



A hierarchy of event-related potential markers of auditory processing in disorders of consciousness



Steve Beukema^{a,b}, Laura E. Gonzalez-Lara^a, Paola Finoia^c, Evelyn Kamau^c, Judith Allanson^c, Srivas Chennu^{d,e}, Raechelle M. Gibson^a, John D. Pickard^c, Adrian M. Owen^a, Damian Cruse^{f,*}

^aThe Brain and Mind Institute, University of Western Ontario, London, ON, Canada

^bMcGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Canada

^cCambridge University Hospitals NHS Foundation Trust, Cambridge, UK

^dSchool of Computing, University of Kent, Chatham Maritime, UK

^eDepartment of Clinical Neurosciences, The University of Cambridge, Cambridge, UK

^fSchool of Psychology, University of Birmingham, Birmingham, UK

ARTICLE INFO

Article history:

Received 9 June 2016

Accepted 3 August 2016

Available online 04 August 2016

Keywords:

Disorders of consciousness

N400

Auditory processing

Vegetative state

Minimally conscious state

ABSTRACT

Functional neuroimaging of covert perceptual and cognitive processes can inform the diagnoses and prognoses of patients with disorders of consciousness, such as the vegetative and minimally conscious states (VS;MCS). Here we report an event-related potential (ERP) paradigm for detecting a hierarchy of auditory processes in a group of healthy individuals and patients with disorders of consciousness. Simple cortical responses to sounds were observed in all 16 patients; 7/16 (44%) patients exhibited markers of the differential processing of speech and noise; and 1 patient produced evidence of the semantic processing of speech (i.e. the N400 effect). In several patients, the level of auditory processing that was evident from ERPs was higher than the abilities that were evident from behavioural assessment, indicating a greater sensitivity of ERPs in some cases. However, there were no differences in auditory processing between VS and MCS patient groups, indicating a lack of diagnostic specificity for this paradigm. Reliably detecting semantic processing by means of the N400 effect in passively listening single-subjects is a challenge. Multiple assessment methods are needed in order to fully characterise the abilities of patients with disorders of consciousness.

© 2016 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The vegetative state (VS; also referred to as unresponsive wakefulness syndrome, Laureys et al., 2010) and minimally conscious state (MCS) are chronic disorders of consciousness that can follow from a severe brain injury. A diagnosis of VS occurs when a patient is considered to be ‘wakeful without awareness’, defined by four broad criteria (Jennett and Plum, 1972; Royal College of Physicians, 2003): 1) evidence for preserved sleep-wake cycles; 2) no evidence of awareness of the self or the environment; 3) no evidence of sustained, reproducible, purposeful, or voluntary response to auditory, tactile, or noxious stimuli; 4) no evidence of language comprehension or expression. A diagnosis of MCS, on the other hand, occurs when minimal but reproducible evidence of awareness is observed.

Clinical judgments of the relative fulfilment of these criteria are currently based on behavioural observations. However, the insensitivity of behavioural assessments of consciousness is well documented, with an estimated misdiagnosis rate of 40% for VS (Andrews et al., 1996; Childs

et al., 1993; Schnakers et al., 2009). Furthermore, in recent years, it has become evident that functional neuroimaging can sometimes provide a clearer picture of the extent to which a given patient fulfils these criteria. Indeed, many examples of covert cognition and consciousness have been reported in patients whose behaviour is nevertheless consistent with the VS (Fernandez-Espejo and Owen, 2013; Stender et al., 2014). In this manuscript, we focus on the fourth criterion above, namely the evidence for the absence of language comprehension.

In two early functional magnetic resonance imaging (fMRI) studies, Coleman et al. (2009, 2007) endeavoured to place patients on a hierarchy of auditory processing abilities: from low-level audition, through speech perceptual processes, to the extraction of meaning (semantics). To accomplish this, patients were presented with sentences containing words that were either semantically ambiguous (e.g. “There were *dates* and *pears* in the fruit bowl”) or relatively less ambiguous (e.g., “There was *beer* and *cider* on the kitchen shelf”). Because resolving this semantic ambiguity required comprehension of the sentence as a whole, the contrast between high and low-ambiguity sentences was considered to index speech *comprehension*, or at least the processing of the meaning of the words. A further contrast was performed between all speech stimuli and a non-speech control stimulus - signal-correlated noise

* Corresponding author.

E-mail address: d.cruse@bham.ac.uk (D. Cruse).

(SCN) - in order to identify what the authors considered to be speech-specific *perceptual* processing of speech. Finally, a contrast of all stimuli to silence was performed to identify early auditory processing. Sixty percent of VS patients (13/22) demonstrated no fMRI activity in any of the contrasts. Nine percent (2/22) showed significant auditory responses, 32% (7/22) showed appropriate activity in the perceptual contrast, and 9% (2/22) in the semantic contrast. Interestingly, the level of auditory processing possessed by these patients (as indexed by the neuroimaging data) was correlated with the patients' behavioural abilities at six-months follow-up ($r = 0.81$), suggesting a potential prognostic value to neural markers of speech processing.

However, few patients can have their language comprehension abilities assessed in this way due to the high cost of MRI, poor scanner availability, and clinical contraindications. Electroencephalography (EEG), however, can reach a greater number of patients than fMRI because it is relatively less expensive and can be performed at the bedside. Moreover, EEG provides a well-studied index of semantic processing in healthy individuals, known as the N400 event-related potential (ERP) - a negative-going potential over centroparietal scalp that peaks around 400 ms post-stimulus. The amplitude of the N400 is primarily sensitive to the context in which a meaningful stimulus occurs. For example, when words are presented in pairs, the second word of the pair (i.e. the target) elicits a larger N400 when the words in the pair are unrelated than when they are related (e.g., *cat-chair* versus *table-chair*). Similarly, when a word is incongruent with a sentence context, it elicits a larger N400 (e.g. "He ate the moist cake with a knife and *potato/fork*"; see Kutas and Federmeier, 2011, for a full review). The difference in the N400 amplitude that is produced by such priming manipulations will be referred to here as the 'N400 effect'.

While the exact linguistic processes indexed by the N400 are still a matter of debate, the sensitivity of the N400 to manipulations of meaning indicate that, at its simplest, the N400 effect reflects the processing of meaning, and therefore provides a neural marker of semantic processing that may not be evident from behaviour. As a result, a number of studies have sought evidence of N400 effects in disorders of consciousness.

Schoenle and Witzke (2004) visually inspected the ERPs elicited by semantically incongruent and congruent sentence-endings and reported evidence of an N400-like deflection in 39% of VS patients. Steppacher et al. (2013; see also Kotchoubey et al., 2005) also reported N400-like effects in continuous wavelet transformed ERP data of 32% of a group of 48 VS patients. Furthermore, the presence of N400-like effects was also correlated with behavioural improvement at long-term follow-up (2–14 years). Rohaut et al. (2015) also observed significant N400 ERP effects in 1/15 VS patients and 5/14 MCS patients, contributing to the evidence that some patients with disorders of consciousness are capable of processing the meaning of speech.

Cruse et al. (2014a) recently demonstrated the importance of the stimuli used to detect statistically significant N400 effects in single-subjects. Specifically, they found that word-pairs taken from associations norms - i.e. pairs in which the target is the word most often produced by a group of healthy participants during free association of a prime - were more likely to lead to reliable single-subject N400 effects when compared with two other common N400 approaches - semantically related word-pairs, and high-cloze sentences. Furthermore, Cruse et al. (2014a) found that repeating words within an assessment - a common practice in N400 studies - altered the observed N400 effects. However, even with a high level of stimulus control, Cruse et al. (2014a) could only detect a statistically reliable N400 effect in 50% of healthy individuals during passive listening. Similarly, Rohaut et al. (2015) employed normatively associated word-pairs and found significant N400 effects in only 42% of healthy controls during passive listening.

Together these results suggest that tests of semantic processing that rely on the priming N400 effect do not have sufficient sensitivity for regular clinical use. Indeed, Cruse et al. (2014a) found that instructing healthy participants to make judgments on the relatedness of the

words in each pair led to an increase in the probability of detecting a significant N400 effect relative to when participants were only instructed to passively listen to the stimuli. This is somewhat unsurprising as relatedness judgments require the participants to attend to the meaning of the stimuli - a manipulation known to increase group-level N400 effects (Bentin et al., 1993). Therefore, even though passive listening to word-pairs does elicit a group level N400 effect, the size of the signal is reduced and therefore is more difficult to detect in single-subjects.

It is plausible, then, that some patients are able to process the meaning of speech, but are unable to complete the demanding task necessary for them to produce positive evidence of an N400 effect. Indeed, patients in the VS are by definition unable to follow commands. It may therefore not be possible to reliably separate semantic processing from an ability to follow commands using an N400 priming paradigm - at least not to the level of sensitivity necessary for its use as a clinical tool.

As a result of the apparent single-subject insensitivity of the priming N400 effect, we sought to identify complementary ERP markers of related speech processes. We employed an approach similar to the fMRI paradigm of Coleman et al. (2009, 2007; Davis et al., 2007) in which a three level hierarchy of auditory processes were investigated. Specifically, we presented participants with speech stimuli taken from the normative-association word-pair task of Cruse et al. (2014a) alongside non-speech noise stimuli (signal-correlated noise). This allowed us to identify 1) *semantic processing* through the (albeit poorly sensitive) classic N400 contrast of related and unrelated targets; 2) *speech perceptual processing* through the contrasts of speech and noise; and 3) *auditory processing* through evidence of auditory evoked potentials.

While there are decades of studies of semantic processing with ERPs, investigations of pre-semantic speech processes (i.e. speech versus noise) are relatively scarce. Within oddball paradigms, in which a rare stimulus occurs within a sequence of repeated stimuli, speech sounds and non-speech sounds have been shown to produce different patterns of mismatch negativity amplitudes, indicating ERP markers of the differential processing of speech and non-speech within 150–250 ms post-stimulus (Jaramillo et al., 1999, 2001). Furthermore, relative to speech sounds (vowels), more positive-going P1 and P2 deflections (~100 and ~200 ms post-stimulus respectively) have been reported in response to noises, while N1 deflections (~70 ms post-stimulus) are more negative-going for vowels relative to noises with primary auditory cortex implicated as the generator of this speech-specific processing (Edmonds et al., 2010). To our knowledge, the exact contrast employed by Coleman et al. (2009, 2007) with fMRI has not been reported with ERPs - i.e. speech versus signal-correlated noise. Nevertheless, we would expect the ERPs elicited by speech to deviate from those elicited by noise early in the epoch, followed by a later deviation according to the meaning of the speech (i.e. the N400 effect).

As this type of speech paradigm has been shown to have potential clinical utility in fMRI (Coleman et al., 2009), here we investigated the potential for ERPs to provide a similar hierarchical assessment of auditory processing in a group of 16 healthy control participants. As a proof of concept, we subsequently applied this method to a group of 16 patients with chronic disorders of consciousness (8 VS, 8 MCS).

2. Materials and methods

2.1. Control participants

Seventeen right-handed, native Canadian English speaking participants were recruited from the Psychology Department's participant resource pool at The University of Western Ontario, or via posters distributed around the University campus. Data from one participant were excluded due to an equipment fault. The remaining sixteen participants ranged in age from 18 to 25 years old (median = 20; 8 female). All participants were compensated with one credit per hour of participation for use towards an undergraduate course requirement, or

alternatively, \$15.00 per hour. The Psychology Research Ethics Board of the University of Western Ontario, Canada, granted ethical approval for this study.

2.2. Patients

A convenience sample of patients were recruited to the University of Western Ontario, Canada, and Cambridge University, England. Of the 20 patients tested, 16 contributed sufficiently clean EEG data for the analyses reported here. Of these 16 patients, 8 fulfilled the CRS-R diagnostic criteria for MCS and 8 for VS. Full details are provided in Tables 1 and 2 and Section 2.3. Ethical approval for the patient study was provided by the Health Sciences Research Ethics Board of the University of Western Ontario, and the National Research Ethics Service of the National Health Service, UK.

2.3. Behavioural diagnosis

The differential diagnosis of VS/MCS was determined by CRS-R assessment (Kalmar and Giacino, 2005). The median number of CRS-R assessments during the week of EEG testing was 4 (range 1–6). Four of the sixteen patients (MCS-2, VS-1, VS-2, VS-5) could only be assessed once within the week of EEG testing. However, the diagnosis on these occasions was consistent with prior CRS-R assessment (2-months to 2-years prior to EEG), and the diagnosis of the referring physician.

2.4. Stimuli and experimental procedure

Word stimuli were identical to those used in an associative priming task reported by Cruse et al. (2014a; see Experiment 2, pp 792–794) because these stimuli were shown to have the highest likelihood of eliciting single-subject N400 effects. Specifically, two hundred word-pairs were selected from Nelson et al.'s (1998) associative norms. From these 200 pairs, 100 of the most strongly associated word-pairs were selected (e.g., bumble-bee) with a mean forward association of 0.81 (SD = 0.05) such that, on average, the target word (i.e. the second word of the pair) was produced by 81% of the participants when asked for the first word that comes to mind following presentation of the prime word (i.e. the first word of the pair). The remaining 100 word-pairs were recombined to create 100 unrelated word-pairs controlled so that phonological, semantic, or associative overlap between the target and any word associated to the prime were minimized. A male, native Canadian-English speaker digitally recorded all word-pairs, and their amplitudes were normalized (mean spoken word length = 638 ms, SD = 138 ms, range = 309–980 ms). For data collected from patients in Cambridge, UK, a native British-English speaker recorded all word-pairs to control for the potential confounding effect of foreign

accents on speech processing (mean spoken word length = 562 ms, SD = 108 ms, range = 339–905 ms). There were no significant differences between the related and unrelated pairs in the spoken length of targets (Canadian English: $t(198) = 1.280, p = 0.203$, two-tailed; British English: $t(198) = 0.289, p = 0.773$, two-tailed) or primes (Canadian English: $t(198) = 0.670, p = 0.505$, two-tailed; British English: $t(198) = 0.879, p = 0.380$, two-tailed). Due to the impact of order effects in priming studies (see Cruse et al., 2014a), no words were repeated across the study. The stimuli were further validated by Cruse et al. (2014a) to show that no significant differences existed in the ERPs elicited by each word category in the absence of priming. These results confirm that any ERP differences found between unrelated and related targets in the experimental condition are due to priming and not other features of the stimuli.

Signal-correlated noise (SCN) stimuli were generated from all words according to Schroeder (1968), in order to produce a non-speech condition that matched the speech stimuli in duration and amplitude envelope. One pair of SCN stimuli was presented between each word-pair. In total, the stimuli consisted of 400 words (100 related word-pairs, 100 unrelated word-pairs), and 400 signal-correlated noise stimuli.

Participants heard all words and all noises without repetition via EARTONE® 3 A Insert Headphones (E-A-R Auditory Systems). Participants were presented with the stimuli in the following repeated pattern: SCN1-SCN2-Prime-Target, where SCN1 was generated from the Target category, and SCN2 from the Prime category. Targets were either related or unrelated to the primes. The order of presentation of word-pairs and SCN-pairs were randomized independently within each participant so that SCN pairs on average were not generated from the same word pair with which they were heard. The stimulus onset asynchrony for all stimuli was 1100 milliseconds. Stimulus delivery was controlled by the Matlab toolbox Psychtoolbox (Brainard, 1997). An anonymous reviewer suggested that the predictable pattern of trials that was employed in this study (i.e. a word-pair always followed a noise-pair) may have contaminated the ERP effects with a consistent baseline shift. To address this concern, we conducted a supplementary study of healthy control participants in which trial order was entirely unpredictable (see Supplementary materials). The perceptual and semantic effects observed in the supplementary study are consistent with those reported in the main manuscript, and therefore indicate that our choice of a predictable trial design does not confound our conclusions.

The experiment was broken up into four blocks of 50 trials (i.e. 50 SCN pairs and 50 word-pairs), after which participants were offered a break. Each block lasted <4-min, with the entire testing protocol lasting <30-min with breaks. In accordance with the standards of behavioural assessment, if a patient exhibited sustained eye closure during testing, the 'Arousal Facilitation Protocol' of the CRS-R was administered during

Table 1

Patient details and EEG results ('+' = significant; '-' = null). CRS-R values are the highest score recorded in the period leading up to and including the research visit. CRS-R values at the time of EEG are reported in parentheses.

Patient ID	Sex	Age	Time post injury	Aetiology	CRS-R	Level 1: auditory	Level 2: perceptual	Level 3: semantic
MCS-1	Male	35	16 y 10 mo	Non-traumatic	13 (6)	+	+	–
MCS-2	Male	33	8 y 2 mo	Traumatic	10 (8)	+	+	–
MCS-3	Male	40	3 y 1 mo	Traumatic	7 (7)	+	+	–
MCS-4	Male	24	1 y 2 mo	Traumatic	17 (17)	+	–	–
MCS-5	Female	55	6 mo	Non-traumatic	12 (7)	+	–	–
MCS-6	Male	44	1 y 1 mo	Traumatic	10 (8)	+	+	–
MCS-7	Male	18	5 mo	Traumatic	10 (9)	+	–	–
MCS-8	Male	30	1 y 2 mo	Traumatic	9 (6)	+	–	+
VS-1	Male	59	5 y 1 mo	Non-traumatic	6 (5)	+	–	–
VS-2	Female	69	3 y 4 mo	Non-traumatic	5 (5)	+	–	–
VS-3	Male	19	3 y 7 mo	Non-traumatic	8 (5)	+	+	–
VS-4	Female	52	6 y 6 mo	Non-traumatic	6 (5)	+	–	–
VS-5	Male	35	3 y 9 mo	Traumatic	5 (5)	+	+	–
VS-6	Female	65	1 y 1mo	Non-traumatic	4 (4)	+	–	–
VS-7	Male	22	6 mo	Traumatic	9 (9)	+	+	–
VS-8	Male	16	9 mo	Non-traumatic	7 (7)	+	–	–

Table 2
CRS-R subscales for each patient.

Patient ID	Auditory	Visual	Motor	Oromotor/verbal	Communication	Arousal
MCS-1	3 – Reproducible movement to command	4 – Object localisation: reaching	2 – Flexion withdrawal	2 – Vocalisation/Oral movement	0 – None	2 – Eye opening without stimulation
MCS-2	1 – Auditory startle	3 – Visual pursuit	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	3 – Attention
MCS-3	1 – Auditory startle	3 – Visual pursuit	0 – None/Flaccid	1 – Oral reflexive movement	0 – None	2 – Eye opening without stimulation
MCS-4	4 – Consistent movement to command	4 – Object localisation: reaching	5 – Automatic motor response	1 – Oral reflexive movement	0 – None	3 – Attention
MCS-5	3 – Reproducible movement to command	2 – Fixation	3 – Localisation to noxious stimulation	2 – Vocalisation/Oral movement	0 – None	2 – Eye opening without stimulation
MCS-6	2 – Localisation to sound	2 – Fixation	2 – Flexion withdrawal	2 – Vocalisation/Oral movement	0 – None	2 – Eye opening without stimulation
MCS-7	2 – Localisation to sound	3 – Visual pursuit	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	2 – Eye opening without stimulation
MCS-8	0 – None	3 – Visual pursuit	2 – Flexion withdrawal	2 – Vocalisation/Oral movement	0 – None	2 – Eye opening without stimulation
VS-1	1 – Auditory startle	1 – Visual startle	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	1 – Eye opening with stimulation
VS-2	1 – Auditory startle	0 – None	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	1 – Eye opening with stimulation
VS-3	2 – Localisation to sound	1 – Visual startle	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	2 – Eye opening without stimulation
VS-4	1 – Auditory startle	0 – None	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	2 – Eye opening without stimulation
VS-5	2 – Localisation to sound	1 – Visual startle	0 – None/Flaccid	1 – Oral reflexive movement	0 – None	1 – Eye opening with stimulation
VS-6	0 – None	0 – None	2 – Flexion withdrawal	1 – Oral reflexive movement	0 – None	1 – Eye opening with stimulation
VS-7	2 – Localisation to sound	1 – Visual startle	2 – Flexion withdrawal	2 – Vocalisation/Oral movement	0 – None	2 – Eye opening without stimulation
VS-8	1 – Auditory startle	1 – Visual startle	1 – Abnormal posturing	2 – Vocalisation/Oral movement	0 – None	2 – Eye opening without stimulation

the task breaks. Specifically, deep muscle pressure was administered in order to prolong the potential vigilance of the patient.

Due to constraints in testing behaviourally non-communicative patients, the lack of sensitivity inherent in passive tasks in the production of the N400 effect (see (Cruse et al., 2014a)), and the desire to reduce motor artifacts, a covert task procedure was implemented. Specifically, all participants were instructed to judge whether each word-pair was related or unrelated by mentally ‘saying’ the word “related” or “unrelated” following the presentation of each pair. Participants were instructed to make this mental judgment quickly, firmly, and efficiently - without debate. All control participants completed this task with their eyes open and fixated on a fixation cross.

2.5. EEG recording procedure

EEG recordings were made using a saline-based (potassium chloride) 129-channel HydroCel Geodesic Sensor Net (EGI Inc., OR, USA). Data were sampled at 250 Hz, referenced to the vertex, with impedances of all channels kept below 50 k Ω . Data were subsequently digitally filtered offline between 0.5 and 25 Hz and epochs created around each stimulus with 100 ms pre-stimulus baseline, and 796 ms post-stimulus. Manual artifact rejection was conducted via visual inspection to remove channels and trials with excessive amplitude variance. Bad channels were interpolated back into the data. All channels were then re-referenced to linked mastoids. Independent Component Analysis (ICA) was used to remove any remaining eye blink and eye movement artifacts using EEGLAB (Delorme and Makeig, 2004), and ERP amplitudes baseline corrected. All pre-processing steps were performed using a combination of custom MATLAB scripts and the open-source toolbox EEGLAB (Delorme and Makeig, 2004).

Across healthy participants, the median number of trials contributing to the analyses were: related targets 95 (range 89–99), unrelated targets 95 (range 89–100), all words 376 (range 356–394), all noises

370.5 (range 355–392). The median number of channels interpolated was 1 (range 0–5).

Across the patient group, the median number of trials contributing to the analyses were: related targets 87 (range 72–93), unrelated targets 86.5 (range 73–96), all words 348 (range 297–383), all noises 340 (range 291–383). The median number of channels interpolated was 9.5 (range 0–34).

2.6. EEG analyses

Three sets of EEG analyses were conducted at increasing levels of a speech processing hierarchy.

2.6.1. Auditory processing

All trial types were considered for this analysis as they all involved the presentation of an auditory stimulus. In order to test for the presence of auditory processing, each time-point post-stimulus was subjected to the topographic consistency test (TCT; Koenig and Melie-García, 2010). The TCT determines whether a significantly consistent scalp distribution of event-related voltages is observed across observations (i.e. trials for single-subject analyses, or participant averages for group analyses). A significantly consistent event-related topography reflects the event-related engagement of a consistent set of brain regions, and therefore a change in EEG activity that is caused by the presentation of an auditory stimulus.

Specifically, at each time-point, the global field power (GFP) - i.e. the standard deviation of voltages across all electrodes - of the grand average ERP is calculated. To test whether this GFP reflects the engagement of a consistent set of neural sources, a randomisation test is performed. Specifically, the electrode labels at each observation that contributed to the grand average ERP are randomly shuffled, averaged, and the resulting GFP recorded. This randomisation is repeated 1000 times to form a surrogate distribution that describes the null hypothesis that

the distribution of voltages in the true data at that time-point are not the result of a consistent set of neural sources across observations. Finally, the p -values at each time-point post-stimulus (0 to 800 ms) were subjected to a False Discovery Rate (FDR) correction ($p < 0.05$, one-tailed) in order to control for multiple comparisons.

2.6.2. Perceptual and semantic processing

As the goal of this research was to identify speech processing at a single-subject level, group-level comparisons were performed on data from healthy participants in order to identify regions of interest (ROIs) for subsequent patient analyses.

For the group-level analyses, the spatiotemporal cluster-mass procedure of the open-source toolbox Fieldtrip was employed (Oostenveld et al., 2011). This procedure uses both parametric and nonparametric statistics to determine significant differences between conditions using spatiotemporal data-points. For group-level analyses, one- or two-tailed dependent samples t -tests were used to compare ERPs across conditions. Spatiotemporally adjacent t -values with p -values < 0.05 were then clustered based on their spatiotemporal proximity. T -values within each cluster were summated, and the largest cluster t retained. Spatiotemporal clusters were defined as at least two statistically significant t -tests for temporally adjacent time-points in the waveform, occurring across at least two spatially neighbouring electrodes (within a 4 cm radius). To correct for multiple comparisons, randomization tests produced 1000 Monte Carlo permutations of the above procedure to test whether the true cluster value occurred by chance (Maris and Oostenveld, 2007). For the perceptual contrast, the ERPs elicited by all words were compared with those elicited by all noises (0–800 ms post-stimulus, two-tailed). For the semantic contrast, the ERPs elicited by unrelated targets were compared with those elicited by related targets. Due to the a priori hypotheses regarding the N400 effect, this contrast was restricted to data from 200 to 800 ms post-stimulus, and was one-tailed (unrelated $<$ related).

Following the group-level analyses, the spatial foci of the significant Perceptual and Semantic effects were identified. Single-subject analyses were then focused on these spatial ROIs. Specifically, electrodes within the identified scalp region were averaged together into one ROI virtual electrode. Voltages were then compared using the same clustering procedure as above, with the exception that clusters are now temporal only, rather than spatiotemporal (see Canales-Johnson et al., 2015 for a similar analysis approach). Analyses in both Perceptual and semantic contrasts were one-tailed in accordance with the directions of the differences from the group analyses.

All ERP comparisons were conducted in Matlab. All other statistical comparisons were conducted with the free software JASP (Love et al., 2015; Morey and Rouder, 2015; Rouder et al., 2009). Equivalent Bayesian comparisons are reported where appropriate. Specifically, to complement the t -tests, the Jeffrey-Zellner-Siow Bayes factor (JZS-BF) tested the strength of the evidence for each observed effect size (Rouder et al., 2009). The default Cauchy distribution with width 0.707 was used as the prior distribution. A JZS-BF between 1/3 and 3 is considered to be only weak/ anecdotal evidence for an effect; from 3 to 10 is substantial evidence; and from 10 to 100 is strong evidence (Jeffreys, 1961). As the Bayes Factor is a ratio of evidence for two hypotheses, the same category descriptions hold for the inverse, i.e. 1/3–1/10 is substantial evidence etc.

2.6.3. Topographic analyses

To test the hypothesis that the Perceptual effect and the Semantic effect did not reflect the activity of entirely overlapping cortical generators, the scalp topographies of these effects were compared using the Global Dissimilarity method (Skrandies, 1990). First, the peaks of each component were visually identified in the GFP of the difference ERPs (i.e. perceptual contrast: all words minus all noises; semantic contrast: unrelated targets minus related targets). Voltages were averaged within

an arbitrary window 100 ms wide, centred at the time-point of peak GFP, and then subjected to the dissimilarity analysis (see Fig. 1).

Global dissimilarity (GD) is the standard deviation of differences between time-windows of interest, after the data have been scaled to the spatial standard deviation (GFP). The significance of the GD values was calculated via randomisation test. Specifically, trial labels were shuffled randomly 1000 times and each time the resulting GD value was recorded. The true GD value was then compared to this distribution in order to calculate a p -value that the true GD occurred by chance ($p < 0.05$).

2.6.4. Source reconstruction

The cortical generators of the ERP effects of interest were estimated with the multiple sparse priors (MSP) distributed source method - an empirical Bayes approach implemented through the MATLAB toolbox SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>). First, grand average difference ERPs were calculated for each contrast - i.e. all words minus all noises, and unrelated targets minus related targets - and re-referenced to the average across all channels. A template cortical mesh with 8196 vertices was co-registered to the template electrode locations. A boundary element forward model (BEM) was calculated, and the inverse solution estimated. Specifically, each contrast and time-window were inverted separately to ensure that the source solution did not overly focus on reducing the error for reconstructing activity outside of the periods of interest (Litvak et al., 2011). ERP data within each time-window of interest were tapered with a Hanning window before inversion in order to focus reconstruction on the sources active at the peaks identified in the GFP - i.e. the centre of the 100 ms-wide window. No prior predictions of source locations were used to constrain the inverse solution. Source power (in arbitrary units) within the 100 ms-wide window of interest was subsequently weighted with a Gaussian and averaged. These averages were then visualised on a template brain and inspected in order to qualitatively describe their correspondence with previous fMRI studies. Source visualisation was performed with the free software MRICron (<http://www.mccauslandcenter.sc.edu/mricron/mricron/>).

3. Results

3.1. Level 1: Auditory processing

3.1.1. Healthy participants

Significantly consistent topographies could be identified in the ERPs of all healthy participants, with a mean onset of 59 ms (SD = 16 ms) and a range of 32 ms–92 ms (Fig. 2). Significant topographies were evident for a mean of 712 ms (SD = 57 ms) with a range of 568 ms–768 ms.

3.1.2. Patients

Significant topographies were evident in the ERPs of all 16 patients. The mean onset time across all patients was 86 ms (SD = 84 ms) and ranged from 0 ms–308 ms (Fig. 3). Patient VS-5 demonstrated an onset at 0 ms, suggesting a myographic startle response. Relative to healthy controls, there was no significant difference in the onset time of the significant topographies across all patients (unequal variance $t(16.12) = 1.251$, $p = 0.229$, two-tailed; Levene's test $F(1) = 11.22$, $p = 0.002$). A Bayesian t -test similarly concluded that the evidence for a difference was not convincing (JZS BF10 = 0.609).

Significant topographies were evident for a mean of 617 ms (SD = 180 ms) with a range of 64 ms–792 ms. This was borderline significantly different from healthy controls in an independent samples t -test (unequal variance $t(17.96) = 2.010$, $p = 0.060$; Levene's test $F(1) = 8.270$, $p = 0.007$), although a Bayesian t -test indicated that this difference was not convincing (JZS BF10 = 1.504, anecdotal evidence).

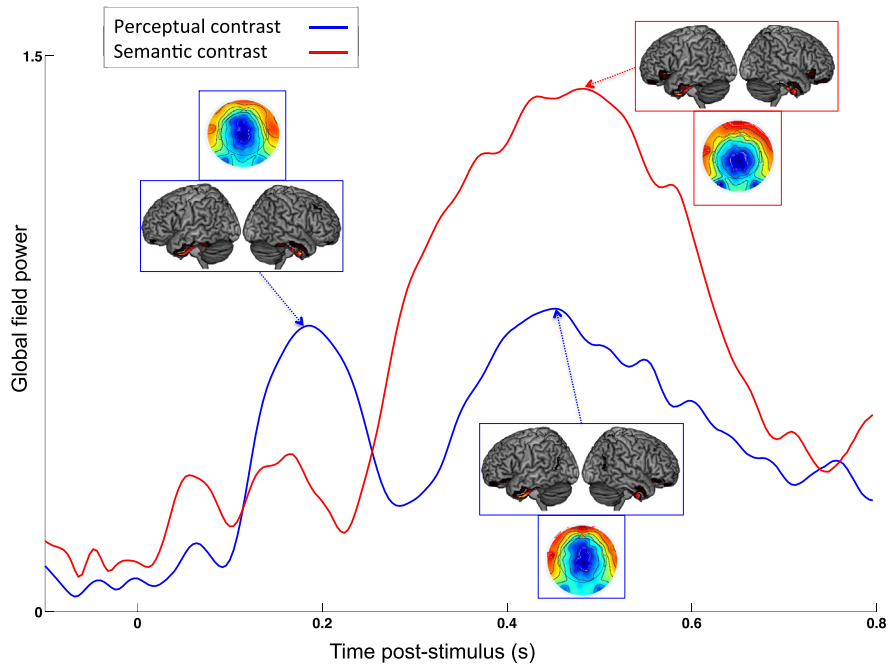


Fig. 1. Global field power of difference ERPs in each contrast (i.e. perceptual: all speech minus all noise; Semantic: all unrelated targets minus related targets). Cortical maps show source estimates of the cortical generators of the ERPs at the respective peaks (see Section 2.6.4). Scalp plots show the distribution of mean voltage differences at the peaks.

3.2. Level 2: perceptual processing

3.2.1. Healthy participants

At the group level, the ERPs elicited by all words were significantly more negative than the ERPs elicited by all noises from 92 ms–796 ms across midline fronto-central-parietal scalp (see Fig. 4). Single-subject analyses were subsequently focused on these scalp sites (see Fig. 5) and constrained to data from 100 ms–800 ms post-stimulus.

Significant single-subject speech-specific effects were evident in the data of 14 out of 16 healthy participants (88%, see Fig. 5) with a mean onset time of 188 ms (SD = 132 ms).

3.2.2. Patients

Significant single-subject Perceptual effects were evident in the data of 7 out of 16 patients (3 VS; 4 MCS) with a mean onset time across patients of 354 ms (SD = 187 ms). Onset time was significantly later than

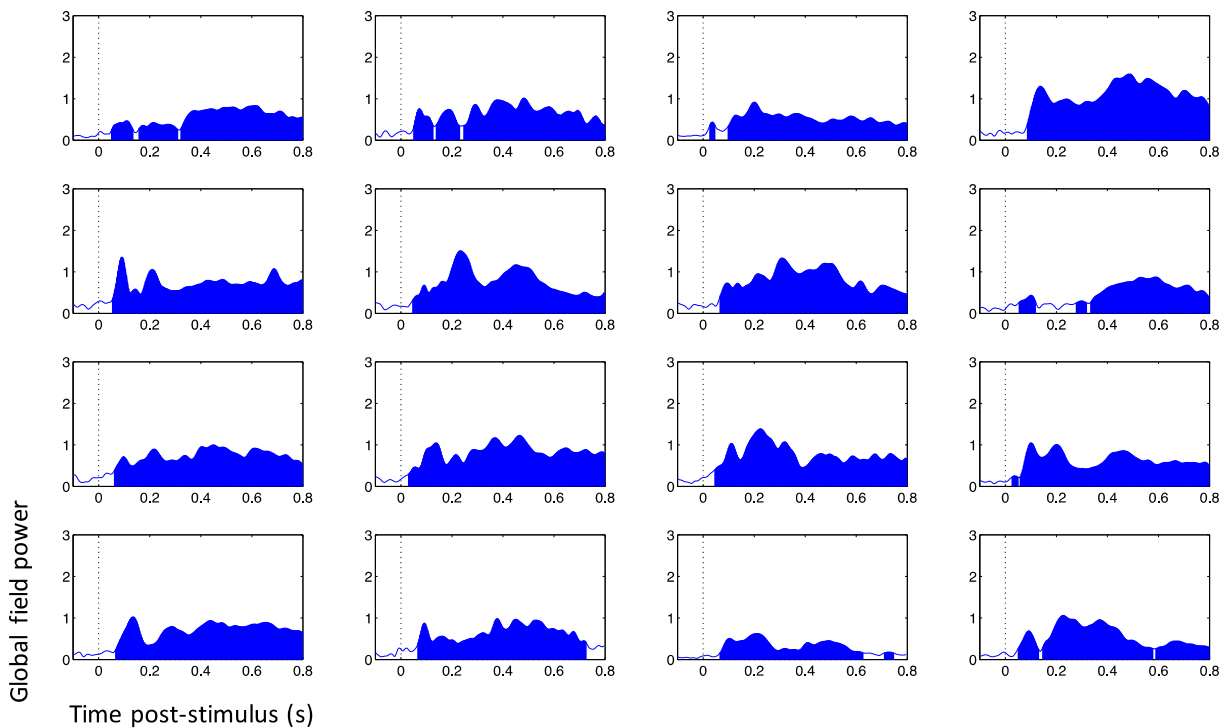


Fig. 2. Healthy control group single subject global field power from all stimuli. Time-points that passed the topographic consistency tests ($p < 0.05$ FDR) are shaded. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

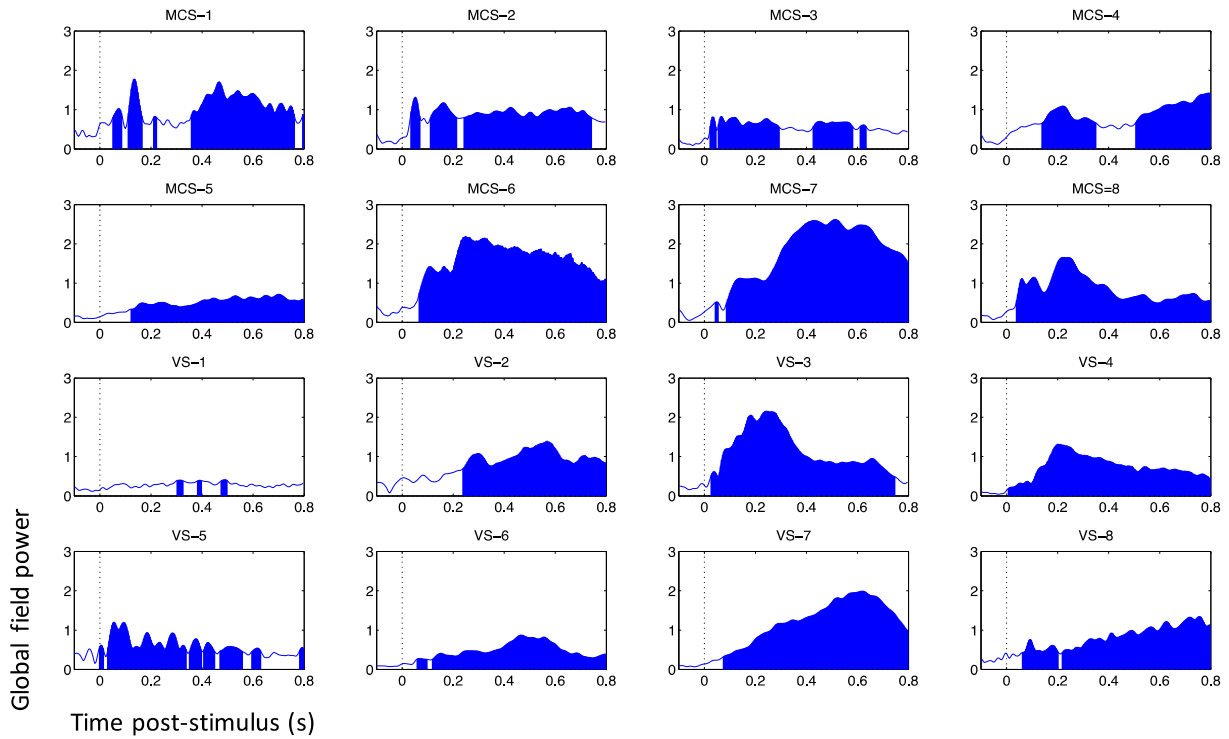


Fig. 3. Patient group single subject global field power from all stimuli. Time-points that passed the topographic consistency tests ($p < 0.05$ FDR) are shaded.

in healthy controls (equal variance $t(19) = 2.368, p = 0.029$, two-tailed; Levene's test $F(1) = 2.441, p = 0.135$), although a Bayesian t test indicated that the evidence for this difference was only marginally convincing (JZS $BF_{10} = 2.454$). The CRS-R total score, CRS-R score on

the day of EEG assessment, and its individual subscales, did not significantly differ between patients exhibiting Perceptual effects and those not (all Mann Whitney $p > 0.122$). Furthermore, the proportions of patients with traumatic/non-traumatic aetiologies did not differ between these groups (Fisher's Exact test $p = 0.315$). All patient ERPs are shown in Fig. 6.

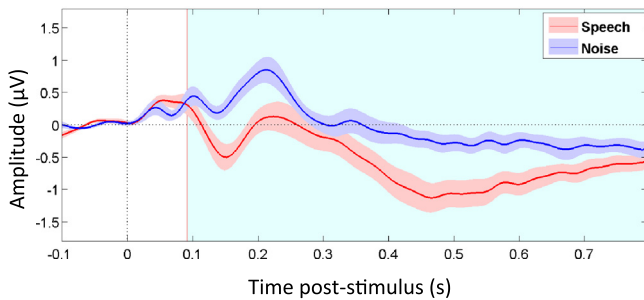
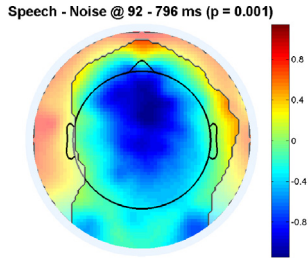


Fig. 4. Grand average ERP effects in the perceptual contrast (all words versus all noises) from 92 to 796 ms post-stimulus. The upper panel shows the spatial distribution of the average voltage across this time-window. The spatial extent of the significant spatiotemporal cluster (i.e. all electrodes that contributed to the cluster) is outlined. The colour bar shows average amplitude differences between ERP categories (speech minus noise) across the temporal extent of the spatiotemporal cluster. The lower panel shows the means of the ERPs within the respective spatial clusters (± 1 standard error). The temporal boundaries of the cluster are shaded in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Level 3: semantic processing

3.3.1. Healthy participants

At the group level, the ERPs elicited by unrelated targets were significantly more negative than those elicited by related targets over centro-parietal scalp from 200 ms–796 ms post-stimulus (see Fig. 7).

Significant single-subject semantic effects were evident in the data of 14 out of 16 healthy participants (88%, see Fig. 8) with a mean onset time of 328 ms (SD = 93 ms).

3.3.2. Patients

One patient (MCS-8) elicited a significant semantic effect ($p = 0.026$; see Fig. 9 for all patient ERPs).

3.4. Healthy topographic analyses

The group-level early and late Perceptual effects identified in the GFP (peaks at 184 ms and 452 ms, Speech minus Noise; see Section 2.6.3 and Fig. 1) exhibited fronto-central scalp distributions, which did not significantly differ from one another ($GD = 0.292, p = 0.558$). Both the early and late Perceptual effects had significantly different scalp distributions to the centro-parietal semantic effect (peak at 484 ms unrelated targets minus related targets; $GD = 0.584, p = 0.001$).

3.5. Source reconstruction

Cortical sources of both the early and late Perceptual effects were primarily estimated to be within bilateral medial and inferior temporal gyri (see Fig. 1). The semantic effect was estimated to originate within

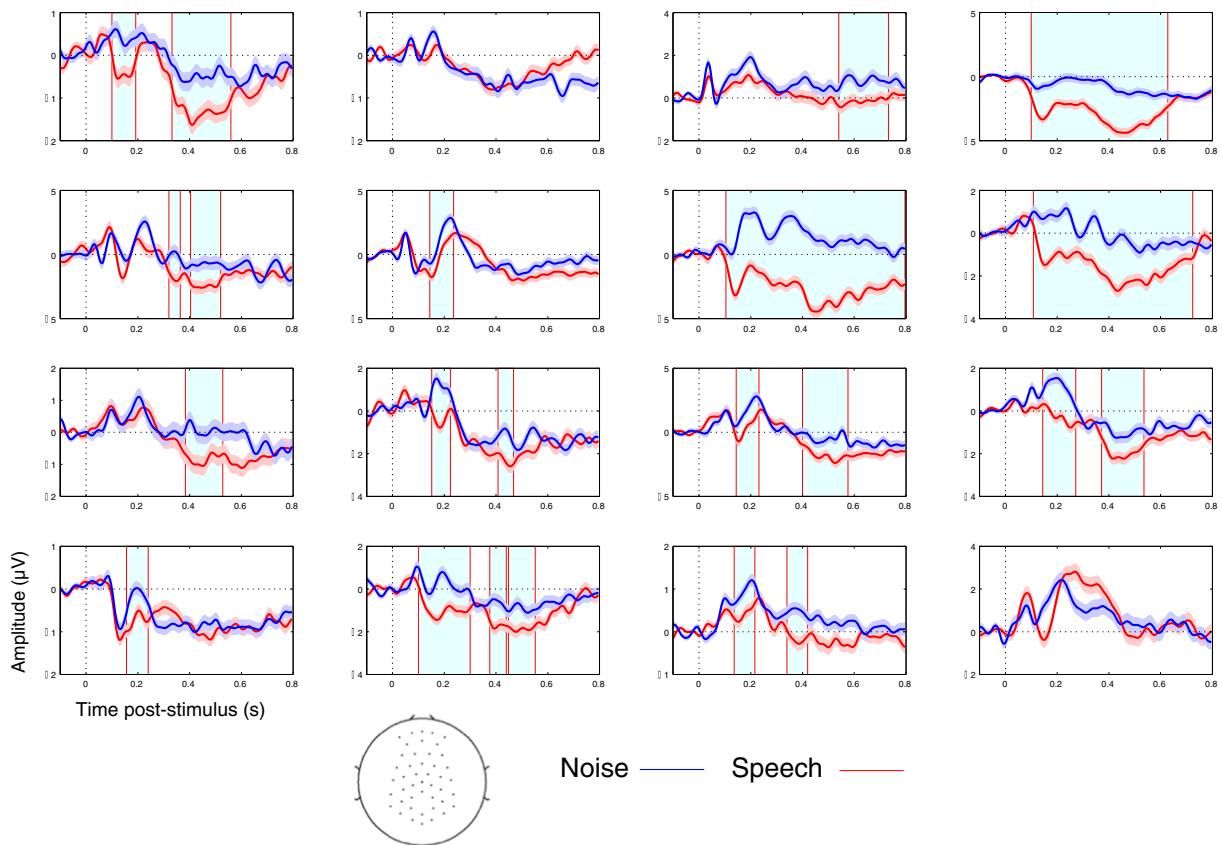


Fig. 5. Healthy control group single subject ERP effects in the perceptual contrast. Each panel shows the average ERPs within the spatial region of interest indicated in the scalp plot (i.e. averaged across trials and electrodes). Significant clusters are shaded in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

bilateral medial and inferior temporal gyri and bilateral inferior frontal gyri (see Fig. 1).

3.6. Differences between VS and MCS

There were no differences between VS and MCS patients in the time of onset of the significant single-subject topographies (VS: $M = 101$ ms, $SD = 112$ ms; MCS: $M = 70$ ms, $SD = 43$ ms; unequal variance $t(8.98) = 0.727$, $p = 0.486$, two-tailed; Levene's test $F(1) = 5.149$, $p = 0.040$; JZS $BF_{10} = 0.511$, anecdotal evidence) nor in the temporal extent of a significant topography (VS: $M = 610$ ms, $SD = 234$ ms; MCS: $M = 624$ ms, $SD = 121$ ms; equal variance $t(14) = 0.156$, $p = 0.879$, two-tailed; JZS $BF_{10} = 0.431$, anecdotal evidence). In fact, tests of the robustness of the JZS Bayes Factors revealed that widening the prior distribution of effect sizes (Cauchy width > 1.25 and > 1 respectively) led to moderate evidence that the time of onset and the temporal extent of significant topographies are the same across VS and MCS patients (i.e. $BF_{10} < 1/3$).

Four out of eight MCS patients and three out of eight VS patients exhibited significant speech vs noise effects. One MCS patient and no VS patients exhibited significant semantic effects. These proportions were not significantly different between VS and MCS.

4. Discussion

The primary aim of this study was to determine the potential for ERPs to provide a hierarchical measure of auditory processing in patients with disorders of consciousness. The highest level of the hierarchy – semantic processing – could be detected in 88% of healthy participants, as could the intermediate ‘Perceptual’ level, while auditory processing could be detected in 100% of healthy participants.

Semantic processing was identified through the classic contrast of related and unrelated targets in a word-pair priming task (Kutas and Federmeier, 2011). Consistent with decades of research on the N400 effect, a significant group-level centro-parietal difference was observed between the ERPs elicited by primed and unprimed targets from 200 to 800 ms post-stimulus (Fig. 7). Furthermore, it was possible to identify this effect in 88% of healthy single-subjects who were instructed to perform covert judgments of the relatedness of the words in each word pair. While the precise cognitive mechanisms indexed by the N400 effect are still a matter of debate (Kutas and Federmeier, 2011), evidence suggests that it reflects the larger semantic prediction error signals elicited by speech that has not been primed (Rabovsky and McRae, 2014). The current study was not designed to speak to the theoretical underpinning of the N400 effect. However, source estimation implicated bilateral medial/inferior temporal and inferior frontal cortex in generating the N400 effect (Fig. 1). This is consistent with previous source estimates (see Lau et al., 2008) as well as theories of language that consider speech comprehension to emerge from top-down, frontal-lobe dependent predictions of incoming auditory information (e.g. McClelland and Elman, 1986). Furthermore, the source estimates of the N400 effect are broadly consistent with the pattern of BOLD activation seen by Coleman et al. (2009, 2007; Davis et al., 2007) in their speech ‘comprehension’ contrast between sentences containing ambiguous and unambiguous words – i.e. left inferior temporal lobe and left inferior frontal gyrus. This suggests that the semantic effect identified with ERPs in this study overlaps with the neural and cognitive mechanisms isolated by the speech comprehension contrast of Coleman et al.’s (2009, 2007) fMRI studies, and therefore that the patient results may be compared across studies.

The perceptual contrast in healthy individuals revealed differences in the processing of speech and noise within 100 ms of stimulus onset

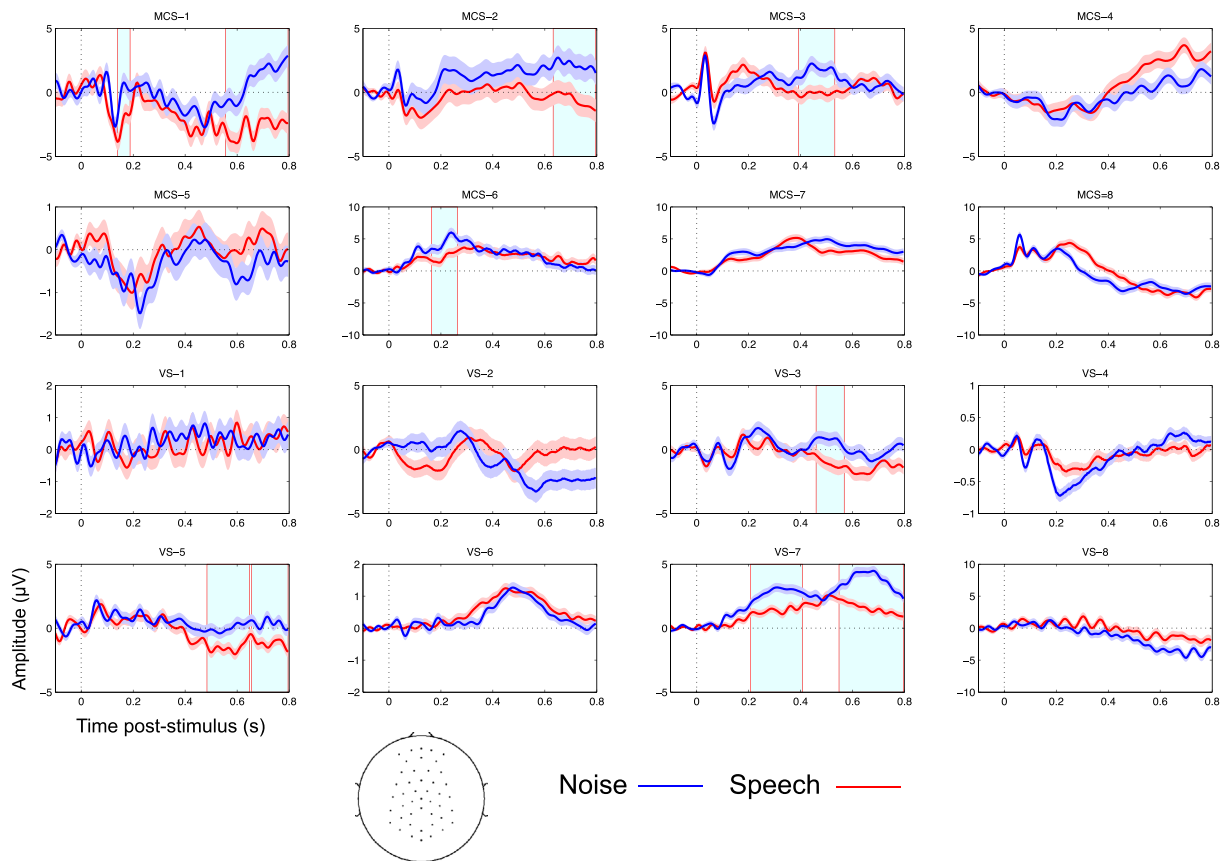


Fig. 6. Patient group single subject ERP effects in the perceptual contrast. Each panel shows the average ERPs within the spatial region of interest indicated in the scalp plot (i.e. averaged across trials and electrodes). Significant clusters are shaded in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 4). The differences in ERPs between these stimuli types exhibited two clear component peaks at 184 ms and 452 ms post-stimulus (Fig. 1). While the scalp distributions of these components did not significantly differ from one another, they both differed from the scalp distribution of the N400 effect (i.e. the semantic contrast). Indeed, source estimation implicated bilateral inferior temporal lobes in the generation of the Perceptual effects, compared to the fronto-temporal cortical generators of the N400 effect (Fig. 1).

The early onset and primarily temporal lobe source estimates of the early Perceptual effect indicate that it reflects differences in sensory processing of speech and noise, perhaps specifically the acoustic-phonetic properties of speech (see Davis et al., 2007, for a similar argument). The late Perceptual effect has a similar morphology and time-course to the N400 effect, with a relatively more frontal scalp distribution. The similarity of this effect to the N400 effect suggests that it reflects cognitive aspects of speech processing, perhaps even semantic processing. Indeed, speech and noise differ not only in terms of their acoustics, but also in more abstract properties. However, its significantly different scalp distribution and source contribution relative to the semantic N400 effect indicates that it does not reflect entirely the same cognitive process. As the late Perceptual effect is produced through the contrast of not only the targets of the priming manipulation, but all words and all noises in the experiment, it is possible that it reflects the access of semantic information, while the N400 effect itself reflects semantic prediction errors that are primarily associated with unrelated targets (Rabovsky and McRae, 2014).

Taken together, the observed ERP effects and source estimations from healthy individuals therefore suggest that this paradigm is capable of identifying a hierarchy of auditory processes in a similar way to previous fMRI studies. Specifically, while speech perception recruits bilateral inferior temporal areas, speech comprehension and the

processing of semantic mismatch additionally recruits bilateral inferior frontal areas. Next, we consider the patient results at each level of the hierarchy.

All patients reported here exhibited evidence of auditory evoked potentials as indexed by a significantly consistent distribution of voltages across the head – i.e. significance in a topographic consistency test (Koenig and Melie-García, 2010; see Figs. 2 and 3). This proportion is in contrast with the 60% of VS patients and 11% of MCS patients who showed no significant differences in a study of fMRI-detected neural activity between silence and auditory stimulation. Indeed, the detection rate in VS patients reported here is significantly higher than previously that reported with fMRI (Fisher's Exact test: $p = 0.004$; (Coleman et al., 2009)). Furthermore, a Bayesian contingency analysis indicated that these data were considerably more likely (54.88 times) under the hypothesis that the detection rates differ between our EEG method and the fMRI method of Coleman et al. (2009), than under a null effect hypothesis (Poisson sampling; Günel and Dickey, 1974; Jamil et al., 2015; Love et al., 2015; Morey and Rouder, 2015). The detection rates did not differ significantly for MCS patients (Fisher's Exact test $p = 1$), and a Bayesian contingency analysis indicated that there was no convincing evidence in favour of either hypothesis (Poisson sampling; BF = 0.511).

It is possible that the reported EEG method for detecting auditory processing (i.e. the topographic consistency test, Koenig and Melie-García, 2010) does not isolate the same processes as the fMRI contrast of Coleman et al. (i.e. all sounds versus silence; Coleman et al., 2009, 2007; Davis et al., 2007), and therefore the detection rates are not necessarily comparable. The topographic consistency test determines the presence of scalp voltage distributions that are significantly consistent across observations (i.e. presentations of auditory stimuli), and therefore provides evidence of auditory evoked potentials, or simple changes

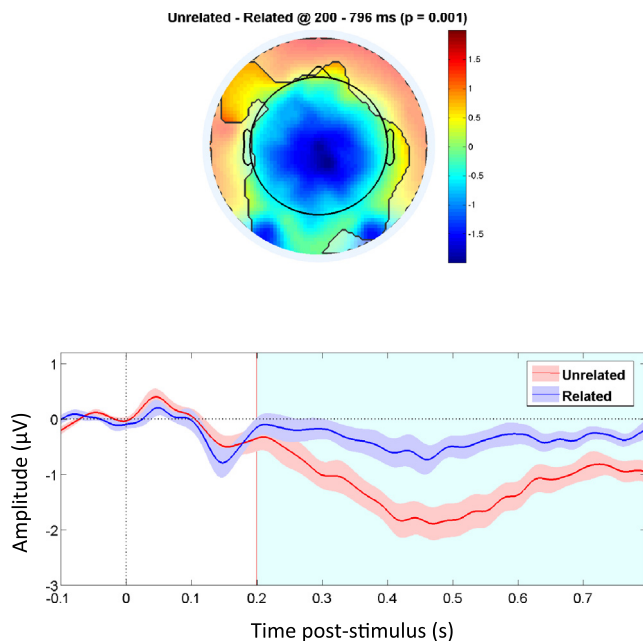


Fig. 7. Grand average ERP effects in the semantic contrast (all unrelated targets < all related targets) from 200 to 796 ms post-stimulus. The upper panel shows the spatial distribution of the average voltage across this time-window. The spatial extent of the significant spatiotemporal cluster (i.e. all electrodes that contributed to the cluster) is outlined. The colour bar shows average amplitude differences between ERP categories (unrelated targets minus related targets) across the temporal extent of the spatiotemporal cluster. The lower panel shows the means of the ERPs within the respective spatial clusters (± 1 standard error). The temporal boundaries of the cluster are shaded in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in cortical processing elicited by an auditory stimulus. Furthermore, this level of processing may not be evident externally. In behavioural assessments, patients VS-6 and MCS-8 did not exhibit evidence of auditory startle – a brainstem mediated reflex (Yeomans and Frankland, 1995) – and yet demonstrated changes in cortical processing after the presentation of an auditory stimulus, thereby hinting at increased sensitivity of this method relative to behavioural examination. As we do not have complementary fMRI data for our healthy participants or patients, it is unclear whether this EEG method is in turn more sensitive than fMRI, or whether they measure qualitatively different processes. Furthermore, the presence of evoked potentials was not diagnostic (i.e. did not differ between VS and MCS; see Section 3.6) in this particular contrast.

Significant Perceptual effects were observed in the data of several patients – 50% (4/8) of MCS patients, and 38% (3/8) of VS patients. Coleman et al. (2009) reported significant effects in a similar fMRI contrast (i.e. speech versus noise) in 47% (9/19) of MCS patients and 23% (5/22) of VS patients, proportions that do not differ significantly from those observed here (Fisher's Exact tests: VS $p = 0.643$, MCS $p = 1$). Bayesian contingency analyses indicated that the evidence was not convincing for either the hypothesis that the proportions differed, or that they did not differ (Poisson sampling; VS BF = 0.905; MCS BF = 0.773).

As with the auditory processing analysis, a significant Perceptual effect in the current paradigm did not differentiate VS and MCS patients. As there are only a small number of patients with significant effects in this contrast, it is not possible to conduct robust comparisons for each diagnosis separately (i.e. 3/8 VS patients and 4/8 MCS patients). However, when comparing all patients who elicited significant Perceptual effects with all patients who did not, regardless of their diagnoses, there were no behavioural differences between the groups (see Section 3.2.2).

At the highest level of our auditory hierarchy, one patient (MCS-8) elicited a statistically significant N400 effect. As can be seen in Figs. 8 and 9, the onset of this effect is somewhat later than in the individual

healthy controls, although the period of time over which it is significant is consistent with healthy single-subject and group statistics. Interestingly, this patient did not produce an auditory startle during behavioural examination. Nevertheless, this patient's EEG shows clear evoked potentials, indicating once again that ERPs can characterise unseen abilities of patients.

However, in the current manuscript, the ERPs of 16 patients with disorders of consciousness were tested for the presence of an N400 effect; one of these tests was significant (i.e. MCS-8). With a statistical threshold of 5% employed in each independent test, it is possible that the 1/16 (6.25%) significant tests reflect a consequence of multiple comparisons, rather than the true presence of an N400 effect. Indeed, the statistical significance of the apparent N400 effect of MCS-8 ($p = 0.026$) does not survive FDR correction across all ERP comparisons in this study (corrected threshold = 0.021; see also Noirhomme et al., 2014). In a clinical context, cross-subject multiple comparisons corrections are not feasible as the number of comparisons will increase with each new patient that is tested. Furthermore, efforts to reduce the rate of false positives will simultaneously reduce the rate of true positives. Clearly, when selecting a statistical threshold against which to judge the evidence for covert cognition, the consequences of a false positive must be weighed against the consequences of a false negative (Cruse et al., 2014b). In the context of the current research data, we would nevertheless conclude that there is insufficient evidence to rule out the possibility that the N400 effect observed in the data of MCS-8 is a consequence of multiple comparisons across patients.

The ability to detect the N400 effect in single-subjects is especially poor during passive listening, with only a ~50% reported hit rate (Cruse et al., 2014a; Rohaut et al., 2015). The healthy participants in the current study were instructed to make covert judgments of relatedness of the word pairs, a manipulation which is known to increase single-subject hit rates, and likely contributed to the higher 88% hit rate here. However, the importance of making judgments of the relatedness of the stimuli on the likelihood of detecting an N400 effect likely leads to a high rate of false negatives in patients with brain injuries. Indeed, even a patient who is capable of processing speech may nevertheless lack the necessary attentional resources to follow the task demands. This data together argue that ERP analyses of the N400 effect may not be a practical means of identifying covert speech comprehension in severely brain-injured individuals.

Multiple analyses methods have been employed to identify ERP effects in disorders of consciousness. For example, visual inspection (Schoenle and Witzke, 2004), traditional and contemporary analyses of ERP voltages (Fischer et al., 2010; Rohaut et al., 2015), analyses of wavelet-transformed data (Kotchoubey et al., 2005; Steppacher et al., 2013), and machine learning (King et al., 2013). Multivariate approaches may be more sensitive than mass univariate approaches (employed here) as the relationships between data-points are specifically taken into account. Indeed, it is evident that the sensitivity to detect the N400 effect will vary across analysis approaches, and the identity of the optimal approach is an empirical question. However, the detection rate of 50% or less for the N400 effect with several methods (Cruse et al., 2014a; Rohaut et al., 2015) when participants are passively-listening, as could be assumed to be the case in patients who are unable to follow the task instructions, suggests that there is more to the reported poor sensitivity than the specifics of the analyses method. Rather, poor sensitivity may also reflect a fundamental paradigmatic issue with the use of priming manipulations to identify semantic processing in passively-listening single-subjects.

When a significant N400 effect is observed, what can be said about the level of awareness of the individual patient? Davis et al. (2007) found that the fMRI markers of speech comprehension ceased to be detectable under only moderate sedation. Nonetheless, semantic priming has been observed under a range of reduced awareness manipulations, such as subliminal priming and inattention (Deacon and Shelley-Tremblay, 2000; Dehaene et al., 2001, 1998), indicating that

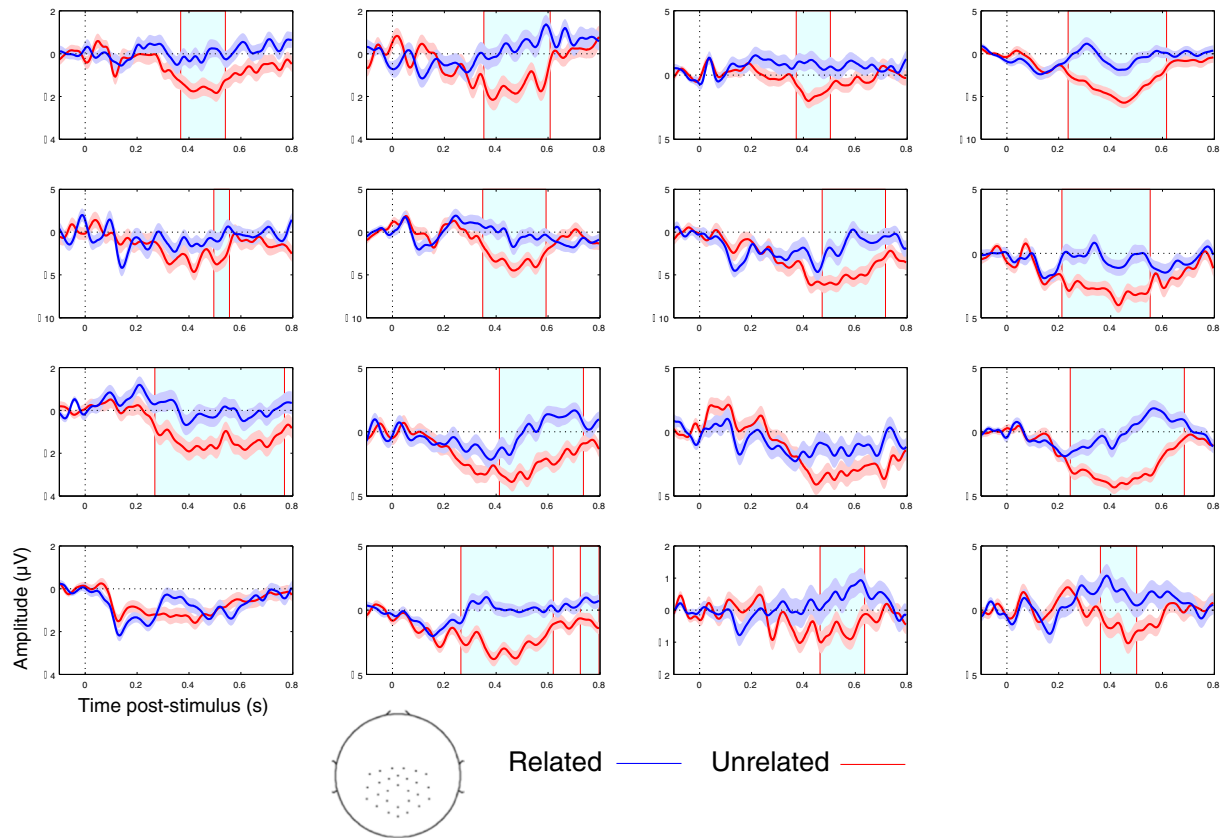


Fig. 8. Healthy control group single subject ERP effects in the semantic contrast. Each panel shows the average ERPs within the spatial region of interest indicated in the scalp plot (i.e. averaged across trials and electrodes). Significant clusters are shaded in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the priming N400 effect is not a reliable marker of awareness of a speech stimulus. Nevertheless, while the individual may not need to be aware of the stimulus in order for semantic priming to occur, perhaps the individual themselves needs to be aware. However, N400 effects and other markers of semantic processing have been reported in sleep (Ibáñez et al., 2009; Kouider et al., 2014), suggesting that an individual does not need to be awake or aware in order to process the meaning of speech. Therefore, the presence of a significant N400 effect in the EEG of a behaviourally non-responsive individual would not be sufficient evidence that the individual was aware. Rohaut et al. (2015) have argued that a later P600 potential indicates awareness of the stimulus in a semantic priming task, as this potential was only seen in MCS patients and fully conscious individuals. We did not see evidence of P600 waveforms in our healthy data, even at the group level (smallest cluster $p > 0.160$, one-tailed analysis of data 500–800 ms post-stimulus, methods as Section 2.6.2). It is unclear why the design of Rohaut et al. (2015) elicited P600s so robustly and ours did not, as this component has been primarily associated with syntactic mismatches, although certain types of semantic mismatches also elicit P600s (see Kuperberg, 2007). Nevertheless, while the presence of an N400 effect is not necessarily a marker of consciousness, its putative fronto-temporal neural generators would suggest the preservation of a relatively complex set of neural processes in a patient who demonstrated such an N400 effect, and may thereby provide an indirect marker of the extent of neural damage.

What then would the absence of a significant N400 effect say about the level of awareness of an individual? As has been described at length elsewhere (Cruse et al., 2014b), the absence of evidence for awareness does not equate to evidence for the absence of awareness. Indeed, the lack of a reliable N400 effect is a null result, and as such is not interpretable. For example, two healthy participants who were demonstrably

conscious and able to understand speech did not elicit statistically reliable N400 effects in this study. Furthermore, patient MCS-1 did not elicit an N400 effect, and yet was able to follow commands with his behaviour, and produced evidence of comprehension of the plot of a short movie clip with fMRI (Naci et al., 2014). This once again highlights the importance of caution in interpreting brain-imaging results that, due to a patient's lack of responsiveness, cannot be verified behaviourally (Cruse et al., 2014b).

5. Conclusion

The accuracy of diagnoses and prognoses of patients with disorders of consciousness may be improved with functional neuroimaging of covert perceptual and cognitive processes (e.g. EEG, ERP, fMRI; Fernandez-Espejo and Owen, 2013). ERPs provide one means with which to identify a hierarchy of auditory processes. The poor sensitivity to detect ERP markers of covert semantic processing – i.e. the N400 effect – in passively listening participants suggests poor clinical utility for semantic priming tasks. Nevertheless, some patients with disorders of consciousness exhibited levels of auditory processing that were not evident in standardised behavioural assessments. These ERP measures did not differentiate VS from MCS, nor do they necessarily indicate consciousness in healthy individuals, suggesting a lack of diagnostic specificity. Indeed, other functional neuroimaging markers fail to distinguish VS from MCS, perhaps due to the diagnostic variability within these groups (Coleman et al., 2009; Fernandez-Espejo et al., 2008; Kotchoubey et al., 2005). Nevertheless, the current ERP measures may provide an indirect marker of the relative preservation of cortical sensory-cognitive processing, and thereby provide prognostic value. However, this is an empirical question worthy of future study.

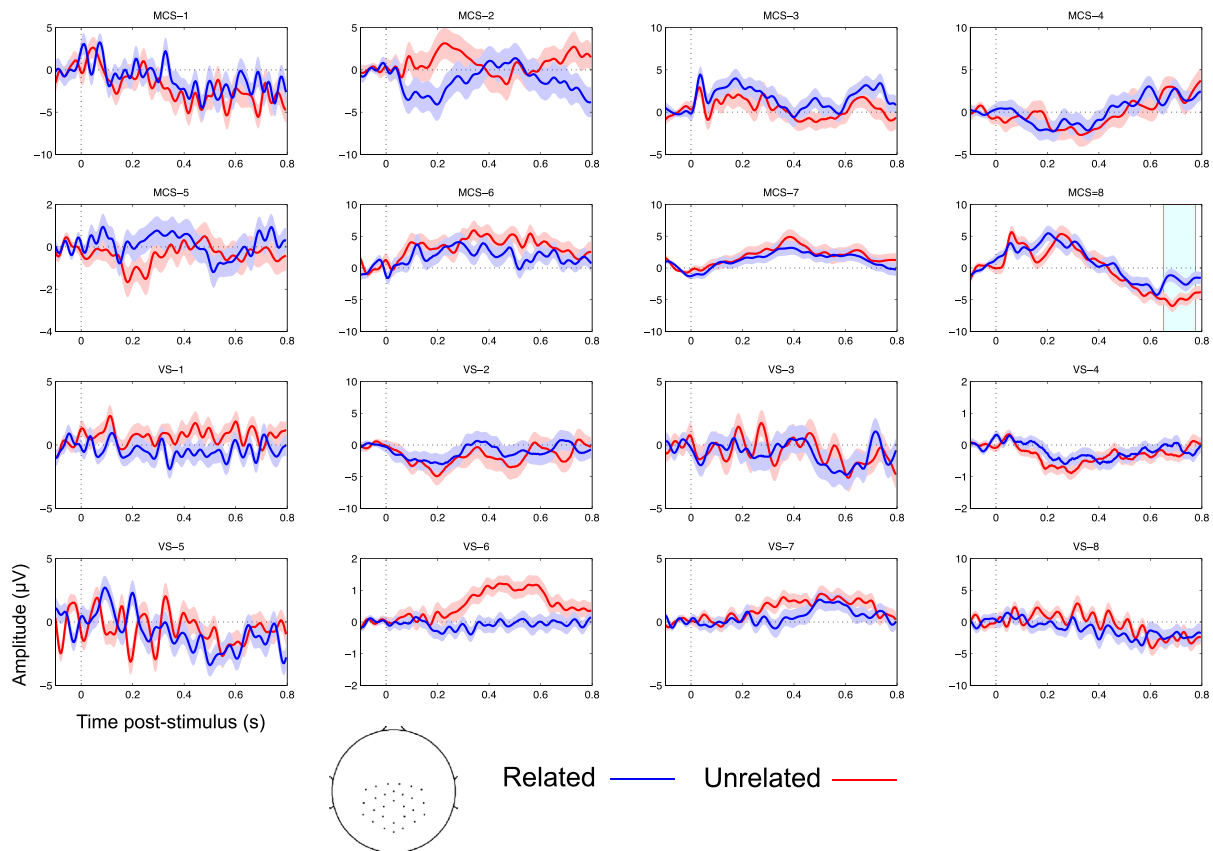


Fig. 9. Patient control group single subject ERP effects in the semantic contrast. Each panel shows the average ERPs within the spatial region of interest indicated in the scalp plot (i.e. averaged across trials and electrodes). Significant clusters are shaded in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Funding

Funding was provided by the Canadian Institutes of Health Research (CIHR) (MFE 123754), Canada Excellence Research Chairs Program (CERC) (215063), National Institute for Health Research (NIHR), Medical Research Council (MRC), and the NIHR Brain Injury Healthcare Technology Cooperative (HTC).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.nicl.2016.08.003>.

References

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

Bentin, S., Kutas, M., Hillyard, S.A., 1993. Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology* 30, 161–169.

Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.

Canales-Johnson, A., Silva, C., Huepe, D., Rivera-Rei, A., Noreika, V., del Carmen Garcia, M., Silva, W., Ciraolo, C., Vaucheret, E., Sedeño, L., Couto, B., Kargieman, L., Baglivo, F., Sigman, M., Chennu, S., Ibáñez, A., Rodríguez, E., Bekinschtein, T.A., 2015. Auditory feedback differentially modulates behavioral and neural markers of objective and subjective performance when tapping to your heartbeat. *Cereb. Cortex* 25, 4490–4503.

Childs, N.L., Mercer, W.N., Childs, H.W., 1993. Accuracy of diagnosis of persistent vegetative state. *Neurology* 43, 1465–1467.

Coleman, M.R., Rodd, J.M., Davis, M.H., Johnsrude, I.S., Menon, D.K., Pickard, J.D., Owen, A.M., 2007. Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain* 130, 2494–2507.

Coleman, M.R., Davis, M.H., Rodd, J.M., Robson, T., Ali, A., Owen, A.M., Pickard, J.D., 2009. Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain* 132, 2541–2552.

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.

Cruse, D., Gantner, I., Soddu, A., Owen, A.M., 2014b. Lies, damned lies, and diagnoses: estimating the clinical utility of assessments of covert awareness in the vegetative state. *Brain Inj.* 28, 1197–1201.

Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.M., Johnsrude, I.S., Matta, B.F., Owen, A.M., Menon, D.K., 2007. Dissociating speech perception and comprehension at reduced levels of awareness. *Proc. Natl. Acad. Sci. U. S. A.* 104, 16032–16037.

Deacon, D., Shelley-Tremblay, J., 2000. How automatically is meaning accessed: a review of the effects of attention on semantic processing. *Front. Biosci.* 5, 82–94.

Dehaene, S., Naccache, L., Le Clec, H.G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.F., Le Bihan, D., 1998. Imaging unconscious semantic priming. *Nature* 395, 597–600.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.-F., Poline, J.-B., Rivière, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.

Edmonds, B.A., James, R.E., Utev, A., Vestergaard, M.D., Patterson, R.D., Krumbholz, K., 2010. Evidence for early specialized processing of speech formant information in anterior and posterior human auditory cortex. *Eur. J. Neurosci.* 32, 684–692.

Fernandez-Espejo, D., Owen, A.M., 2013. Detecting awareness after severe brain injury. *Nat. Rev. Neurosci.* 14, 801–809.

Fernández-Espejo, D., Junqué, C., Vendrell, P., Bernabeu, M., Roig, T., Bargalló, N., Mercader, J.M., 2008. Cerebral response to speech in vegetative and minimally conscious states after traumatic brain injury. *Brain Inj.* 22 (11), 882–890.

Fischer, C., Luauté, J., Morlet, D., 2010. Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. *Clin. Neurophysiol.* 121, 1032–1042.

Günel, E., Dickey, J., 1974. Bayes factors for independence in contingency tables. *Biometrika* 61, 545–557.

Ibáñez, A.M., Martín, R.S., Hurtado, E., López, V., 2009. ERPs studies of cognitive processing during sleep. *Int. J. Psychol.* 44, 290–304.

Jamil, T., Ly, A., Morey, R.D., Love, J., Marsman, M., Wagenmakers, E.J., 2015. Default “Günel and Dickey” Bayes Factors for Contingency Tables (In submission).

Jaramillo, M., Alku, P., Paavilainen, P., 1999. An event-related potential (ERP) study of duration changes in speech and non-speech sounds. *Neuroreport* 10, 3301.

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., Alho, K., 2001. Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cogn. Brain Res.* 12, 459–466.

- Jeffreys, H., 1961. *Theory of Probability*. Oxford University Press, Oxford, UK.
- Jennett, B., Plum, F., 1972. Persistent vegetative state after brain damage. A syndrome in search of a name. *Lancet* 1, 734–737.
- Kalmar, K., Giacino, J.T., 2005. The JFK coma recovery scale—revised. *Neuropsychol Rehabil* 15, 454–460.
- King, J.R., Faugeras, F., Gramfort, A., Schurger, A., Karoui, E.I., Sitt, J.D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T.A., Cohen, L., Naccache, L., Dehaene, S., 2013. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage* 83, 726–738.
- Koenig, T., Melie-García, L., 2010. A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topogr.* 23, 233–242.
- Kotchoubey, B., Lang, S., Mezger, G., Schmalohr, D., Schneck, M., Semmler, A., Bostanov, V., Birbaumer, N., 2005. Information processing in severe disorders of consciousness: vegetative state and minimally conscious state. *Clin. Neurophysiol.* 116, 2441–2453.
- Kouider, S., Andriillon, T., Barbosa, L.S., Goupil, L., Bekinschtein, T.A., 2014. Inducing task-relevant responses to speech in the sleeping brain. *Curr. Biol.* 24 (18), 2208–2214.
- Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res.* 1146, 23–49.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.
- Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.
- Laureys, S., Celesia, G.G., Cohadon, F., Lavrijsen, J., Leon-Carrion, J., Sannita, W.G., Szabon, L., Schmutzhard, E., von K.R., W., Zeman, A., Dolce, G., 2010. Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. *BMC Med* 8, 68.
- Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., Friston, K., 2011. EEG and MEG data analysis in SPM8. *Comput. Intell. Neurosci.* 2011, 1–32.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A.J., Ly, A., Gronau, Q.F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J.N., Morey, R.D., Wagenmakers, E.J., 2015. *JASP (Version 0.7.1)* [Computer Software].
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.
- McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86.
- Morey, R.D., Rouder, J.N., 2015. *BayesFactor (Version 0.9.11-3)* [Computer Software].
- Naci, L., Cusack, R., Anello, M., Owen, A.M., 2014. A common neural code for similar conscious experiences in different individuals. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14277–14282.
- Nelson, D.L., McEvoy, C.L., Schreiber, T.A., 1998. The University of South Florida Word Association, Rhyme, and Word Fragment Norms. URL: <http://web.usf.edu/FreeAssociation/> (accessed 4.22.14).
- Noirhomme, Q., Lesenfants, D., Gomez, F., Soddu, A., Schrouff, J., Garraux, G., Luxen, A., Phillips, C., Laureys, S., 2014. Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. *NeuroImage: Clinical* 4, 687–694.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9.
- Physicians, T.R.C.O., 2003. The vegetative state: guidance on diagnosis and management. *Clin Med* 3, 249–254.
- Rabovsky, M., McRae, K., 2014. Simulating the N400 ERP component as semantic network error: insights from a feature-based connectionist attractor model of word meaning. *Cognition* 132, 68–89.
- 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.
- Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237.
- Schnakers, C., Vanhauzenhuysse, A., Giacino, J.T., Ventura, M., Boly, M., Majerus, S., Moonen, G., Laureys, S., 2009. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurol.* 9, 35.
- Schoenle, P.W., Witzke, W., 2004. How vegetative is the vegetative state? Preserved semantic processing in VS patients – evidence from N 400 event-related potentials. *NeuroRehabilitation* 19, 329–334.
- Schroeder, M.R., 1968. Reference signal for signal quality studies. *J. Acoust. Soc. Am.* 44, 1735–1736.
- Skrandies, W., 1990. Global field power and topographic similarity. *Brain Topogr.* 3, 137–141.
- Stender, J., Gosseries, O., Bruno, M.-A., Charland-Verville, V., Vanhauzenhuysse, A., Demertzi, A., Chatelle, C., Thonnard, M., Thibaut, A., Heine, L., Soddu, A., Boly, M., Schnakers, C., Gjedde, A., Laureys, S., 2014. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet* 384, 514–522.
- Steppacher, I., Eickhoff, S., Jordanov, T., Kaps, M., Witzke, W., Kissler, J., 2013. N400 predicts recovery from disorders of consciousness. *Ann. Neurol.* 73, 594–602.
- Yeomans, J.S., Frankland, P.W., 1995. The acoustic startle reflex: neurons and connections. *Brain Res. Rev.* 21 (3), 301–314.