journal of Neurotrauma*, *28*(4), 527–534.

Iyer, V. N., Mandrekar, J. N., Danielson, R. D., Zubkov, A. Y., Elmer, J. L., & Wijdicks, E. F. (2009). Validity of the FOUR score coma scale in the medical intensive care unit. *Mayo Clinic Proceedings*, *84*(8), 694–701.

Khanal, K., Bhandari, S., Shrestha, N., Acharya, S., & Marhatta, M. (2016). Comparison of outcome predictions by the Glasgow coma scale and the Full Outline of UnResponsiveness score in the neurological and neurosurgical patients in the Intensive Care Unit. *Indian Journal of Critical Care Medicine*, *20*(8), 473-476.

Laureys, S., Bodart, O., & Gosseries, O. (2014). The Glasgow Coma Scale: Time for critical reappraisal? *The Lancet Neurology*, *13*(8), 755–757.

Laureys, S., Celesia, G. G., Cohadon, F., Lavrijsen, J., León-Carrión, J., Sannita, W. G., ... Dolce, G. (2010). Unresponsive wakefulness syndrome: A new name for the vegetative state or apallic syndrome. *BMC Medicine*, *8*, 2–5.

Mah, R. L., & Connolly, J. F. (2018). *Electrophysiological markers of variations in perceptual and cognitive processing in coma.*

Mah, R. L., Connolly, J. F., Hamielec, C., & Fox-Robichaud, A. E. (2018). *Advancing prognostication by the detection of fluctuating states of consciousness in coma as reflected by an electrophysiological response.*

Piarulli, A., Bergamasco, M., Thibaut, A., Cologan, V., Gosseries, O., & Laureys, S. (2016, Sep 01). EEG ultradian rhythmicity differences in disorders of consciousness during wakefulness. *Journal of Neurology*, *263*(9), 1746–1760.

Schnakers, C., Giacino, J., Kalmar, K., Piret, S., Lopez, E., Boly, M., ... Laureys, S. (2006). Does the four score correctly diagnose the vegetative and minimally conscious states? *Annals of Neurology*, *60*(6), 744-745.

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., Majerus, S., ... Laureys, S. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC neurology*, *9*(1), 1.

Sitt, J. D., King, J.-R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A.,

... Naccache, L. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*, *137*(8), 2258–2270.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and imparied conciousness. A practical scale. *Lancet*, *2*(7872), 81–84.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, *79*(1), 61–78.

Wijdicks, E. F., Bamlet, W. R., Maramattom, B. V., Manno, E. M., & McClelland, R. L. (2005). Validation of a new coma scale: The FOUR score. *Annals of Neurology*, *58*(4), 585–593.

6 Summary and Conclusions

The first aim of this thesis was to develop and test an ERP stimulation battery to be used in the prediction of coma outcomes. Through this process, we sought to determine the best paradigms to elicit specific components of interest, which have previously been shown to have high positive predictive power. As has been shown by Fischer et al. (1999), one of these predictive components is the MMN. It is often taken for granted that this component will always be generated unless there is some specific pathology that interferes with it, however we have shown evidence that the detectability of this component can change over time. Our second aim was to determine whether the MMN is a reliable indicator of emergence out of coma, which methods of detection are optimal, and whether a prognostic test using the MMN would need to be used multiple times to ensure an informative result. The third aim was to assess the utility of spectral entropy, and specifically the wavelet decomposition methodology presented in Piarulli et al. (2016), as a means of determining a patient's level of consciousness for the purpose of targeting the administration of the ERP stimulation battery.

Through the validation of the paradigms on healthy control populations, and then the application to several patient cases, we found evidence that the MMN

appears to wax and wane over time in a sample of comatose patients. We further assessed the utility of an analysis method to determine when to begin stimulation with the battery. This chapter will serve to unite the findings of the studies presented in this thesis, reflect upon their significance, and set out a future research plan that follows from this foundational work.

The structure of this chapter is as follows: first will be a summary of the preceding chapters, their findings, and a discussion of their individual contributions. Following this will be a brief discussion of the broader significance of the findings of the studies presented in this thesis, some general conclusions of the thesis, and finally some avenues for research that stem from this thesis.

6.1 Summary of results

This thesis consists of three major topics discussed over four studies. In this section, we will outline the major findings of each topic.

Determining which paradigms elicit strong responses in absence of explicit attention

In Chapter 2, six paradigms designed to elicit the MMN, P300, and N400 ERP components were compared first between a group of younger adults and a group of older adults. Both groups were instructed not to pay attention to the sounds they heard. This comparison allowed us to examine the effect of age on the paradigms' ability to elicit strong ERP responses.

In comparing these two groups of adults, we found an effect of age on the ERPs that were elicited. Specifically, MMN peaks from older adults were later and had reduced amplitudes compared to those from younger adults. This is consistent with findings from other studies reporting similar amplitude reductions in the

MMN. A reduction of amplitude is also seen was the P300 elicited by the SON and the N400 elicited by semantically incongruent words in the semantic violation sentences and word-word priming pairs.

To better examine the effect of attention, a third group of participants received stimulation from four of the six paradigms. These paradigms were ones eliciting the P300 and N400, and were expected to be influenced by attention. The participants were first instructed to ignore the stimuli, and then to press a button in response to specific stimuli. In every paradigm, the active response condition was associated with an increase in the number of participants who showed an appropriate ERP, as well as a general increase in the ERP amplitude. In some cases, this amplitude difference was quite large. For instance, the SON paradigm generated P300 peaks on average around 5 μV in the passive condition, and on average around 15 μV in the active condition.

There was also an effect of the strength of the stimulus context. In the case of the MMN, larger peaks were found when the deviants were rare rather than being a deviation from a pattern. For the P300, the categorical difference of the SON being embedded within a sequence of tones produced larger peaks than when the SON was embedded within a list of other names. With the N400, better performance was seen with semantically incongruent words embedded in carrier sentences to help build a semantic expectancy than when the words followed isolated related words, as was the case with the word-word priming.

In the end, two paradigms were chosen that best elicit the three ERP components of interest in all situations. For the MMN, the auditory oddball paradigm was the only one of the two paradigms to even elicit a strong MMN. Additionally, the auditory oddball paradigm with tones as standard sounds was able to elicit robust P300's in almost every participant to the SON, whereas the

more traditional SON paradigm using only words and names was only able to elicit the P300 in some participants and with an amplitude that was quite small. Finally, for the N400, semantic violation sentences were chosen over the word-word priming in part because of the increase in performance seen with the stronger contexts of the sentences, and in part, because of the enhancing effect of attention on both paradigms. For paradigms eliciting the N400, one must consider a trade-off between those using a global discourse context which may elicit more robust responses in healthy populations, and the diminished context processing capacities of patients with traumatic brain injuries who have disorders of consciousness.

Evaluating methods of detecting a waxing and waning MMN in comatose patients

Chapters 3 and 4 examined four comatose patients who were receiving auditory stimulation over the course of about 24 hours. With these patients, the main objective was to evaluate five methods of detecting the MMN. These methods were: visual inspection by a trained electrophysiologist, a serial t-test method (as in Marchand, D'Arcy, & Connolly, 2002), the topographic consistency test (as in Koenig & Melie-García, 2010), a spatiotemporal clustering analysis (as in Maris & Oostenveld, 2007), and the computation of the Jeffrey-Zellner-Siow Bayes factor (JZS-BF) (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

In all four cases, the visual inspection of average waveforms by a trained electrophysiologist gave the best results. While this method does have some shortcomings, such as the need for many trials to construct an average waveform, and relative subjectivity, it still remains the best performing method thus far.

The serial t-test, with the parameters that were used in the analysis, did not perform very well. It indicated several stimulus blocks that contained significant

differences between the standard and deviant waveforms within the MMN window, however, most of these were concluded to be likely false positives. This was especially evident when the signal had noise contamination, where the differences in the conditions could wholly be attributed to the noise. As is mentioned in Chapter 4, narrowing the relevant time window and increasing the number of consecutive significant points could help increase the false positive rejection rate.

The topographic consistency test also performed poorly, especially with patients with recordings from lower density arrays. It did, however, show some positive results consistent with the visual inspection of the MMN. This suggests that the method is capable of accurately detecting the MMN when the data are of high quality. The primary shortcoming of this method is that the data have to have a good signal-to-noise ratio and be recorded with sufficient electrode density. For instance, if a patient is not able to tolerate a higher density array held by a cap (for instance, patients who have undergone a craniotomy) or has an unusual scalp topography (either because of deformations of the skull or from other medical devices like staples or extraventricular drains), this method will not function correctly, and may result in an increase of false positives. The topographic consistency test is also very sensitive to external noise since it is correlating activity across electrodes. This makes it difficult to use in a critical care environment.

On the other hand, the spatiotemporal cluster-based analysis performed much better, especially as a confirmatory test to use alongside visual inspection. Almost always when there was a positive visual result, there was also a corresponding positive cluster result. However, this method is susceptible to the same density shortcomings as the topographic consistency test, as it correlates activity

in the spatial dimension as well as the temporal dimension. This means that if a low number of electrodes is used, the result can be heavily biased by external noise. It is also susceptible to non-cortical noise, as any structured signal that is detected across electrodes will be correlated through the analysis, leading to spurious results.

The method that appeared to work the best regardless of electrode density and external noise was the Bayesian t-tests. This method produced results that were confirmatory for the visual inspection. When the MMN was prominent in the waveform, the test would show "strong evidence" for the MMN. When the MMN was less obvious, but still present visually, the test would show "anecdotal" to "substantial" evidence of the MMN.

In summary, the serial t-test and topographic consistency tests did not provide reliable confirmations of visual detection of the MMN in the clinical sample examined. Both the spatiotemporal clustering method and the Bayesian t-tests provided more reliable confirmations of the MMN after visual inspection.

An important and unexpected finding was that the MMN appeared to vary over recording time points in the signals of all patients. The MMN has previously been found in various states of consciousness, but its stability over time in patients with altered consciousness has not been assessed. The current evidence suggests that the MMN does vary only in its detectability. It may also vary in its presence. Additional testing is necessary to both confirm this effect in other patients and to determine whether the MMN is merely less detectable during certain periods, or whether it is absent altogether. This does suggest that when ERP assessments are used on patients, they should be repeated over a period of time and during periods when the patient is likely more aware or conscious.

Unclear results of wavelet decomposition of spectral entropy in coma-

tose patients

In Chapter 5, we applied an analysis method based on wavelet decomposition of spectral entropy. This method has been previously used on data collected from patients who were diagnosed as minimally conscious or in a vegetative state Piarulli et al. (2016). This method had been used both to better classify the patients into their diagnostic groups as well as to show that there was an ultradian rhythm with a 70 minute period that was present in those patients who were minimally conscious.

In applying this analysis methodology to comatose patients, we were able to show distinct patterns in each patient tested and compare those to those previously reported. One patient whose recordings were completely contaminated by noise, resulting in the method not producing an interpretable output. Another patient showed no primary spectral entropy peak period. This pattern was more similar to the findings from patients who were vegetative. A third patient showed a primary spectral entropy peak period of about 70 minutes, which was the same as that previously reported in minimally conscious patients.

Much like some of the earlier methods of detecting the MMN, spectral entropy analysis also appears to be extremely sensitive to external noise. If the data being collected were only of a patient silently resting, and if there was any extraneous noise, the data may not be able to be used for any other purpose. In the case of the data presented in this thesis, there were other event-related time locking events to use to further extract information about the cortical signals, which allowed an ERP analysis to be conducted. If the sole use of the data collected was for use in a spectral entropy analysis, then it must be of high quality from the outset. Both reducing external electrical noise and ensuring the electrode site preparations have been done well are necessary for this, as

well as constantly monitoring the quality of the incoming data and re-preparing electrode sites as necessary.

The upside to all of this preparation is that this method shows promise as a way to target the delivery of the cognitive assessment battery. Since the tests will produce better results when the patient is more aware, being able to have a more objective measure of their level of consciousness will increase the likelihood of stimulating them while they are responsive. Furthermore, by reducing the amount of external noise collected, the ERP analysis becomes easier to perform, and the results more clear to see.

6.2 Implications and Contributions

Recommendations on paradigms for eliciting prognostic ERPs

At the end of Chapter 2, two paradigms were proposed as the best choices to elicit prognostic ERPs from the original cohort of six. The first was the auditory oddball paradigm with the subject's own name embedded in a sequence of tones. This test combined elements from Holeckova, Fischer, Giard, Delpuech, and Morlet (2006) and Fischer, Dailler, and Morlet (2008), resulting in a paradigm capable of eliciting both the MMN and the P300, both of which have previously been shown to have high positive predictive power in clinical cases.

The novel auditory oddball paradigm has the added benefit of being able to elicit both components within one testing block, which enables the quick acquisition of data. This is especially important when dealing with patients whose level of consciousness may be fluctuating or be inconsistent.

The second paradigm chosen was one with semantic violation sentences. This paradigm performed better in eliciting the N400 in healthy participants when

compared to the word-word priming paradigm. While the proportion of healthy participants showing the N400 while not paying attention was just over 60%, the proportion of participants increased by another 10%when they paid attention to the stimuli. This underscores the necessity of reminding all patients to attend to the stimuli, even if they might not appear capable.

The use of semantic violation sentences has been criticized over tasks that use a global discourse context. While discourse paradigms may work well with fully capable, healthy individuals, patients with traumatic brain injuries are potentially more likely to be unable to maintain the contextual information in memory for long enough to generate a meaningful response. Balancing the potentially reduced sensitivity of the semantic violation sentence paradigm for the complexity of the global discourse context paradigm is one that must be made with the patient's abilities in mind.

Recommendations on methods for confirming presence of the MMN

Throughout chapters 3 and 4, four methods of confirming the presence of a visually identified MMN were evaluated within the context of a small clinical population. Of these four methods, three performed poorly when the data were recorded from an electrode array of low density (*i.e.* when the number of electrodes was below 16). This suggests that, whenever possible, recordings should be done with as many electrodes as possible. However, when the patient also has an irregular physiology or their scalp contains metal, some electrodes may need to be excluded from analysis or higher density recordings may not be possible altogether.

Two methods performed poorly due to external noise that was captured in the signal. While this remains an issue when collecting data in a critical care environment, there are methods of mitigation. Wherever possible, acquisition

equipment should be placed away from other equipment and electrode leads be run away from other sensor leads from the patient. Turning off overhead fluorescent lighting also is helpful in reducing high frequency noise which can bias the analysis results.

In the end, two methods were recommended to confirm a visually identified MMN: the spatiotemporal cluster analysis and Bayesian t-tests. While these methods still perform better when data quality and electrode density are high, they were still able to confirm the presence of a MMN in most cases.

These tests also lend themselves to easy adaptation for use in automatic stimulus delivery and test interpretation, as they can generate binary significance results. This can be helpful for use in a clinical setting where clinicians may not want to, or cannot meaningfully, interpret a numerical result. At the same time, if clinicians find a numerical value better for a more subjective interpretation, both methods can produce a useful significance value that shows the confidence in or amount of evidence for the presence of the MMN.

New evidence for the dynamic nature of the MMN

Generally, the MMN has been thought of as a component that is stable and does not change over periods of time. At least in the case of healthy control subjects, if the MMN is present at one point, it will continue to be present at another later time. Prior work from Tzovara et al. (2013) has showed evidence suggesting that the MMN can return to a patient over a period of several days. This was indexed by the changes in the ability of a classification algorithm to properly identify single trials in an auditory oddball paradigm between recording sessions.

In chapters 3 and 4, we have presented further evidence that the MMN has a dynamic nature. In all patients presented in these chapters, the MMN was found to vary in its detectability over short periods of time. Specifically, within

periods of less than 24 hours, the MMN was initially not detectable, then became detectable for at least one stimulation block, and then was again not detectable. This suggests that, at least for these patients, if the MMN were being used as a prognostic measure, it may have been missed simply because of when the test occurred. The corollary to this is that prognostic tests or models should account for this temporal fluctuation in detectability and should be run multiple times during periods when the patient is most likely to be aware or in a higher state of consciousness.

Recommendations for use of spectral entropy in prognostication

Finally, in Chapter 5, we presented evidence that spectral entropy, and specifically the application of a wavelet decomposition method, can be useful for prognostic testing of comatose patients. This method showed that one patient had results similar to those of a previously reported VS/UWS group, and that another had results similar to those of a previously reported MCS group. These patients were all behaviourally identical–they had no spontaneous eye opening or movement, were not able to respond to any commands, or verbalize. They did, however, have differential results when comparing their spectral entropy. This would suggest that neurologically, patients who present as comatose may actually have some level of covert consciousness. Being better able to differentiate between levels of performance in comatose patients throughout their recovery enables clinicians to refine their treatment plans and start rehabilatory treatments sooner.

As was pointed out by Piarulli et al. (2016) in regards to the use of spectral entropy analysis in VS/UWS/MCS patients, the constant monitoring of spectral entropy in comatose patients should further be done as a means of better targeting the prognostic tests. If it is the case that these patients have fluctuating

levels of consciousness that are not shown behaviourally, then being able to detect periods of increased consciousness would be beneficial to reduce the number of patients whose best performance may be missed. Spectral entropy based time windows may also be used for other behavioural and functional neuroimaging tests which partly rely on the level of consciousness of the patient.

6.3 Topics for further research

There were several related topics of research that remained unanswered from the results of the studies in this thesis.

1. Determine the usefulness of the N400 as either a marker of coma emergence or functional state after emergence

Even though the stimulation battery involved the elicitation of the N400, and data from the patients reported in this thesis were collected, the data were not examined for the N400.

A first step in evaluating the usefulness of the N400 as a prognostic indicator would be to develop an appropriate method of processing these data. As the complexity of the elicitation paradigms is higher than those of the MMN and P300, the method of analysis would have to account for this.

Specifically, the number of trials for each condition in each block is significantly lower for the N400 paradigms than for the MMN/P300 paradigms. Also, the components can have lower amplitudes and less well defined peaks for the N400 paradigms. This means that the signal can be harder to isolate from the background noise, and could lead to a higher false negative rate. Ensuring the analysis methods are able to accurately pick out the component from the noisy patient data is crucial to properly determining the usefulness of these

paradigms.

The previously collected patient dataset will be useful in developing the analysis method, although without additional data, it will be less useful in determining whether the N400 is a marker of emergence or functional state. This is because of the small sample size and high number of patients whose life support was removed before an accurate outcome could be determined.

2. Application of spectral entropy awareness metrics to patient stimulation

Chapter 5 presented data that showed that spectral entropy analysis could be applied to comatose patients and showed that at least some of them may have had a level of consciousness similar to minimally conscious patients in another study. Piarulli et al. (2016) had suggested that monitoring vegetative or minimally conscious patients' spectral entropy levels over time could be useful in deciding when to administer assessment tests to ensure a higher level of consciousness at the time of testing. While the methodology does require some amount of intent and preparation to produce clean, usable data, the work done to reduce external noise serves to improve the quality of the data used in the assessments.

To apply this methodology to target assessment periods in coma, the first step would be to constantly monitor the signal from the patient to determine their baseline entropy level. Until the utility of spectral entropy can be established in this patient group, it would be suggested to continue the use of the stimulation battery as normal, but identify periods of increased entropy. Wherever possible, stimulus delivery would be prioritized to these periods, as opposed to periods of silent rest or for nurse/patient interactions. In either case, periods of increased spectral entropy can also help inform the method of ERP analysis by identifying

periods when ERPs may be easier to isolate.

3. Further data collection and use of the recommended paradigms

One of the major shortcomings of all of the studies presented here is that the sample size is small. Another factor that reduces the amount of usable data were the patients who had fewer electrodes recorded.

While some very interesting results have been found in the few patients who have been tested, increasing the number of cases will only help to strengthen these results. Specifically, determining whether the MMN does truly vary in presence in a signal, or if it is simply more difficult to detect will require a much larger sample size.

As was alluded to earlier, properly testing the N400 paradigms for their utility will also require a larger clinical sample, because the component is more difficult to elicit generally, but also because of the lower number of trials in each pass of the paradigm.

In all stages of this work, the use of other, better ERP detection algorithms will be vital for better extracting the components from the noise. This work has already started (see Tzovara et al., 2013 or Armanfard, Komeili, Reilly, Mah, & Connolly, 2016).

In conclusion, through the development of a framework for the extended monitoring of the levels of consciousness in unresponsive patients, and through the application of this framework on several patients, we have demonstrated variability of the MMN over the period of a day, and have shown that spectral entropy can be applied to the care of comatose patients. These preliminary results lend themselves to a very broad plan of further research, both in better characterizing the MMN in this patient population and in developing prognostic tools for predicting more granular patient outcomes.

References

Armanfard, N., Komeili, M., Reilly, J. P., Mah, R., & Connolly, J. F. (2016). Automatic and continuous assessment of ERPs for mismatch negativity detection. In *2016 38th annual international conference of the ieee engineering in medicine and biology society (embs)* (pp. 969–972). doi: 10.1109/EMBC.2016.7590863

Fischer, C., Dailler, F., & Morlet, D. (2008). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical Neurophysiology*, *119*(10), 2224–30.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., & Salord, F. (1999). Mismatch negativity and late auditory evoked potentials in comatose patients. *Clinical Neurophysiology*, *110*(9), 1601–1610.

Holeckova, I., Fischer, C., Giard, M.-H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. *Brain Research*, *1082*(1), 142–152.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, *23*(3), 233–242.

Marchand, Y., D'Arcy, R. C., & Connolly, J. F. (2002, nov). Linking neurophysiological and neuropsychological measures for aphasia assessment. *Clinical Neurophysiology*, *113*(11), 1715–1722.

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

Piarulli, A., Bergamasco, M., Thibaut, A., Cologan, V., Gosseries, O., & Laureys, S. (2016, Sep 01). EEG ultradian rhythmicity differences in disorders of consciousness during wakefulness. *Journal of Neurology*, *263*(9), 1746–1760.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225-237.

Tzovara, A., Rossetti, A. O., Spierer, L., Grivel, J., Murray, M. M., Oddo, M., & De Lucia, M. (2013). Progression of auditory discrimination based on neural decoding predicts awakening from coma. *Brain*, *136*(1), 81–89.