EEG ASSESSMENT OF DISORDERED CONSCIOUSNESS

EEG ASSESSMENT OF DISORDERED CONSCIOUSNESS: A FRAMEWORK AND A CASE STUDY

BY

NATALIA LAPINSKAYA, H. B. Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science

McMaster University MASTER OF SCIENCE (2018) Neuroscience
Hamilton, Ontario

TITLE: EEG assessment of disordered consciousness: a framework and a case study

AUTHOR: Natalia Lapinskaya, H. B. Sc. (University of Toronto, Canada)

SUPERVISOR: Dr. J. Connolly

NUMBER OF PAGES: ix, 47

LAY ABSTRACT

Assessing cognitive abilities in disorders of consciousness such as coma currently relies on assessments of overt behaviour, such as the ability to follow commands or react to a stimulus. Neuroimaging has shown that absence of overt behavior does not necessarily indicate absence of covert cognition, raising questions about behaviour-only assessment. This study describes a neuroimaging testing battery aimed at evaluating a hierarchy of cognitive functions without the need for a behavioural response by measuring brain activity driven by auditory stimulation. This battery was administered to a comatose patient over a 24-hour period during two recording sessions one week apart, as well as to a sample of healthy young adults. The results show that changes in the patient's condition between testing sessions was accompanied by detectable and quantifiable change in their stimulus-driven brain activity. The results also suggest fluctuations in the patient's ability to produce detectable responses over the course of 24 hours, which in turn suggests that repeated testing is necessary for a complete evaluation. Overall, neuroimaging provides a promising avenue for non-behavioral assessment of cognition, which will greatly benefit a population whose physical faculties may be compromised.

ABSTRACT

Assessing cognitive abilities in disorders of consciousness (DOC) relies on assessments of overt behaviour, such as the ability to follow commands. Neuroimaging has shown that absence of overt behavior does not necessarily indicate absence of covert cognition, raising questions about behaviour-only assessment. Several electroencephalographic (EEG) markers of higher cognitive functions (event-related potentials; ERPs) have shown the potential to differentiate between DOC states, as well as predict awakening and condition upon emergence. However, no one ERP has emerged with sensitivity and specificity high enough to be widely accepted, showing that further investigation is needed. More recently, evidence has emerged for fluctuations of ERP detectability in DOC over the course of several hours, and for prognostic power of changes in ERP presentation between testing sessions. This investigation builds on such findings towards improving evaluation of cognition in DOC. A testing battery combining several well-known auditory ERPs was administered to a comatose patient over a 24-hour period during two recording sessions one week apart, as well as to a sample of healthy young adults. The patient scored 3 and 6 on the Glasgow Coma Scale (GCS) during the first and second session, respectively. The results show that changes in GCS score were accompanied by changes in ERP detectability. The results also suggest detectability fluctuations over the course of 24 hours, which in turn suggests that repeated testing is necessary for complete evaluation. Future work should validate these findings with a larger sample; additionally, establishing population norms for single-subject prevalence, latency, and amplitude of ERPs would improve confidence in interpreting patient results. With the current understanding of both healthy and DOC ERPs, detecting ERP presence may contribute to a positive DOC prognosis with a degree of confidence, but caution must be exercised in making negative prognoses or high-stakes care decisions based on ERP absence.

ACKNOWLEDGMENTS

I gratefully acknowledge the guidance of my advisor, Dr. John Connolly, in undertaking this research. I am grateful to Dr. Alison Fox-Robichaud for her perspective as a clinician and for being our link to the Hamilton General Hospital. I am grateful to Dr. Jim Reilly and Dr. Sue Becker for productive methodological discussions. I am also grateful to Dr. Ellen Lau, who has been and continues to be a role model and an extremely positive academic influence.

I am grateful to the nurses and staff of the Intensive Care Unit at the Hamilton General Hospital, particularly Lisa Klotz, without whom a large part of this research would not have been possible (and for the company and blankets through the night shifts).

I wish to thank my lab mates at the LMBLab and fellow members of ARiEAL, particularly Rober Boshra, Richard Mah, Dan Schmidtke and Narcisse Torshizi, for shared technical expertise, companionship, camaraderie, and blurring the lines between friends and colleagues. I am also grateful to our volunteer research assistants for their contributions to data collection – in particular Nitish Dhingra, who shared the round-the-clock data collection at the Hamilton General. Special thanks to Sandra Murphy of the Neuroscience graduate program and Chia-Yu Lin of ARiEAL– they keep these gears turning, we'd get nowhere without them.

I also wish to thank the friends who've been on this academic journey with me – Leigh Ann, Stefanie, Angele, Julie, Lara, Anton, Allyson, and all the DC and Toronto folks, thanks for keeping me balanced.

I thank my family for their unwavering support through this process, and my partner Andrew, who is my rock, my challenger, and my support, in this undertaking and in all those to come.

I am most of all grateful to the patient described in the thesis case study, and her family for their permission to conduct this research. I am honored by your trust, and hope that this work may prove useful to those who seek to improve patient care.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

ANOVA – analysis of variance

CHT – concurrent hierarchical tracking

CRS-R - coma recovery scale revised

DOC - disorders of consciousness

ECog – electrocorticography

EEG - electroencephalography

ERP - event-related potential

GCS - Glasgow coma scale

ICU - intensive care unit

LIS - locked-in syndrome

MCS – minimally conscious state

MEG - magnetoencephalography

MMN - mismatch negativity

SOA - stimulus onset asynchrony

SON - subject's own name

TBI – traumatic brain injury

VS – vegetative state

1. Introduction

Accurate assessment and prognosis in disorders of consciousness (DOC) is important for a number of practical and ethical reasons. Assessing cognitive abilities in DOC relies on assessments of overt behaviour, such as the ability to follow commands or react to a stimulus. Neuroimaging has shown that absence of overt behavior does not necessarily indicate absence of covert cognition, raising questions about behaviour-only assessment.

Electroencephalography (EEG) offers a relatively inexpensive, portable, non-invasive and non-disruptive way to evaluate covert cognition in absence of overt behaviour. EEG correlates of higher physiological and cognitive functions such as attention and memory have been used to differentiate between DOC states, and have emerged as potential predictors of awakening and condition upon emergence. Most commonly studied are event-related potentials (ERPs), several of which can be elicited independently of instruction, task, or even attention, making them appropriate for populations whose faculties may be compromised. ERPs have been shown to be able to distinguish DOC states and predict patient outcomes (e.g. 1-5). However, no one ERP has yet emerged with sensitivity and specificity high enough to be widely accepted (6-8), and reports of ERP behaviour in DOC can be inconsistent (e.g. 1, 3, 4), showing that while this is a promising avenue for improving DOC diagnosis, further work is still needed.

More recently, evidence has emerged that ERP detectability in DOC patients can fluctuate over the course of several hours (e.g. case study described in 9), and that

changes in ERP presentation between testing sessions carry prognostic power (10). The current study builds on such findings with the aim of improving evaluation of covert cognition in DOC and potentially addressing the inconsistencies in existing literature.

A testing battery that combines several well-known EEG measures, including three ERPs testing a hierarchy of cognitive functions, and one recently discovered phenomenon that has intriguing possibilities with regard to quantifying the comprehension of spoken language, was administered to a comatose patient over a 24-hour period during two recording sessions one week apart. The testing paradigms were validated with a sample of healthy controls. This paper describes the patient case study, the results of the control investigation, and discusses the utility of such a battery in a clinical setting.

1.1 The concept of consciousness and the mechanisms of DOC

In discussing DOC, it is useful to first consider the definition of consciousness. Some define it as awareness, and others make an explicit distinction between the concepts, stating that while consciousness is the capacity for subjective experience, awareness is merely one manifestation of that capacity (11). Yet others distinguish between awareness of oneself and awareness of one's surroundings, positing that one need not exhibit them simultaneously (e.g. advanced dementia), and imply that one form of awareness is sufficient to be considered conscious (11, 12). Separate from the concept of awareness is the concept of arousal (here used synonymously with wakefulness or alertness), which is relevant to both conscious and unconscious states (13). Giacino et al define this concept as a continuum between asleep and awake

states, operationalized by the intensity of the stimulus required to interrupt sleep and duration of wakefulness once sleep has been interrupted (11). Some recent evidence suggests that DOC patients undergo short-term changes along that continuum that cause variability in their response to stimuli (9, 14). Neurologically, the processes that are thought to underpin consciousness, awareness, and arousal span multiple brain regions. Reticular formation, a large network of nuclei and projections originating in the brainstem and extending to thalamus, cerebral cortex, and spinal cord, is thought to play a central role (11, 15).

Disorders of consciousness originate from damage to cerebral cortex, underlying white matter, or brainstem structures (16). The cause of damage can be traumatic (e.g. physical trauma to the head) or non-traumatic (e.g. stroke, anoxia, tumors, infections). DOC states are categorized as coma, vegetative state/unresponsive wakefulness syndrome (VS/UWS), and minimally conscious state (MCS). Coma results from severe, diffuse bihemispheric lesions and/or brain stem injury that usually involves the upper two-thirds of the brainstem (17). It usually persists for a maximum of one month, in absence of metabolic, toxic, or infectious complications (18). The neurobehavioral criteria for coma are: absence of eye opening, purposeful movement, verbal communication, inability to follow commands or sustain visual pursuit if the eyes are held open (19). It is also characterized by an absence of sleep/wake cycles as detectable by EEG (17). Coma can subsequently progress into vegetative state, also referred to as unresponsive wakefulness syndrome. Neurologically, VS differs from coma in the relative preservation (or restoration) of brainstem functionality in the presence of widespread bihemispheric pathology (16,17). A period of coma always

precedes VS in traumatic brain injury, but VS may be observed without coma in metabolic and dementing disorders (19). The neurobehavioral criteria of VS are similar to those of coma, except that the patient should exhibit eye opening, either spontaneous or in response to a stimulus; additionally, sleep/wake cycles may be evident from EEG (19). VS may progress into minimally conscious state, which is differentiated by reproducible, even if minimal and fluctuating, purposeful behaviour or evidence of meaningful interaction with the environment, as well as preserved sleep-wake cycles (19, 20). Behaviourally, these criteria can present as one or more of the following: following simple commands, gestural or verbal yes/no responses regardless of accuracy, intelligible verbalization, displaying affect (e.g. laughing, crying) appropriate to the stimulus, or sustained visual pursuit or fixation (20). Emergence from MCS is characterized by consistent demonstration of functional interactive communication or functional use of at least two objects (20).

Behavioural assessment of DOC is commonly done using the Glasgow Coma Scale, which assigns a score based on the patient's eye, motor, and verbal responses (illustrated in Table 1, (21)). Coma is classified as a score of 3-8 (19). The JFK Coma Recovery Scale (now revised, CRS-R) was developed for the purpose of characterizing cognitive functioning in post-coma states, and has been used to differentiate VS from MCS (22). CRS-R is comprised of 6 sub-scales assessing auditory, visual, motor, oromotor, communicative, and arousal capabilities and ranges from purely reflexive activity to cognitively mediated behaviour (22). Glasgow Outcome Scale is used to categorize DOC outcomes as either death, persistent vegetative state, severe disability,

moderate disability, or low disability/good recovery, scored as 1, 2, 3, 4 and 5 respectively (23).

Parameter	**Score**					
	1	**2**	**3**	**4**	**5**	**6**
Eye Opening	None	To pain	To voice	Spontaneous, open with blinking		
Verbal	None	Unintelligible	Inappropriate; Words	Confused; Sentences	Oriented	
Motor	None	Abnormal extension (decerebrate)	Abnormal flexion (decorticate)	Flexion withdrawal from pain	Purposeful localization to painful stimulus	Follows movement commands

Table 1. Scoring of the Glasgow Coma Scale.

Prognosis of DOC is difficult since there is a need to account for multiple clinical variables, but several generalizable trends have been documented. DOC caused by trauma has a better prognosis with regard to awakening than that resulting from non-traumatic (especially hypoxic-ischemic) causes (24). The longer the patient remains in a DOC state, the lower the probability of awakening: for example, once the patient reaches 3 months in non-traumatic or 12 months in traumatic VS, the Multi-Society Task Force on PVS (1994) puts the probability of awakening at <1% (24). There is evidence to suggest that MCS has more favourable outcomes in terms of survival, regaining consciousness, and recovering cognitive faculties than VS, for both traumatic and non-traumatic etiologies (25, 26), which stresses the importance of accurate diagnosis.

Neuroimaging assessments of covert cognition, such as those described in the present study, fall under the assessment criterion of “meaningful interaction with the

environment". Currently, this capacity is evaluated behaviourally. The ability to evaluate it an absence of overt behaviour would be an asset to DOC diagnosis.

1.2 Event-related potentials

Event-related potentials (ERPs) measurable at the scalp are the summation of a large number of post-synaptic neuronal potentials that occur in response to a specific event or stimulus (27). ERP investigations most often focus on visual or auditory methods of stimulation; in DOC specifically the stimulation is commonly auditory. There are several well-documented ERP components – patterns of activity in response to a particular type of stimulus – that are linked to a number of cognitive faculties, such as memory, attention, executive control, language comprehension, and others. A great advantage of ERPs is the ability to measure neural activity on the order of milliseconds. In fact, ERPs are most sensitive to processes that take 2 seconds or less (27).

ERPs have been documented in healthy adults for several decades. They've also been documented in children, older adults, individuals with neuropsychiatric disorders (see 28, 29 for a review), individuals with stroke (e.g. 30) and severe TBI (e.g. 31). While ERP components have been elicited using varying stimulus paradigms, guidelines have been put forth to standardize elicitation and detection methods (e.g. 29). Testing batteries combining multiple ERPs have also been described (e.g. 32). Here, a testing battery similar to that described in (32) is utilized to elicit several well-documented ERP components, whose general characteristics and reports of DOC elicitation are described separately in the following sections.

1.2.1 Mismatch Negativity (MMN)

MMN was first reported in 1978 by Risto Naatanen and colleagues (33). It is elicited in response to an auditory deviation in a stream of repeating auditory stimuli. Since its discovery, it has been replicated in numerous studies, and is interpreted to be related to short-term memory. It has shown good predictive power with regard to awakening from coma (1, 3), and has also been shown to have some prognostic value with regard to linguistic ability upon emergence (2). An interesting recent investigation showed that change in MMN presentation between testing sessions - specifically, improvement or deterioration of standard/deviant discrimination – accurately predicted survival in 21 of 30 DOC patients, with 100% positive predictive value (10). While MMN is consistently reported to have high specificity as a predictor of awakening (i.e. a consistently high percentage of those who exhibited MMN regained consciousness), there has not been a consensus regarding its sensitivity (i.e. of the total number of those who regained consciousness, an inconsistent percentage exhibited MMN). For example, while Kane et al. report 89.7% sensitivity (and 100% specificity) for predicting awakening in sub-GCS 8 coma (3), Fischer and colleagues report 31.6% sensitivity (and 90.9% specificity) (1). This presents a problem: if a patient shows MMN, it's possible to say with some confidence that awakening is likely; however, an absence of MMN adds little information to the prognosis.

This discrepancy in specificity can be explained by several factors. First, MMN may not be detectable in a percentage of healthy, unimpaired adults: while some studies do report 100% control elicitation of MMN (e.g. 34, n=15), others do not (e.g. 35, n = 17, 82% detection; both studies use pure tone stimuli). Second, reports of its test-

retest reliability for the same healthy adult individual have varied from 0.3 to 0.8 (35-37). While individual variation and test-retest reliability of MMN has not been studied extensively in impaired populations, there is some evidence for even greater variability than in healthy adults. For example, Lew et al report MMN test-retest reliability of TBI patients (initial GCS 3-12) who were fully recovered at the time of testing as 0.21 across two testing sessions 2-60 days apart (36). More recently, Mah et al report a cycling of MMN presence over a 48-hour period in a GCS 4 TBI patient, comatose at the time of testing (9). Third, MMN can be affected by factors such as levels of vigilance (e.g. disappearing at certain stages of sleep), age (children and elderly show less stable or diminished effects), and neurodegenerative conditions (e.g. diminished effects in Alzheimer's) (see 28 for a review); as well as sedatives such as propofol (38). This body of work implies that while there may well be individuals with a naturally undetectable MMN, a patient could exhibit a false negative depending on time and circumstances of testing.

Measures can be taken to reduce such false negatives. Detectability and reliability can be improved with the type of deviant - duration deviant has been reported to have the highest reliability (37, 39) and studies reporting 100% control elicitation use it as well (e.g. 34) - as well as shorter inter-stimulus interval (ISI) (40), greater number of trials, and processing techniques to maximize signal-to-noise ratio (35).

Altogether, this suggests that MMN stimuli must include a duration deviant, must be presented with a short ISI and in a large number of trials, recorded over a long period of time to account for potential cycling of vigilance, and looked at in combination with other measures if one wishes to accurately assess a clinical state.

1.2.2 The P300 effect (P300)

This effect was first reported by Sam Sutton and colleagues in 1965 (41). A number of theories have emerged with regard to exact cognitive faculties it engages (42), however, it is reliably elicited by an oddball paradigm, where a stream of standard stimuli is interrupted by a particularly salient deviant. It has therefore been described as a salience-orienting response (42). P300 and MMN are both deviant-related responses, but P300 is elicited by a deviant that is made particularly salient, either actively through a task (e.g. counting deviant tones), or passively by difference in acoustics (e.g. dog barks in a stream of tones), or emotional significance (e.g. subject's own name among other proper names). Two subcomponents of P300 have been identified, termed P3a and P3b (42). P3a is characterized by frontal activity and is thought to be driven by task-related attentional processing, while P3b activity is temporo-parietal and is tied to attention and memory (42). P3a and P3b are distinguished by latency as well as topography, with P3a typically peaking earlier. In DOC, however, topography and latency may be distorted. That complicates separating the subcomponents, therefore, here they will be discussed together as P300. The current study employs two P300 paradigms, aimed at eliciting each subcomponent.

P300 was first demonstrated in coma by Reuter and colleagues in 1989 (43), and has since been used extensively in investigating disorders of consciousness. In the active category, P300 has been successfully elicited in Locked-In Syndrome (LIS) with a task of counting deviant tones in a stream of standards, and has been proposed as a way to differentiate LIS from vegetative state (VS) and minimally conscious state (MCS) (44). Reports of task-based P300 elicitation in DOC proper are inconsistent. While the

effect has been reported in MCS and VS both, a number of studies report no difference between the two states (e.g. 4, 5; see also 44 for review). However, other studies report task-based P300 to be able to distinguish between DOC states with significance (e.g. 45). It's important to note, however, that (45) used subject's own name as the salient deviant, while (4) and (5) used words and tones, respectively.

Passive paradigms which include subject's own name (SON) have been used extensively in DOC. While some report different behaviors in MCS and VS, and LIS and DOC, (e.g. 46), others report the opposite (e.g. 47, no difference in response to SON in MCS vs VS). This paradigm's power in differentiating between DOC states proper has not yet been established. Its strength appears to be predicting awakening: some report it to have greater sensitivity and specificity than MMN (48). Another advantage of this passive paradigm is that it isn't predicated on the patient's ability to understand or follow instructions, as the name itself is salient enough to involuntarily capture attention, especially if spoken by a familiar voice (49, 50).

This phenomenon is susceptible to similar issues as MMN. Some healthy adults will fail to show a significant effect, especially in passive tone-only paradigms (44,48). It is also affected by conditions such as schizophrenia, some forms of dementia and mood disorders, and previous instances of TBI even if TBI did not result in hospitalization (see 29 for review). It is also interesting to note that DOC patients have been shown to have one of the MMN or P300 effects but not the other (48), prompting calls for examining both in the same population as a matter of regular practice (7).

The current study employs both an active and a passive P300 paradigm. The active, task-based paradigm (aimed at eliciting P3a) is potentially useful to distinguish

DOC from LIS, and the passive SON-based paradigm (aimed at eliciting P3b) has a potentially wider applicability. The task-based paradigm consists of of standard and deviant tones with instructions to count the deviants (similar to (44)). The passive SON-based paradigm is novel. It aims to maximize passive salience of the deviant by making it stand out through both acoustics and meaning, and consists of subject's own name spoken by a familiar voice among pure tones.

1.2.3 The N400 effect (N400)

This effect was first reported by Marta Kutas and Steven Hillyard in 1980 (51). It is linked to recognizing semantic incongruity, and, as such, requires the preservation of high-level linguistic abilities as well as working memory (52). It has been reported in comatose patients, however, its presence seems to depend on the etiology of coma and the site of injury if the cause is trauma-related. It has been reported absent in those with temporal lesions, and those with anoxic damage are consistently less responsive than those with vascular lesions (53-56). While it has not been evaluated as a means of prognosis in coma, it's shown some predictive power in recovery from UWS and MCS (57). Steppacher et al report N400 as a better predictor of recovery from UWS and MCS than P300, as well as having a strong correlation with long term (2-14 years) outcomes, making this effect a promising avenue to explore (57). This is intuitive - if the functions required for high-level linguistic processing are preserved, it is likely that there is less overall damage.

Several paradigms have been used to successfully elicit this effect in DOC. Word pairs (*bumble* + ***bee*** vs *bumble* + ***cat***) have been used (e.g. 54), as well as sentences (*I take my coffee with cream and* ***sugar*** vs *I take my coffee with cream and* ***socks***) (e.g.

57). It is interesting to note that DOC patients may show a response to one type of paradigm but not the other, with only a very small number responding to both (4). Further, N400 does not necessarily overlap with any of the previously described effects, and it is entirely possible to have a patient showing N400 but nothing else (4). N400 is affected by conditions such as Alzheimer's disease, focal seizure disorders, developmental disabilities, and language disorders (see 29 for a review). Also, as the other ERPs described above, N400 may fail to appear even in healthy adults, especially in passive listening (e.g. 58, 59). While the methodology of those investigations has been criticized for potentially lacking sufficient context strength of the stimuli to reliably elicit the effect, other works have not reported universal elicitation – for example, Mah & Connolly report elicitation in 4 of 13 healthy controls in a passive listening sentence paradigm (32). Finally, any subject tested must have sufficient proficiency in the language of presentation to understand full sentences and recognize abnormalities in meaning.

Nevertheless, successful elicitation in DOC, potential prognostic power, and non-overlap with other ERPs make N400 a useful addition to a testing battery. The current study employs a sentence-based paradigm, as those have been reported to elicit the greatest amplitude of the effect (52, 60), with instructions to take conscious note of incongruent sentences.

1.3 Concurrent Hierarchical Tracking (CHT) effect

This effect was reported in early 2016 by Nai Ding and colleagues (61). It is elicited by rhythmically presented phrases and sentences (but not similarly presented

lists of unconnected words, or phrases/sentences in a language unknown to the listener), and is hypothesized to track semantic and/or grammatical linguistic structure. Despite its recent discovery, it has been replicated in healthy adults using MEG (61), EEG (62-64), and ECog (61).

Due to its recent discovery, this effect has not been studied in DOC, and its behavior on an individual level is not well understood. The one study for which individual subject results are available reports the percentages of detection at sentence, phrase, and word levels as 63%, 44%, and 100% of subjects, respectively (63), despite all existing studies, including (63), reporting significant group-level effects. As with the ERP effects described above, it is reasonable to assume that certain conditions (e.g. language impairments) would diminish this response, and that there may be a percentage of healthy adults in whom the effect is not reliably detectable. Further, as with the N400 effect, a certain degree of proficiency in the language of presentation is necessary. Sleep has also been reported to disrupt this effect (62).

CHT presents interesting possibilities for coma work. It may quantifiably differentiate between phrasal and sentential levels of language comprehension, which can give a fine-grained picture of a patient's language abilities. Comprehension of simple phrases and sentences also requires a lesser linguistic proficiency than semantic manipulations (such as the N400 effect), which may make CHT useful for a multilingual population. The current study uses a replication of the paradigm in (64), which used monosyllabic words that combine into 2-word phrases and 4-word sentences.

1.4 Single subject significance

1.4.1 ERPs

DOC studies have employed several methods for determining the presence of an effect on an individual patient level, and the choice of detection method has generated discussion within the field (e.g. 65, 66). Visual inspection remains a very common approach, either to make a decision about the presence of an ERP or to select data for subsequent statistical testing. Statistical methods frequently are (or based on) a sample-by-sample t-test (e.g. 35, 38, 44, 46, 72), bootstrapping (e.g. 5, 48), or continuous wavelet transform (e.g. 56, 57). The choice of detection method is not a trivial one, since it has been shown to impact results, which may in turn explain the sensitivity/specificity discrepancies in existing reports. Gabriel et al have compared the performance of several methods in detecting MMN in a set of 27 healthy controls, including sample-by-sample t-test, cross-correlation, continuous wavelet transform, multivariate analysis, and visual inspection alone (65). They found that continuous wavelet transform was the only method to detect MMN in 100% of their subjects, and rates of detection ranged from 100% to 44% among the 6 methods investigated (65). Their findings raise an important consideration for DOC analysis: method matters.

The most straightforward way to determine which method is “best” is to compare its results to the objective truth. If we take as objective truth that every normal-hearing, conscious, neurotypical person must exhibit auditory ERPs, the “best” method is the one with the highest rate of detection. However, it is not a given that the true ERP prevalence among healthy populations is 100%. Investigating the rates of false positives - for example, testing the presence of ERPs in absence of stimuli (such as recording

from a subject placed in a quiet room) - will go a long way towards shedding light on the possibility of over-detection. Moving away from a binary yes/no judgment regarding effect presence would eliminate the issue altogether. DOC deviation from the norms could be quantified instead (e.g. number of standard deviations from mean amplitude) and used in evaluation. That would require establishing norms for controls matched for age and, if applicable, non-DOC medical history. It would be impractical to collect a control sample for every DOC patient tested, so it may be beneficial to establish population norms based on age groups, and quantify the effects of non-DOC conditions.

For the purposes of DOC, the most valuable detection method is the one that gives the best prognostic value. Such a method has not yet been established. One way to expedite achieving that goal is to re-analyze large datasets with several methods to determine which set of results is the most convergent with the eventual outcome. The resources required for such an undertaking are not trivial, and would require a collaborative approach to analysis.

The analysis reported in the present case study is a repeated-measures ANOVA. As with any method, there are advantages (consideration for multiple testing of the same patient) and drawbacks (dimensionality reduction by averaging across a time window and multiple electrodes). Bearing the above discussion in mind, this dataset should be available for re-analysis with other methods.

Several features are common to DOC data analysis irrespective of the detection method ultimately selected. DOC patients can exhibit ERPs that are atypical, yet still significant. For example, Kane et al report a “healthy-like” (i.e. comparable with healthy volunteers in the same study) response in only 2 of 53 patients (3). While ERPs have

defined latency windows, numerous studies employed expanded time windows when determining ERP presence in DOC (e.g. 1, 48). Significance is evaluated in a small subset of electrodes, commonly Fz, Cz, and Pz (e.g. 44, 46, 48), or even one electrode (e.g. 1 (Fz), 3 (Pz)), likely because the recordings are commonly done with a small electrode set. Minimal amplitude thresholds are usually not set. If a significance test is performed on every sample or at every timepoint, significance must be maintained over a continuous time window (e.g. 35, 46, 48).

1.4.1 CHT

CHT effect is measured as the EEG response power at target frequencies, which correspond to the presentation rate of words, phrases, and sentences. The only currently available investigation to report subject-level results employed local spectral F-test to evaluate the response at target frequencies (63). That investigation relied on a single-session design, whereas in the patient case study described here multi-session repeated measures were taken from the same participant. The results of both repeated measures ANOVA and local spectral F-test are presented here. The latter followed the implementation described by the original authors (63), with one important modification. The null hypothesis for this analysis is taken to be that the EEG response power at frequencies of interest is the same as that at other frequencies. Ding et al make the assumption that the normalized power at non-target frequencies follows a chance distribution, and estimate it by pooling the power over all non-target frequencies over all the subjects. Therefore, the significance test compares an individual subject's response power at a target frequency to the pooled power over all non-target frequencies over all subjects.

However, it is not a given that normalized response power follows chance distribution for every subject, even after the data is detrended and filtered. Testing conditions and individual physiological differences may affect the EEG response, and such factors would affect both the target and non-target frequencies. Therefore, it is more conservative to keep an entirely within-subject design, and compare each individual's response at target frequencies only to their own response at non-target frequencies. Further, Ding et al. pool the non-target response power over all non-target frequencies, despite applying a 0.1-25Hz filter to the original data. This undermines their own assumption of non-target response power being distributed at chance. It is possible that they do limit the range of non-target frequencies, and simply omitted such a statement from the description of methods. Regardless, the current study limits the non-target frequency range to those not affected by the bandpass filter.

2. Methods

2.1 Participants

2.1.1 Patient

A 28 year old female patient was recruited from the intensive care unit (ICU) at the Hamilton General Hospital following a motor vehicle accident. The patient suffered a C4/C5 vertebrae fracture and severe spinal cord injury, which resulted in quadriplegia. Neurological injuries included hypoxic-ischemic injury and bilateral vertebral artery dissection in addition to Grade III diffuse axonal injury (brainstem). The patient initially

scored 3 on the GCS assessment. GCS score was assessed daily by the ICU bedside nurse.

The patient was selected due to the low GCS score (<8) in absence of sedatives. Prior medical history was not available. Consent was solicited from the patient's father. The recording sessions took place 11 and 18 days after initial admission. The patient exhibited a GCS score of 3 (no eye, verbal, or motor response) at the first session, and GCS 6 (spontaneous eye opening, grimacing to pain, and no verbal response) at the second, both assessed by the bedside nurse.

Following the recording sessions, the GCS score fluctuated between 4 and 6. The patient was moved between ICU and step-down unit several times, and stabilized enough to remain in the step-down unit 4 months after initial admission. The outcome diagnosis was vegetative state, which persisted until the patient's death from complications of septic shock 12 months after initial admission.

The study protocol for controls and patients was approved by the Hamilton Integrated Research Ethics Board.

2.1.2 Controls

The purpose of recruiting a sample of healthy controls was to provide a working dataset which could serve as an immediate reference for the processing, analysis, and interpretation of the data collected from the patient. Healthy young adults were recruited from the undergraduate population at McMaster University. In total, 18 unique participants are included in the present report (92% female), mean age 20.3 years. A subset of 6 participants was tested on each of the 6 paradigms. All were native speakers of English, right-handed as assessed by the Edinburgh Handedness Inventory

(67), with no history of neurological or hearing disorders as self-reported on a screening questionnaire.

2.2 Stimuli

All auditory stimuli were normalized to 80dB SPL (except intensity deviants as described below), and were presented binaurally through earbud inserts. All experiments were created and presented using Presentation software (68). Speech synthesis for the CHT paradigms was done using Neospeech software (69). Natural speech was recorded from a single native English speaking female volunteer. A short recorded introduction was included at the beginning of testing, and each paradigm was preceded by its own pre-recorded instructions as described below.

2.2.1 MMN

MMN was elicited using an oddball paradigm described in (70). This paradigm included 2400 tones at a regular 450-msec stimulus onset asynchrony (SOA), with 82% standard tones (50ms, 1000 Hz) and three types of deviant tone (6% each): a duration deviant (125ms), a frequency deviant (1200Hz), and an intensity deviant (+10 dB SPL) (70). It was preceded by a recorded instruction to listen passively.

2.2.2 P300

P300 was elicited using two paradigms. The first was a modification of the SON-based paradigm described in (48). The modified paradigm included 2500 stimuli, with 96% standard tones (50ms, 1000 Hz) and 4% SON stimulus, all presented at 1000ms SOA. The patient's name was recorded as spoken by her father using a Voice Recorder

application on a cellular phone, and normalized to 80 dB SPL via Praat software. This was not logistically feasible for controls, therefore, their names were digitally synthesized using Neospeech software. Pronunciation accuracy was confirmed with the participant prior to testing. The SON-based paradigm included instructions to count the occurrences of SON, since it's been reported that active attention to the SON stimulus improved P300 elicitation (e.g. 32).

The second paradigm was an exact replication of the MMN paradigm described above, with an addition of verbal instructions to count the deviants.

2.2.3 N400

This effect was elicited by auditory presentation of sentences ending in a congruent or incongruent word, (e.g. *I take my coffee with cream and* ***sugar/socks****.*) 160 sentences with highest Cloze probabilities (0.99-0.89) as reported in (71) were selected. Of those, 80 were kept intact, and 80 had their final words replaced with something obviously incongruent. They were recorded for auditory presentation by a female native speaker of English, and presented in a randomized order with 1200-1400ms jittered interval between them. This paradigm was preceded by instructions to take active note of sentences that were odd in meaning.

2.2.4 CHT

The paradigm reported in (61) and (64) was replicated. 60 four-word sentences with the structure [adjective/pronoun – noun – verb – noun] (e.g. *red cat climbs trees*) were synthesized via Neospeech with an American English female voice. The words were presented in equal intervals of 320ms with no acoustic gaps between sentences

and no prosodic cues. Sentences were presented in 30 randomized blocks of 12 with 1000ms interval between blocks.

A reduced speed version of this paradigm was also presented, with words presented every 520ms. (61) and (64) included an active button-press task to ensure attention to stimuli. To accommodate DOC patients, in the current study the task was replaced with instructions to count sentences.

2.3 Electrophysiological Recording

2.3.1 Patient

Recordings were made at 11 days and 18 days after admission, at the patient's bedside in the ICU, over a 24-hour period with breaks for patient care as needed. The stimuli were presented in 2.5 hour blocks, structured as follows: 10 minutes of rest + MMN + CHT original speed + N400 + 10 minutes of rest + P300-task + CHT reduced speed + P300-SON, with 1 minute of silence between the paradigms (see Figure 1). The patient remained positioned horizontally throughout. 15.5 hours of data were successfully recorded in the first session, and 19 hours in the second.

The data was collected using a BioSemi Active Two system (Ag/AgCl electrodes), using 16 electrodes in the 10-20 configuration: F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, Oz. in addition to three external reference electrodes (placed on the tip of the nose, and both mastoid bones). Electrooculographic activity was recorded from electrodes at the outer canthus of the left eye, and the forehead above the left eye (see Figure 1). The signal was digitized at 1024hz sampling rate, with online bandpass filtering at 0.1Hz-100Hz.

2.3.2 Controls

Testing was conducted at the Language, Memory and Brain Lab at McMaster University. As with the patient, auditory stimuli were delivered binaurally using Presentation Software, and the EEG signal was digitized at 1024Hz sampling rate and 0.1Hz-100Hz bandpass filtering. Participants were seated in front of a computer screen displaying a fixation cross. Continuous EEG recordings were made with a 64-channel system (BioSemi ActiveTwo, Ag/AgCl electrodes) according to the International 10-20 system. Activity was also recorded at the same five reference electrodes as with the patient. Due to time limitations imposed on working with undergraduate participants, they were tested on subsections of the full paradigm, as follows: 1) MMN + P300-task, 2) P300-SON, and 3) N400 + CHT (both versions), with each recording session lasting approximately 1 hour. Controls received exactly the same instructions as the patient, with no additional input from the experimenters.

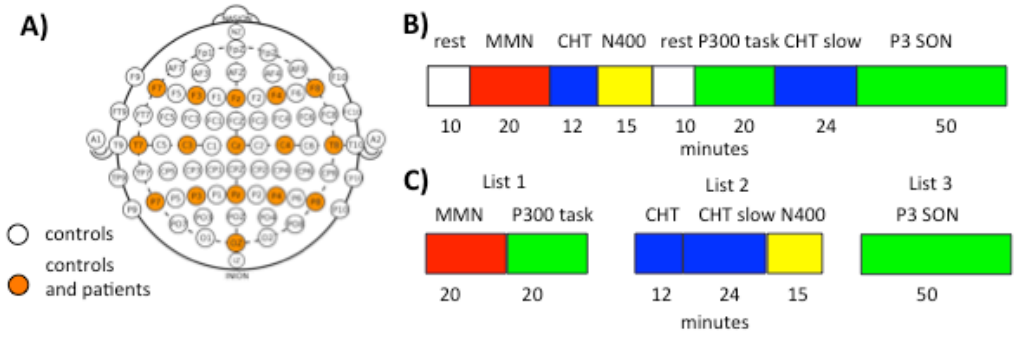

Figure 1. A) Electrode placement in controls and patient. B) Full testing battery administered to patient; total time 2h 41min. C) Paradigms as administered to controls.

2.4 Data Processing

2.4.1 ERPs

The data was filtered offline with a bandpass of 0.1-10Hz, with an 8th order zero-phase Butterworth IIR filter. Ocular artifacts were removed through ICA decomposition. Components whose amplitude was 10 or more times higher over frontal than posterior sites were discarded. The external reference electrodes frequently became dislodged during the first recording session (GCS 3), therefore the data from that session was referenced to a Global reference. The Global reference (computed as an average of all scalp electrodes), however, is not an ideal choice for low-density recordings, as undersampling of scalp activity can result in a biased average (73). Therefore, data from the second (GCS 6) session was referenced to a summation of the two mastoid references. Mastoid references were used for control data. The recording was segmented as follows. MMN and P300-task segments began 100ms before stimulus onset and ended 500ms after. N400 and P300-SON segments began 100ms before the stimulus and ended 1000ms after. All segments were baselined to the 100ms pre-stimulus interval. A local DC detrend was performed. Segments containing artifacts were rejected through an automatic search with the following parameters: voltage steps of 50uV/ms, maximum allowable difference of 200uV within a 200ms interval. In total, 1.1% of data was rejected across all controls, 3.4% for patient session 1, and 1.1% for patient session 2.

2.4.2 CHT

Data processing followed the procedure described in (63). Data was filtered 0.1-25Hz. Ocular artifact rejection and referencing was the same as for the ERPs. The first sentence of each trial was omitted during segmentation. An average of trials was computed for each participant, and response was pooled over all channels. Power spectrum of the pooled average was computed using DFT, with a resolution of 0.017Hz. The power spectrum was then normalized in order to remove a 1/f trend usually present in EEG data. Each frequency bin was normalized by surrounding 1Hz (0.5Hz on each side).

2.5 Statistical Analysis and Significance Determination

2.5.1 ERPs

ERP analysis generally followed the procedure described for a similar testing paradigm by Mah & Connolly (32). ERP amplitudes were evaluated within a topographic region of interest (ROI) selected through visual inspection to be best representative of activity (ROIs listed in Table 2 and illustrated in Fig. 2 and 3). ROIs were defined as per the NeuroElectroMagnetic Onthologies Consortium definitions (74). The ROIs for the controls and patient were evaluated separately.

Within each ROI, mean amplitude was computed within a 50ms window centered on the ERP peak. While traditionally, a winder window of 100 or 200ms might be used, the selection of a 50ms window here was motivated by the patient's response, whose ERP peaks appeared narrower than that of controls. The peak width was then kept consistent between controls and the patient. For controls, the peak was detected as the

point of greatest amplitude in an ERP-canonical time window: 100-250ms post-stimulus for MMN, 250-400ms for P300, and 300-500ms for N400. ERP search windows are often extended in DOC patients, as their injury may impede stimulus processing (e.g. 48). Here, any latency beginning with that of the control time window was considered. For the patient, the ERP amplitude was evaluated using repeated measures ANOVA, using *Test Run* (6 levels for GCS 3, 7 for GCS 6) and *Condition* (2 levels) as factors. Significant main effects and interaction effects was followed with Tukey's Honestly Significant Difference test to determine which test runs showed a significant difference between conditions. For controls, the difference was evaluated in each participant separately using a two-tailed Welch's t-test.

	MMN	**P300-task**	**P300-SON**	**N400**
Controls	Mid-frontal	Mid-central	Mid-central	Mid-central
PT-GCS3	Left-frontoparietal Left-centroparietal	Left-frontoparietal	Mid-frontal	Mid-central
PT-GCS6	Right-frontoparietal	Left-frontal	Mid-frontal Left-central	Left-central

Table 2. Topographic regions of interest in controls and patient.

2.5.2 CHT

Repeated measures ANOVA was used to evaluate difference between target and non-target frequencies, with *Run* (6 or 7 levels) and *Frequency* (4 levels: non-target, sentence, phrase, word) as factors, with Tukey HSD followup of significant main effects or interactions.

To remain faithful to the methodology of the only currently existing work to investigate this effect on an individual level (63), the results of local spectral F-test (F(1,1404)) (75) are also presented, comparing the response power at three target frequencies (sentence, phrase, word; 0.78Hz, 1.56Hz, 3.12Hz (original speed) or 0.48Hz, 0.96Hz, 1.92Hz (reduced speed), respectively) to pooled power over non-target frequencies in the range of 0.1-25Hz, for each individual subject and each individual patient test run.

3. Results

3.1 Controls

3.1.1 ERPs

All individual controls showed an expected statistically significant response to the Duration deviant in the MMN paradigm (individual Welch's t-test statistics: t between 3.8 and 11.7, all $p<0.01$). Five of six showed a significant response to the other MMN deviants, all P300-task deviants, and P300-SON (all $p<0.05$). Four of six showed a significant N400 effect (all $p<0.05$), with one nearing significance at $t(61)=1.8$, $p=0.07$. P300-SON effect had the greatest average amplitude, and Duration deviant elicited the largest MMN and P300-task responses.

Paradigm	Individual presence	Mean amplitude (uV)
MMN Duration	6/6	-5.43(1.19)
MMN Frequency	5/6	-3.21(1.72)
MMN Intensity	5/6	-2.80(3.22)
P300 Duration	5/6	14.32(5.09)
P300 Frequency	5/6	11.25(5.05)
P300 Intensity	5/6	7.27(3.89)
P300 SON	5/6	15.16(4.39)
N400	4/6	-3.52(1.56)

Table 3. Individual prevalence of ERP effects and mean amplitudes within the selected ROI.

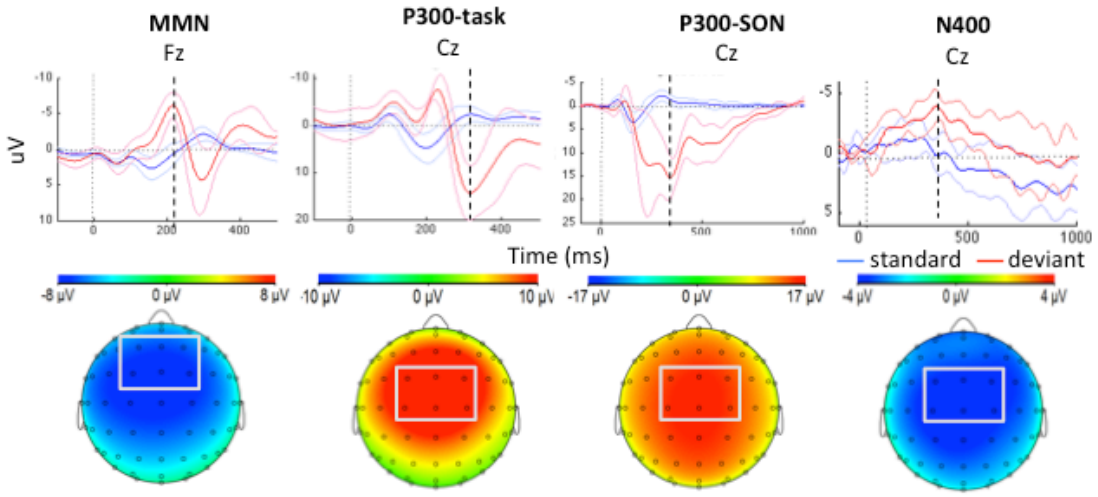

Figure 2. Control ERP results. The ROIs selected for analysis are outlined. The labels Standard and Deviant refer to the following (respectively). MMN, P300-task: standard tones, duration deviant. P300-SON: standard tones, SON deviant. N400: congruent sentences, incongruent sentences.

3.1.2 CHT

CHT effect was considered to be present if the response power at frequencies of interest differed from that at surrounding frequencies, as evaluated by local spectral F-test ($F(1,1404)$, $p<0.05$, FDR-corrected). All participants showed a word-level effect at reduced presentation speed, and 4 of 6 at original presentation speed. Phrase and sentence effects were present for 1 of 6 participants at each presentation speed. At reduced speed, one more participant showed phrase and sentence effect at $p=0.09$.

	Paradigm	Individual presence	Power (|fft|)
CHT-original	**Word**	4/6	0.0558 (0.0245)
	Phrase	1/6	0.0150 (0.0130)
	Sentence	1/6	0.0267 (0.0157)
CHT-slow	**Word**	6/6	0.0493 (0.0217)
	Phrase	1/6	0.0503 (0.0196)
	Sentence	1/6	0.1376 (0.0322)

Table 4. Individual prevalence of CHT effect and mean power pooled over all channels.

3.2 Patient

3.2.1 ERPs

A repeated measures ANOVA was conducted separately for each testing session, with factors *Run* (levels: 1-6 or 1-7) and *Condition* (levels: Standard, Deviant). At GCS 3, a main effect of *Condition* were observed for MMN-Intensity ($F(1,12574) = 4.18$, $p<0.05$) and P300-SON ($F(1,11429) = 7.78$, $p<0.05$). However, followup Tukey HSD comparisons did not show any significant condition differences within individual

testing runs. At GCS 6, a main effect of *Condition* was observed for P300-SON ($F(1,17352) = 18.59$, $p<0.001$). Followup Tukey HSD comparisons showed a significant ($p<0.05$) condition difference during Runs 1, 2, 3, 4, and 7. The SON results are illustrated in Figure 3. The topographic distribution of the response was more frontal than might be expected from canonical P3b.

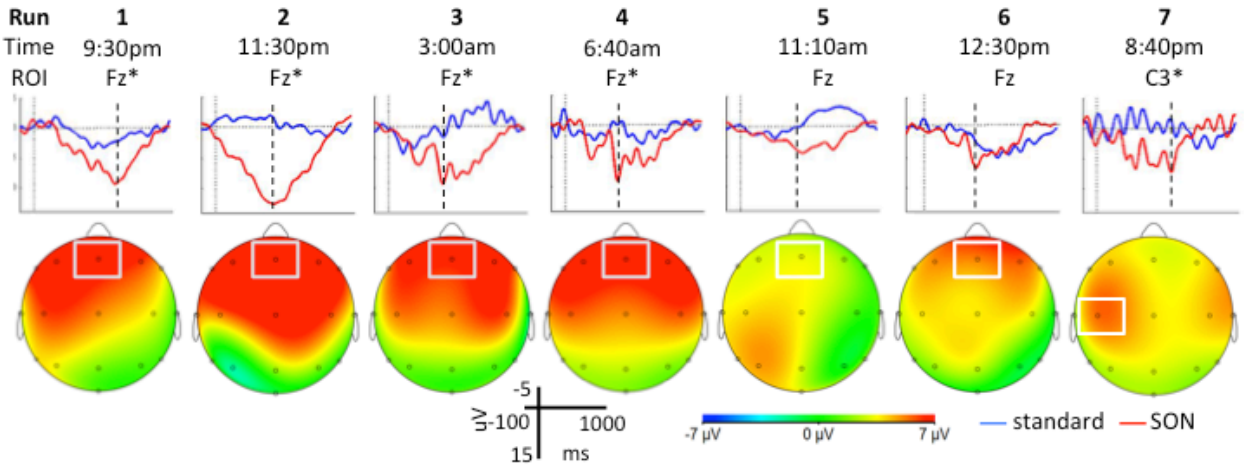

Figure 3. Patient's P300 responses to the SON deviant. ROIs selected for analysis are outlined. Significance at $p<0.05$ denoted by *.

3.2.2 CHT

The results of the repeated measures ANOVA and local spectral F-test were generally in agreement. At GCS 3, a main effect of *Frequency* was observed at original presentation speed as determined by ANOVA ($F(3, 720) = 8.04$, $p<0.01$). Tukey HSD followup showed a significant difference between non-target frequencies and the sentence condition during Run 1 ($p<0.05$) , but the absence of word- or phrase-level effects complicates interpretation. Local spectral F-test did not evaluate any of the effects seen at GCS 3 as significant.

Patient's GCS 6 session showed a significant main effect of *Frequency* ($F(3, 840) = 2.75$, $p<0.05$) and a significant *Run* Frequency* interaction ($F(18, 840) = 2.59$, $p<0.01$) at original presentation speed. Tukey HSD followup revealed a significant difference between non-target frequencies and word contition during Run 2 ($p<0.01$). Local spectral F-test showed a significant effect at word level at original presentation speed at Runs 1 and 2 ($F(1, 1404) = 4.35$ and 6.06 respectively, $p<0.05$ for both). At reduced speed, a *Run*Condition* interaction was significant ($F(18, 812) = 6.66$, $p<0.01$), and Tukey HSD followup revealed a word-level effect at Run 2 ($p<0.01$). Local spectral F-test did not return any significant effects for GCS 6 at reduced presentation speed.

The power at frequencies of interest, where the effect was present, was comparable to the control average for the original speed, but less than the control average for reduced speed. The results are illustrated in Figure 4.

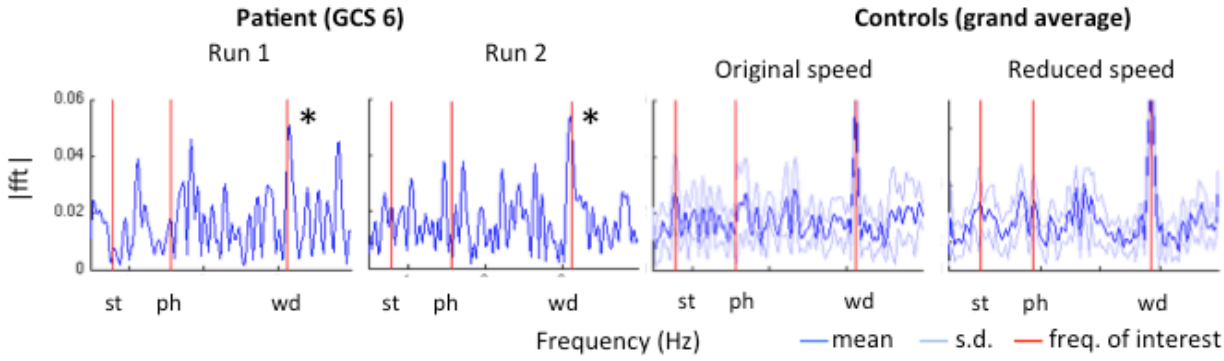

Figure 4. Concurrent Hierarchical Tracking results for the patient and controls. Frequencies of interest: st – sentence, ph – phrase, wd – word. * = significance at $p<0.05$.

4. Discussion

The present study was undertaken with a goal of improving upon current methods of EEG detection of covert cognition. Based on the results of both the patient case study and the control investigation, several such improvements are feasible. In addition, some light can be shed on the inconsistencies among existing reports. The results also show that clinical applications of ERP paradigms would benefit from establishing normative parameters, such as a range of healthy latencies and amplitudes, to assist with interpreting patient results. Finally, it would appear that even with a battery that addresses several methodological concerns in the current testing practices, making a prognosis based on ERP results is not straightforward. This is in part due to the difficulty of relating the results of 24-hour testing to single-session results, which comprise the majority of existing work. The testing battery described here proposes novel elements - multi-hour testing of multiple ERPs - and, with continuing data collection, may prove more accurate in DOC prognosis and diagnosis than single-session testing.

4.1 Paradigm Performance

All ERP paradigms showed expected results for the majority of controls. MMN results varied by type of deviant, with duration eliciting the largest amplitude of response. This is consistent with previous investigations, suggesting that duration is an inherently more salient characteristic than sound intensity or frequency. Therefore, to minimize testing time and resource use, it may be beneficial to use only a duration deviant in clinical applications of this paradigm. The response elicited to the task-based

P300 paradigm was more centrally distributed than one might expect of a canonical P3a. However, it was still more frontal than the response to the SON-based paradigm, preserving the difference reported in previous studies. P300-task response did not vary by type of deviant, with duration again eliciting the largest amplitude of response, again suggesting reducing redundancy by choosing only one deviant type in clinical applications.

A P300 response to SON was reliably elicited in 5 of 6 controls, and was the only effect to reach statistical significance for the patient. The control responses differed from those to the task-based paradigm in latency and topographical distribution, showing that despite the presence of task to count occurrences of SON, the response was not P3a-like. The paradigm used in the current study has not been previously reported in DOC or healthy control literature, so it is difficult to draw general conclusions from a small sample size. However, if this trend holds for larger subject samples, this paradigm can prove very useful. It combines several types of passive salience: acoustic, due to length and complexity of the SON stimulus as compared to simple tones; conceptual, since a word with meaning is contrasted against tones; and emotional, since one's own name is highly recognizable, especially if spoken by a familiar voice. Additionally, the SON is made actively salient by the instruction to count its occurrences. This combination of active and passive saliency is versatile enough to be used with multiple levels of awareness, and, unlike other language-based paradigms, does not depend on language proficiency or comprehension. The drawback, however, is the difficulty in identifying the exact aspect of the paradigm which is eliciting a response.

The N400 paradigm produced the expected results in the majority of controls, peaking at 360ms post target word. The difference between conditions persisted until at least 1000ms post target word.

The CHT effect was robust at the word level for single subjects, but not at the phrase or sentence levels. This is somewhat consistent with the only existing study where the results are reported for single subjects, which reports sentence, phrase, and word effects in 63%, 44%, and 100% of subjects, respectively (63). The current study showed effects in a lesser proportion of subjects (17%, 17%, and 67% at the original presentation speed, and 17%, 17%, and 100% at reduced speed). One possibility is that this is due to a methodological choice in significance determination. Ding et al compared single-subject frequencies of interest to non-target frequencies pooled across all frequencies and all subjects, while the current study keeps a purely within-subject design. This possibility was investigated by replicating the analysis described in (63). While the *p*-values resulting from that pooled-subject frequency comparison were smaller, the difference was negligible for the purposes of establishing significance. The proportion of subjects exhibiting significant effects remained the same. Therefore, it appears that both pooled and single subject non-target frequency distributions are very similar for the purpose of approximating a chance distribution. Employing pooled-subject frequency analysis would at best contribute to marginal significance.

Curiously, group-level effects are present in the Ding et al study at all three levels despite inconsistent single-subject response. This may be due to a low number of trials (n=22 in (63), n=30 here). Averaging over a greater number of trials attenuates the noise at non-target frequencies while preserving the signal at target frequencies (if it is

indeed present), which is what may be responsible for group-level effects despite incomplete subject-level response. Before considering clinical use, subject-level findings should be investigated further, including an optimal number of trials and appropriate significance detection.

4.2 Normative Data

In DOC literature, detecting single-subject ERP presence carries a measure of subjectivity in selecting latency, amplitude, peak width, and topographical distribution within which an ERP could be deemed plausibly present. This is not inherently problematic - after all, clinical evaluations and diagnoses are made subjectively by physicians based on the information available. However, even experienced researchers can have opinions inconsistent with one another, and detection parameters vary in the current DOC literature and can widely deviate from those used for healthy adults. For example, for P300, a phenomenon reported to peak no later than 600ms post-stimulus in healthy controls (76), Fischer et al use a search window of 250-1000ms post-stimulus for DOC patients (48). That time window would encompass several known ERP effects observed in healthy subjects - P300, N400, and P600. There are also commonly no minimum amplitude requirements for an effect to count. Statistical tests provide a more objective measure of effect presence than visual inspection alone, but many of them still require an *a priori* decision about time windows or topographical regions of interest. Even with less subjective approaches, such as automated pattern recognition (e.g. machine learning), there is still a need for a normative pattern (e.g. “training set”) to which the patient data would be compared. Multiple studies report age, alertness,

attention, past non-DOC causing injury (e.g. mTBI), and neurological and psychiatric conditions (e.g. Alzheimer's, schizophrenia) as factors influencing latency and amplitude. Finally, the actual prevalence of detectable ERPs among healthy individuals will greatly influence their clinical utility and the interpretation of DOC results. All of the above strongly suggests that investigating ERPs among the general non-DOC population is an important step in using them in DOC work.

Typical ranges for ERP latencies are often reported as common knowledge in the field, but have never been systematically established in the general population. The reports of ERP amplitudes vary, and no minimal threshold currently exists. Minimal peak width, that is, how long the difference between conditions should minimally last, is also not established. Below is a survey of the parameters reported in healthy adults, which shows that the degree of consensus regarding them varies by paradigm.

For MMN, the detection window has been reported ranging from as early as 75ms (1) to as late as 250ms (78) post-stimulus, with duration deviants peaking later (52). The amplitude of the difference between the standard and the deviant has been reported as low as 1.5uV (40) and as high as 5.6uV (79), with a number of reports in the 2-3uV range (1, 38, 40, 52). For P300, the detection window has been given as early as 180ms post stimulus (80) and lasting until as late as 600ms (76), and reported amplitude has ranged from 2.1uV (80) to 15.7uV (81). Tone-based stimuli appear to elicit earlier and more modest responses than SON-based stimuli. Raw millivolt values for auditory N400 are less frequently reported, but are commonly under 5uV, and reported detection windows start as early as 200ms and end as late as 1000ms post stimulus (58, 60, 83, 84).

Establishing population norms and quantifying how common non-DOC conditions affect ERP responses is important for interpreting patient results, and is the most logical and non-controversial way to establish standardized ERP detection parameters. These norms must be established with standardized paradigms (e.g. 29). In establishing of these norms, it will also become apparent if it's feasible to implement paradigms and analyses with near universal detection ability - or what percentage of people without DOC truly lack these responses. In contrast to a binary presence/absence judgment, patient results can then be interpreted according to their difference from the population norms, e.g. the number of standard deviations from the population means.

4.3 Patient Case Study

When discussing the patient's results, the considerations are twofold: obtaining information relevant to prognosis, and characterizing the patient's state at the time of testing. ERPs describe here index faculties that fall in the broad neurobehavioral category of "meaningful interaction with the environment", though the extent of that interaction varies by ERP. Behaviour in this category distinguishes VS from MCS (19, 20), which is relevant for diagnosis and prognosis both; ERPs themselves are also reported to carry prognostically relevant information. Irrespective of prognosis, characterizing the patient's abilities and awareness may impact the behaviour of their family and caregivers.

The patient described in the present study suffered a hypoxic-ischemic neurological injury in addition to diffuse axonal injury, with post-coma outcome of VS followed by death 12 months after initial admission. No significant ERP response was

observed at either GCS 3 or GCS 6, except a P300 component to SON at GCS 6. The absences are not surprising: ERP abolition is often correlated with a poor outcome, and is particularly expected if the coma etiology includes anoxia (see 6 for review). The distribution of the P300 component was frontal, with greater expression in the left hemisphere. The frontal distribution is more suggestive of a P3a, rather than P3b, subcomponent, which is tied to orienting and novelty rather than recognition proper. However, the P300-task paradigm was explicitly aimed at producing a P3a-like response, and the fact that such a response was absent suggests that there is a property of the P300-SON paradigm that is key for elicitation. That property may be emotional salience of a familiar voice, the degree of difference between standard and deviant stimulus, or indeed true recognition. Future investigations into this paradigm may provide clarification.

A P300 response to SON has been demonstrated in DOC of various etiologies (e.g. 47). In that investigation, 2 of 18 anoxic DOC patients demonstrated a globally central novelty P3 response to SON, one diagnosed as MCS, other VS. Of 9 non-anoxic (TBI, stroke, encephalitis) DOC patients, 5 showed the novelty P3 response, and 2 (one VS, one MCS) of those 5 in turn showed an additional late parietal component that the authors interpret at P3b (47). The time of testing ranged from 8 to 59 months after coma onset (47). Therefore, there is predecent for observing P300-SON in long-term VS. There is also precedent for observing P300 in absence of other ERPs (e.g. 4, 48). In these respects, the patient's results are consistent with existing research.

How do these results reflect the patient's outcome? A challenge to making a prognosis on the basis of multi-session multi-ERP 24-hour testing is the need to

interpret both effect presence and absence. Here, P300-SON was present during one of the two testing sessions, while MMN, P300-task, and N400 were entirely absent. Comparatively little work is available in DOC on multi-ERP testing batteries such as the one described here, but the results that are reported show that ERPs differ in their predictive power. For example, Kotchoubey and colleagues used multiple paradigms aiming to elicit MMN, P300, and N400 in a group of 98 DOC patients, and discovered that only MMN correlated with outcomes 6 months post testing (4). On the other hand, novelty P300 to SON has been found to have good sensitivity and specificity (0.71 and 0.85, respectively) as a predictor of awakening and good predictive value (0.81 positive, 0.76 negative); all except specificity higher than MMN (48). Here, however, it is the absence of ERPs that seems to be more reflective of the eventual outcome – if we make the (reasonable) assumption that less typical results reflect worse pathology.

What information can we infer from these results about the patient's cognition? The SON paradigm used here employs several levels of salience: acoustic, linguistic, and emotional. There is no converging evidence for the patient's ability to perform tasks, so any effect is likely elicited passively, due to salience or novelty. It has been argued that frontal novelty P300 is an orienting response, while a more posterior P3b-like component is tied to salience and recognition (47). If the patient's response seen here is interpreted as a frontal, novelty-based component, then it is evidence, at the least, for her ability to distinguish acoustic differences and exercise short-term memory to store the common and uncommon stimuli. However, topography-based distinctions in patients with brain damage may not be reliable, therefore, it's not possible to be certain how much emotional salience contributed to the observed response.

The idea that a DOC patient's vigilance fluctuates with time is rather intuitive - after all, this holds true for healthy individuals throughout a 24-hour cycle. While this is a new area in DOC research, evidence is emerging that it's possible to detect these fluctuations with EEG paradigms (e.g. 9). In the present case study, P300-SON showed a pattern of disappearance and reappearance. At GCS 6, it was present for every run except the time window of 10:00am-4:00pm. The latency of the peak fluctuated, with longer latency at 4:30am and 4:15pm. Longer latency implies longer processing time, which could be brought about by diminished awareness or fatigue. The effect was most prominent in the mid-frontal region for runs 1-4, and left-central region for run 7. Changes in the region of prominence are not without precedent – for example, this is true for the MMN response described in the case study in (9).

Together with other emerging evidence, the results of the patient case study strongly suggest that testing DOC patients at multiple timepoints must be investigated further. Fluctuating detectability of ERPs may explain inconsistent reports of their prognostic power in existing literature. A 24-hour period appears to be sufficient to observe response cycling, but this should be confirmed by future investigations.

4.4 Clinical Utility

The current investigation was undertaken with the aim of contributing to the development of a clinically useful tool. What are the features of such a tool? Several, partially based on Cruse et al's discussions of neuroimaging assessments of covert consciousness (85), are discussed below as pertains to the testing battery presented in the current study.

1. What is the true positive rate for healthy controls? Are false positives possible?

As discussed above, none of the effects included in the battery are universally present in single subjects, despite these subjects being, by all accounts, healthy neurotypical adults conscious at the time of testing. True and false positives are important for interpreting patient results: if we observe an ERP, how confident are we that the cognitive operation in question is truly taking place?

The ERP detection rates for healthy young adults are generally high, though data is not readily available for all ERPs. EEG MMN has been reported in 80%-100% of participants (1, 34). P300 in response to tones has been reported in 100% (44; active task) to 78% (44; no task) of participants. P300 elicitation with linguistic stimuli has been reported in the 90%+ range (86, 87), though single-subject data for mixed SON/tone paradigms such as the one discussed here is not readily available. The variance in existing reports can be addressed by standardizing elicitation paradigms and procedures for determining single subject significance. More importantly, however, is investigating ERPs in a diverse population. It has been shown that ERPs are affected by factors such as age and non-DOC medical conditions. Future work should establish detection rates for, at the least, several age groups, to determine if ERP testing is appropriate for all segments of the patient population.

Is it possible for these effects to be seen without the requisite process to have taken place (i.e. false positives)? This question has not been directly addressed by the ERP literature. It would be beneficial to evaluate ERP paradigm performance in those who by all accounts should be unable to produce a response (e.g. deaf subjects,

unequivocally dead subjects), or simply attempt detecting auditory ERPs in absence of stimuli (e.g. while subject sits in a quiet room).

2. What are the true and false negative rates?

This is an important consideration in patient care, because we must be confident in interpreting a null result. If we fail to elicit an ERP, is it absent due to brain damage or does the patient belong to that segment of the population that does not produce a detectable ERP?

How many of those who don't show a response are truly incapable of producing one (true negative)? The only way to evaluate the true negative rate is to test ERP paradigms with those who, by other available measures, cannot be expected to respond to a given paradigm (e.g. death, deafness, lack of language skills if the paradigm is language-based). Some language-based work has been done - for example, the CHT effect at phrase and sentence levels is not detectable in a language unknown to the listener (61). However, as with false positives, this question has not been directly addressed in literature.

How many of those who are not producing one should by all accounts be capable (i.e. false negative)? False negative rates can be quantified as the proportion of healthy, awake, normal-hearing, neurotypical subjects who fail to respond.

3. Do the results converge with other measures? What is the prognostic value?

Existing diagnostic measures differentiate multiple DOC states, and those states have different prognoses, with a better outlook for MCS than VS (88). According to

several retrospective studies, diagnosis by clinical consensus has a 37%-43% chance of being wrong (89-91). Investigations of covert consciousness often cite these figures as motivation: this chance is too high given the stakes, and must be improved with new mesures. Therefore, convergence with existing diagnostic measures may be of secondary importance. On the other hand, convergence with the outcome (in other words, prognostic power) is not.

Reports of prognostic power of ERPs are inconsistent. For example, Fischer at al report SON-based P300 as a better outcome predictor than MMN (48), while Kane et al report the opposite (3). In the current study, SON-based P300 would have yielded a prognosis inconsistent with the outcome (as of 12 months post-admission). The current study, along with other emerging evidence (e.g. (9)) suggests that single-instance testing may yield false negatives, which may explain the discrepancy in literature. Specificity in predicting awakening is reported to be quite high for some ERPs based on single-instance testing (e.g. 90%-100% for MMN (3,6)), meaning that it's possible to make a positive prognosis based on their presence.

6. What are the stakes for the outcomes of this testing, and are the above rates acceptable given the stakes?

The stakes in testing coma patients for covert awareness can be quite high. Therefore, conclusions must be drawn on the basis of converging evidence. Some of the effects described in the present work are more robust than others. In particular, duration-based MMN, and P300-SON (provided the trends described here hold for larger populations) have comparatively high true positive and comparatively low false

negative rates. If a patient fails to show a response, this has, conservatively estimating, an 80% chance of being an accurate reflection of their ability for MMN, and close to 100% chance for P300-SON - provided that patient retained their hearing. Other paradigms are less robust. Despite a high proportion of single-subject responses, task-based P300 is predicated on understanding and following task instructions. Existing CHT paradigms may have a high rate of false negatives, and therefore no treatment decisions should be based on a null result. However, there are calls for making no treatment decisions at all based on null results for any EEG test of covert awareness (63).

4.5 Limitations

Is it not possible to make generalizable statements about DOC patients from a single case study. These results should be treated as one piece of evidence among many in order to draw robust conclusions. The interpretation of the results would benefit from the patient's medical history prior to DOC, since it is known that non-DOC conditions affect the presence and presentation of ERPs. There was a degree of subjectivity in using visual inspection to select subsets of data for the purposes of analysis. This can be mitigated by employing multiple raters or a more agnostic statistical analysis that does not require a priori data selection. Selection of reference electrode(s) during EEG data processing can impact the end results. For low-density recordings, an external reference site such as the nose or the mastoids has been argued to be better than a Global average. The patient's GCS 3 session was referenced to a Global average, which may have introduced artifacts into the data. As the testing

paradigms described here have been validated with healthy controls in a number of existing works, the primary purpose of the control sample in this investigation was to provide a working dataset which could serve as an immediate reference for the analysis and interpretation of the data collected from the patient. This purpose may have been better served by recruiting controls with a linguistic, and, if applicable, medical background similar to the patient's. An obvious limitation of the control investigation is that the testing was administered in a single session rather than 24-hour multisession recording. In the future, it would be very beneficial to obtain such a dataset from healthy controls.

4.6 Ethical Implications

A patient's prognosis for functional recovery from a DOC state can influence care decisions, including life-sustaining treatment. Any improvement to DOC diagnosis and prognosis is unequivocally beneficial.

However, before diagnostic and prognostic power of ERPs has been fully investigated and ascertained, the implications of the information they provide must be considered carefully. ERPs discussed here assess several cognitive functions, ranging from basic audition to advanced language processing. In the case study described, a reliable response was elicited to the patient's own name. Such a finding may impact the decision-making and bedside behaviour of the patient's family. Care must be taken, therefore, to convey in clear layman-appropriate manner that despite observing a response, we cannot be certain the patient *recognizes* or *understands* her name or her father's voice. Further, such information must be presented in the context of prognostic

value it possesses, that is, observing an ERP response does not guarantee improvement or emergence from the DOC state (as is illustrated in the present case study).

ERPs are a promising avenue for evaluation of cognition in absence of overt behaviour. However, inconsistencies in existing reports with regard to prognostic sensitivity and specificity highlight the need for standardization of stimulus paradigms and detection criteria. Recording and elicitation guidelines have been proposed (e.g. 29), and a field-accepted set of standards may well emerge in the future, which will likely increase clinicians' confidence in using ERPs diagnostically. At present, ERP evidence should be approached as research in progress, and should not serve as basis for decisions as significant as withdrawal of life support.

4.7 Future Directions

There are several avenues for future research to enhance the evaluation of covert consciousness in DOC. Interpreting patients' results would greatly benefit from establishing control norms for prevalence and presentation of ERPs, particularly with standardized testing paradigms. Age and non-DOC conditions affect ERP responses, so it would be beneficial to establish norms based on age groups, and quantify the effects of non-DOC conditions. Investigating false positives and false negatives among ERPs would also be an asset to clinical investigations. Finally, it would be very beneficial to obtain a control sample tested in the exact same manner as the patient - weekly sessions, 24-hour round-the-clock administration of the entire battery. This would also contribute to understanding and quantifying ERP stability, test-retest

reliability, habituation, and natural fluctuations of vigilance in the presence of continuous stimuli (e.g. would it be possible to fall asleep?).

Given the variation of single-subject detection procedures employed for ERPs in DOC work, this data should be made available for re-analysis with multiple methods in future work investigating an optimal detection approach.

The majority of experimental paradigms used in this study have been validated in prior work. The P300-SON paradigm, however, is novel yet promising. Future research could validate it with a larger control sample. Similarly, the CHT effect should be investigated further before being considered for inclusion in a clinical testing battery.

5. Conclusions

This investigation was undertaken with an eye towards developing a clinical tool for evaluating covert cognition in DOC patients who don't exhibit overt behaviour, and improving upon the methods currently used for this purpose. A testing battery was designed to elicit several well-known EEG responses, and was administered to a DOC patient over 24 hours at GCS 3 and GCS 6, as well as 18 healthy controls. The results suggest fluctuations in the patient's ability to produce detectable responses, which in turn suggests that repeated testing is necessary to fully evaluate a patient's abilities. This may also explain inconsistencies in the existing literature, which largely relies on single-session testing to make diagnostic and prognostic inferences. The results also suggest that maximizing the relevant property of experimental manipulations, such as maximizing the salience of own name in a P300-SON paradigm, may be beneficial in

DOC work, even at the expense of knowing which exact aspect of the manipulation is eliciting the response.

Future work should validate the finding of fluctuating detectability. Establishing population norms for single-subject prevalence, latency, and amplitude of ERPs would improve confidence in interpreting results among patients. In the current state of ERP knowledge in both healthy individuals and DOC patients, we may make a positive prognosis based on ERP presence with a degree of confidence, but must exercise caution in making negative prognoses or high-stakes care decisions based on ERP absence. However, given the documented rates of DOC misdiagnosis, further developing cognitive assessment tools that don't rely on overt behaviour will be of great benefit.

REFERENCES

1. Fischer C, Morlet D, Bouchet P, Luaute J, Jourdan C, Salord F. Mismatch negativity and late auditory evoked potentials in comatose patients. Clinical neurophysiology. 1999 Sep 1;110(9):1601-10.
2. Kane NM, Butler SR, Simpson T. Coma outcome prediction using event-related potentials: P3 and mismatch negativity. Audiology and Neurotology. 2000 May 19;5(3-4):186-91.
3. Kane NM, Curry SH, Rowlands CA, Manara AR, Lewis T, Moss T, Cummins BH, Butler SR. Event-related potentials—neurophysiological tools for predicting emergence and early outcome from traumatic coma. Intensive Care Medicine. 1996 Jan 1;22(1):39-46.
4. Kotchoubey B, Lang S, Mezger G, Schmalohr D, Schneck M, Semmler A, Bostanov V, Birbaumer N. Information processing in severe disorders of consciousness: vegetative state and minimally conscious state. Clinical Neurophysiology. 2005 Oct 1;116(10):2441-53.
5. Chennu S, Finoia P, Kamau E, Monti MM, Allanson J, Pickard JD, Owen AM, Bekinschtein TA. Dissociable endogenous and exogenous attention in disorders of consciousness. NeuroImage: Clinical. 2013 Jan 1;3:450-61.
6. Morlet D, Fischer C. MMN and novelty P3 in coma and other altered states of consciousness: a review. Brain topography. 2014 Jul 1;27(4):467-79.
7. Daltrozzo J, Wioland N, Mutschler V, Kotchoubey B. Predicting coma and other low responsive patients outcome using event-related brain potentials: a meta-analysis. Clinical neurophysiology. 2007 Mar 31;118(3):606-14.
8. Fischer C, Luauté J, Adeleine P, Morlet D. Predictive value of sensory and cognitive evoked potentials for awakening from coma. Neurology. 2004 Aug 24;63(4):669-73.
9. Mah RL, Connolly JF, Fox-Robichaud AE, Hamielec C. Electrophysiological markers of variations in perceptual and cognitive processing in coma. Society for Neuroscience, 2016
10. Tzovara A, Rossetti AO, Spierer L, Grivel J, Murray MM, Oddo M, De Lucia M. Progression of auditory discrimination based on neural decoding predicts awakening from coma. Brain. 2012 Nov 11;136(1):81-9.
11. Giacino JT. Disorders of Consciousness: Differential Diagnosis and Neuropathologic Features. In Seminars in Neurology 1997 Jun (Vol. 17, No. 02, pp. 105-111). © 1997 by Thieme Medical Publishers, Inc..
12. Howsepian AA. The 1994 Multi-Society Task Force consensus statement on the persistent vegetative state: a critical analysis. Issues L. & Med.. 1996;12:3.
13. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: the state of the science. Nature Reviews Neurology. 2014 Feb;10(2):99.
14. Landsness E, Bruno MA, Noirhomme Q, Riedner B, Gosseries O, Schnakers C, Massimini M, Laureys S, Tononi G, Boly M. Electrophysiological correlates of behavioural changes in vigilance in vegetative state and minimally conscious state. Brain. 2011 Aug 1;134(8):2222-32.
15. Augustine JR. Human neuroanatomy. Academic Press; 2008.
16. Zollman FS, editor. Manual of traumatic brain injury management. Demos Medical Publishing; 2011 Mar 24.
17. Plum F, Posner JB. The diagnosis of stupor and coma. Oxford University Press, USA; 1982.
18. Bernat JL. Chronic disorders of consciousness. The Lancet. 2006 Apr 8;367(9517):1181-92.

19. Giacino JT, Zasler ND, Whyte J, Katz DI, Glen M, Andary M. Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. Archives of Physical Medicine and Rehabilitation. 1995 Feb 1;76(2):205-9.
20. Giacino JT, Ashwal S, Childs N, Cranford R, Jennett B, Katz DI, Kelly JP, Rosenberg JH, Whyte JO, Zafonte RD, Zasler ND. The minimally conscious state: definition and diagnostic criteria. Neurology. 2002 Feb 12;58(3):349-53.
21. Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. The Lancet. 1974 Jul 13;304(7872):81-4.
22. Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: Measurement characteristics and diagnostic utility. Archives of physical medicine and rehabilitation. 2004 Dec 1;85(12):2020-9.
23. Jennett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. The Lancet. 1975 Mar 1;305(7905):480-4.
24. Multi-Society Task Force on PVS. Medical aspects of the persistent vegetative state. New England Journal of Medicine. 1994 May 26;330(21):1499-508.
25. Giacino JT. The vegetative and minimally conscious states: consensus-based criteria for establishing diagnosis and prognosis. NeuroRehabilitation. 2004 Jan 1;19(4):293-8.
26. Luaute JM, Maucort-Boulch DM, Tell L, Quelard F, Sarraf T, Iwaz J, Boisson D, Fischer C. Long-term outcomes of chronic minimally conscious and vegetative states. Neurology. 2010 Jul 20;75(3):246-52.
27. Luck SJ. Event-related potentials. APA handbook of research methods in psychology. 2012 Apr;1:523-46.
28. Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical neurophysiology. 2007 Dec 1;118(12):2544-90.
29. Duncan CC, Barry RJ, Connolly JF, Fischer C, Michie PT, Näätänen R, Polich J, Reinvang I, Van Petten C. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clinical Neurophysiology. 2009 Nov 1;120(11):1883-908.
30. Connolly JF, D'Arcy RC, Newman RL, Kemps R. The application of cognitive event-related brain potentials (ERPs) in language-impaired individuals: review and case studies. International Journal of Psychophysiology. 2000 Oct 1;38(1):55-70.
31. Connolly JF, Mate-Kole CC, Joyce BM. Global aphasia: an innovative assessment approach. Archives of physical medicine and rehabilitation. 1999 Oct 1;80(10):1309-15.
32. Mah RL, Connolly JF. A framework for the extended monitoring of levels of cognitive function in unresponsive patients. PloS one. 2018 Jul 19;13(7):e0200793.
33. Näätänen R, Gaillard AW, Mäntysalo S. Early selective-attention effect on evoked potential reinterpreted. Acta psychologica. 1978 Jul 1;42(4):313-29.
34. Morlet D, Fischer C (2001) The mismatch negativity (MMN) recorded in comatose patients actually discloses mismatch processes. EPIC XIII 41 (3):199
35. Bishop DV, Hardiman MJ. Measurement of mismatch negativity in individuals: A study using single‐trial analysis. Psychophysiology. 2010 Jul;47(4):697-705.

36. Lew HL, Gray M, Poole JH. Temporal stability of auditory event-related potentials in healthy individuals and patients with traumatic brain injury. Journal of Clinical Neurophysiology. 2007 Oct 1;24(5):392-7.
37. Tervaniemi M, Lehtokoski A, Sinkkonen J, Virtanen J, Ilmoniemi RJ, Näätänen R. Test–retest reliability of mismatch negativity for duration, frequency and intensity changes. Clinical Neurophysiology. 1999 Aug 1;110(8):1388-93.
38. Simpson TP, Manara AR, Kane NM, Barton RL, Rowlands CA, Butler SR. Effect of propofol anaesthesia on the event-related potential mismatch negativity and the auditory-evoked potential N1. British journal of anaesthesia. 2002 Sep 1;89(3):382-8.
39. Pekkonen E, Rinne T, Näätänen R. Variability and replicability of the mismatch negativity. Clinical Neurophysiology. 1995 Nov 1;96(6):546-54.
40. Escera C, Yago E, Polo MD, Grau C. The individual replicability of mismatch negativity at short and long inter-stimulus intervals. Clinical Neurophysiology. 2000 Mar 1;111(3):546-51.
41. Sutton S, Braren M, Zubin J, John E. Evoked potential correlates of stimulus uncertainty. Science. 1965;150:1187–1188.
42. Polich J. Updating P300: an integrative theory of P3a and P3b. Clinical neurophysiology. 2007 Oct 31;118(10):2128-48.
43. Reuter BM, Linke DB, Kurthen M. Kognitive Prozesse bei Bewusstlosen? Eine Brain-Mapping-Studie zu P300. Arch Psychol 1989;141:155–73.
44. Lugo ZR, Quitadamo LR, Bianchi L, Pellas F, Veser S, Lesenfants D, Real RG, Herbert C, Guger C, Kotchoubey B, Mattia D. Cognitive processing in non-communicative patients: what can event-related potentials tell us?. Frontiers in Human Neuroscience. 2016;10.
45. Schnakers C, Perrin F, Schabus M, Majerus S, Ledoux D, Damas P, Boly M, Vanhaudenhuyse A, Bruno MA, Moonen G, Laureys S. Voluntary brain processing in disorders of consciousness. Neurology. 2008 Nov 11;71(20):1614-20.
46. Perrin F, Schnakers C, Schabus M, Degueldre C, Goldman S, Brédart S, Faymonville ME, Lamy M, Moonen G, Luxen A, Maquet P. Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome. Archives of neurology. 2006 Apr 1;63(4):562-9.
47. Fischer C, Luaute J, Morlet D. Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. Clinical neurophysiology. 2010 Jul 1;121(7):1032-42.
48. Fischer C, Dailler F, Morlet D. Novelty P3 elicited by the subject's own name in comatose patients. Clinical neurophysiology. 2008 Oct 31;119(10):2224-30.
49. Wood N, Cowan N. The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? J Exp Psychol Learn Mem Cogn 1995;21:255–60.
50. Holeckova I, Fischer C, Giard MH, Delpuech C, Morlet D. Brain responses to a subject's own name uttered by a familiar voice. Brain research. 2006 Apr 12;1082(1):142-52.
51. Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science. 1980;207:203–5.
52. Lau EF, Phillips C, Poeppel D. A cortical network for semantics:(de) constructing the N400. Nature Reviews Neuroscience. 2008 Dec 1;9(12):920-33.

53. Rämä P, Relander-Syrjänen K, Öhman J, Laakso A, Näätänen R, Kujala T. Semantic processing in comatose patients with intact temporal lobes as reflected by the N400 event-related potential. Neuroscience letters. 2010 Apr 26;474(2):88-92.
54. Kotchoubey B, Daltrozzo J, Wioland N, Mutschler V, Lutun P, Birbaumer N, Jaeger A. Semantic processing in a coma patient. Grand Rounds. 2005;5:37-41.
55. Daltrozzo J, Wioland N, Mutschler V, Lutun P, Calon B, Meyer A, Pottecher T, Lang S, Jaeger A, Kotchoubey B. Cortical information processing in coma. Cognitive and Behavioral Neurology. 2009 Mar 1;22(1):53-62.
56. Kotchoubey B. Apallic syndrome is not apallic: is vegetative state vegetative?. Neuropsychological Rehabilitation. 2005 Jul 1;15(3-4):333-56.
57. Steppacher I, Eickhoff S, Jordanov T, Kaps M, Witzke W, Kissler J. N400 predicts recovery from disorders of consciousness. Annals of neurology. 2013 May 1;73(5):594-602.
58. Cruse, D., Beukema, S., Chennu, S., Malins, J.G., 2014a. The reliability of the N400 in single subjects: implications for patients with disorders of consciousness. NeuroImage: Clinical 4, 788–799.
59. Rohaut, B., Faugeras, F., Chausson, N., King, J.R., El I., K., Cohen, L., Naccache, L., 2015. Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. Neuropsychologia 66, 279–292.
60. Connolly JF, Phillips NA. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. Journal of Cognitive Neuroscience. 1994 Jul;6(3):256-66.
61. Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. Nature neuroscience. 2016 Jan 1;19(1):158-64.
62. Makov S, Sharon O, Ding N, Ben-Shachar M, Nir Y, Golumbic EZ. Sleep disrupts high-level speech parsing despite significant basic auditory processing. Journal of Neuroscience. 2017 Aug 9;37(32):7772-81.
63. Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (eeg). Frontiers in human neuroscience. 2017 Sep 28;11:481.
64. Hirzel M, Lapinskaya N, Lidz J, & Lau E. Syntactic constituency rate effects in EEG. Society for the Neurobiology of Language, 2017
65. Gabriel D, Muzard E, Henriques J, Mignot C, Pazart L, Andre´-Obadia N, et al. Replicability and impact of statistics in the detection of neural responses of consciousness. Brain. 2016; 139(6):e30
66. Naccache L, Sitt J, King JR, Rohaut B, Faugeras F, Chennu S, et al. Reply: Replicability and impact of statistics in the detection of neural responses of consciousness. Brain. 2016; 139(6):e31.
67. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971 Mar 1;9(1):97-113.
68. Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com, Version 18.0
69. NeoSpeech, Santa Clara, CA, http://neospeech.com/
70. Todd J, Michie PT, Schall U, Karayanidis F, Yabe H, Näätänen R. Deviant matters: duration, frequency, and intensity deviants reveal different patterns of mismatch negativity reduction in early and late schizophrenia. Biological psychiatry. 2008 Jan 1;63(1):58-64.

71. Block CK, Baldwin CL. Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. Behavior research methods. 2010 Aug 1;42(3):665-70.
72. Marchand, Y., D'Arcy, R. C., & Connolly, J. F. (2002). Linking neurophysiological and neuropsychological measures for aphasia assessment. Clinical Neurophysiology, 113(11), 1715-1722.
73. Dien J. Issues in the application of the average reference: Review, critiques, and recommendations. Behavior Research Methods, Instruments, & Computers. 1998 Mar 1;30(1):34-43.
74. NEMO Consortium. NEMO Technical Report TR-2012-008: Scalp Regions of Interest (ROI). Neuro ElectroMagnetic Ontologies; 2012. Available from: http://nemo.nic.uoregon.edu/wiki/File:NEMO-TR- 2012-008_ROIdefinitions.docx.
75. Bahmer A, Baumann U. Recording and online analysis of auditory steady state responses (ASSR) in Matlab. Journal of neuroscience methods. 2010 Mar 15;187(1):105-13.
76. Polich J. Clinical application of the P300 event-related brain potential. Physical Medicine and Rehabilitation Clinics. 2004 Feb 1;15(1):133-61.
78. Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical neurophysiology. 2007 Dec 1;118(12):2544-90.
79. Todd J, Whitson L, Smith E, Michie PT, Schall U, Ward PB. What's intact and what's not within the mismatch negativity system in schizophrenia. Psychophysiology. 2014 Apr 1;51(4):337-47.
80. Tavakoli P, Varma S, Campbell K. Highly relevant stimuli may passively elicit processes associated with consciousness during the sleep onset period. Consciousness and cognition. 2017 Nov 2.
81. Wagner P, Röschke J, Grözinger M, Mann K. A replication study on P300 single trial analysis in schizophrenia: confirmation of a reduced number of 'true positive' P300 waves. Journal of psychiatric research. 2000 May 1;34(3):255-9.
83. Beukema S, Gonzalez-Lara LE, Finoia P, Kamau E, Allanson J, Chennu S, Gibson RM, Pickard JD, Owen AM, Cruse D. A hierarchy of event-related potential markers of auditory processing in disorders of consciousness. NeuroImage: Clinical. 2016 Feb 1;12:359-71.
84. Hagoort P, Brown CM. ERP effects of listening to speech: semantic ERP effects. Neuropsychologia. 2000 Oct 1;38(11):1518-30.
85. Cruse D, Gantner I, Soddu A, Owen AM. Lies, damned lies and diagnoses: Estimating the clinical utility of assessments of covert awareness in the vegetative state. Brain injury. 2014 Aug 1;28(9):1197-201.
86. Connolly JF, Major A, Allen S, D'Arcy RC. 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. 1999 Aug 1;21(4):444-64.
87. D'Arcy RC, Connolly JF. An event-related brain potential study of receptive speech comprehension using a modified Token Test. Neuropsychologia. 1999 Dec 1;37(13):1477-89.
88. Giacino JT, Ashwal S, Childs N, Cranford R, Jennett B, Katz DI, Kelly JP, Rosenberg JH, Whyte JO, Zafonte RD, Zasler ND. The minimally conscious state: definition and diagnostic criteria. Neurology. 2002 Feb 12;58(3):349-53.

89. Childs NL, Mercer WN, Childs HW: Accuracy of diagnosis of persistent vegetative state. Neurology. 1993, 43 (8): 1465-1467.
90. Andrews K, Murphy L, Munday R, Littlewood C: Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. BMJ. 1996, 313 (7048): 13-16.
91. Schnakers C, Vanhaudenhuyse A, Giacino J, Ventura M, Boly M, Majerus S, Moonen G, Laureys S. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. BMC neurology. 2009 Dec;9(1):35.