



(51) International Patent Classification:

A61B 5/00 (2006.01)

(21) International Application Number:

PCT/CN2020/092186

(22) International Filing Date:

25 May 2020 (25.05.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: CENTER FOR EXCELLENCE IN BRAIN SCIENCE AND INTELLIGENCE TECHNOLOGY, CHINESE ACADEMY OF SCIENCES [CN/CN]; 320 Yue Yang Road, Shanghai 200031 (CN).

(72) Inventors: WANG, Liping; 320 Yue Yang Road, Shanghai 200031 (CN). GUI, Peng; 320 Yue Yang Road, Shanghai 200031 (CN). JIANG, Yuwei; 320 Yue Yang Road, Shanghai 200031 (CN).

(74) Agent: SHANGHAI PATENT &amp; TRADEMARK LAW

OFFICE, LLC; 435 Guiping Road, Shanghai 200233 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

(54) Title: A SYSTEMATIC DEVICE AND SCHEME TO ASSESS THE LEVEL OF CONSCIOUSNESS DISORDER BY USING LANGUAGE RELATED BRAIN ACTIVITY

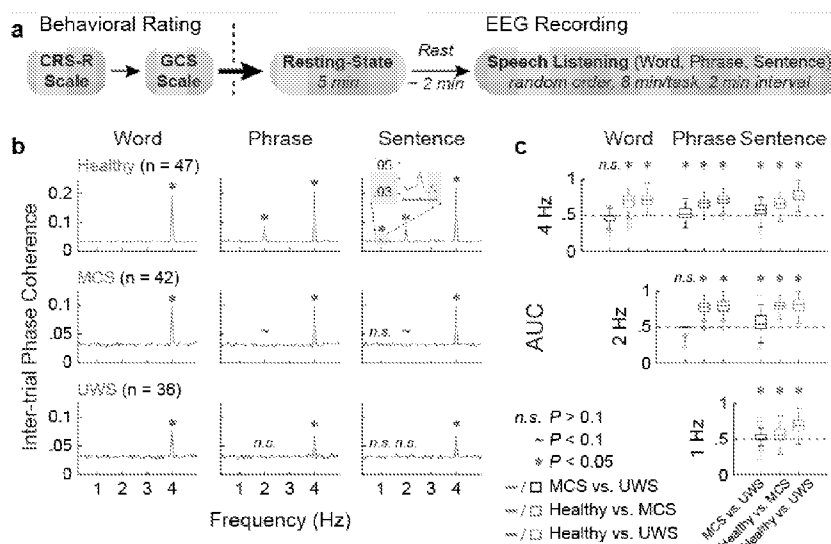


FIG. 2

(57) Abstract: A novel language paradigm is adopted which elicited rhythmic brain responses tracking the single-word, phrase and sentence rhythms in speech, to examine whether bedside electroencephalography (EEG) recordings can help inform diagnosis and prognosis. EEG-derived neural signals, including both speech-tracking responses and temporal dynamics of global brain states, were associated with behavioral diagnosis of consciousness. Crucially, multiple EEG measures in the language paradigm were robust to predict future outcomes in individual patients. Thus, EEG-based language assessment provides a novel and reliable approach to objectively characterize and predict states of consciousness and to longitudinally track individual patients' language processing abilities at the bedside.

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,  
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

**Published:**

— *with international search report (Art. 21(3))*

## A SYSTEMATIC DEVICE AND SCHEME TO ASSESS THE LEVEL OF CONSCIOUSNESS DISORDER BY USING LANGUAGE RELATED BRAIN ACTIVITY

### BACKGROUND OF THE INVENTION

5 Assessing residual consciousness and language abilities in unresponsive patients is a challenge for cognitive neuroscience and a major clinical concern. Every year, thousands of patients, due to severe brain injuries, lose their communication abilities and fall into different clinical conditions ranging from coma, to unresponsive wakefulness syndrome (UWS), minimally conscious state (MCS). The clinical diagnostic assessment of patients' conditions  
10 is mainly based on motor and oro-motor non-reflex behavior at the bedside <sup>1</sup>. In particular, UWS patients present moments of arousal, during which they open their eyes and produce complex behavior reflexes, but they show no clear signs of intentional behavior <sup>2</sup>. By contrast, MCS patients present some intentional behaviors but seem unable to establish any long-lasting functional communication <sup>3</sup>. The subtle difference between patients in a MCS  
15 and those with UWS, often compounded by the presence of additional deficits, can lead to a high rate of misdiagnosis <sup>4</sup>.

Recent work has shown that the level of consciousness can be indicated by various dynamic features of EEG signals, such as the amplitude and latency of auditory evoked responses <sup>5</sup>, spectral power <sup>6</sup>, and signal complexity and functional connectivity, for instance  
20 assessed using as weighted symbolic mutual information <sup>7</sup>. Indeed, the major consciousness theories claim that consciousness is characterized by a dynamic process of self-sustained, coordinated brain activity that constantly evolves, rather than being a static brain function <sup>8</sup>. Accordingly, a series of recent studies have reported that the dynamics of resting-state activity in functional magnetic resonance imaging (fMRI) may provide a specific cortical  
25 signature of the loss of consciousness <sup>9</sup>. Specifically, during loss of consciousness, brain activity is largely restricted to a dynamic pattern dominated by the structural connectivity. In contrast, conscious states are characterized by a more complex pattern of brain activity with long-distance (e.g., frontal-parietal) interactions. Similar to those seen in fMRI signals, the dynamic patterns in EEG can be described as "microstates", which are defined as global  
30 patterns of scalp potential topographies that dynamically vary over time in an organized manner <sup>10</sup>. That is, resting-state or task-related EEG can be described by a limited number of scalp potential topographies (maps) that remain stable for 60 to 120 ms before rapid transition to a different topography that remains stable again <sup>11</sup>. Given its temporal resolution, the pattern of EEG microstates seems likely to provide a better index of such fast dynamics  
35 and therefore better reflect the level of consciousness in patients with DOC, but this has yet

to be experimentally tested.

Auditory oddball paradigms have commonly been used in EEG studies, to detect the residual consciousness in DOC patients<sup>12</sup>. In these paradigms, although subjects may be instructed to count the number of times they hear a specific target sound<sup>13</sup> or a violation of temporal regularities, such paradigms rely primarily on an assessment of sensory responses at several hierarchical levels. Using active paradigms (e.g., mental imagery of playing tennis), some patients with DOC were found to respond to commands, which requires greater cognitive abilities<sup>14</sup>. Rather than using pure tones as auditory stimuli, several studies have attempted to develop reliable language paradigms in order to detect neural signatures of semantic processing<sup>15, 16</sup>, as natural language stimuli might be easier for patients to attend to. While cortical responses to natural speech in unresponsive patients in neuroimaging studies have provided evidence that natural language stimuli activated more auditory cortical regions than the scrambled auditory stimuli<sup>15-17</sup>, EEG results have been variable. Most of these EEG studies examined the N400 component or inter-subject correlation of neural signals in response to the narrative content of natural speech, and found no or weak differences between UWS and MCS patient groups<sup>7, 13, 16, 18, 19</sup>. The use of EEG in active linguistic paradigms that combine stimulus-evoked activity and dynamic brain states for assisting diagnosis and prognosis of DOC remains unexplored.

In the present study, we adopted a novel hierarchical auditory linguistic sequence paradigm that included three levels of processing, respectively at the single-word, phrase and sentence levels<sup>20</sup>, to compare the EEG features between activity in resting passive task and that during active tasks (three language conditions). Our aim was to assess the depth of language processing in disorders of consciousness, and to separate two distinct possibilities concerning this depth. First, the higher the consciousness level, the deeper the processing level of linguistic stimuli may be. This hypothesis may seem plausible given that the integration of multiple words into phrases and sentences calls for late, sustained and integrative brain activity which is typically found associated with conscious processing<sup>8</sup>. Second, alternatively, much language processing may remain possible in the absence of consciousness, as attested by a variety of masking and inattention paradigms in normal subjects, including the observation of brain responses to syntactic and semantic violations under non-conscious conditions<sup>16, 21, 22</sup>. Even in this case, it may still prove clinically useful to assess the depth of unconscious processing of linguistic stimuli in patients, as this may be predictive of their recovery.

Our paradigm allows us to combine both speech-tracking activity and dynamic pattern of brain states. We first evaluate, in normal participants, whether the detection of

hierarchical structure in the sequences requires top-down cognitive resources and is modulated by attention. Moving to patients, we then employ a multivariate approach, integrating both speech-tracking activities and temporal dynamics of global brain states, to evaluate the depth of linguistic processing in DOC patients and the value of EEG measures for diagnosis and prognosis of consciousness. We train and validate classification algorithms that use EEG-derived metrics and clinic diagnosis as inputs, and attempt to predict the clinical outcomes of individual patients.

## SUMMARY OF THE INVENTION

The purpose of the present invention is to provide a novel systematic device and scheme to assess the level of consciousness disorder by using language related brain activity.

In the first aspect, the present invention provides a method of assessing the level of consciousness disorder, the method comprising: using a hierarchical linguistic processing paradigm to test consciousness (such as residual consciousness) in subjects with disorders of consciousness.

In one preferred embodiment, the method comprising:

- (1) Stimulating the subjects;
- (2) Collecting the electroencephalogram (EEG) data of the patients;
- (3) Analysing the phase coherence, multivariate pattern and/or brain microstate;
- (4) Diagnosis and prediction.

In another preferred embodiment, the electroencephalogram data is pre-processed using BrainVision Analyzer; or the electroencephalogram data is pre-processed in the EEGLAB toolbox for the brain state analysis; preferably, the electroencephalogram is resting electroencephalogram.

In another preferred embodiment, the single-trial electroencephalogram data is transformed into the frequency domain, preferably the Discrete Fourier Transform is used for the transformation.

In another preferred embodiment, binary classifiers are used to discriminate different subject groups.

In another preferred embodiment, the LDA is trained for pairwise classifications at each target frequency under each task; preferably, the decoding is implemented as follows: 1) the input features are the ITPC values at all EEG channels; 2) for each comparison, a majority of subjects is randomly chosen as training set, while the other subjects as testing set; 3) a 3 to 10-fold cross-validation is applied on the training set, that is, for each fold, the classifier is fit on the majority subjects and validated on the other training set; 4) finally, the

classification performance is computed as the sum of the Area Under the Received Operative Curve, based on the probabilistic classification of the independent testing set; 5) the steps 2 to 4 are repeated 40~400 times to produce the mean classification AUC for these two groups at the each frequency for each condition.

5 In another preferred embodiment, the brain state analysis is performed using MicrostateAnalysis.

In another preferred embodiment, classification analysis is used to identify the consciousness states of individuals; or the patients for whom we had behavioral measurements more than 100 days after EEG assessments.

10 In another preferred embodiment, the consciousness in patients with disorders of consciousness can be reflected by the strength of speech-tracking responses, especially neural tracking of higher-level linguistic structures, i.e., phrases and sentences.

In another preferred embodiment, expanded the behavioral measurements by including the total score and six sub-scores and using the same model to search for the optimal CRS-R  
15 feature combination for prognosis; preferably, the six sub-scores are auditory, visual, motor, oromotor, communication and arousal; more preferably, using the visual subscale alone.

In another preferred embodiment, the features from both CRS-R and EEG are combined and submitted to construct the model.

20 In another preferred embodiment, the stimuli is Word stimuli, Phrase stimuli and/or Sentence stimuli.

In another preferred embodiment, the stimuli is performed by different language hierarchy (such as, increasing language hierarchy) or language paradigms; preferably, in the sentence condition, 2~5 levels of semantic hierarchies are used, for example, including single-word frequency at 4 Hz, phrasal frequency at 2 Hz, and sentential frequency at 1 Hz;  
25 preferably, the phrase condition only included word and phrasal levels; preferably, the word condition only included the 4 Hz word frequency.

In another aspect, the present invention provides a systematic device for assessing the level of consciousness disorder, comprising the measuring instruments, softwares or  
30 programs for the method of any one of the above methods.

In a preferred embodiment, the systematic device comprising:

Language generator;

Electroencephalogram measuring instrument;

measuring instruments, softwares or programs for analysing phase coherence;

35 measuring instruments, softwares or programs for analysing multivariate pattern; and/or

measuring instruments, softwares or programs for analysing brain microstate.

In another aspect, the present invention provides a computer system for assessing the level of consciousness disorder, comprising:

5 a device (or measuring instruments, softwares or programs) for stimulating the subjects;

a device (or measuring instruments, softwares or programs)for collecting the electroencephalogram (EEG) data of the patients;

10 a device (or measuring instruments, softwares or programs) for analysing the phase coherence, multivariate pattern and/or brain microstate;

a device (or measuring instruments, softwares or programs) for diagnosis and prediction.

In a preferred embodiment, the electroencephalogram data is pre-processed using BrainVision Analyzer; or the electroencephalogram data is pre-processed in the EEGLAB  
15 toolbox for the brain state analysis; preferably, the electroencephalogram is resting electroencephalogram.

In a preferred embodiment, the single-trial electroencephalogram data is transformed into the frequency domain, preferably the Discrete Fourier Transform is used for the transformation.

20 In a preferred embodiment, binary classifiers are used to discriminate different subject groups.

In a preferred embodiment, the LDA is trained for pairwise classifications at each target frequency under each task; preferably, the decoding is implemented as follows: 1) the input features are the ITPC values at all EEG channels; 2) for each comparison, a majority of  
25 subjects is randomly chosen as training set, while the other subjects as testing set; 3) a 3 to 10-fold cross-validation is applied on the training set, that is, for each fold, the classifier is fit on the majority subjects and validated on the other training set; 4) finally, the classification performance is computed as the sum of the Area Under the Received Operative Curve, based on the probabilistic classification of the independent testing set; 5) the steps 2  
30 to 4 are repeated 40~400 times to produce the mean classification AUC for these two groups at the each frequency for each condition.

In a preferred embodiment, the brain state analysis is performed using MicrostateAnalysis.

35 In a preferred embodiment, classification analysis is used to identify the consciousness states of individuals; or the patients for whom we had behavioral measurements more than

100 days after EEG assessments.

In a preferred embodiment, the residual consciousness in patients with disorders of consciousness can be reflected by the strength of speech-tracking responses, especially neural tracking of higher-level linguistic structures, i.e., phrases and sentences.

5 In a preferred embodiment, expanded the behavioral measurements by including the total score and six sub-scores and using the same model to search for the optimal CRS-R feature combination for prognosis; preferably, the six sub-scores are auditory, visual, motor, oromotor, communication and arousal; more preferably, using the visual subscale alone.

10 In a preferred embodiment, the features from both CRS-R and EEG are combined and submitted to construct the model.

In a preferred embodiment, the stimuli is Word stimuli, Phrase stimuli and/or Sentence stimuli; preferably, in the sentence condition, 2~5 levels of semantic hierarchies are used, for example, including single-word frequency at 4 Hz, phrasal frequency at 2 Hz, and sentential frequency at 1 Hz; preferably, the phrase condition only included word and phrasal levels; 15 preferably, the word condition only included the 4 Hz word frequency.

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 Paradigm and neural tracking of hierarchical linguistic structures. a, Illustration of stimuli presented in the word, phrase, and sentence conditions. b, Schematic of the 20 attentional experiment in normal participants, which required either attending to the auditory sequence or to concurrent visual stimuli. c, Group-averaged inter-trial phase coherence under four conditions in the attentive study (n = 22 subjects). Attend to: one-sided paired-sample *t*-test;  $t_{\text{Word-4Hz}}(21) = 10.11$ ,  $P_{\text{Word-4 Hz}} = 8 \times 10^{-10}$ ;  $t_{\text{Sentence-1Hz}}(21) = 6.11$ ,  $P_{\text{Sentence-1 Hz}} = 2.3 \times 10^{-6}$ ;  $t_{\text{Sentence-2Hz}}(21) = 7.26$ ,  $P_{\text{Sentence-2 Hz}} = 1.9 \times 10^{-7}$ ;  $t_{\text{Sentence-4Hz}}(21) = 11.1$ ,  $P_{\text{Sentence-4 Hz}} = 1.4 \times 10^{-10}$ ; Ignore: one-sided paired-sample *t*-test;  $t_{\text{Word-4Hz}}(21) = 8.22$ ,  $P_{\text{Word-4 Hz}} = 2.67 \times 10^{-8}$ ;  $t_{\text{Sentence-1Hz}}(21) = 4$ ,  $P_{\text{Sentence-1 Hz}} = 3.2 \times 10^{-4}$ ;  $t_{\text{Sentence-2Hz}}(21) = 4.14$ ,  $P_{\text{Sentence-2 Hz}} = 2.3 \times 10^{-4}$ ;  $t_{\text{Sentence-4Hz}}(21) = 8.58$ ,  $P_{\text{Sentence-4 Hz}} = 1.3 \times 10^{-8}$ . d, Comparison of inter-trial phase coherence values between attend to and ignore conditions (n = 22 subjects). Attend to vs. Ignore: two-sided paired-sample *t*-test;  $t_{\text{Word-4Hz}}(21) = 0.42$ ,  $P_{\text{Word-4Hz}} = 0.678$ ;  $t_{\text{Sentence-4Hz}}(21) = 1.58$ ,  $P_{\text{Sentence-4Hz}} = 0.128$ ,  $t_{\text{Sentence-2Hz}}(21) = 5.41$ ,  $P_{\text{Sentence-2 Hz}} = 2.3 \times 10^{-5}$ ,  $t_{\text{Sentence-1Hz}}(21) = 2.66$ ,  $P_{\text{Sentence-1 Hz}} = 0.015$ . Colored dots represent individual subjects. Black dots represent mean values. Error bars represent S.E.M. All panels: *n.s.*,  $P > 0.05$ ; \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ . 25 30

Fig. 2 Procedure and auditory-evoked brain activity in the clinical study. a, Schematic procedure in patients with DOC. In the same study day, after behavioral ratings, the EEG 35 recording started with a 5-minute resting-state, followed by a short-term rest and three

linguistic auditory blocks in a random order across subjects. b, Group-averaged inter-trial phase coherence in the word, phrase and sentence conditions. Data from the healthy control (n = 47), MCS (n = 42), and UWS (n = 36) groups are plotted from top to bottom. One-sided paired-sample *t*-test. Word condition:  $t_{4\text{Hz-Healthy}}(46) = 8.24$ ,  $P_{4\text{Hz-Healthy}} = 1.3 \times 10^{-10}$ ;  $t_{4\text{Hz-MCS}}(41) = 5.52$ ,  $P_{4\text{Hz-MCS}} = 2.1 \times 10^{-6}$ ;  $t_{4\text{Hz-UWS}}(35) = 3.78$ ,  $P_{4\text{Hz-UWS}} = 5.8 \times 10^{-4}$ . Phrase condition:  $t_{4\text{Hz-Healthy}}(46) = 8.78$ ,  $P_{4\text{Hz-Healthy}} = 2.1 \times 10^{-11}$ ;  $t_{4\text{Hz-MCS}}(41) = 5.36$ ,  $P_{4\text{Hz-MCS}} = 3.5 \times 10^{-6}$ ;  $t_{4\text{Hz-UWS}}(35) = 3.87$ ,  $P_{4\text{Hz-UWS}} = 4.5 \times 10^{-4}$ ;  $t_{2\text{Hz-Healthy}}(46) = 7.25$ ,  $P_{2\text{Hz-Healthy}} = 3.8 \times 10^{-9}$ ;  $t_{2\text{Hz-MCS}}(41) = 1.70$ ,  $P_{2\text{Hz-MCS}} = 0.097$ ,  $t_{2\text{Hz-UWS}}(35) = 0.15$ ,  $P_{2\text{Hz-UWS}} = 0.881$ . Sentence condition:  $t_{4\text{Hz-Healthy}}(46) = 8.35$ ,  $P_{4\text{Hz-Healthy}} = 9.1 \times 10^{-11}$ ;  $t_{4\text{Hz-MCS}}(41) = 5.40$ ,  $P_{4\text{Hz-MCS}} = 2.2 \times 10^{-6}$ ;  $t_{4\text{Hz-UWS}}(35) = 3.93$ ,  $P_{4\text{Hz-UWS}} = 3.8 \times 10^{-4}$ ;  $t_{2\text{Hz-Healthy}}(46) = 6.62$ ,  $P_{2\text{Hz-Healthy}} = 3.4 \times 10^{-8}$ ;  $t_{2\text{Hz-MCS}}(41) = 1.88$ ,  $P_{2\text{Hz-MCS}} = 0.068$ ,  $t_{2\text{Hz-UWS}}(35) = 0.39$ ,  $P_{2\text{Hz-UWS}} = 0.702$ ;  $t_{1\text{Hz-Healthy}}(46) = 4.48$ ,  $P_{1\text{Hz-Healthy}} = 4.9 \times 10^{-5}$ ,  $t_{1\text{Hz-MCS}}(41) = 0.58$ ,  $P_{1\text{Hz-MCS}} = 0.567$ ,  $t_{1\text{Hz-UWS}}(35) = -0.61$ ,  $P_{1\text{Hz-UWS}} = 0.546$ . c, Decoding performance of the multivariate pattern analysis using phase coherence at 1, 2, and 4 Hz with LDA (n = 100 permutations). One-sided *t*-test, comparing with 0.5: MCS vs. UWS,  $t_{\text{Word-4Hz}}(99) = -4.31$ ,  $P_{\text{Word-4Hz}} = 0.99998$ ;  $t_{\text{Phrase-4Hz}}(99) = 2.75$ ,  $P_{\text{Phrase-4Hz}} = 3.6 \times 10^{-3}$ ,  $t_{\text{Phrase-2Hz}}(99) = -2.75$ ,  $P_{\text{Phrase-2Hz}} = 0.996$ ,  $t_{\text{Sentence-4Hz}}(99) = 6.52$ ,  $P_{\text{Sentence-4Hz}} = 1.5 \times 10^{-9}$ ,  $t_{\text{Sentence-2Hz}}(99) = 4.8$ ,  $P_{\text{Sentence-2Hz}} = 2.4 \times 10^{-6}$ ,  $t_{\text{Sentence-1Hz}}(99) = 2.02$ ,  $P_{\text{Sentence-1Hz}} = 0.029$ ; Healthy vs. MCS,  $t_{\text{Word-4Hz}} = 17.63$ ,  $P_{\text{Word-4Hz}} = 1.3 \times 10^{-32}$ ,  $t_{\text{Phrase-4Hz}}(99) = 17.45$ ,  $P_{\text{Phrase-4Hz}} = 2.8 \times 10^{-32}$ ,  $t_{\text{Phrase-2Hz}}(99) = 30.04$ ,  $P_{\text{Phrase-2Hz}} = 7.6 \times 10^{-52}$ ,  $t_{\text{Sentence-4Hz}}(99) = 17.25$ ,  $P_{\text{Sentence-4Hz}} = 6.7 \times 10^{-32}$ ,  $t_{\text{Sentence-2Hz}}(99) = 38.61$ ,  $P_{\text{Sentence-2Hz}} = 8.7 \times 10^{-62}$ ,  $t_{\text{Sentence-1Hz}}(99) = 6.31$ ,  $P_{\text{Sentence-1Hz}} = 4 \times 10^{-9}$ ; Healthy vs. UWS:  $t_{\text{Word-4Hz}}(99) = 23.75$ ,  $P_{\text{Word-4Hz}} = 5.6 \times 10^{-43}$ ,  $t_{\text{Phrase-4Hz}}(99) = 25.79$ ,  $P_{\text{Phrase-4Hz}} = 5 \times 10^{-46}$ ,  $t_{\text{Phrase-2Hz}}(99) = 31.06$ ,  $P_{\text{Phrase-2Hz}} = 3.8 \times 10^{-53}$ ,  $t_{\text{Sentence-4Hz}}(99) = 31.25$ ,  $P_{\text{Sentence-4Hz}} = 2.2 \times 10^{-53}$ ,  $t_{\text{Sentence-2Hz}}(99) = 34.71$ ,  $P_{\text{Sentence-2Hz}} = 1.6 \times 10^{-57}$ ,  $t_{\text{Sentence-1Hz}}(99) = 17.36$ ,  $P_{\text{Sentence-1Hz}} = 4.1 \times 10^{-32}$ . Boxes represent interquartile ranges (IQR), central bars indicate the median, and whiskers indicate  $1.5 \times \text{IQR}$ . “+” symbols indicate outliers. Horizontal dashed lines indicate the chance level. All panels: *n.s.*,  $P > 0.1$ ;  $\sim$ ,  $P < 0.1$ ; \*,  $P < 0.05$ .

Fig. 3 Global patterns of brain states. a, Mean accuracy of cross-validation (CV) criterion and mean global explained variance (GEV) varied with numbers of brain state maps, showing that the optimal number of clustering maps across subjects was four. b, Group-averaged brain state maps of healthy controls in task conditions (averaged across all three task conditions, n = 47 subjects), sorted by the brain state probability. Colors represent the relative potential distribution. c, The probability distribution of four maps in the three groups for each of the four conditions. MANOVA: Resting,  $F(6,200) = 5.176$ ,  $P = 5.7 \times 10^{-5}$ ; Word,  $F(6,240) = 13.543$ ,  $P = 3.1 \times 10^{-13}$ ; Phrase,  $F(6,240) = 14.258$ ,  $P = 6.9 \times 10^{-14}$ ; Sentence,

$F(6,240) = 12.56$ ,  $P = 2.6 \times 10^{-12}$ . Lines indicate mean values, and shade areas represent S.E.M. d,  $\Delta C_p$  (Probability-weighted spatial correlation difference) between A-P and L-R maps for each group and condition. One-way ANOVA, Bonferroni corrected: Resting,  $P_{\text{Healthy-MCS}} = 3.7 \times 10^{-11}$ ,  $P_{\text{Healthy-UWS}} = 2.4 \times 10^{-9}$ ; Word,  $P_{\text{Healthy-MCS}} = 2.9 \times 10^{-19}$ ,  $P_{\text{Healthy-UWS}} = 3.8 \times 10^{-22}$ ; Phrase,  $P_{\text{Healthy-MCS}} = 2.1 \times 10^{-15}$ ,  $P_{\text{Healthy-UWS}} = 9.4 \times 10^{-21}$ ,  $P_{\text{MCS-UWS}} = 0.037$ ; Sentence,  $P_{\text{Healthy-MCS}} = 1.7 \times 10^{-15}$ ,  $P_{\text{Healthy-UWS}} = 9.9 \times 10^{-22}$ ,  $P_{\text{MCS-UWS}} = 0.014$ . Colored dots represent individual subjects. Black dots represent mean values. Error bars represent S.E.M. e, Comparison of  $\Delta C_p$  between the MCS and UWS groups. The difference matched well along with the linguistic hierarchy from the resting-state to sentence condition. The red line indicates statistical significance between patient groups in each condition. One-way ANOVA, Bonferroni corrected:  $P_{\text{Resting}} = 1$ ,  $P_{\text{Word}} = 0.33$ ,  $P_{\text{Phrase}} = 0.037$ ,  $P_{\text{Sentence}} = 0.014$ . Panels c - e:  $n_{\text{Healthy-Resting}} = 34$ ,  $n_{\text{Healthy-Task}} = 47$ ,  $n_{\text{MCS-Resting}} = 41$ ,  $n_{\text{MCS-Task}} = 42$ ,  $n_{\text{UWS-Resting}} = 30$ ,  $n_{\text{UWS-Task}} = 36$ . All panels: *n.s.*,  $P > 0.05$ ; \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ .

Fig. 4 Duration and occurrence of brain state maps. a, Duration. *Left panel*: the duration of the L-R map for healthy controls ( $n_{\text{Resting}} = 34$ ,  $n_{\text{Task}} = 47$ ), patients with MCS ( $n_{\text{Resting}} = 41$ ,  $n_{\text{Task}} = 42$ ), and patients with UWS ( $n_{\text{Resting}} = 30$ ,  $n_{\text{Task}} = 36$ ). One-way ANOVA, Bonferroni corrected: Resting,  $P_{\text{Healthy-MCS}} = 5.6 \times 10^{-10}$ ,  $P_{\text{Healthy-UWS}} = 5.8 \times 10^{-10}$ ,  $P_{\text{MCS-UWS}} = 1$ ; Word,  $P_{\text{Healthy-MCS}} = 2.7 \times 10^{-8}$ ,  $P_{\text{Healthy-UWS}} = 8.2 \times 10^{-13}$ ,  $P_{\text{MCS-UWS}} = 0.083$ ; Phrase,  $P_{\text{Healthy-MCS}} = 1.1 \times 10^{-8}$ ,  $P_{\text{Healthy-UWS}} = 1.05 \times 10^{-14}$ ,  $P_{\text{MCS-UWS}} = 0.016$ ; Sentence,  $P_{\text{Healthy-MCS}} = 1.1 \times 10^{-9}$ ,  $P_{\text{Healthy-UWS}} = 1.1 \times 10^{-15}$ ,  $P_{\text{MCS-UWS}} = 0.017$ . Boxes represent IQR, central dots indicate the median, and whiskers indicate  $1.5 \times \text{IQR}$ . Colored dots indicate outliers. *Right panel*: the duration of the L-R map on the first (circle) and last (solid) EEG recording for the recovered (+ve: outcome-positive;  $n_{\text{Resting}} = 7$ ,  $n_{\text{Task}} = 8$ ) and unrecovered (-ve: outcome-negative;  $n_{\text{Resting}} = 5$ ,  $n_{\text{Task}} = 7$ ) patients. Colored lines indicate the first and last EEG recording of individual subjects. Friedman tests: Resting,  $\chi^2_{+ve} = 3.57$ ,  $P_{+ve} = 0.059$ ,  $\chi^2_{-ve} = 1.8$ ,  $P_{-ve} = 0.18$ ; Word,  $\chi^2_{+ve} = 8.0$ ,  $P_{+ve} = 0.005$ ,  $\chi^2_{-ve} = 0.14$ ,  $P_{-ve} = 0.705$ ; Phrase,  $\chi^2_{+ve} = 8.0$ ,  $P_{+ve} = 0.005$ ,  $\chi^2_{-ve} = 3.57$ ,  $P_{-ve} = 0.059$ ; Sentence,  $\chi^2_{+ve} = 8.0$ ,  $P_{+ve} = 0.005$ ,  $\chi^2_{-ve} = 0.14$ ,  $P_{-ve} = 0.705$ . b, Comparison of the L-R map duration between MCS and UWS groups in the four conditions. The increase in the differences of L-R map duration between MCS and UWS groups paralleled the increase in linguistic hierarchy. The red line indicates statistical significance between patient groups of L-R map duration in each condition. c, Comparison of the difference of L-R map duration between the first and last recording in the four conditions in the recovered and non-recovered patients. d-f, Occurrence of the A-P map, with the same format as a-c. For A-P differences in d and e, one-way ANOVA, Bonferroni corrected: Resting,  $P_{\text{Healthy-MCS}} = 1.1 \times 10^{-12}$ ,  $P_{\text{Healthy-UWS}} = 1.2 \times 10^{-11}$ ,  $P_{\text{MCS-UWS}} = 1$ ; Word,  $P_{\text{Healthy-MCS}} =$

$6.1 \times 10^{-15}$ ,  $P_{\text{Healthy-UWS}} = 7.1 \times 10^{-19}$ ,  $P_{\text{MCS-UWS}} = 0.156$ ; Phrase,  $P_{\text{Healthy-MCS}} = 3.4 \times 10^{-15}$ ,  $P_{\text{Healthy-UWS}} = 8.4 \times 10^{-21}$ ,  $P_{\text{MCS-UWS}} = 0.028$ ; Sentence,  $P_{\text{Healthy-MCS}} = 3.6 \times 10^{-16}$ ,  $P_{\text{Healthy-UWS}} = 5.4 \times 10^{-21}$ ,  $P_{\text{MCS-UWS}} = 0.063$ . For Followup differences in d and f, Friedman tests: Resting,  $\chi^2_{+ve} = 3.57$ ,  $P_{+ve} = 0.059$ ,  $\chi^2_{-ve} = 0.2$ ,  $P_{-ve} = 0.655$ ; Word,  $\chi^2_{+ve} = 8.0$ ,  $P_{+ve} = 0.005$ ,  $\chi^2_{-ve} = 1.29$ ,  $P_{-ve} = 0.257$ ; Phrase,  $\chi^2_{+ve} = 8.0$ ,  $P_{+ve} = 0.005$ ,  $\chi^2_{-ve} = 1.29$ ,  $P_{-ve} = 0.257$ ; Sentence,  $\chi^2_{+ve} = 8.0$ ,  $P_{+ve} = 0.005$ ,  $\chi^2_{-ve} = 0.14$ ,  $P_{-ve} = 0.705$ . All panels: *n.s.*,  $P > 0.05$ ; \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .

Fig. 5 Diagnosis and outcome prediction. a, Classification and prediction procedures using the multiple EEG features (for the details, see *Methods*). b, The confusion matrix of diagnosed consciousness classification generated by the LDA. The classifier was trained on the EEG metrics derived from the sentence task. The pie chart shows the mismatch between clinical diagnosis of patients and outcomes. c, The performance of outcome prediction using CRS-R total-scores. Chi-squared test: AUC = 70%,  $\chi^2 = 7.2$ ,  $P = 0.016$ , accuracy = 74%. d, The prognostic validity of the model using CRS-R total-scores. The classifier was trained on the dataset of 38 patients from c with cross validation, and then tested the generalization ability on the new dataset of 25 patients. Chi-squared test: AUC = 31%,  $\chi^2 = 4.6$ ,  $P = 0.049$ , accuracy = 28%. e, The performance of outcome prediction using EEG (*left panel*, chi-squared test: AUC = 77%,  $\chi^2 = 11.5$ ,  $P = 9.2 \times 10^{-4}$ , accuracy = 76%) and the comparison of prediction and actual outcome of individual patients (*right panel*). The dots/diamonds above the threshold (gray line, prediction score = 0.1) represent the patients with predicted positive outcomes (+ve), while the others represent those with predicted negative outcomes (-ve). f, The prognostic validity of the model using EEG. The classifier was trained on the old dataset e, and generalized to the new dataset. Chi-squared test: AUC = 80%,  $\chi^2 = 8.8$ ,  $P = 0.005$ , accuracy = 80%. f shows the same format as e. g: group;  $\omega$ : normalization coefficients.

Fig. 6 The flowchart showing patients selection in data analysis.

Fig. 7 Individual ITPC responses to hierarchical linguistic structures in Healthy controls ( $n = 47$ ), MCS ( $n = 42$ ) and UWS ( $n = 36$ ) patients at three task levels. In each inset, the dots in the left represent the ITPC values from individual subject at target frequencies (1/2/4 Hz), while the dots in the right represent the individual mean value at its respective neighbours. Solid black dots represent grand mean values. *n.s.*,  $P > 0.1$ ;  $\sim$ ,  $P < 0.1$ ; \*,  $P < 0.001$ ; one-sided paired-sample *t*-test: see legend of Fig. 2b for precise statistical values.

Fig. 8 ITPC responses to hierarchical linguistic structures in individual patients at three task levels. Each bar denotes the response from one subject. Red dots indicate the significance (exact  $P < 0.05$ ; one-sided exact test, the statistical significance (exact  $P$ ) of the

ITPC response at a target frequency is the probability that the target frequency response differs from the null distribution, which consisted of responses at all non-target frequencies, see *Methods*).

Fig. 9 Brain state maps of healthy controls and patients in all four task conditions.

5 Number of subjects:  $n_{\text{Healthy-Resting}} = 34$ ,  $n_{\text{Healthy-Task}} = 47$ ,  $n_{\text{MCS-Resting}} = 41$ ,  $n_{\text{MCS-Task}} = 42$ ,  $n_{\text{UWS-Resting}} = 30$ ,  $n_{\text{UWS-Task}} = 36$ . *Top row*: Template Anterior-Posterior (A-P) and Left-Right (L-R) maps obtained from healthy controls. *Bottom three rows*: Original four maps (Maps A, B, C, and D) of each group in each condition.

Fig. 10 Duration and occurrence of brain state maps. **a**, The duration of the A-P Map for the healthy control ( $n_{\text{Resting}} = 34$ ,  $n_{\text{Task}} = 47$ ), MCS ( $n_{\text{Resting}} = 41$ ,  $n_{\text{Task}} = 42$ ), and UWS ( $n_{\text{UWS}} = 30$ ,  $n_{\text{Task}} = 36$ ) groups in all four task conditions. **b**, The occurrence of the L-R Map for healthy control, MCS, and UWS groups. Note that there were no differences between the three groups. Boxes represent IQR, central dots indicate the median, and whiskers indicate  $1.5 \times \text{IQR}$ . Colored dots indicate outliers. **c**, The duration of the L-R map for healthy controls (gray;  $n_{\text{Resting}} = 34$ ,  $n_{\text{Task}} = 47$ ), recovery patients (+ve, green;  $n_{\text{Resting}} = 11$ ,  $n_{\text{Task}} = 19$ ), and non-recovery MCS (-ve MCS, blue;  $n_{\text{Resting}} = 30$ ,  $n_{\text{Task}} = 23$ ) and non-recovered UWS (-ve UWS, red;  $n_{\text{Resting}} = 30$ ,  $n_{\text{Task}} = 36$ ) patients in all four conditions. One-way ANOVA, Bonferroni corrected: Healthy vs. +ve,  $P_{\text{Resting}} = 0.001$ ,  $P_{\text{Word}} = 0.018$ ,  $P_{\text{Phrase}} = 0.013$ ,  $P_{\text{Sentence}} = 0.008$ ; +ve vs. -ve MCS,  $P_{\text{Word}} = 0.012$ ,  $P_{\text{Phrase}} = 0.009$ ,  $P_{\text{Sentence}} = 0.002$ . **d**, The occurrence of the A-P map for healthy controls, recovery patients, non-recovery MCS patients, and non-recovery UWS patients. One-way ANOVA, Bonferroni corrected: Healthy vs. +ve,  $P_{\text{Resting}} = 6.8 \times 10^{-5}$ ,  $P_{\text{Word}} = 8.1 \times 10^{-7}$ ,  $P_{\text{Phrase}} = 5.7 \times 10^{-7}$ ,  $P_{\text{Sentence}} = 1.8 \times 10^{-7}$ ; +ve vs. -ve MCS,  $P_{\text{Word}} = 0.046$ ,  $P_{\text{Phrase}} = 0.048$ ,  $P_{\text{Sentence}} = 0.026$ . Panel c and d: colored dots represent individual subjects. Black dots represent mean values. Error bars represent S.E.M. All panels: 25 \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $p < 0.001$ .

Fig. 11 Correlation between the volumes of brain injury and the  $\Delta\text{Cp}$ . **a**, The comparison of volumes of brain injury between MCS ( $n = 17$ ) and UWS ( $n = 10$ ) patients. The black dots and error bars denote the mean value and S.E.M.  $t_{25} = 0.64$ ,  $P = 0.53$ , two-tailed two-sample  $t$ -test. **b**, The correlation between  $\Delta\text{Cp}$  and the volumes of brain injury in three task conditions. Pearson's correlation test (two-tailed),  $n_{\text{MCS}} = 17$ ,  $n_{\text{UWS}} = 10$ . **c**, An example patient (Patient 7): the MRI data and maps of the stroke patient without brain damage. **d**, An example patient (Patient 17): the MRI data and maps of the TBI patient with large brain damage. **e**, The comparison of  $\Delta\text{probability}$  between the stroke patient without brain damage and the TBI patient with brain damage, as shown in C and D. P1: Patient 7, P2: Patient 17. **f**, 35  $\Delta\text{Cp}$ , the same format as e. **g**, The MRI data and brain states of a stroke patient with brain

damage (an example patient, Patient 2). The orange box indicates the first EEG recording in unrecovered state. The green box indicates the last EEG recording in recovery state. **h**, The comparison of  $\Delta$ probability in Patient 2 between the first EEG recording in unrecovered state and the last EEG recording in recovery state. **i**,  $\Delta C_p$ , the same format as **h**. **W**: Word, **P**: Phrase, **S**: Sentence. **F**: First recording, **L**: Last recording. The percentage under each spatial map indicates the probability of each map.

Fig. 12 Diagnosis and outcome prediction using SVM. **a**, The confusion matrix of diagnosed consciousness classification generated by the cross-validated SVM. The feature combinations we used were [ $\Delta C_p$  + Duration<sub>L-R</sub> + Occurrence<sub>A-P</sub> + ITPC<sub>1Hz</sub> + ITPC<sub>2Hz</sub> + ITPC<sub>4Hz</sub>] for Sentence task. **b**, The performance of outcome prediction on training data using SVM classifier with the best feature combinations. *Left*: Outcome prediction accuracies by EEG on 38 EEG recordings (15 outcome-positive patients). *Right*: Comparison of individual predictions and actual outcomes. The patients with UWS are shown to the left of dashed line, and the patients with MCS are shown to the right. The dots above the threshold (gray line, prediction score = 0.3) represent the patients with predicted positive outcomes, while the others represent those with predicted negative outcomes. The actual outcome-negative patients are marked by orange dots, and the actual outcome-positive patients are marked by green diamonds. Solid green diamonds represent the outcome in patients that regained wakefulness. The feature combinations we used were: [ $\Delta$ Probability + Duration<sub>L-R</sub> + Transition<sub>A-P</sub>] for Word condition, [ $\Delta$ Probability + Occurrence<sub>A-P</sub> + Duration<sub>L-R</sub> + Transition<sub>L-R</sub> + ITPC<sub>4Hz</sub>] for Phrase condition, [Occurrence<sub>A-P</sub> + Duration<sub>L-R</sub> + Transition<sub>L-R</sub> + ITPC<sub>1Hz</sub> + ITPC<sub>2Hz</sub>] for Sentence condition.

Fig. 13 Comparison of EEG-based and CRS-based classifiers for diagnosis and outcome prediction. **a**, Performance of clinical diagnosis using the CRS-R total-score. The classification model (LDA) was trained on the first 38 patients with cross validation (*Left*) and then tested (without retraining) on a novel dataset of 25 patients (*Right*). **b**, *Left*: Performance of outcome prediction using the optimal CRS-R sub-score (Visual subscale). *Right*: The prognostic validity of this model. The classifier was trained with cross-validation on the first dataset of 38 patients, then tested for generalization on a new dataset of 25 patients. **c**, Comparison of the prediction performance for models with the same number of features, based either on CRS-R (7 features, as in **b**) or the EEG recording under the word condition (7 features: 1 ITPC and 6 microstates). The optimal feature combination was  $\Delta$ Probability + Occurrence<sub>A-P</sub>. **d**, To test whether the superior EEG generalization ability was due to the larger number of EEG features used in the model, we then ran another way of selecting features by merely using the features with the first two highest weights in the

model, and compared their performance of generalization. For the model using EEG see **c**, and for CRS-R, the best two features were Visual and Arousal subscales. **e**, Comparison of the performance of outcome prediction, using a standard LDA without feature selection, using all 7 features under the word condition of EEG versus all 7 features of the CRS-R scores (1 total-score and 6 sub-scores: auditory, visual, motor, oromotor, communication and arousal). Generalizations using EEG recorded during the other two task conditions (phrase: AUC = 89%,  $\chi^2 = 13.1$ ,  $P = 5.5 \times 10^{-4}$ , accuracy = 84%; sentence: AUC = 93%,  $\chi^2 = 13.1$ ,  $P = 5.5 \times 10^{-4}$ , accuracy = 84%; chi-squared test) showed similar results as that during the word condition.

Fig. 14 Comparisons of performance of outcome prediction using EEG versus EEG plus CRS-R scores. *Upper*: The confusion matrix of outcome prediction by EEG scores. *Lower*: The confusion matrix of outcome prediction by the combination of EEG and CRS-R scores.

Fig. 15 Multiple CRS-R ratings across time. Each inset indicates one patient. **a**, Individual patients ( $n = 15$ ). Within each inset, every blue dot indicates one CRS-R rating, and the gray line indicates the GLM fitting of all ratings in the entire period. Day 0 and red vertical dashed lines indicate the day of first EEG recording. **b**, The comparison of CRS-R scores between the EEG recording day and the day within a week (on average within 2.67 days). Colored lines indicate the ratings of individual patients. Black line indicates the mean. No significant difference was found between the two ratings ( $n = 15$ ,  $t_{14} = 0.899$ ,  $P = 0.384$ ; two-sided paired-sample  $t$ -test).

## DETAILED DESCRIPTION OF THE INVENTION

We adopted a hierarchical linguistic processing paradigm to test residual consciousness in patients with DOC. We demonstrated that two EEG-derived neural signals, speech-tracking activity and global dynamic pattern, were associated with the behavioral diagnosis of consciousness and clinical outcomes. This correlation significantly increased along with the increase in language hierarchy. Furthermore, the multiple EEG measurements were robust enough for the prediction of behavioral diagnosis and future outcomes in individual patients. This therefore represents a novel approach for clinical use of EEG measures in the diagnosis and prognosis of consciousness in patients with DOC.

In the past decade, neuroimaging and electrophysiological approaches have been used to examine states of consciousness in unresponsive patients, including fMRI<sup>14, 23, 31, 32</sup>, PET<sup>28</sup>, EEG<sup>5-7, 18</sup>, and multimodal imaging<sup>33, 34</sup>. fMRI has a number of limitations, including a high cost, lack of portability, and impossibility of bedside clinical testing, whereas high-density EEG is more feasible to deploy at the patient's bedside and to help track individual patients

longitudinally.

Previous fMRI experiments have shown that the brain spontaneously generates a dynamic series of constantly changing activity and functional connectivity between brain regions<sup>35, 36</sup>, which contributes to efficient exchanges between neural populations; this suggests that the neural correlates of consciousness could be found in temporally evolving dynamic processes<sup>9, 37</sup>. Our study investigated the similar discrete and dynamic states in EEG (referred to as “microstates”) and proposed that the dynamic pattern of scalp potentials reflects the momentary state of global neural activity, which might correspond to the changes in consciousness over time<sup>9, 23</sup>. Specifically, we showed that the global brain activity, in particular the duration of L-R maps and the occurrence of A-P maps during the language tasks, can significantly differentiate between UWS and MCS patients at both the group and individual levels. Furthermore, the capacity to discriminate UWS and MCS patients by examining the dynamics of brain states increased with task hierarchy, from resting, word, phrase, to sentence conditions. The results lend support to the idea that those EEG states are not meaningless recurrent patterns, but may indeed separate sensory versus higher-level cognition functions and are associated with the level of consciousness and task performance (e.g., the hierarchical level of language processing)<sup>38, 39</sup>.

Despite some promising neuroimaging studies<sup>14, 32</sup>, there have been few reliable active EEG paradigms for assisting diagnosis and prognosis of DOC<sup>19, 28, 40</sup>. Here, we present the first EEG evidence that speech-tracking neural responses and cortical dynamic patterns are directly associated with multiple-levels of speech processing in patients with DOC. We found that phrase- and sentence-level responses disappear in UWS patients, suggesting no preservation of deeper-level linguistic processing once consciousness is lost (in agreement with previous studies of anesthesia<sup>41</sup> and sleep<sup>42</sup>). While these responses drastically reduce in MCS patients, multivariate and brain-state analyses indicated that linguistic structures continue to modulate neural processing, suggesting some degree of deeper-level processing in MCS patients. Increased variability of cortical dynamics in MCS patients (Figs. 3c and 4) could explain why phrase- and sentence-rate responses diminished in the phase-coherence spectrum in this group (Fig. 2b). Recently, it has been debated whether phrase- and sentence-rate responses reflect neural entrainment to mentally constructed syntactic structures<sup>43</sup> or semantic properties of individual words<sup>44</sup>. Our study, however, did not intend to distinguish semantic and syntactic processing and employed the phrase- and sentence-rate responses as general measures of higher-level linguistic processing. The current result that word-rate remains in UWS and MCS patients while the phrase- and sentence-rate responses diminish actually further confirms that the phrase- and sentence-rate

responses reflect deeper levels of speech processing than the word-rate response. In the future, a combination of multiple EEG paradigms, including the present paradigm as well as syntactic and semantic violation paradigms<sup>16, 17, 45</sup>, could facilitate the assessment of language comprehension abilities in individual patients.

Our results demonstrate the diagnostic potential of EEG speech responses. The diagnosed state classification and future outcome prediction model showed that both speech-tracking responses (e.g., ITPC at 2 and 4 Hz) and global dynamic patterns (e.g., Occurrence<sub>A-P</sub> and  $\Delta C_p$ ) were required for the best decoding and prediction accuracy. This suggests that combining different analytical methods could deliver better diagnostic and predictive capabilities. Note that different from the most of previous studies showing the prediction of the recovery from coma<sup>29</sup>, our study demonstrated the recovery prediction of patients from different stages (e.g. UWS, MCS, EMCS etc.). However, we should also note that we lacked a detailed set of consecutive behavioral measurement for each patient during the recovery period. Thus, although we selected the patients whose CRS-R outcome was obtained more than 100 days after EEG assessments, our research does not indicate that EEG signals can precisely predict clinical outcome 100 days ahead of behavior. Based on previous studies<sup>46, 47</sup>, 6 months after DOC onset, 17% non-traumatic UWS will recover consciousness, and 67% posttraumatic UWS will recover consciousness. A systematic evaluation of whether and by how many days the EEG model can anticipate on behavioral recovery requires further investigation. Nevertheless, there are several reasons why our results engender confidence in the clinical use of EEG at the bedside. First, the current paradigm has been demonstrated to be useful for multiple languages, including Chinese, English, and Hebrew<sup>20, 42</sup>, indicating that it can be applied clinically in different language environments. Second, EEG ITPC and global brain states can theoretically be measured using fewer electrodes (e.g., sixteen channels<sup>48</sup>) and with cheaper EEG systems, while still preserving discriminative power and clinical utility. Third, the time-frame of the paradigm is also feasible for daily bedside examinations at hospital or at home, since the experiment lasted for less than 20 minutes when including only two linguistic conditions (word and sentence).

However, there are also several limitations of the current study that should be noted. First, the speech stimuli were synthesized using software; prior studies have demonstrated that personal and meaningful stimuli elicit more robust and reliable responses in brain-injured patients<sup>49</sup>. Future work could therefore use more personalized speech stimuli, e.g., on topics that are familiar to the patient. Second, the phrase- and sentence-level processing could potentially be enhanced when the speech rate is slowed down, especially for DOC patients. Also note that the ITPC signals at 1, 2 and 4 Hz of the healthy controls

recorded in the noisy hospital environment were significantly reduced compared to laboratory recordings. Improvements in data acquisition systems and recording conditions may provide higher EEG data quality. Third, although some patients had been diagnosed multiple times using EEG recordings, and the diagnostic accuracy of the model seems very high and it is possible that the model misdiagnosed some patients due to fluctuations in levels of consciousness over time. Indeed, clinicians are required to conduct multiple behavioral assessments of consciousness before they achieve a stable and accurate diagnosis. For example, there is evidence that the number of assessments has a significant effect on clinical diagnosis within two weeks, and even within a day<sup>50</sup>, and a higher responsiveness during behavioral assessment was found in the morning compared to the afternoon<sup>30</sup>. Similarly, the level of consciousness that is evident from a patient's EEG is likely to fluctuate both within and across days. Future work may therefore require multiple EEG sessions from the early stage of coma to the follow-up recovery.

## METHODS

### Participants

The study protocol was approved by the Ethical Committee of the Huashan Hospital of Fudan University (approval number: HIRB-2014-281), and informed consent was obtained from all healthy subjects and caregivers of all patients. All patients were native Mandarin Chinese speakers. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications<sup>18, 34, 51</sup>.

Twenty-seven healthy subjects participated in the first EEG study investigating how top-down attention modulates speech-tracking activity (experimental details shown in the below description; 15 males; mean age = 23.9 years; range = 20 to 30 years). Five participants were excluded from the final analyses due to poor quality of EEG data (for 22 subjects used in the final analysis: 12 males; mean age = 23.73 years; range = 20 to 30 years).

For the clinical study, the analyses were based on usable EEG data acquired from patients with DOC (disorders of consciousness) during July 2016 and June 2019. All patients had been diagnosed with MCS or UWS according to the Chinese versions of the CRS-R<sup>52, 53</sup> and GCS<sup>54</sup>. EEG was recorded in patients that had not received sedation (mostly midazolam) within the previous 24 hours, in order to minimize the influence of drugs on spontaneous brain activity and arousal levels.

First, a total of 93 patients were recruited from July 2016 to October 2018. Since some patients were followed up several times, the final dataset included 133 resting-state

recordings from 89 patients and 132 task-related recordings from 92 patients. Due to contamination of EEG data from environmental noise and extreme body movements, 62 resting-state recordings from 50 patients and 54 task-related recordings from 43 patients were discarded (Fig.6). Data from a final total of 70 patients were used in the present study, and the etiologies of these patients were stroke (36, 51.43%), traumatic brain injury (31, 44.29%), and anoxia (3, 4.29%) (see Table 1 for details). Thus, the final dataset included 71 resting-state recordings from 54 patients with DOC (48 males; mean age = 49.3 years; range = 17 to 75 years, 41 MCS recordings and 30 UWS recordings), and 78 linguistic task-related recordings from 60 patients with DOC (52 males; mean age = 47.8 years; range = 9 to 68 years). The number of recordings varied between patients.

To test the external validity and generalization ability of the prediction model, another EEG dataset acquired from October 2018 to June 2019 was included (Fig.6, 25 recordings, 12 MCS and 13 UWS; 13 males; mean age = 39.9 years; range = 18 to 69 years; detailed information in Table 2).

We also recruited 61 healthy volunteers from local communities as a control group (20 males; mean age = 31.3 years; range = 22 to 65 years). After checking the data quality, there were 34 usable resting-state recordings (8 males; mean age = 25.8 years; range = 22 to 50 years) and 47 usable task-related recordings (17 males; mean age = 31.1 years; range = 22 to 58 years).

**Table 1.** Detailed demographic and clinical information of patients recruited during year 2016-2018.

| Group | Etiology | ID | Sex | Age (years) | GCS | CRS-R | DOC month |
|-------|----------|----|-----|-------------|-----|-------|-----------|
| MCS   | Anoxia   | 12 | M   | 25          | 8   | 7     | 70        |
|       | Stroke   | 1  | F   | 47          | 9   | 8     | 46.1      |
|       |          | 2  | M   | 65          | 10  | 8     | 6.4       |
|       |          | 2  | M   | 65          | 9   | 8     | 7.4       |
|       |          | 2  | M   | 65          | 10  | 11    | 11        |
|       |          | 2  | M   | 66          | 10  | 12    | 14.7      |
|       |          | 2  | M   | 66          | 10  | 16    | 21.6      |
|       |          | 2  | M   | 67          | 10  | 16    | 36.9      |
|       |          | 3  | M   | 37          | 8   | 5     | 53.5      |
|       |          | 7  | M   | 39          | 8   | 6     | 21.2      |
|       |          | 9  | F   | 36          | 9   | 8     | 13        |
|       |          | 18 | F   | 47          | 6   | 6     | 11.4      |
|       |          | 18 | F   | 47          | 8   | 8     | 20.8      |
|       |          | 18 | F   | 47          | 8   | 5     | 25.2      |
|       |          | 20 | F   | 12          | 7   | 5     | 8         |
|       |          | 20 | F   | 12          | 10  | 9     | 15        |
|       |          | 27 | F   | 62          | 10  | 9     | 7.1       |

|     |        |     |   |    |    |    |      |
|-----|--------|-----|---|----|----|----|------|
|     |        | 29  | M | 68 | 10 | 9  | 5.5  |
|     |        | 30  | M | 45 | 10 | 9  | 0.5  |
|     |        | 33  | F | 47 | 8  | 5  | 8.6  |
|     |        | 33  | F | 48 | 11 | 12 | 13.5 |
|     |        | 39  | F | 66 | 11 | 14 | 9.3  |
|     |        | 46  | M | 50 | 9  | 7  | 29   |
|     |        | 50  | M | 54 | 10 | 7  | 36   |
|     |        | 56  | M | 59 | 8  | 7  | 3    |
|     |        | 68  | M | 64 | 9  | 8  | 7    |
|     |        | 74  | M | 55 | 6  | 7  | 6    |
| TBI |        | 5   | M | 63 | 10 | 11 | 4    |
|     |        | 11  | M | 44 | 10 | 7  | 8.6  |
|     |        | 11  | M | 44 | 10 | 9  | 14.6 |
|     |        | 11  | M | 43 | 10 | 9  | 17   |
|     |        | 17  | M | 48 | 8  | 5  | 4.4  |
|     |        | 17  | M | 48 | 8  | 7  | 5.4  |
|     |        | 17  | M | 49 | 9  | 8  | 14.4 |
|     |        | 22  | M | 24 | 9  | 5  | 22.6 |
|     |        | 28  | F | 30 | 8  | 5  | 1    |
|     |        | 28  | F | 30 | 15 | 22 | 4.1  |
|     |        | 32  | M | 42 | 11 | 17 | 7.3  |
|     |        | 32  | M | 42 | 11 | 17 | 9.5  |
|     |        | 35  | M | 56 | 8  | 7  | 1.5  |
|     |        | 43  | F | 17 | 10 | 6  | 3    |
|     |        | 45  | M | 34 | 11 | 12 | 9    |
|     |        | 66  | M | 58 | 9  | 8  | 180  |
|     |        | 76  | M | 9  | 6  | 7  | 4    |
|     |        | 77  | F | 48 | 14 | 20 | 2    |
|     |        | 77  | F | 48 | 14 | 20 | 5    |
|     |        | 77  | F | 48 | 15 | 22 | 13   |
|     |        | 78  | M | 44 | 8  | 8  | 4.5  |
|     |        | 80  | F | 61 | 6  | 7  | 6    |
|     |        | 80  | F | 61 | 6  | 7  | 6.3  |
|     |        | 80  | F | 61 | 11 | 16 | 10.6 |
|     |        | 82  | F | 57 | 6  | 7  | 6    |
|     |        | 82  | F | 57 | 6  | 7  | 6.2  |
|     |        | 89  | M | 36 | 9  | 8  | 5    |
|     |        | 96  | M | 26 | 9  | 9  | 7    |
|     |        | 97  | M | 54 | 10 | 9  | 6.3  |
|     |        | 101 | M | 47 | 8  | 7  | 7.5  |
| UWS | Anoxia | 57  | M | 60 | 7  | 4  | 180  |
|     |        | 86  | F | 38 | 8  | 5  | 10.5 |
|     | Stroke | 3   | M | 37 | 8  | 5  | 49   |
|     |        | 9   | F | 35 | 6  | 3  | 6    |
|     |        | 18  | F | 52 | 8  | 4  | 8.3  |
|     |        | 49  | M | 42 | 6  | 4  | 5.5  |
|     |        | 52  | M | 63 | 7  | 4  | 7.5  |
|     |        |     |   |    |    |    |      |

|     |     |   |    |    |   |      |
|-----|-----|---|----|----|---|------|
|     | 52  | M | 64 | 6  | 4 | 16.2 |
|     | 54  | M | 55 | 9  | 7 | 36   |
|     | 54  | M | 56 | 9  | 7 | 44.7 |
|     | 55  | F | 39 | 8  | 6 | 10   |
|     | 58  | M | 66 | 6  | 4 | 54   |
|     | 58  | M | 66 | 6  | 4 | 62.7 |
|     | 59  | F | 63 | 8  | 6 | 60   |
|     | 61  | M | 62 | 9  | 6 | 4    |
|     | 62  | M | 48 | 8  | 5 | 4    |
|     | 64  | M | 75 | 9  | 5 | 42   |
|     | 65  | M | 58 | 6  | 5 | 57   |
|     | 67  | M | 56 | 6  | 3 | 4    |
|     | 69  | M | 75 | 6  | 3 | 2    |
|     | 70  | M | 57 | 6  | 4 | 2    |
|     | 71  | F | 37 | 8  | 4 | 7    |
|     | 72  | M | 31 | 3  | 1 | 5    |
|     | 79  | M | 47 | 6  | 5 | 6.5  |
|     | 87  | M | 39 | 3  | 2 | 17.5 |
|     | 90  | F | 50 | 8  | 6 | 3    |
|     | 99  | F | 63 | 8  | 5 | 4.5  |
| TBI | 21  | M | 19 | 8  | 3 | 1.2  |
|     | 21  | M | 19 | 8  | 3 | 1.4  |
|     | 25  | M | 48 | 8  | 4 | 4.5  |
|     | 32  | M | 41 | 6  | 4 | 1    |
|     | 43  | F | 19 | 9  | 6 | 14   |
|     | 44  | M | 37 | 9  | 5 | 10   |
|     | 47  | M | 68 | 10 | 8 | 7    |
|     | 48  | M | 62 | 7  | 4 | 7    |
|     | 48  | M | 63 | 6  | 5 | 15.8 |
|     | 63  | M | 66 | 9  | 6 | 15   |
|     | 73  | M | 48 | 5  | 4 | 2.5  |
|     | 75  | F | 26 | 8  | 7 | 4    |
|     | 84  | M | 55 | 6  | 4 | 52   |
|     | 88  | M | 54 | 9  | 6 | 16   |
|     | 89  | M | 36 | 5  | 3 | 2    |
|     | 98  | F | 62 | 8  | 5 | 8.6  |
|     | 100 | M | 58 | 9  | 6 | 3    |

Rec., recording; GCS, Glasgow Coma Scale; CRS-R, JFK Coma Recovery Scale-Revised; DOC month, duration of disorder of consciousness in month; M, male; F, female; MCS, minimally conscious state; UWS, unresponsive wakefulness syndrome.

### Blinding

- 5 Both the attentional and clinical studies were double-blind (both subject and experimenter blinding was instituted). Both the subjects who participated in the studies and experimenters who collected data didn't know the real purpose of this research. Both the interviewees and interviewers obtaining the follow-up assessment were blinded to the

behavioral and EEG-based diagnosis in the hospital. In addition, the experimenters who collected EEG data were not involved in data analyses.

Specifically, in the clinical study, the behavioral measurement and EEG & MRI data analyses were conducted by two separated groups of experimenters with a completely  
 5 blinded fashion. P.G and Y.J. from the Institute of Neuroscience analyzed and modeled the EEG & MRI data were blinded with the measurements of patient outcomes. X.W., D.Z. and their colleagues from the Huashan Hospital performed the clinical evaluation of patients' behavior did not involve in the EEG and MRI data analyses.

## 10 Behavioral classification and evaluation of consciousness

Patients were categorized into a minimally conscious state (MCS) group and an unresponsive wakefulness syndrome (UWS) group according to CRS-R based behavioral evaluations<sup>55, 56</sup>.

Patients diagnosed with UWS were awake but showed no behavioral signs of  
 15 consciousness. They could open their eyes, had basic reflexes, and woke up or fell asleep at various intervals. Patients diagnosed with MCS had partial preservation of consciousness, with the presence of subtle but reproducible signs of consciousness. Moreover, patients with MCS can be subcategorized into two distinct subgroups based on the complexity of their behaviors. MCS+ describes high-level behavioral responses, and is determined by the  
 20 presence of command following, intelligible verbalization, or non-functional communication. MCS- describes low-level behavioral responses, and is determined by the presence of visual pursuit, localization of noxious stimulation, or contingent behavior related with environmental stimuli (such as smiling or crying in respond to the linguistic or visual content of emotional stimuli). Finally, patients who were able to functionally communicate and/or  
 25 use different objects were diagnosed with EMCS, i.e., emerged from MCS. The classification and behavioral ratings were performed by experienced doctors on the same day of testing, usually before EEG recording.

### Chinese materials

THE CHINESE MATERIALS IN THE AUDITORY STREAM IN THE TWO STUDIES WERE  
 30 ADAPTED FROM PREVIOUS WORK<sup>20, 53</sup>; AUDITORY STIMULI HAD ONE (WORD), TWO (WORD AND PHRASAL), OR THREE (WORD, PHRASAL, AND SENTENTIAL) LINGUISTIC LEVELS. THE SPEECH MATERIALS WERE SYNTHESIZED USING A FREE ONLINE TEXT-TO-SPEECH ENGINE ([HTTP://AI.BAIDU.COM/TECH/SPEECH/TTS](http://ai.baidu.com/tech/speech/tts)).

### Word stimuli

35 First, fifty 4-word sentences with an np-vp structure were predefined (Table 4). The 200

words were submitted to the online text-to-speech engine individually to generate their pronunciations. Subsequently, twenty 64-word sequences were generated using these words. For each sequence, 64 unique words were randomly selected from the 200-word speech library, concatenated in a row, and adjusted manually for their relative temporal spacing. Finally, the duration of each word was adjusted to 250 ms. Thus, the duration of each sequence was 16 s.

**Table 4.** Chinese Materials.

| <i>4-word sentences (np-vp)</i> |      |      |      | <i>2-word phrases (np)</i> |    |    |    |
|---------------------------------|------|------|------|----------------------------|----|----|----|
| 老牛耕地                            | 外公看报 | 军队撤退 | 青草发芽 | 老牛                         | 外公 | 军队 | 青草 |
| 朋友请客                            | 轮船起航 | 小孩哭泣 | 和尚念经 | 朋友                         | 轮船 | 小孩 | 和尚 |
| 厨师做饭                            | 渔民划船 | 农民种菜 | 蜘蛛结网 | 厨师                         | 渔民 | 农民 | 蜘蛛 |
| 电脑开机                            | 冰雪融化 | 鸭子游泳 | 祖父下棋 | 电脑                         | 冰雪 | 鸭子 | 祖父 |
| 游客爬山                            | 灯泡发光 | 蝴蝶飞舞 | 老师讲课 | 游客                         | 灯泡 | 蝴蝶 | 老师 |
| 树木生长                            | 婴儿啼哭 | 剪刀生锈 | 太阳落山 | 树木                         | 婴儿 | 剪刀 | 太阳 |
| 汽车拐弯                            | 士兵战斗 | 猎豹奔跑 | 鲸鱼喷水 | 汽车                         | 士兵 | 猎豹 | 鲸鱼 |
| 公司开张                            | 学生写字 | 英雄救火 | 法官判案 | 公司                         | 学生 | 英雄 | 法官 |
| 医生看病                            | 导游讲解 | 母鸡下蛋 | 护士打针 | 医生                         | 导游 | 母鸡 | 护士 |
| 绵羊吃草                            | 熊猫睡觉 | 行人过街 | 叛徒告密 | 绵羊                         | 熊猫 | 行人 | 叛徒 |
| 开水沸腾                            | 骏马奔驰 | 女孩跳舞 | 飞机降落 | 开水                         | 骏马 | 女孩 | 飞机 |
| 小偷逃走                            | 乌龟爬行 | 苹果成熟 |      | 小偷                         | 乌龟 | 苹果 |    |
| 兄弟吵架                            | 观众鼓掌 | 狮子打架 |      | 兄弟                         | 观众 | 狮子 |    |

### Phrase stimuli

For the fifty 4-word sentences used in the word condition, 50 noun phrases were chosen to form the phrase library (Table 4). Thirty-two phrases were randomly selected from the library and connected for each 16 s phrase sequence. A total of 20 sequences were generated. To avoid liaison in phrasal pronunciation, the speech of every phrase sequence was synthesized at the word but not the phrasal level.

### Sentence stimuli

Similar to the sequences used in the word condition, 16 sentences were randomly chosen from the fifty 4-word sentences (np-vp) and concatenated together to form a 16 s sentence sequence.

### Organization of stimuli in the tasks

In the sentence condition, three levels of semantic hierarchies were used, including single-word frequency at 4 Hz, phrasal frequency at 2 Hz, and sentential frequency at 1 Hz. The phrase condition only included word and phrasal levels. The word condition only included the 4 Hz word frequency.

For each condition of every recording, 30 sequences were randomly selected from respective 20 pre-synthesized candidates and connected to form a 480 s speech stream,

without any additional blanks between them.

### Experimental design of the attentional study

The tasks were conducted in a sound-attenuated chamber and performed using the Psychtoolbox in MATLAB (R2015b, The MathWorks Inc., USA).

5 The attentional experiment involved a full factorial design with two factors: attention (two levels, attend to or ignore) and linguistic condition (two levels, word or sentential stimuli). Thus, there were four blocks in total, with different task conditions (attend to word: attend to word audio while ignoring the simultaneous visual task; attend to sentence: attend to sentential audio while ignoring the simultaneous visual task; ignore word: attend to visual task while ignoring simultaneous word audio; ignore sentence: attend to visual while ignoring simultaneous sentential audio). Subjects were asked to either attend to or ignore a visual attention task in separate blocks, with the simultaneous presentation of 8 minute Chinese speech material (Fig. 1b). The auditory stream was adapted from previous work<sup>20, 53</sup> and consisted of Chinese monosyllabic words, which had either one (word) or three (word, 10 phrase, and sentence) linguistic levels (Fig. 1a and Table 4).

The auditory stream started 20 s after the onset of the first visual trial in each block and ended before the offset of the last visual trial, and was delivered through two loudspeakers next to the monitor  $\sim 80$  cm away from the subjects' ears at  $\sim 65$  dB SPL. In general, the auditory stream in each condition was composed of thirty 16 s long Chinese sequences with 20 no noticeable gap between them. The audio in each block was played for 8 minutes without a break.

Visual stimuli were presented on a 23 inch LCD monitor  $\sim 60$  cm from the subjects. A fixation cross was presented at the start of visual trials for 1.5 s, followed by a statement composed by a shape, a greater than or smaller than sign, and a number. After 4 s, a shape 25 matrix was presented. The shape matrix consisted of a random number ( $24 \pm 2$ ) of five shapes (isosceles right triangle, equilateral triangle, square, pentagon, and hexagon) in four colors (blue, green, yellow, and magenta). In the visual attention condition, the subjects needed to respond within 12 s of the presentation of this matrix by pressing the left or right arrow key to indicate whether the prospective statement was correct (the number of a specific shape in the matrix was greater or smaller than the given number). The assignments of the 30 keys (agree or disagree with the cue) were counterbalanced across subjects. A 2.5 s visual feedback was given as soon as the response was made to indicate whether the response was correct or incorrect. The next trial began after a 3–6.1 s inter-trial interval. There were 32 trials in each block, which lasted  $\sim 10$  min in total.

35 In the visual ignore condition, the shape matrix was always presented for 7 s in each trial

since the subjects did not need to respond. The subjects were asked to attend to the audio while ignoring the visual trials. After each block, subjects were asked to decide whether the words/sentences in a testing list had been played or not.

The order of the four task blocks was randomized and counterbalanced across subjects. EEG data were collected continuously and segmented into 16 s epochs. To obtain clean data, we excluded trials with noise, extreme movement, and eye-blinks. The mean trials used in attend to word, attend to sentence, ignore word and ignore sentence condition were 28.7, 28.2, 28.8, and 27.4, respectively.

Note that the analysis of behavioral performance of the visual task showed no significant difference between the attend to sentence and attend to word conditions (accuracy:  $75.71 \pm 1.79\%$  vs.  $76.56 \pm 2.52\%$ ,  $p = 0.796$ ; reaction time:  $6.99 \pm 0.24$  vs.  $6.84 \pm 0.33$  seconds,  $p = 0.518$ ; paired-sample  $t$ -test).

### Experimental design of the clinical study

Tests were conducted in hospital wards or similar places, and performed using the Psychtoolbox in MATLAB (R2015b, The MathWorks Inc., USA).

First, a 5-minute resting EEG was measured at the beginning of each recording session. After a 2-minute rest period, three blocks were then presented. These blocks corresponded to three 8-minute Mandarin Chinese audio sequences with different semantic levels: word, phrase, and sentence conditions (Fig. 2a). Prior to each task block, a brief introduction was played to instruct the subject to be quite and listen carefully, which was also synthesized using the same online text-to-speech engine. To reduce environmental noise, the acoustic stimuli were delivered through headphones at about 65 dB SPL, over which participants wore an additional pair of sound shielding earmuffs.

The order of task conditions was randomized and counterbalanced across subjects, controlled by a random function in MATLAB. In addition, the order of stimuli in each task condition was also shuffled across subjects.

### EEG recording

In the attentional study, data was collected using a 64-channel EEG recording system (actiCHamp, Brain Products GmbH, Germany). In the clinical study, a 257-channel system (GES 300 or GES 400, Electrical Geodesics, Inc., USA) and a 257-channel electrode cap (HCGSN 257-channel net cap, Electrical Geodesics, Inc., USA) were used. EEG signals were referenced online to the FCz (the attentional study) or Cz (the clinical study) electrode. The impedance of all electrodes was kept below 5 k $\Omega$  (the attentional study) or 20 k $\Omega$  (the clinical study). The EEG signals were sampled at 1000 Hz.

### Criteria of EEG data quality

Considering the noisy recording environment and abundant artifacts caused by patients' involuntary movements, we first examined the proportion of bad duration during each recording by applying the following criteria on each data channel (for the ITPC analysis, the number of data channel was 257; for the brain state analysis, the number was 204, since electrodes placed on the cheeks and neck were excluded firstly):

- 1) Gradient criterion: the instant voltage change exceeds the maximal allowed step, which is 30  $\mu\text{V}/\text{ms}$ ;
- 2) Max–Min criterion: the absolute difference between the maximal and minimal voltage within every 200 ms sliding window exceeds 120  $\mu\text{V}$ , and the sliding step is 10 ms;
- 3) Amplitude criterion: the absolute voltage value exceeds 100  $\mu\text{V}$ ;
- 4) Low activity criterion: the absolute difference between the maximal and minimal voltage within every 100 ms sliding window is smaller than 1  $\mu\text{V}$ , and the sliding step is 10 ms.

Data points that met any of these criteria were marked as artifacts, and the 200 ms period both before and after each artifact were marked as bad intervals. Channels with bad duration longer than 20% of the total recording length were marked as bad channels. Recordings in which the number of bad channels exceeded 70 were discarded.

### **Data pre-processing**

For the ITPC analysis, EEG data was pre-processed using BrainVision Analyzer (2.0.1, Brain Products, GmbH, Germany) as follows: data was bandpass filtered (0.1-40 Hz) with a notch filter (50 Hz) firstly, channels were semi-automatically inspected and bad ones were interpolated; data was then re-referenced to the common average of signals from all EEG channels; an independent component analysis (ICA) was performed to remove blinks and eye movements; finally, data was segmented to 16 s epochs and down-sampled to 50 Hz.

For the brain state analysis, EEG data was pre-processed in the EEGLAB toolbox (Version 14.1.1), as follows: the electrodes placed on the cheeks and on the neck were removed firstly; data of the maintained 204 electrodes were bandpass filtered (0.2–40 Hz); channels were semi-automatically inspected and bad channels were interpolated before and after ICA; an ICA was performed to remove blinks and eye movements; data was segmented into 2 s epochs and bad epochs were manually removed; finally, data was re-referenced and bandpass filtered again (2–20 Hz).

### **Phase coherence analysis and multivariate pattern analysis**

The single-trial EEG data was transformed into the frequency domain using the Discrete Fourier Transform (DFT) without additional smoothing windows. The DFT coefficient was denoted as  $X_k(f)$  for the  $k^{\text{th}}$  trial ( $k = 1, 2, \dots, n$ ) and the phase information was  $A_k(f) = \angle X_k(f)$ . The inter-trial phase coherence (ITPC) is defined as:

$$\text{ITPC}(f) = 1/n \sum_{k=1}^n (\cos(A_k(f)))^2 + 1/n \sum_{k=1}^n (\sin(A_k(f)))^2$$

Binary classifiers were used to discriminate different subject groups. Since there were three groups (healthy controls, patients with MCS, and patients with UWS), the LDA was trained for pairwise classifications at each target frequency under each task. The decoding was implemented as follows: 1) the input features were the 257 ITPC values at all EEG channels; 2) for each comparison (ITPC values of two groups at one frequency in one task condition), 4/5 subjects was randomly chosen as training set, while the other 1/5 as testing set; 3) a 5-fold cross-validation was applied on the training set, that is, for each fold, the classifier was fit on 4/5 subjects and validated on 1/5 of the training set; 4) finally, the classification performance was computed as the sum of the Area Under the Received Operative Curve (AUC), based on the probabilistic classification of the independent testing set; 5) the steps 2 to 4 were repeated 100 times to produce the mean classification AUC for these two groups at the each frequency for each condition.

### Brain state analysis

Brain state analysis was performed using MicrostateAnalysis (Version 0.3, software free at <http://www.thomaskoenig.ch/index.php/software/microstates-in-eeglab/>). For each condition, we computed EEG map topographies at the time of global field power peaks at individual level, disregarding map polarity, and identified the predominant brain state maps using k-means clustering. Four maps were selected as the optimal number of brain states, which was determined using cross-validation criterion and global explained variance. According to the best assessments of global explained variance and stability, we defined the group-averaged maps using the healthy controls as template maps of each condition.

With the template maps, we analyzed brain state probability, mean duration, mean occurrence, and mean transition probability of the healthy controls and patient groups. To summarize the spatial information of the four predominant brain states in a single subject, we calculated a *probability-weighted spatial correlation difference*,  $\Delta C_p$ . The template maps were further classified into the two following categories: the A-P map, which was created by averaging template maps 'A' and 'B'; and the L-R map, which was created by averaging template maps 'C' and 'D'. For each subject, spatial correlation of each given map corresponds to the spatial Pearson's correlation between the given tested map and the template maps (A-P and L-R map) averaged from healthy subjects. The difference in the spatial correlation with the two template maps ( $\Delta C$ ) indicates the similarity of the four maps in each patient compared to the healthy controls<sup>26</sup>.

Each difference of spatial correlation corresponds to the spatial Pearson's correlation, which was calculated as follows:

$$\Delta C = \frac{\sum_i^n (I_i \cdot \mathbf{V}_{AP,i})}{\sqrt{\sum_i^n I_i^2} \cdot \sqrt{\sum_i^n \mathbf{V}_{AP,i}^2}} - \frac{\sum_i^n (I_i \cdot \mathbf{V}_{LR,i})}{\sqrt{\sum_i^n I_i^2} \cdot \sqrt{\sum_i^n \mathbf{V}_{LR,i}^2}}$$

$\Delta C$  is the difference of spatial correlation of the two template maps,  $n$  is the number of electrodes,  $I$  is the measured voltage of individual map,  $\mathbf{V}_{AP}$  is the measured voltage of A-P template map and  $\mathbf{V}_{LR}$  is the measured voltage of L-R template map,  $i$  is the electrode  $i$ .

Accordingly, we calculated the probability-weighted spatial correlation difference as follows:

$$\Delta C_p = \sum_{k=1}^4 \Delta C_k \cdot \rho_k$$

where  $\rho$  is the probability of a given map (Fig. 3c, Map A, B, C and D), and  $k$  denotes the map  $k$ .

### Diagnosis and Prediction analysis

We first used classification analysis to identify the consciousness states of individuals. The exclusion criteria for the classification dataset were as follows: (1) patients with a DOC duration shorter than 3 months; (2) patients that had received deep brain stimulation in the last 120 days; and (3) patients with an unstable level of consciousness caused by an unexpected disease. After exclusion, data from a final total of 47 healthy controls, 31 patients with MCS, and 30 patients with UWS were included. These feature combinations were used to train three-class LDA classifiers to discriminate between healthy controls, patients with MCS, and patients with UWS. There were in total 893 possible feature combinations when using the EEG metrics from all three levels of the language task (the sentence condition contains 3 ITPC metrics and 6 brain state metrics, which produces  $2^9 - 1 = 511$  feature combinations; the phrase and word condition produces 255 and 127 feature combinations respectively). For each classification, to avoid model overfitting, only one out of 893 feature combinations was selected and submitted to the model. All steps were cross-validated (leave-one-subject-out). A classifier with regularization first searched for the optimal feature combination within each task, and calculated the classification probability for each individual subject. To avoid model overfitting, only those selected feature combinations were entered in the final LDA. The regularized version of LDA was used by estimating covariance matrices. Cross-validation relied on the leave-one-subject-out method with 108 permutations. Considering the bias effect of unequal class sizes in the LDA classification, we did not rely on uneven prior probabilities for class sizes, but assumed that all classes have the same number of

samples<sup>57, 58</sup>. We thus randomly chose 29 samples from the healthy controls and MCS group individually, in order to match the sample number of the UWS group. This sampling process was repeated 2000 times. For the model fitting of each feature combination, the accuracy of the classification was averaged over the 2000 permutations. The mean accuracies of classifications  
 5 allowed us to determine the optimal feature combination. For individual subject, under the optimal feature combination, the maximum probability during the 2000 permutations decided which group a given subject was classified to.

Subsequently, to predict patient outcome, we selected the patients for whom we had behavioral measurements more than 100 days after EEG assessments. Each patient was first  
 10 labeled as showing a positive or negative outcome. The clinical diagnoses of patients could be subcategorized into four different subclasses with proposed ascending levels of consciousness, namely UWS/VS, MCS-, MCS+ and EMCS<sup>12, 55</sup>. Here, positive outcome was defined as any advance in the transition of clinical categorization during follow-up, while negative outcome was defined as stasis or retrogress in the transition. Data from a total of 38  
 15 patients were used in the prediction analysis, including 15 positive-outcome patients (10 MCS and 5 UWS) who became fully awakened or exhibited significantly improved behavioral signs in the follow-up measurements, and 23 negative-outcome patients (7 MCS and 16 UWS). Similarly, the process of consciousness state classification was used to classify the outcomes of individuals. The classifier was built using three classes – healthy  
 20 controls, positive-outcome patients, and negative-outcome patients – that corresponded to normalized coefficients ( $\omega$ ) of 1 (healthy control), 0.5 (positive-outcome), and 0 (negative-outcome). Under the optimal feature combination, we applied the normalized coefficients to classification probabilities ( $P$ ), and then defined the weighted sum as the predicted score ( $\phi = \sum P \bullet \omega$ ) of task-single prediction. As the task-single predictive scores  
 25 varied across task conditions for individual subjects, the task-mean prediction was used and defined as the average of three scores from the three task conditions.

The external validation (generalization ability) of the classifier was examined on the new dataset (25 patients), which contained 15 positive-outcome patients (5 MCS and 10 UWS) and 10 negative-outcome patients (7 MCS and 3 UWS) (Fig. 5d, f and Fig.13). The  
 30 classifier (LDA) for outcome prediction using EEG metrics was first trained on the dataset of 38 patients with the cross-validation procedure within the dataset, and then tested on the new dataset of 25 patients. The similar analysis procedures for outcome prediction and generalization were performed using behavioral features: CRS-R total-scores and 6 subscales (1, Auditory [0-4]; 2, Visual [0-5]; 3, Motor [0-6]; 4, Oromotor [0-3]; 5, Communication  
 35 [0-2]; 6, Arousal [0-3]). For the CRS-R total-score classifier, we also computed chance

performance by repeating the same generalization 100 times using shuffled outcome labels of the testing dataset.

The direct comparisons of outcome prediction and its generalization between EEG and CRS-R scores were also examined by using LDA without searching for the optimal feature combinations (Fig.13e). The input features for training the two-class classifier were values of the EEG or CRS-R (total-score and 6 subscales) metrics. The labels corresponding to each subject (samples) were either outcome-positive or negative. We used 5-fold cross-validation in all the task conditions, with random samples allocated to folds stratified by labelled class.

We plotted the ROC curves of predicted scores to carry out AUC measurements, which were used to estimate the abilities of task-single and task-mean to prognosticate outcomes. The optimal threshold for prognosticating outcomes was determined by the point with maximal sum of sensitivity and specificity on the ROC curve. The corresponding predictive threshold was equal to 0.1 after normalization (Figs. 5e, f). Patients with predicted scores that were higher than the threshold were identified as positive-outcome. The prediction accuracy was calculated by comparing the predicted labels of patients and their actual outcome in the follow-up diagnosis.

### **MRI data acquisition and extracting the volume of brain injury**

MRI structure images were acquired from the twenty-seven patients (16 males; mean age = 44.6 years; range = 9 to 68 years) on the same day as the EEG recordings. MRI data were collected by a 3 Tesla MRI scanner (n = 21, Siemens Magnetom Verio, Germany, using turbo spin-echo sequence), or a 1.5 Tesla MRI scanner (n = 6, GE Signa EXCITE Twinspeed Zoom, USA, using a fast spin-echo sequence). By checking the structural contrast, head motion of images and reliability of lesion detection<sup>59</sup>, we identified hyperintensity of brain lesions in T2WI data from 21 patients, hypointensity of brain lesions in FLAIR data from 4 patients, and hypointensity of brain lesions in T1WI data from 2 patients. The areas of the lesions were drawn on each image manually under doctors' instruction. The volume of the lesions was calculated by summing all voxels within the lesion areas across all slices, and multiplied by the voxel size<sup>60, 61</sup>. We then evaluated the correlation between  $\Delta C_p$  and lesion volume using Pearson's correlation. The analyses of MRI data were performed using MATLAB (2017b, MathWorks, USA) and ITK-SNAP (3.8).

### **Statistics**

For the ITPC analyses, the significance tests were applied to individual subject and group subjects respectively. At the individual level, the one-sided exact test was recruited. For ITPC between 0.2 and 5 Hz, 77 frequencies were used in total (1/16 Hz for each bin). The null hypothesis is that, the response phase is not synchronized to the stimulus and the ITPC

at the target frequency is not significantly larger than those in neighboring frequencies. Thus, the statistical significance (exact  $P$ ) of the response at a target frequency is the probability that the target frequency response differs from the null distribution (non-target frequencies; numbers of non-target frequencies within subject under the three conditions: 76 frequencies for word, 75 for phrase, and 74 for sentence). At the group level, the chance-level phase coherence for each target frequency is the average of its neighboring non-target frequencies (4 bins on each side of each target frequency, which is equivalent to 0.25 Hz). The statistical significance is the difference between the response at a target frequency and the response at its neighbors (one-sided paired-sample  $t$ -test). For classification results, one-sided one-sample  $t$ -tests were applied to examine the significances of decoding performance, comparing with the chance level of 0.5.

In the brain state analysis, for the probability of 4 maps, the main effect of group under each condition was examined using MANOVA. A repeated measures ANOVA was used to evaluate the group effect in each task condition, in which EEG metrics and Task (3 levels: word, phrase, and sentence) were treated as repeated measures while Group (3 levels: healthy control, MCS, and UWS) was the between-subjects factor. For pairwise comparisons between the three groups, one-way ANOVA tests (Bonferroni corrected) were applied to all EEG metrics in each condition. In addition, for follow-up patients, Friedman tests were used to test the changes in brain state parameters between their first and last recordings. Chi-squared tests (Fisher's exact test) were used to estimate the statistical significance of the match between the classified/predicted labels and the diagnosed labels in the classification and prediction analyses. All data distributions were assumed to be normal, but this was not formally tested.

### **Inter-trial phase coherence (speech-tracking activity)**

We constructed hierarchical linguistic structures using an isochronous, 4 Hz sequence of Chinese words that were independently synthesized (Fig. 1, see *Methods*). The auditory sequences included three linguistic levels, that is, monosyllabic words, 2-word phrases, and 4-word sentences. Monosyllabic words were presented at constant rates (250 ms per word), which means that the corresponding neural tracking of words, phrases, and sentences could be tracked at distinct frequencies (4, 2, and 1 Hz respectively; Fig. 1a).

We first tested whether top-down attention is required for the EEG responses to hierarchical language structures. We recruited 22 healthy human subjects to perform an attentional task. The subjects either attended to the auditory stimuli (the attend to condition), which were either word lists or sentence sequences, or performed a visual task while the auditory stimuli were simultaneously presented and ignored (the ignore condition; Fig. 1b,

*Methods*). In both the attend to and ignore conditions, we found a significant and compatible 4-Hz response in the inter-trial phase coherence (ITPC) spectrum, for both the word list and sentence conditions, compared with baseline (Fig. 1c). However, the ITPC values at 1 and 2 Hz in the sentence conditions were significantly weakened after the attention was shifted to the visual stimuli (attend to *vs.* ignore,  $P_{1\text{Hz}} = 0.015$  and  $P_{2\text{Hz}} = 2.3 \times 10^{-5}$ ; paired-sample *t*-test; Figs. 1c, d). Thus, these results indicate the automaticity of processes underlying single-word tracking (4 Hz), and partially attentional modulation of neural processing of higher-level linguistic structures, i.e., phrases and sentences (respectively at 2 and 1 Hz).

Based on these results, our hypothesis was that residual consciousness in patients with DOC could be reflected by the strength of speech-tracking responses, especially neural tracking of higher-level linguistic structures, i.e., phrases and sentences (measured by ITPC). To test this hypothesis, we examined brain responses to sentence sequences in 42 patients with MCS, 36 patients with UWS, and 47 healthy controls (see patients' details in Table 1 and patients selection in Fig.6). As shown in Fig. 2a, after a clinical diagnosis using the Coma Recovery Scale-Revised (CRS-R) and Glasgow Coma Scale (GCS) (the categorization of MCS and UWS was done after the CRS-R diagnosis), a 5-minute resting-state EEG was first recorded at the start of each recording session. After a 2-minute rest period, three 8-minute blocks containing Mandarin Chinese speech stimuli with different linguistic levels (word lists, phrase sequences, and sentence sequences) were presented. The results showed a progressive increase in the strength of EEG ITPC that matched the increasing level of behavioral responsiveness as quantified by the CRS-R, from UWS to MCS, alongside the healthy control group for comparison. Specifically, word-level tracking, measured by the 4-Hz ITPC, is significant in the healthy control, MCS, and UWS groups ( $P_{4\text{Hz-Healthy}} = 1.3 \times 10^{-10}$ ;  $P_{4\text{Hz-MCS}} = 2.1 \times 10^{-6}$ ;  $P_{4\text{Hz-UWS}} = 5.8 \times 10^{-4}$ ; paired-sample *t*-test; Fig. 2b *left* and Fig.7). Phrase-level tracking, measured by the 2-Hz ITPC, is significant in the healthy control group, marginally significant in the MCS group, and not significant in the UWS group ( $P_{2\text{Hz-Healthy}} = 3.8 \times 10^{-9}$ ;  $P_{2\text{Hz-MCS}} = 0.097$ ;  $P_{2\text{Hz-UWS}} = 0.881$ ; paired-sample *t*-test; Fig. 2b *middle* and Fig.7). Sentence-level tracking measured by the 1-Hz ITPC, is significant in the healthy control group, but not significant in MCS or UWS group ( $P_{1\text{Hz-Healthy}} = 4.9 \times 10^{-5}$ ;  $P_{1\text{Hz-MCS}} = 0.567$ ;  $P_{1\text{Hz-UWS}} = 0.546$ ; paired-sample *t*-test; Fig. 2b *right* and Fig.7).

It is worth noting that, although there were no group-level significant differences in 1 or 2-Hz ITPC between the MCS and UWS groups (Fig. 2b), some differences at the individual level were apparent (Fig.8). Namely, eleven patients with MCS and four patients with UWS exhibited significant ITPC at 1 or 2 Hz, which could indicate residual consciousness in these patients. Indeed, six (5 MCS and 1 UWS) of these fifteen (40.0 %) patients showed

significant improvement of clinical diagnosis 100 days after the EEG recordings (outcome predictions also see the classification results). We then applied multivariate pattern analysis on the 1, 2, and 4 Hz ITPC for all available electrodes to classify the patient groups (linear discriminant analysis, LDA; see *Methods*). Figure 2c shows successful decoding of patient groups. Even MCS and UWS groups were significantly distinguishable, particularly when listening to sentences.

### Temporal dynamics of global brain states

The brain is inherently active in a regular manner at both rest and during cognitive tasks, and this dynamic pattern has been proposed to be the neural signature of consciousness<sup>9, 23</sup>. We hence evaluated the second hypothesis that residual consciousness can be characterized by monitoring the dynamic patterns of brain states, in that these brain dynamics would be associated with different cognitive states. To this aim, we quantified the spatial and temporal dynamics of brain activity in healthy controls and patients by examining the properties (e.g., probability, occurrence, duration, and transition) of the global pattern of scalp potential topographies (also referred as “microstates”)<sup>11, 24</sup> in conditions with four increasing levels.

Group-level clustering identified an optimum of four clusters across groups and conditions, which reached the highest cross-validation criterion and explained approximately 80% of variance (Fig. 3a). The spatial configurations of the four maps in healthy controls (Fig. 3b) were highly consistent with the four maps described in previous studies<sup>25, 26</sup>. We then labeled and sorted the four sets of maps according to the appearing probability of brain states at resting state in the healthy controls. Specifically, map A showed a fronto-central maximum, map B showed a symmetric frontal to occipital orientation, map C showed a left occipital to right frontal, and map D showed a right occipital to left frontal orientation (Fig. 3b; the group-averaged brain states of each group in each condition are shown in Fig.9).

Previous studies using simultaneous EEG and fMRI recordings have suggested that brain states A and B are more closely related to the attention and saliency networks, as their corresponding blood-oxygen-level-dependent (BOLD) activations were located in the anterior cingulate cortex and parietal-frontal areas, and that states C and D are related to the auditory and visual sensory networks, as their corresponding BOLD signals were located in bilateral temporal and extrastriate visual areas<sup>11, 27</sup>. We thus predicted that higher levels of consciousness would be associated with a higher probability of the activation of high-level cognitive neural networks, that is, maps A and B (anterior-posterior maps, defined as the A-P map, Fig. 3b). In parallel, we predicted that reduced consciousness would involve a greater relative contribution of lower-level sensory areas, corresponding to maps C and D (left-right maps, defined as the L-R map, Fig. 3b). Multivariate analysis of variance (MANOVA)

showed that, for the four conditions, the healthy controls demonstrated a pattern of a high probability of the A-P map and a low probability of the L-R map (Fig. 3c). The patient groups showed the opposite pattern, with a low probability of the A-P map and a high probability of the L-R map (Fig. 3c).

5       Next, we examined the difference of the dynamic of brain states between the MCS and UWS groups. For each individual patient, we quantified the probability-weighted spatial correlation difference between the A-P and L-R maps ( $\Delta C_p$ , see *Methods*). This difference reflected the spatial similarity between the maps of patient group and template maps derived from the healthy control group, and was used as an index of residual consciousness. We  
10       found a progressive increase in the difference of  $\Delta C_p$  between the MCS and UWS group as the level of the linguistic hierarchy increased from resting to word, phrase and sentence conditions (Figs. 3d, e). Specifically, at the phrase and sentence levels (but not at resting or word levels), the MCS group showed a significantly higher  $\Delta C_p$  than the UWS group ( $P_{\text{Phrase}} = 0.037$ ;  $P_{\text{Sentence}} = 0.014$ ; one-way ANOVA, Bonferroni corrected; Fig. 3e). This was  
15       indicative of an increased probability of the A-P map (frontal-parietal networks) and a decreased probability of the L-R map (sensory networks). Furthermore, the difference in  $\Delta C_p$  between MCS and UWS was significantly larger in the phrase and sentence condition than that in the resting and word condition (phrase *vs.* word,  $t_{76} = 2.29$ ,  $P = 0.03$ ; sentence *vs.* word,  $t_{76} = 3.14$ ,  $P = 0.002$ ; two-sided two-sample *t*-test). Thus, positive and high  $\Delta C_p$   
20       potentially may indicate residual consciousness.

We then investigated whether the probability difference of the maps was due to the duration (how long the map remained stable) or the frequency of occurrence (how many times the map occurred in one second) of each map. We found that the probability differences between the MCS and UWS groups in the phrase and sentence conditions could  
25       be attributed to a shorter duration of the L-R map, thought to reflect sensory networks ( $P_{\text{Phrase}} = 0.016$ ;  $P_{\text{Sentence}} = 0.017$ ; one-way ANOVA, Bonferroni corrected; Figs. 4a, b), and a higher occurrence of the A-P map, putatively associated with frontal-parietal networks ( $P_{\text{Phrase}} = 0.028$ ;  $P_{\text{Sentence}} = 0.063$ ; one-way ANOVA, Bonferroni corrected; Figs. 4d, e) in the MCS group relative to the UWS group. Importantly, the increase of the significant difference  
30       between the MCS and UWS groups matched the increasing linguistic level of conditions (Figs. 4b, e). No significant differences were found in the duration of the A-P map or in the occurrence of the L-R map (Fig.10).

If the duration and/or occurrence of maps indeed reflect the strength of residual consciousness in patients, we should observe corresponding changes before and after their  
35       recovery. In the sub-population of subjects who had multiple EEG recordings (12 out of 54

patients for resting-state recordings, 15 out of 60 patients for recordings during the linguistic tasks), the duration of the L-R map became shorter (Figs. 4a, c) and the occurrence of the A-P map increased along with recovery (Figs. 4d, f). Such changes were not observed in the non-recovery patients (Figs. 4a, c, d, f). The distribution of the A-P map duration and L-R map occurrence in healthy controls, recovery and non-recovery patients with MCS and UWS are shown in Fig.10.

To exclude the possibility that the different spatial maps in the two patient groups arose from a difference in brain injury, we then analyzed the volumes of brain damage in the 27 patients (17 MCS and 10 UWS) from the database who had received a structural magnetic resonance imaging scan on the same day as the EEG recordings. The results showed that there was no significant difference in the injured volume between the MCS and UWS patients (Fig.11a), and no significant correlation between the volume of brain injury and the  $\Delta C_p$  at the all three levels of task (Fig.11b). For example, the patient 07 with no brain damage has a low value of  $\Delta C_p$  (Fig.11c), by contrast, the patient 17 with a high volume of brain injury shows a relative higher  $\Delta C_p$  (Fig.11d). Furthermore, if we assume that the size of brain damage would not change dramatically during the recovery period, the observed changes in spatial maps before and after recovery (Fig. 4) also help rule out the possibility that the brain states measured with spatial maps purely reflect the underlying brain injury. The change of spatial maps of an example patient (ID: 2) is shown in Fig.11g-i.

## Individual diagnosis and prediction

Our multiple measurements of brain activity allowed us to move from group-level analyses and attempt personalized diagnoses and predictions. We first trained a three-class LDA classifier with leave-one-subject-out cross-validation for the diagnosis of individual control, MCS and UWS subjects who had at least 3 months duration of DOC. The inputs to the classifier were the multiple EEG measurements, including the ITPC (3 features: ITPC values at 1, 2, and 4 Hz) and global dynamic patterns of brain activity (6 features:  $\Delta Probability$ ,  $\Delta C_p$ ,  $Occurrence_{A-P}$ ,  $Duration_{L-R}$ ,  $Transition_{A-P}$  and  $Transition_{L-R}$ ). A classifier with regularization first searched for the optimal feature combination within each task, and calculated the classification probability for each individual subject (Fig. 5a, detailed information of feature selection see *Methods*). Then, to avoid model overfitting, only those selected feature combinations were entered in the final LDA. All steps were cross-validated (leave-one-subject-out). Figure 5b plots the confusion matrix generated by the LDA. The best classification was found in the sentence condition with  $\Delta C_p$ ,  $Transition_{A-P}$ ,  $ITPC_{1Hz}$ , and  $ITPC_{2Hz}$  as input EEG features. A chi-squared test was used to estimate the classifier's performance, which was highly significant ( $\chi^2 = 95.84$ ,  $P = 7.6 \times 10^{-20}$ , accuracy = 75%). The

decoder categorized healthy control, MCS, and UWS subjects with 89%, 58%, and 70% accuracy, respectively, all well above the chance level of 33%. The high decoding accuracy was confirmed by another discriminative classifier, support vector machine (SVM), with 96%, 65%, and 73% accuracy for healthy control, MCS and UWS subjects (Fig.12a).

Although a proportion of patients with UWS were classified as patients with MCS (30%, 9 of 30), it is possible that these patients had some degree of consciousness that was not detected by the CRS-R in the behavioral assessment. Interestingly, a greater proportion (33.3%, 3 of 9) of such potentially misdiagnosed patients with UWS had positive outcomes (fully awakened or exhibited improved behavior after the EEG recording), as compared with patients for which both the CRS-R and EEG classifier agreed on a diagnosis of UWS (9.5%, 2 of 21). Conversely, the classifier also diagnosed some patients with MCS as UWS (12 of 31). A lower proportion (25%, 3 of 12) of those potentially misdiagnosed MCS patients had positive outcomes as compared with patients diagnosed as MCS by both the CRS-R and EEG measurements (44.4%, 8 of 18). However, we should note that the number of patients in these groups might be too small to generate sufficient power for statistical analysis of between-group differences in these proportions.

We also examined whether EEG recordings could predict the subsequent recovery of consciousness in individual patients. Thirty-eight patients with multiple measurements were included in this analysis. Based on the CRS-R total-score  $\geq 6$  months after DOC onset<sup>28</sup>, 15 of them had a positive outcome, denoted as +ve (for patients selection see Fig.6). We used the same cross-validated method as above to construct an LDA classifier, this time aiming to separate the 15 outcome-positive and 23 outcome-negative patients. The results showed that while the CRS-R total-score could partially predict outcomes (AUC = 70%,  $\chi^2 = 7.2$ ,  $P = 0.016$ , chi-squared test; Fig. 5c), the prediction using the EEG measurements was better (AUC = 77%,  $\chi^2 = 11.5$ ,  $P = 9.2 \times 10^{-4}$ , chi-squared test; Fig. 5e *left*; Sensitivity: EEG vs. CRS-R,  $P = 0.07$ , McNemar's test). The best EEG predictive ability was achieved by the mean performance (see *Methods*) of the word, phrase, and sentence conditions with 87% sensitivity and 70% specificity, resulting in correct prediction for 13 of 15 outcome-positive and 16 of 23 outcome-negative patients. The best feature combinations for the prediction at each task level were  $\Delta\text{Probability} + \text{Occurrence}_{\text{A-P}}$  for the word condition,  $\text{Occurrence}_{\text{A-P}} + \text{ITPC}_{4\text{Hz}}$  for the phrase condition, and  $\Delta\text{Cp} + \text{Occurrence}_{\text{A-P}}$  for the sentence condition (Table 3). We then examined the prognostic ability within individual patients by calculating the predictive scores within UWS and MCS patients who had CRS-R based clinical outcomes. The prediction results demonstrated a high prediction accuracy in both groups (81% UWS and 71% MCS; individual prediction scores are shown in Fig. 5e *right*). Thus, prediction

using EEG measures was better than using behavioral observations alone, either using a single CRS-R measure in our study, or using multiple CRS-R measurements in another study<sup>29</sup>.

It is worth noting that, although the number of subjects in the dataset for prediction model was relatively small (38 patients), the best feature combination for the classification only consisted of 2 or 3 features, suggesting that it was unlikely the model was overfitted. Nevertheless, to ensure the reliability of our results, we validated the same dataset using a different classifier, SVM, with cross-validation. This additional analysis confirmed our results showing a significant predictive accuracy from EEG (Fig.12b,  $\chi^2 = 15.4$ ,  $P = 1.6 \times 10^{-4}$ , accuracy = 82%, chi-squared test).

Most importantly, to further test the external validity and the generalization ability of our models, we tested them (without retraining) on a new dataset consisting of 25 additional patients (12 MCS and 13 UWS, Table 2). The classifier (LDA) trained with the previous dataset using the same feature combinations, showed a high predictive accuracy in both outcome-positive and -negative groups across the two sample sets (Fig. 5f; *left*,  $\chi^2 = 8.8$ ,  $P = 0.005$ , accuracy = 80%, chi-squared test; individual prediction scores in Fig. 5f *right*). However, similar generalization using the classifier trained with the CRS-R total-score showed a much lower predictive accuracy (Fig. 5d;  $\chi^2 = 4.6$ ,  $P = 0.049$ , accuracy = 28%, chi-squared test). As a control, the classification accuracy was 50% when we shuffled the outcome label of the testing dataset; and the classifier for diagnosis (i.e. MCS versus UWS), once trained with CRS-R scores on the first 38 patients, successfully generalized to the new dataset (Fig.13a,  $\chi^2 = 21.3$ ,  $P = 2.7 \times 10^{-6}$ , accuracy = 96%, chi-squared test).

Could the higher predictive accuracy with EEG measurements be due to the larger number of EEG features used? To evaluate this, we expanded the behavioral measurements by including the total score and six sub-scores (e.g. auditory, visual, motor, oromotor, communication and arousal) and using the same model to search for the optimal CRS-R feature combination for prognosis. With cross-validation on the first 38 patients, the best prediction was obtained using the visual subscale alone (Fig.13b *left*), but again that classifier failed to generalize to the new dataset of 25 patients (Fig.13b *right*). Direct comparison of prediction performance between CRS-R or the EEG recording under the word condition (both have seven features) yielded a similar result (Fig. 13c,d). Furthermore, we examined the outcome prediction using a standard LDA model without feature selection before classification. The results confirmed the superior generalization ability using EEG metrics than using CRS-R features (Fig.13e).

Finally, the confusion matrix generated by the CRS-R scores showed that an important

proportion of outcome-positive patients (53%, 8 of 15; 3 MCS and 5 UWS) was mis-predicted as outcome-negative (Fig. 5c). Compared with the CRS-R, our EEG task-based assessments could potentially contribute to more accurate diagnoses, as the classifier that we constructed using EEG measurements was able to significantly predict future CRS-R dichotomized outcomes (Fig. 5e): six of those eight patients were classified as outcome-positive by the EEG-based classifier. Furthermore, we combined the features from both CRS-R and EEG and submitted them to the model. The results showed that, at the word and phrase levels, the classification accuracies by using both EEG and CRS-R were slightly higher than those using only EEG signals (classification accuracies at all levels see Fig.14 and Table 3). The best feature combination in the classification indeed contained the CRS-R, which might indicate that the prediction model could benefit from the use of both EEG and CRS-R. Taken together, these results confirm the prognostic ability of EEG metrics recorded during language listening and suggest that these EEG measures may complement demographic and behavioral diagnoses in the clinic.

The outcome of the patients in the prediction was based on a single CRS-R, which could lead to misdiagnosis due to the variability of the level of consciousness over time (e.g. morning vs. afternoon within a day and across different days, see reference<sup>30</sup>). Although we examined a subset of patients (n = 15) who had multiple diagnoses before and after EEG recordings and found relatively stable CRS-R scores across days at both group and individual levels (Fig.15), repetitive behavioral measurements (CRS-R) is likely to be important for accurate diagnosis and prediction, and a systematic comparison of the predictive value with multiple EEG measurements and multiple CRS-R scores remains to be explored.

**Table 2.** Detailed demographic and clinical information of new-collected patients recruited during year 2018-2019.

| Group | Etiology | ID  | Sex | Age (years) | GCS | CRS-R | DOC month |
|-------|----------|-----|-----|-------------|-----|-------|-----------|
| MCS   | Anoxia   | 102 | M   | 20          | 8   | 9     | 9.5       |
|       |          | 115 | M   | 20          | 7   | 8     | 6.5       |
|       | Stroke   | 33  | F   | 48          | 12  | 13    | 26.9      |
|       |          | 78  | M   | 48          | 13  | 18    | 18.9      |
|       |          | 104 | M   | 65          | 9   | 9     | 12.2      |
|       |          | 104 | M   | 64          | 6   | 8     | 5.2       |
|       |          | 104 | M   | 64          | 9   | 9     | 6.2       |
|       |          | 118 | M   | 30          | 8   | 8     | 7.1       |
|       | TBI      | 109 | M   | 52          | 8   | 8     | 4.3       |
|       |          | 109 | M   | 52          | 8   | 8     | 10.4      |
|       |          | 112 | M   | 45          | 8   | 7     | 9.2       |
|       |          | 114 | M   | 32          | 8   | 8     | 18.1      |
| UWS   | Anoxia   | 102 | M   | 19          | 8   | 6     | 6.4       |

|        |     |   |    |   |   |      |
|--------|-----|---|----|---|---|------|
|        | 117 | M | 52 | 7 | 5 | 19.6 |
|        | 117 | M | 52 | 9 | 6 | 21.5 |
|        | 119 | M | 18 | 7 | 5 | 8.3  |
| Stroke | 111 | F | 31 | 8 | 5 | 5.3  |
|        | 113 | M | 46 | 8 | 6 | 5.2  |
|        | 113 | M | 46 | 8 | 6 | 6.2  |
|        | 116 | F | 69 | 8 | 6 | 5.5  |
|        | 118 | M | 29 | 8 | 5 | 6.2  |
| TBI    | 96  | M | 27 | 8 | 6 | 11.1 |
|        | 103 | M | 46 | 4 | 1 | 5.6  |
|        | 109 | M | 52 | 8 | 6 | 7    |
|        | 110 | F | 30 | 8 | 6 | 19.1 |

### Example 1. Inter-trial phase coherence (speech-tracking activity)

We constructed hierarchical linguistic structures using an isochronous, 4 Hz sequence of Chinese words that were independently synthesized (Fig. 1, see *Methods*). The auditory sequences included three linguistic levels, that is, monosyllabic words, 2-word phrases, and 4-word sentences. Monosyllabic words were presented at constant rates (250 ms per word), which means that the corresponding neural tracking of words, phrases, and sentences could be tracked at distinct frequencies (4, 2, and 1 Hz respectively; Fig. 1a).

We first tested whether top-down attention is required for the EEG responses to hierarchical language structures. We recruited 22 healthy human subjects to perform an attentional task. The subjects either attended to the auditory stimuli (the attend to condition), which were either word lists or sentence sequences, or performed a visual task while the auditory stimuli were simultaneously presented and ignored (the ignore condition; Fig. 1b, *Methods*). In both the attend to and ignore conditions, we found a significant and compatible 4-Hz response in the inter-trial phase coherence (ITPC) spectrum, for both the word list and sentence conditions, compared with baseline (Fig. 1c). However, the ITPC values at 1 and 2 Hz in the sentence conditions were significantly weakened after the attention was shifted to the visual stimuli (attend to *vs.* ignore,  $P_{1\text{Hz}} = 0.015$  and  $P_{2\text{Hz}} = 2.3 \times 10^{-5}$ ; paired-sample *t*-test; Figs. 1c, d). Thus, these results indicate the automaticity of processes underlying single-word tracking (4 Hz), and partially attentional modulation of neural processing of higher-level linguistic structures, i.e., phrases and sentences (respectively at 2 and 1 Hz).

Based on these results, our hypothesis was that residual consciousness in patients with DOC could be reflected by the strength of speech-tracking responses, especially neural tracking of higher-level linguistic structures, i.e., phrases and sentences (measured by ITPC). To test this hypothesis, we examined brain responses to sentence sequences in 42 patients with MCS, 36 patients with UWS, and 47 healthy controls (see patients' details in Table 1

and patients selection in Fig.6). As shown in Fig. 2a, after a clinical diagnosis using the Coma Recovery Scale-Revised (CRS-R) and Glasgow Coma Scale (GCS) (the categorization of MCS and UWS was done after the CRS-R diagnosis), a 5-minute resting-state EEG was first recorded at the start of each recording session. After a 2-minute rest period, three 8-minute blocks containing Mandarin Chinese speech stimuli with different linguistic levels (word lists, phrase sequences, and sentence sequences) were presented. The results showed a progressive increase in the strength of EEG ITPC that matched the increasing level of behavioral responsiveness as quantified by the CRS-R, from UWS to MCS, alongside the healthy control group for comparison. Specifically, word-level tracking, measured by the 4-Hz ITPC, is significant in the healthy control, MCS, and UWS groups ( $P_{4\text{Hz-Healthy}} = 1.3 \times 10^{-10}$ ;  $P_{4\text{Hz-MCS}} = 2.1 \times 10^{-6}$ ;  $P_{4\text{Hz-UWS}} = 5.8 \times 10^{-4}$ ; paired-sample  $t$ -test; Fig. 2b *left* and Fig.7). Phrase-level tracking, measured by the 2-Hz ITPC, is significant in the healthy control group, marginally significant in the MCS group, and not significant in the UWS group ( $P_{2\text{Hz-Healthy}} = 3.8 \times 10^{-9}$ ;  $P_{2\text{Hz-MCS}} = 0.097$ ;  $P_{2\text{Hz-UWS}} = 0.881$ ; paired-sample  $t$ -test; Fig. 2b *middle* and Fig.7). Sentence-level tracking measured by the 1-Hz ITPC, is significant in the healthy control group, but not significant in MCS or UWS group ( $P_{1\text{Hz-Healthy}} = 4.9 \times 10^{-5}$ ;  $P_{1\text{Hz-MCS}} = 0.567$ ;  $P_{1\text{Hz-UWS}} = 0.546$ ; paired-sample  $t$ -test; Fig. 2b *right* and Fig.7).

It is worth noting that, although there were no group-level significant differences in 1 or 2-Hz ITPC between the MCS and UWS groups (Fig. 2b), some differences at the individual level were apparent (Fig.8). Namely, eleven patients with MCS and four patients with UWS exhibited significant ITPC at 1 or 2 Hz, which could indicate residual consciousness in these patients. Indeed, six (5 MCS and 1 UWS) of these fifteen (40.0 %) patients showed significant improvement of clinical diagnosis 100 days after the EEG recordings (outcome predictions also see the classification results). We then applied multivariate pattern analysis on the 1, 2, and 4 Hz ITPC for all available electrodes to classify the patient groups (linear discriminant analysis, LDA; see *Methods*). Figure 2c shows successful decoding of patient groups. Even MCS and UWS groups were significantly distinguishable, particularly when listening to sentences.

## Example 2. Temporal dynamics of global brain states

The brain is inherently active in a regular manner at both rest and during cognitive tasks, and this dynamic pattern has been proposed to be the neural signature of consciousness<sup>9, 23</sup>. We hence evaluated the second hypothesis that residual consciousness can be characterized by monitoring the dynamic patterns of brain states, in that these brain dynamics would be associated with different cognitive states. To this aim, we quantified the spatial and temporal

dynamics of brain activity in healthy controls and patients by examining the properties (e.g., probability, occurrence, duration, and transition) of the global pattern of scalp potential topographies (also referred as “microstates”) <sup>11, 24</sup> in conditions with four increasing levels.

Group-level clustering identified an optimum of four clusters across groups and conditions, which reached the highest cross-validation criterion and explained approximately 80% of variance (Fig. 3a). The spatial configurations of the four maps in healthy controls (Fig. 3b) were highly consistent with the four maps described in previous studies <sup>25, 26</sup>. We then labeled and sorted the four sets of maps according to the appearing probability of brain states at resting state in the healthy controls. Specifically, map A showed a fronto-central maximum, map B showed a symmetric frontal to occipital orientation, map C showed a left occipital to right frontal, and map D showed a right occipital to left frontal orientation (Fig. 3b; the group-averaged brain states of each group in each condition are shown in Fig.9).

Previous studies using simultaneous EEG and fMRI recordings have suggested that brain states A and B are more closely related to the attention and saliency networks, as their corresponding blood-oxygen-level-dependent (BOLD) activations were located in the anterior cingulate cortex and parietal-frontal areas, and that states C and D are related to the auditory and visual sensory networks, as their corresponding BOLD signals were located in bilateral temporal and extrastriate visual areas <sup>11, 27</sup>. We thus predicted that higher levels of consciousness would be associated with a higher probability of the activation of high-level cognitive neural networks, that is, maps A and B (anterior-posterior maps, defined as the A-P map, Fig. 3b). In parallel, we predicted that reduced consciousness would involve a greater relative contribution of lower-level sensory areas, corresponding to maps C and D (left-right maps, defined as the L-R map, Fig. 3b). Multivariate analysis of variance (MANOVA) showed that, for the four conditions, the healthy controls demonstrated a pattern of a high probability of the A-P map and a low probability of the L-R map (Fig. 3c). The patient groups showed the opposite pattern, with a low probability of the A-P map and a high probability of the L-R map (Fig. 3c).

Next, we examined the difference of the dynamic of brain states between the MCS and UWS groups. For each individual patient, we quantified the probability-weighted spatial correlation difference between the A-P and L-R maps ( $\Delta C_p$ , see *Methods*). This difference reflected the spatial similarity between the maps of patient group and template maps derived from the healthy control group, and was used as an index of residual consciousness. We found a progressive increase in the difference of  $\Delta C_p$  between the MCS and UWS group as the level of the linguistic hierarchy increased from resting to word, phrase and sentence conditions (Figs. 3d, e). Specifically, at the phrase and sentence levels (but not at resting or

word levels), the MCS group showed a significantly higher  $\Delta C_p$  than the UWS group ( $P_{\text{Phrase}} = 0.037$ ;  $P_{\text{Sentence}} = 0.014$ ; one-way ANOVA, Bonferroni corrected; Fig. 3e). This was indicative of an increased probability of the A-P map (frontal-parietal networks) and a decreased probability of the L-R map (sensory networks). Furthermore, the difference in  $\Delta C_p$  between MCS and UWS was significantly larger in the phrase and sentence condition than that in the resting and word condition (phrase vs. word,  $t_{76} = 2.29$ ,  $P = 0.03$ ; sentence vs. word,  $t_{76} = 3.14$ ,  $P = 0.002$ ; two-sided two-sample  $t$ -test). Thus, positive and high  $\Delta C_p$  potentially may indicate residual consciousness.

We then investigated whether the probability difference of the maps was due to the duration (how long the map remained stable) or the frequency of occurrence (how many times the map occurred in one second) of each map. We found that the probability differences between the MCS and UWS groups in the phrase and sentence conditions could be attributed to a shorter duration of the L-R map, thought to reflect sensory networks ( $P_{\text{Phrase}} = 0.016$ ;  $P_{\text{Sentence}} = 0.017$ ; one-way ANOVA, Bonferroni corrected; Figs. 4a, b), and a higher occurrence of the A-P map, putatively associated with frontal-parietal networks ( $P_{\text{Phrase}} = 0.028$ ;  $P_{\text{Sentence}} = 0.063$ ; one-way ANOVA, Bonferroni corrected; Figs. 4d, e) in the MCS group relative to the UWS group. Importantly, the increase of the significant difference between the MCS and UWS groups matched the increasing linguistic level of conditions (Figs. 4b, e). No significant differences were found in the duration of the A-P map or in the occurrence of the L-R map (Fig.10).

If the duration and/or occurrence of maps indeed reflect the strength of residual consciousness in patients, we should observe corresponding changes before and after their recovery. In the sub-population of subjects who had multiple EEG recordings (12 out of 54 patients for resting-state recordings, 15 out of 60 patients for recordings during the linguistic tasks), the duration of the L-R map became shorter (Figs. 4a, c) and the occurrence of the A-P map increased along with recovery (Figs. 4d, f). Such changes were not observed in the non-recovery patients (Figs. 4a, c, d, f). The distribution of the A-P map duration and L-R map occurrence in healthy controls, recovery and non-recovery patients with MCS and UWS are shown in Fig.10.

To exclude the possibility that the different spatial maps in the two patient groups arose from a difference in brain injury, we then analyzed the volumes of brain damage in the 27 patients (17 MCS and 10 UWS) from the database who had received a structural magnetic resonance imaging scan on the same day as the EEG recordings. The results showed that there was no significant difference in the injured volume between the MCS and UWS patients (Fig.11a), and no significant correlation between the volume of brain injury and the

$\Delta C_p$  at the all three levels of task (Fig.11b). For example, the patient 07 with no brain damage has a low value of  $\Delta C_p$  (Fig.11c), by contrast, the patient 17 with a high volume of brain injury shows a relative higher  $\Delta C_p$  (Fig.11d). Furthermore, if we assume that the size of brain damage would not change dramatically during the recovery period, the observed changes in spatial maps before and after recovery (Fig. 4) also help rule out the possibility that the brain states measured with spatial maps purely reflect the underlying brain injury. The change of spatial maps of an example patient (ID: 2) is shown in Fig.11g-i.

### Example 3. Individual diagnosis and prediction

Our multiple measurements of brain activity allowed us to move from group-level analyses and attempt personalized diagnoses and predictions. We first trained a three-class LDA classifier with leave-one-subject-out cross-validation for the diagnosis of individual control, MCS and UWS subjects who had at least 3 months duration of DOC. The inputs to the classifier were the multiple EEG measurements, including the ITPC (3 features: ITPC values at 1, 2, and 4 Hz) and global dynamic patterns of brain activity (6 features:  $\Delta \text{Probability}$ ,  $\Delta C_p$ ,  $\text{Occurrence}_{A-P}$ ,  $\text{Duration}_{L-R}$ ,  $\text{Transition}_{A-P}$  and  $\text{Transition}_{L-R}$ ). A classifier with regularization first searched for the optimal feature combination within each task, and calculated the classification probability for each individual subject (Fig. 5a, detailed information of feature selection see *Methods*). Then, to avoid model overfitting, only those selected feature combinations were entered in the final LDA. All steps were cross-validated (leave-one-subject-out). Figure 5b plots the confusion matrix generated by the LDA. The best classification was found in the sentence condition with  $\Delta C_p$ ,  $\text{Transition}_{A-P}$ ,  $\text{ITPC}_{1\text{Hz}}$ , and  $\text{ITPC}_{2\text{Hz}}$  as input EEG features. A chi-squared test was used to estimate the classifier's performance, which was highly significant ( $\chi^2 = 95.84$ ,  $P = 7.6 \times 10^{-20}$ , accuracy = 75%). The decoder categorized healthy control, MCS, and UWS subjects with 89%, 58%, and 70% accuracy, respectively, all well above the chance level of 33%. The high decoding accuracy was confirmed by another discriminative classifier, support vector machine (SVM), with 96%, 65%, and 73% accuracy for healthy control, MCS and UWS subjects (Fig.12a).

Although a proportion of patients with UWS were classified as patients with MCS (30%, 9 of 30), it is possible that these patients had some degree of consciousness that was not detected by the CRS-R in the behavioral assessment. Interestingly, a greater proportion (33.3%, 3 of 9) of such potentially misdiagnosed patients with UWS had positive outcomes (fully awakened or exhibited improved behavior after the EEG recording), as compared with patients for which both the CRS-R and EEG classifier agreed on a diagnosis of UWS (9.5%, 2 of 21). Conversely, the classifier also diagnosed some patients with MCS as UWS (12 of

31). A lower proportion (25%, 3 of 12) of those potentially misdiagnosed MCS patients had positive outcomes as compared with patients diagnosed as MCS by both the CRS-R and EEG measurements (44.4%, 8 of 18). However, we should note that the number of patients in these groups might be too small to generate sufficient power for statistical analysis of between-group differences in these proportions.

We also examined whether EEG recordings could predict the subsequent recovery of consciousness in individual patients. Thirty-eight patients with multiple measurements were included in this analysis. Based on the CRS-R total-score  $\geq 6$  months after DOC onset<sup>28</sup>, 15 of them had a positive outcome, denoted as +ve (for patients selection see Fig.6). We used the same cross-validated method as above to construct an LDA classifier, this time aiming to separate the 15 outcome-positive and 23 outcome-negative patients. The results showed that while the CRS-R total-score could partially predict outcomes (AUC = 70%,  $\chi^2 = 7.2$ ,  $P = 0.016$ , chi-squared test; Fig. 5c), the prediction using the EEG measurements was better (AUC = 77%,  $\chi^2 = 11.5$ ,  $P = 9.2 \times 10^{-4}$ , chi-squared test; Fig. 5e *left*; Sensitivity: EEG vs. CRS-R,  $P = 0.07$ , McNemar's test). The best EEG predictive ability was achieved by the mean performance (see *Methods*) of the word, phrase, and sentence conditions with 87% sensitivity and 70% specificity, resulting in correct prediction for 13 of 15 outcome-positive and 16 of 23 outcome-negative patients. The best feature combinations for the prediction at each task level were  $\Delta\text{Probability} + \text{Occurrence}_{\text{A-P}}$  for the word condition,  $\text{Occurrence}_{\text{A-P}} + \text{ITPC}_{4\text{Hz}}$  for the phrase condition, and  $\Delta\text{Cp} + \text{Occurrence}_{\text{A-P}}$  for the sentence condition (Table 3). We then examined the prognostic ability within individual patients by calculating the predictive scores within UWS and MCS patients who had CRS-R based clinical outcomes. The prediction results demonstrated a high prediction accuracy in both groups (81% UWS and 71% MCS; individual prediction scores are shown in Fig. 5e *right*). Thus, prediction using EEG measures was better than using behavioral observations alone, either using a single CRS-R measure in our study, or using multiple CRS-R measurements in another study<sup>29</sup>.

It is worth noting that, although the number of subjects in the dataset for prediction model was relatively small (38 patients), the best feature combination for the classification only consisted of 2 or 3 features, suggesting that it was unlikely the model was overfitted. Nevertheless, to ensure the reliability of our results, we validated the same dataset using a different classifier, SVM, with cross-validation. This additional analysis confirmed our results showing a significant predictive accuracy from EEG (Fig.12b,  $\chi^2 = 15.4$ ,  $P = 1.6 \times 10^{-4}$ , accuracy = 82%, chi-squared test).

Most importantly, to further test the external validity and the generalization ability of

our models, we tested them (without retraining) on a new dataset consisting of 25 additional patients (12 MCS and 13 UWS, Table 2). The classifier (LDA) trained with the previous dataset using the same feature combinations, showed a high predictive accuracy in both outcome-positive and -negative groups across the two sample sets (Fig. 5f; *left*,  $\chi^2 = 8.8$ ,  $P = 0.005$ , accuracy = 80%, chi-squared test; individual prediction scores in Fig. 5f *right*). However, similar generalization using the classifier trained with the CRS-R total-score showed a much lower predictive accuracy (Fig. 5d;  $\chi^2 = 4.6$ ,  $P = 0.049$ , accuracy = 28%, chi-squared test). As a control, the classification accuracy was 50% when we shuffled the outcome label of the testing dataset; and the classifier for diagnosis (i.e. MCS versus UWS), once trained with CRS-R scores on the first 38 patients, successfully generalized to the new dataset (Fig.13a,  $\chi^2 = 21.3$ ,  $P = 2.7 \times 10^{-6}$ , accuracy = 96%, chi-squared test).

Could the higher predictive accuracy with EEG measurements be due to the larger number of EEG features used? To evaluate this, we expanded the behavioral measurements by including the total score and six sub-scores (e.g. auditory, visual, motor, oromotor, communication and arousal) and using the same model to search for the optimal CRS-R feature combination for prognosis. With cross-validation on the first 38 patients, the best prediction was obtained using the visual subscale alone (Fig.13b *left*), but again that classifier failed to generalize to the new dataset of 25 patients (Fig.13b *right*). Direct comparison of prediction performance between CRS-R or the EEG recording under the word condition (both have seven features) yielded a similar result (Fig. 13c,d). Furthermore, we examined the outcome prediction using a standard LDA model without feature selection before classification. The results confirmed the superior generalization ability using EEG metrics than using CRS-R features (Fig.13e).

Finally, the confusion matrix generated by the CRS-R scores showed that an important proportion of outcome-positive patients (53%, 8 of 15; 3 MCS and 5 UWS) was mis-predicted as outcome-negative (Fig. 5c). Compared with the CRS-R, our EEG task-based assessments could potentially contribute to more accurate diagnoses, as the classifier that we constructed using EEG measurements was able to significantly predict future CRS-R dichotomized outcomes (Fig. 5e): six of those eight patients were classified as outcome-positive by the EEG-based classifier. Furthermore, we combined the features from both CRS-R and EEG and submitted them to the model. The results showed that, at the word and phrase levels, the classification accuracies by using both EEG and CRS-R were slightly higher than those using only EEG signals (classification accuracies at all levels see Fig.14 and Table 3). The best feature combination in the classification indeed contained the CRS-R, which might indicate that the prediction model could benefit from the use of both EEG and

CRS-R. Taken together, these results confirm the prognostic ability of EEG metrics recorded during language listening and suggest that these EEG measures may complement demographic and behavioral diagnoses in the clinic.

The outcome of the patients in the prediction was based on a single CRS-R, which could lead to misdiagnosis due to the variability of the level of consciousness over time (e.g. morning vs. afternoon within a day and across different days, see reference<sup>30</sup>). Although we examined a subset of patients ( $n = 15$ ) who had multiple diagnoses before and after EEG recordings and found relatively stable CRS-R scores across days at both group and individual levels (Fig.15), repetitive behavioral measurements (CRS-R) is likely to be important for accurate diagnosis and prediction, and a systematic comparison of the predictive value with multiple EEG measurements and multiple CRS-R scores remains to be explored.

**Table 3.** The comparisons between outcome prediction by EEG and by both EEG and CRS-R rating in different task conditions.  $n+ve = 15$ ,  $n-ve = 23$ ; chi-squared tests.

|           |             | Optimal feature combination                                                        | AUC | Accuracy | $\chi^2$ | $P$                  |
|-----------|-------------|------------------------------------------------------------------------------------|-----|----------|----------|----------------------|
| Word      | EEG         | $\Delta$ Probability + Occurrence <sub>A-P</sub>                                   | 74% | 74%      | 8.2      | 0.007                |
|           | EEG + CRS-R | Occurrence <sub>A-P</sub> + CRS-R                                                  | 82% | 82%      | 14.6     | $2 \times 10^{-4}$   |
| Phrase    | EEG         | Occurrence <sub>A-P</sub> + ITPC <sub>4Hz</sub>                                    | 75% | 76%      | 9.5      | 0.005                |
|           | EEG + CRS-R | $\Delta$ Probability + Occurrence <sub>A-P</sub> + Duration <sub>L-R</sub> + CRS-R | 78% | 79%      | 11.6     | 0.001                |
| Sentence  | EEG         | $\Delta$ Cp + Occurrence <sub>A-P</sub>                                            | 74% | 74%      | 8.2      | 0.007                |
|           | EEG + CRS-R | $\Delta$ Cp + Occurrence <sub>A-P</sub>                                            | 74% | 74%      | 8.2      | 0.007                |
| Task Mean | EEG         | -                                                                                  | 79% | 76%      | 11.5     | $9.4 \times 10^{-4}$ |
|           | EEG + CRS-R | -                                                                                  | 79% | 76%      | 11.5     | $9.4 \times 10^{-4}$ |

## REFERENCES

1. Giacino, J.T., Fins, J.J., Laureys, S. & Schiff, N.D. Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol* **10**, 99-114 (2014).
2. Bruno, M.A., Gosseries, O., Ledoux, D., Hustinx, R. & Laureys, S. Assessment of  
5 consciousness with electrophysiological and neurological imaging techniques. *Curr Opin Crit Care* **17**, 146-151 (2011).
3. Giacino, J.T., *et al.* The minimally conscious state: definition and diagnostic criteria. *Neurology* **58**, 349-353 (2002).
4. Owen, A.M. The Search for Consciousness. *Neuron* **102**, 526-528 (2019).
- 10 5. Strauss, M., *et al.* Disruption of hierarchical predictive coding during sleep. *Proc Natl Acad Sci U S A* **112**, E1353-1362 (2015).
6. Goldfine, A.M., Victor, J.D., Conte, M.M., Bardin, J.C. & Schiff, N.D. Determination of awareness in patients with severe brain injury using EEG power spectral analysis. *Clin Neurophysiol* **122**, 2157-2168 (2011).
- 15 7. Sitt, J.D., *et al.* Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* **137**, 2258-2270 (2014).
8. Dehaene, S. & Changeux, J.P. Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200-227 (2011).
9. Barttfeld, P., *et al.* Signature of consciousness in the dynamics of resting-state brain activity. *Proc Natl Acad Sci U S A* **112**, 887-892 (2015).
- 20 10. Lehmann, D., Ozaki, H. & Pal, I. EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalogr Clin Neurophysiol* **67**, 271-288 (1987).
11. Michel, C.M. & Koenig, T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *Neuroimage* **180**, 577-593 (2018).
- 25 12. Laureys, S. & Schiff, N.D. Coma and consciousness: paradigms (re)framed by neuroimaging. *Neuroimage* **61**, 478-491 (2012).
13. Schnakers, C., *et al.* Voluntary brain processing in disorders of consciousness. *Neurology* **71**, 1614-1620 (2008).
14. Owen, A.M., *et al.* Detecting awareness in the vegetative state. *Science* **313**, 1402 (2006).
- 30 15. Braiman, C., *et al.* Cortical Response to the Natural Speech Envelope Correlates with Neuroimaging Evidence of Cognition in Severe Brain Injury. *Curr Biol* **28**, 3833-3839 e3833 (2018).
16. Rohaut, B., *et al.* Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia* **66**, 279-292 (2015).
- 35 17. Coleman, M.R., *et al.* Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain* **130**, 2494-2507 (2007).
18. Faugeras, F., *et al.* Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia* **50**, 403-418 (2012).
19. Engemann, D.A., *et al.* Robust EEG-based cross-site and cross-protocol classification of states  
40 of consciousness. *Brain* **141**, 3179-3192 (2018).
20. Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* **19**, 158-164 (2016).
21. Kiefer, M. The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Brain Res Cogn Brain Res* **13**, 27-39 (2002).
- 45 22. Batterink, L. & Neville, H.J. The human brain processes syntax in the absence of conscious awareness. *J Neurosci* **33**, 8528-8533 (2013).
23. Demertzi, A., *et al.* Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci Adv* **5**, eaat7603 (2019).
- 50 24. Khanna, A., Pascual-Leone, A., Michel, C.M. & Farzan, F. Microstates in resting-state EEG: current status and future directions. *Neurosci Biobehav Rev* **49**, 105-113 (2015).

25. Koenig, T., *et al.* Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. *Neuroimage* **16**, 41-48 (2002).
26. Brodbeck, V., *et al.* EEG microstates of wakefulness and NREM sleep. *Neuroimage* **62**, 2129-2139 (2012).
- 5 27. Britz, J., Van De Ville, D. & Michel, C.M. BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *Neuroimage* **52**, 1162-1170 (2010).
28. Stender, J., *et al.* Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet* **384**, 514-522 (2014).
29. Scarpino, M., *et al.* Prognostic value of post-acute EEG in severe disorders of consciousness, using American Clinical Neurophysiology Society terminology. *Neurophysiol Clin* **49**, 317-327  
10 (2019).
30. Wannez, S., *et al.* The repetition of behavioral assessments in diagnosis of disorders of consciousness. *Ann Neurol* **81**, 883-889 (2017).
31. Wu, X., *et al.* Intrinsic Functional Connectivity Patterns Predict Consciousness Level and  
15 Recovery Outcome in Acquired Brain Injury. *J Neurosci* **35**, 12932-12946 (2015).
32. Monti, M.M., *et al.* Willful modulation of brain activity in disorders of consciousness. *N Engl J Med* **362**, 579-589 (2010).
33. Chennu, S., *et al.* Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain* **140**, 2120-2132 (2017).
- 20 34. Di Perri, C., *et al.* Neural correlates of consciousness in patients who have emerged from a minimally conscious state: a cross-sectional multimodal imaging study. *Lancet Neurol* **15**, 830-842 (2016).
35. Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L. & Breakspear, M. Time-resolved resting-state brain networks. *Proc Natl Acad Sci U S A* **111**, 10341-10346 (2014).
- 25 36. Hutchison, R.M., *et al.* Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* **80**, 360-378 (2013).
37. Tagliazucchi, E., *et al.* Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *J R Soc Interface* **13**, 20151027 (2016).
38. Cojan, Y., *et al.* The brain under self-control: modulation of inhibitory and monitoring cortical  
30 networks during hypnotic paralysis. *Neuron* **62**, 862-875 (2009).
39. Barttfeld, P., *et al.* Distinct patterns of functional brain connectivity correlate with objective performance and subjective beliefs. *Proc Natl Acad Sci U S A* **110**, 11577-11582 (2013).
40. Claassen, J., *et al.* Detection of Brain Activation in Unresponsive Patients with Acute Brain Injury. *N Engl J Med* **380**, 2497-2505 (2019).
- 35 41. Davis, M.H., *et al.* Dissociating speech perception and comprehension at reduced levels of awareness. *Proc Natl Acad Sci U S A* **104**, 16032-16037 (2007).
42. Makov, S., *et al.* Sleep Disrupts High-Level Speech Parsing Despite Significant Basic Auditory Processing. *J Neurosci* **37**, 7772-7781 (2017).
43. Ding, N. & Jin, P. Low-frequency Neural Activity Reflects Rule-based Chunking during  
40 Speech Listening. *bioRxiv*, 742585 (2019).
44. Frank, S.L. & Yang, J. Lexical representation explains cortical entrainment during speech comprehension. *PLoS One* **13**, e0197304 (2018).
45. Steppacher, I., *et al.* N400 predicts recovery from disorders of consciousness. *Ann Neurol* **73**, 594-602 (2013).
46. Giacino, J.T., *et al.* Practice guideline update recommendations summary: Disorders of consciousness: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology* **91**, 450-460 (2018).
- 50 47. Giacino, J.T., *et al.* Comprehensive systematic review update summary: Disorders of consciousness: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation

- Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology* **91**, 461-470 (2018).
48. Lehmann, D., *et al.* EEG microstate duration and syntax in acute, medication-naïve, first-episode schizophrenia: a multi-center study. *Psychiatry Res* **138**, 141-156 (2005).
- 5 49. Perrin, F., Castro, M., Tillmann, B. & Luaute, J. Promoting the use of personally relevant stimuli for investigating patients with disorders of consciousness. *Front Psychol* **6**, 1102 (2015).
50. Cortese, M.D., *et al.* Coma recovery scale-r: variability in the disorder of consciousness. *BMC Neurol* **15**, 186 (2015).
- 10 51. Edlow, B.L., *et al.* Early detection of consciousness in patients with acute severe traumatic brain injury. *Brain* **140**, 2399-2414 (2017).
52. Giacino, J.T., Kalmar, K. & Whyte, J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* **85**, 2020-2029 (2004).
53. Ding, N., *et al.* Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Front Hum Neurosci* **11**, 481 (2017).
- 15 54. Teasdale, G. & Jennett, B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* **2**, 81-84 (1974).
55. Bruno, M.A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G. & Laureys, S. From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *J Neurol* **258**, 1373-1384 (2011).
- 20 56. Wannez, S., *et al.* Prevalence of coma-recovery scale-revised signs of consciousness in patients in minimally conscious state. *Neuropsychol Rehabil* **28**, 1350-1359 (2018).
57. Büyüköztürk, Ş. & Çokluk Bökeoğlu, Ö. Discriminant Function Analysis: Concept and Application. *Eğitim Araştırmaları - Eurasian Journal of Educational Research* **8**, 73-92 (2008).
- 25 58. Dixon, S.J., *et al.* Application of classification methods when group sizes are unequal by incorporation of prior probabilities to three common approaches: Application to simulations and mouse urinary chemosignals. *Chemometrics and Intelligent Laboratory Systems* **99**, 111-120 (2009).
59. Geurts, B.H., Andriessen, T.M., Goraj, B.M. & Vos, P.E. The reliability of magnetic resonance imaging in traumatic brain injury lesion detection. *Brain Inj* **26**, 1439-1450 (2012).
- 30 60. Pialat, J.B., *et al.* Evolution of lesion volume in acute stroke treated by intravenous t-PA. *J Magn Reson Imaging* **22**, 23-28 (2005).
61. Rosso, I.M., *et al.* Reduced amygdala volumes in first-episode bipolar disorder and correlation with cerebral white matter. *Biol Psychiatry* **61**, 743-749 (2007).

## CLAIMS

1. A method of assessing the level of consciousness disorder, the method comprising:  
using a hierarchical linguistic processing paradigm to test consciousness (such as residual  
5 consciousness) in subjects with disorders of consciousness.

2. The method of claim 1, wherein the method comprising:

(1) Stimulating the subjects;

(2) Collecting the electroencephalogram (EEG) data of the patients;

10 (3) Analysing the phase coherence, multivariate pattern and/or brain microstate;

(4) Diagnosis and prediction.

3. The method of claim 2, wherein the stimuli is Word stimuli, Phrase stimuli and/or  
Sentence stimuli; or

15 wherein the stimuli is performed by different language hierarchy (such as, increasing  
language hierarchy) or language paradigms;

preferably, in the sentence condition, 2~5 levels of semantic hierarchies are used, for  
example, including single-word frequency at 4 Hz, phrasal frequency at 2 Hz, and sentential  
frequency at 1 Hz;

20 preferably, the phrase condition only included word and phrasal levels;

preferably, the word condition only included the 4 Hz word frequency.

4. The method of claim 2, wherein the single-trial electroencephalogram data is  
transformed into the frequency domain, preferably the Discrete Fourier Transform is used for  
25 the transformation; and/or

binary classifiers are used to discriminate different subject groups; and/or

the LDA is trained for pairwise classifications at each target frequency under each task;  
preferably, the decoding is implemented as follows: 1) the input features are the ITPC values  
at all EEG channels; 2) for each comparison, a majority of subjects is randomly chosen as  
30 training set, while the other subjects as testing set; 3) a 3 to 10-fold cross-validation is  
applied on the training set, that is, for each fold, the classifier is fit on the majority of  
subjects and validated on the other training set; 4) finally, the classification performance is  
computed as the sum of the Area Under the Received Operative Curve, based on the  
probabilistic classification of the independent testing set; 5) the steps 2 to 4 are repeated  
35 40~400 times to produce the mean classification AUC for these two groups at the each

frequency for each condition.

5. The method of claim 2, wherein the brain state analysis is performed using MicrostateAnalysis.

5

6. The method of claim 2, wherein classification analysis is used to identify the consciousness states of individuals; or the patients for whom we had behavioral measurements more than 100 days after EEG assessments.

10 7. The method of claim 2, wherein the consciousness in patients with disorders of consciousness can be reflected by the strength of speech-tracking responses, especially neural tracking of higher-level linguistic structures, i.e., phrases and sentences.

15 8. The method of claim 2, wherein expanded the behavioral measurements by including the total score and six sub-scores and using the same model to search for the optimal CRS-R feature combination for prognosis; preferably, the six sub-scores are auditory, visual, motor, oromotor, communication and arousal; more preferably, using the visual subscale alone.

20 9. The method of claim 2, wherein the features from both CRS-R and EEG are combined and submitted to construct the model.

25 10. The method of claim 2, wherein the electroencephalogram data is pre-processed using BrainVision Analyzer; or the electroencephalogram data is pre-processed in the EEGLAB toolbox for the brain state analysis; preferably, the electroencephalogram is resting electroencephalogram.

30 11. A systematic device for assessing the level of consciousness disorder, comprising the measuring instruments, softwares or programs for the method of any one of claims 1 to 10.

30

12. The systematic device of claim 11, wherein it comprising:

Language generator;

Electroencephalogram measuring instrument;

measuring instruments, softwares or programs for analysing phase coherence;

35 measuring instruments, softwares or programs for analysing multivariate pattern; and/or

measuring instruments, softwares or programs for analysing brain microstate.

13. A computer system for assessing the level of consciousness disorder, comprising:

a device (or measuring instruments, softwares or programs) for stimulating the  
5 subjects;

a device (or measuring instruments, softwares or programs)for collecting the  
electroencephalogram (EEG) data of the patients;

a device (or measuring instruments, softwares or programs) for analysing the phase  
coherence, multivariate pattern and/or brain microstate;

10 a device (or measuring instruments, softwares or programs) for diagnosis and  
prediction.

14. The computer system of claim 13, wherein the stimuli is Word stimuli, Phrase  
stimuli and/or Sentence stimuli; or

15 wherein the stimuli is performed by different language hierarchy (such as, increasing  
language hierarchy) or language paradigms;

preferably, in the sentence condition, 2~5 levels of semantic hierarchies are used, for  
example, including single-word frequency at 4 Hz, phrasal frequency at 2 Hz, and sentential  
frequency at 1 Hz;

20 preferably, the phrase condition only included word and phrasal levels;

preferably, the word condition only included the 4 Hz word frequency.

15. The computer system of claim 13, wherein the single-trial electroencephalogram  
data is transformed into the frequency domain, preferably the Discrete Fourier Transform is  
25 used for the transformation; and/or

binary classifiers are used to discriminate different subject groups; and/or

the LDA is trained for pairwise classifications at each target frequency under each task;  
preferably, the decoding is implemented as follows: 1) the input features are the ITPC values  
at all EEG channels; 2) for each comparison, a majority of subjects is randomly chosen as  
30 training set, while the other subjects as testing set; 3) a 3 to 10-fold cross-validation is  
applied on the training set, that is, for each fold, the classifier is fit on the majority subjects  
and validated on the other training set; 4) finally, the classification performance is computed  
as the sum of the Area Under the Received Operative Curve, based on the probabilistic  
classification of the independent testing set; 5) the steps 2 to 4 are repeated 40~400 times to  
35 produce the mean classification AUC for these two groups at the each frequency for each

condition.

16. The computer system of claim 13, wherein the brain state analysis is performed using MicrostateAnalysis.

5

17. The computer system of claim 13, wherein classification analysis is used to identify the consciousness states of individuals; or the patients for whom we had behavioral measurements more than 100 days after EEG assessments.

10 18. The computer system of claim 13, wherein the residual consciousness in patients with disorders of consciousness can be reflected by the strength of speech-tracking responses, especially neural tracking of higher-level linguistic structures, i.e., phrases and sentences.

15 19. The computer system of claim 13, wherein expanded the behavioral measurements by including the total score and six sub-scores and using the same model to search for the optimal CRS-R feature combination for prognosis; preferably, the six sub-scores are auditory, visual, motor, oromotor, communication and arousal; more preferably, using the visual subscale alone.

20 20. The computer system of claim 13, wherein the features from both CRS-R and EEG are combined and submitted to construct the model.

25 21. The computer system of claim 13, wherein the electroencephalogram data is pre-processed using BrainVision Analyzer; or the electroencephalogram data is pre-processed in the EEGLAB toolbox for the brain state analysis; preferably, the electroencephalogram is resting electroencephalogram.

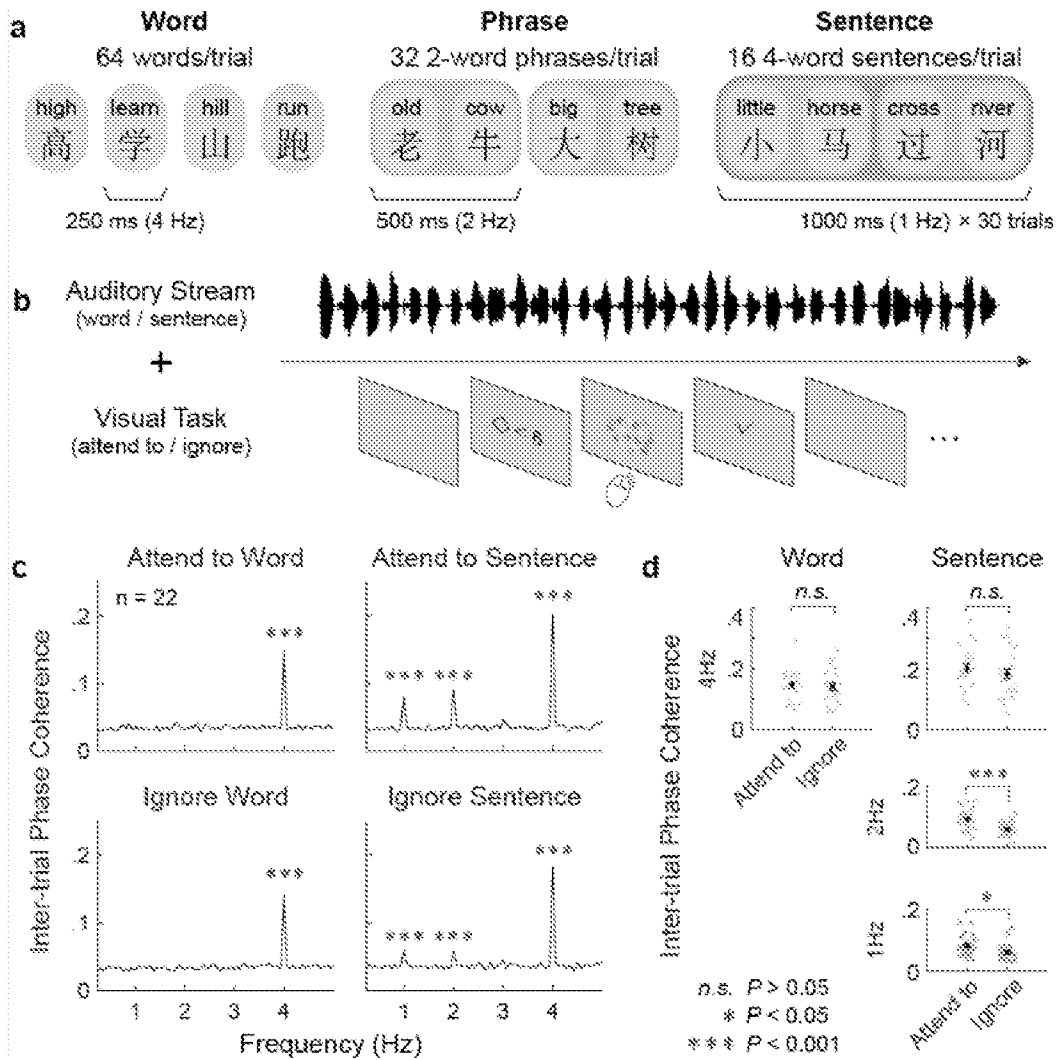


FIG. 1

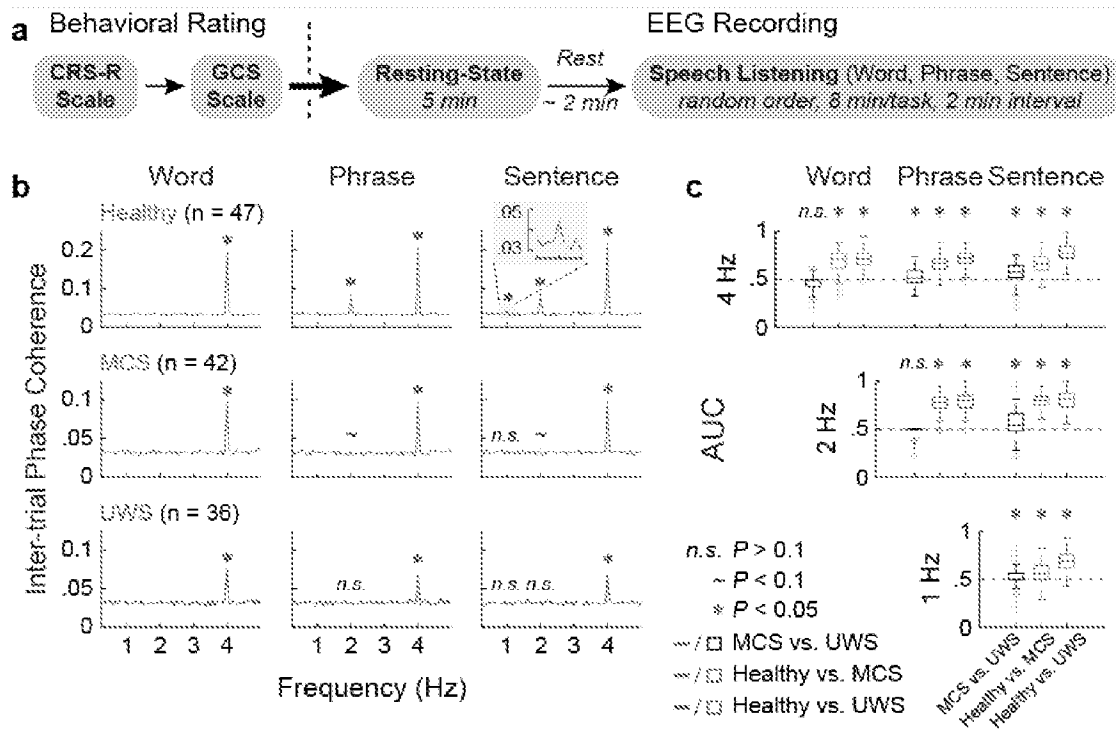


FIG. 2

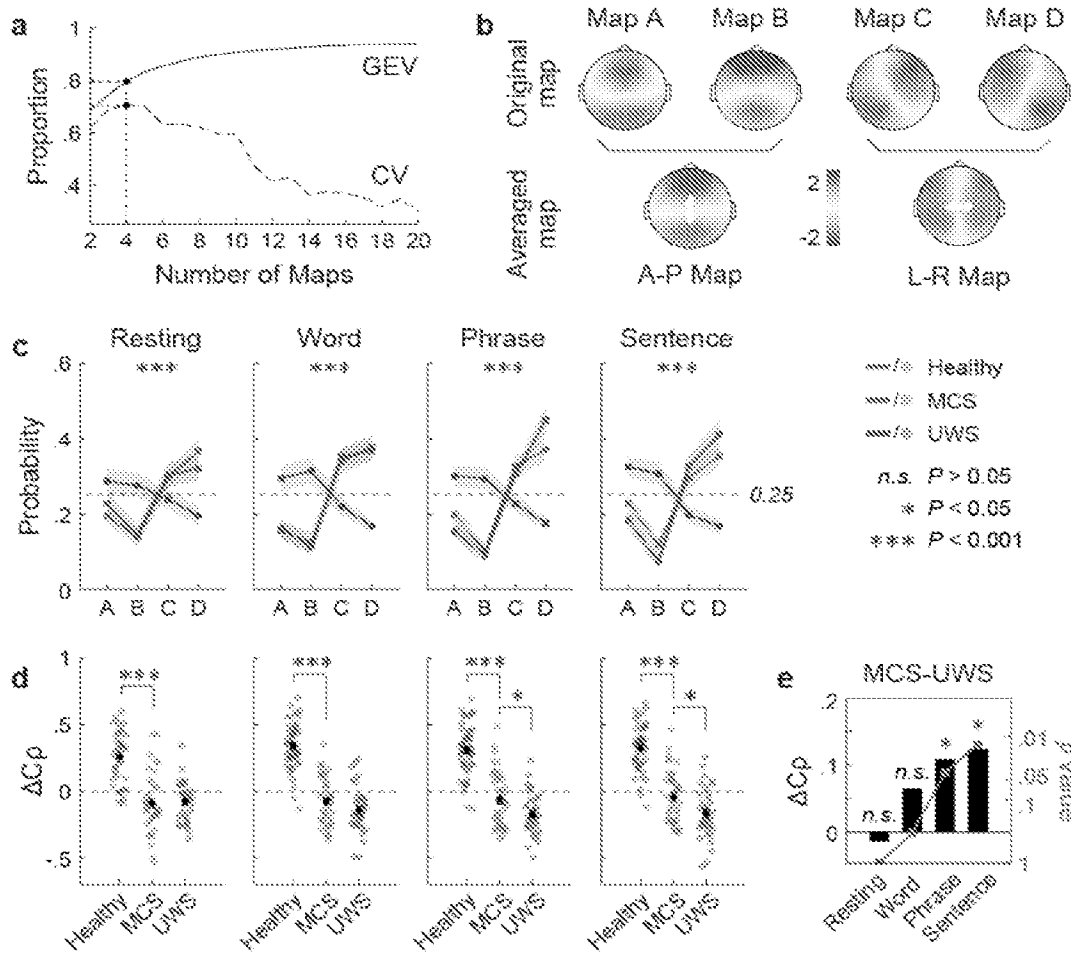


FIG. 3

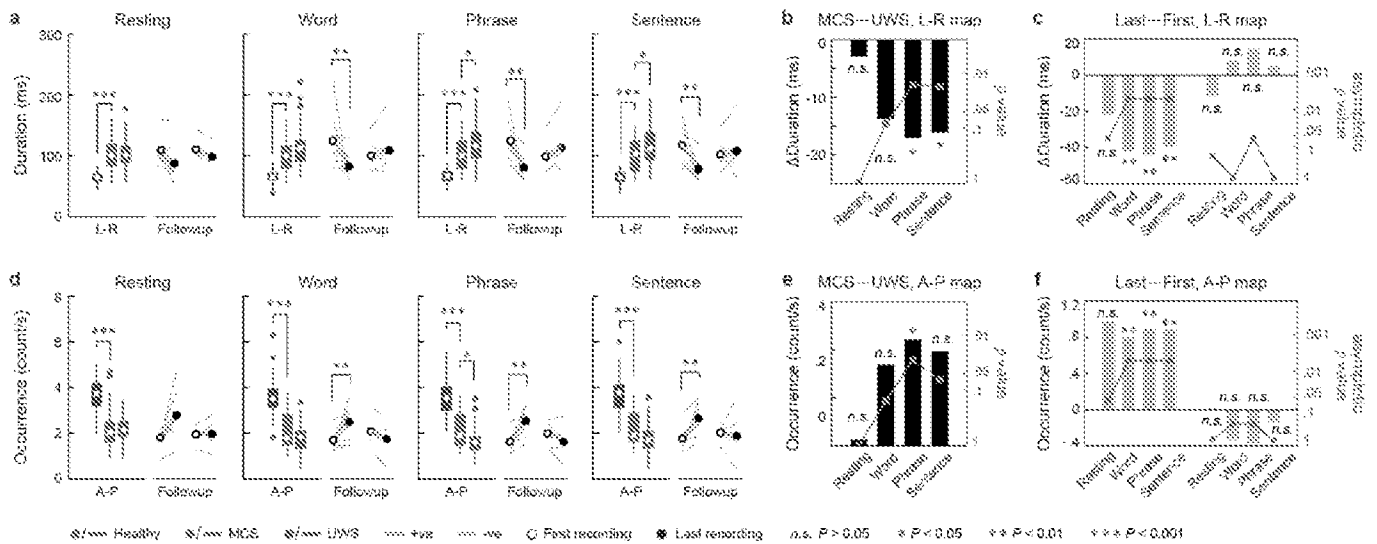


FIG. 4

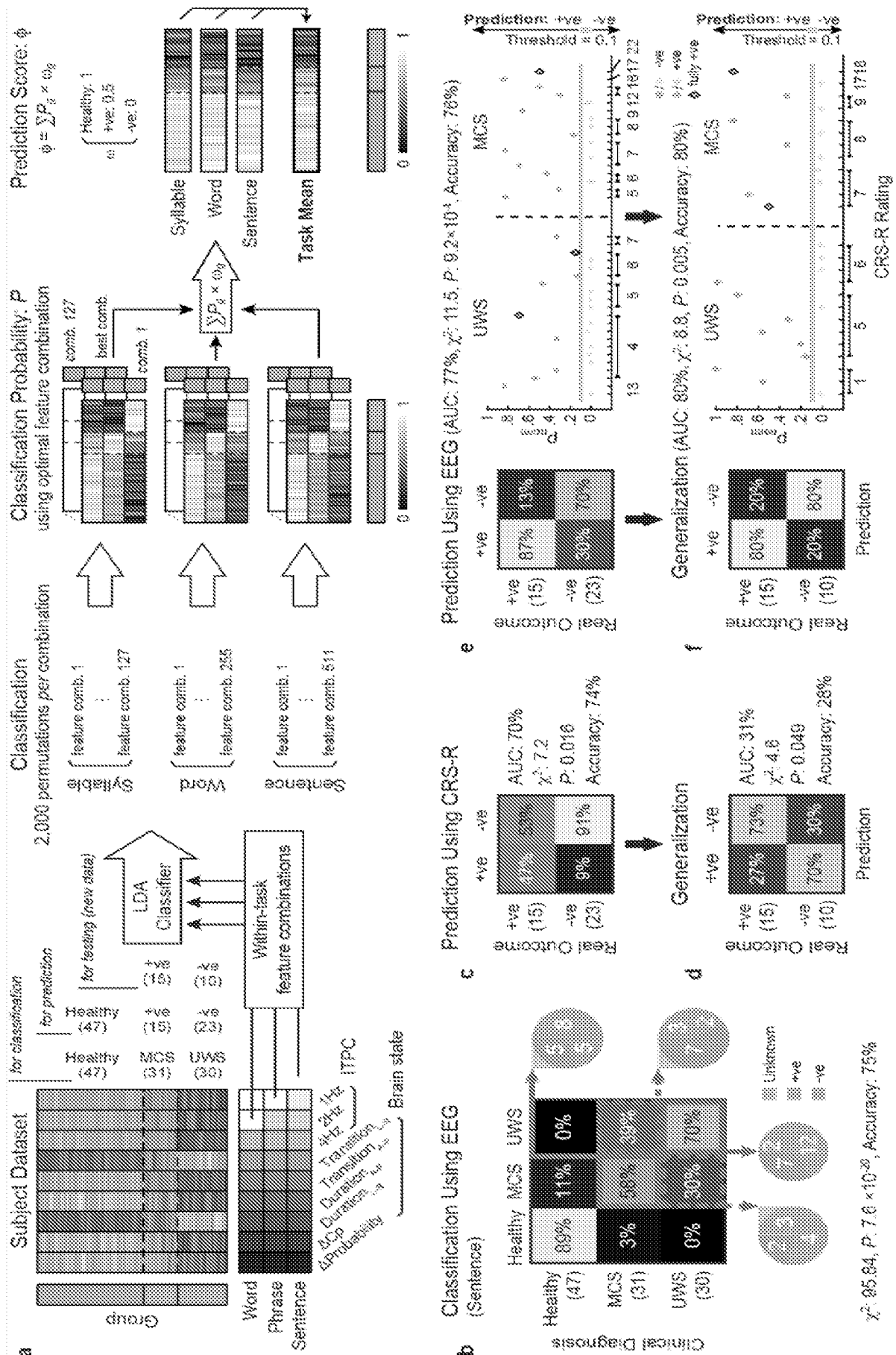


FIG. 5

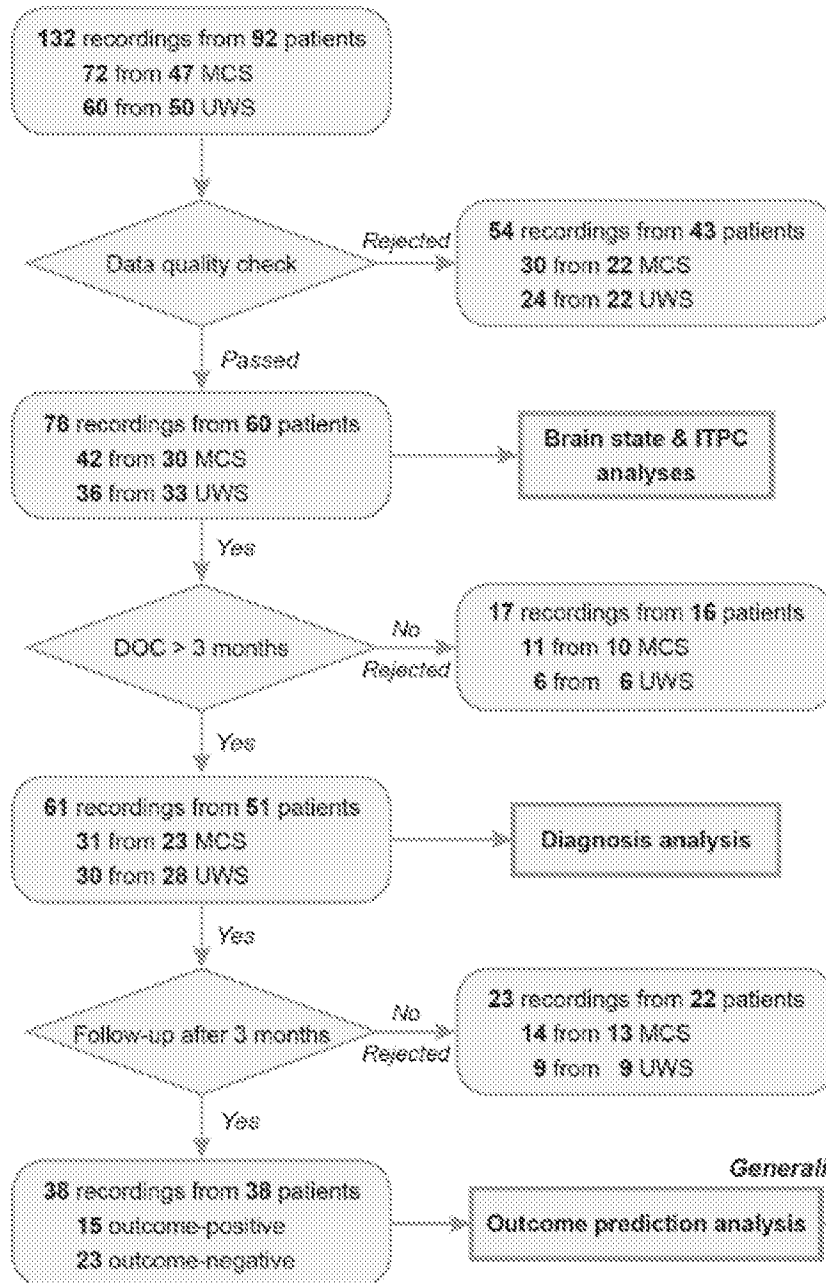
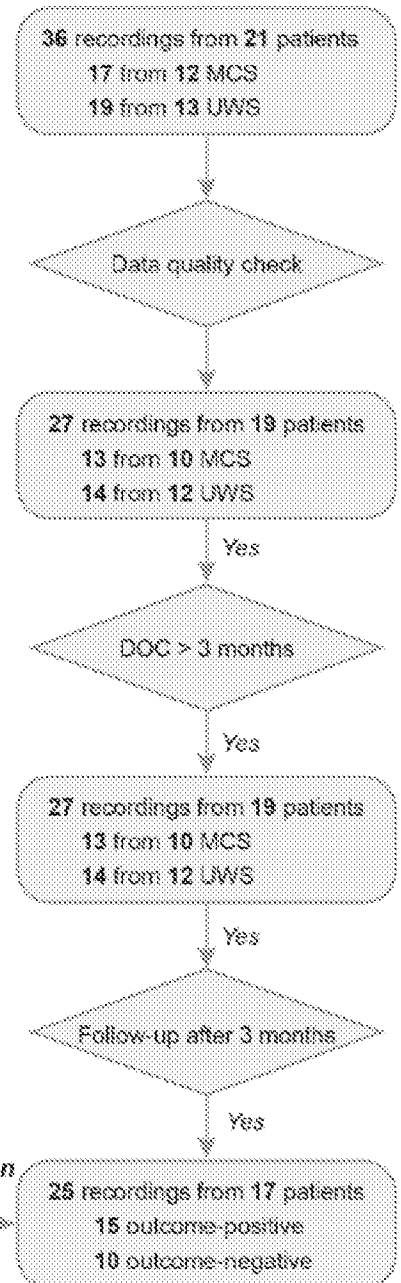
**Dataset 2016.7 - 2018.10****Dataset 2018.10 - 2019.6**

FIG. 6

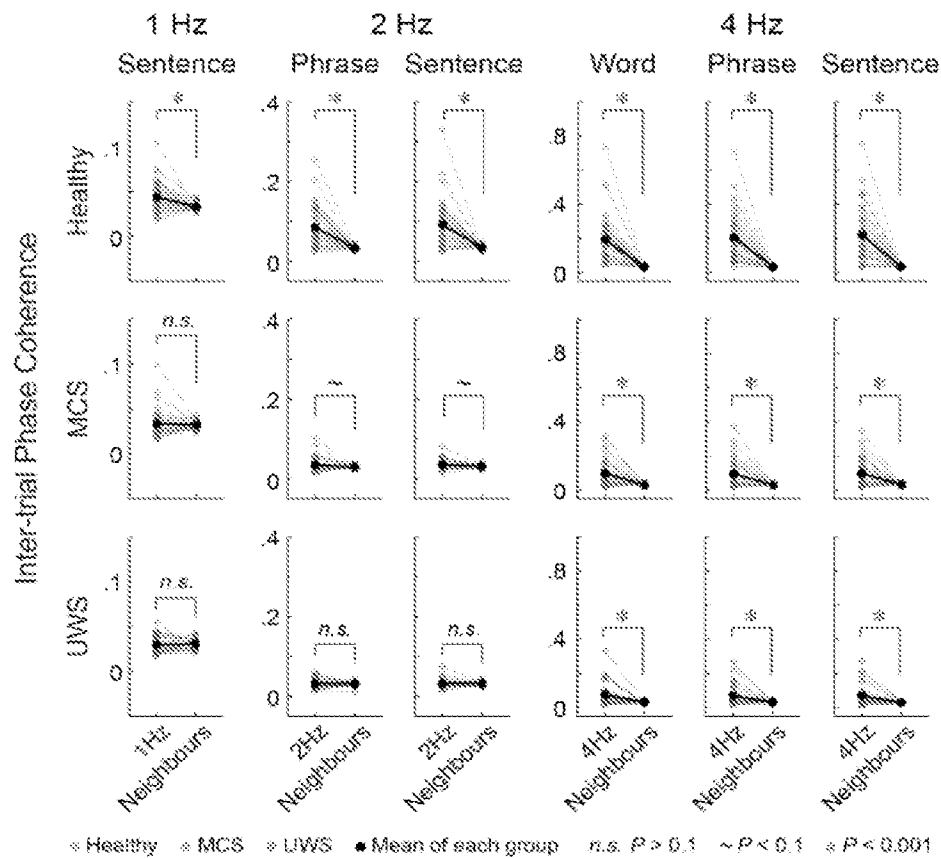


FIG. 7

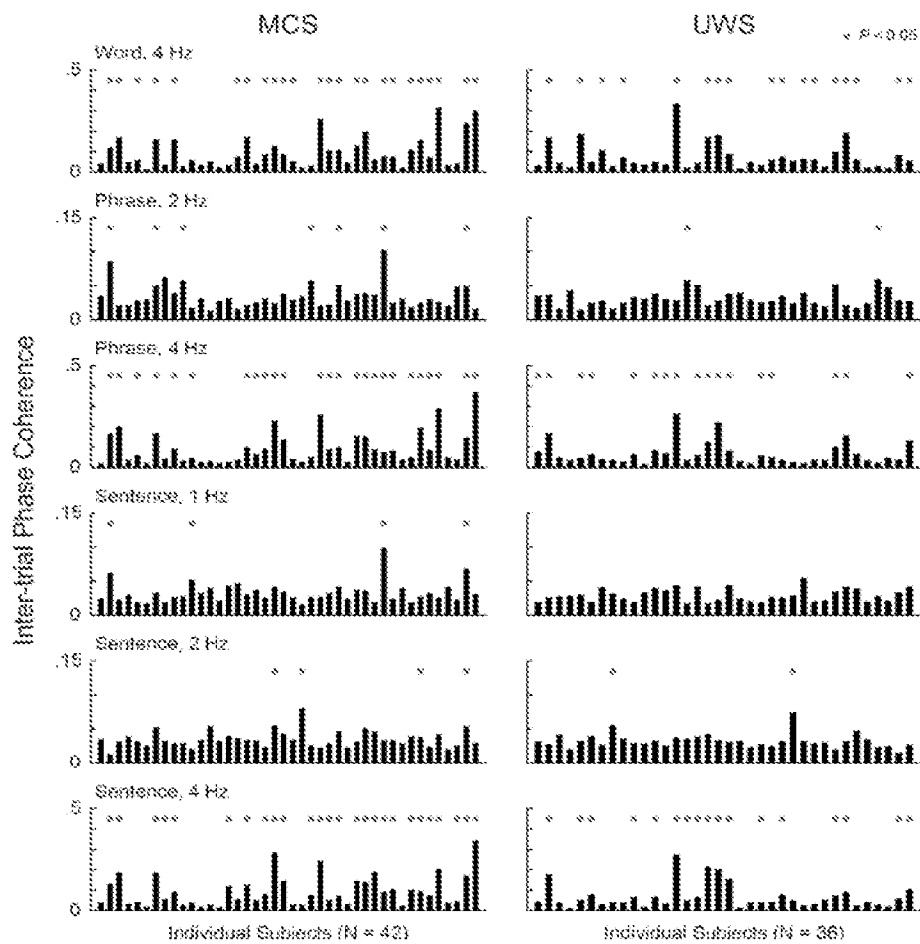


FIG. 8

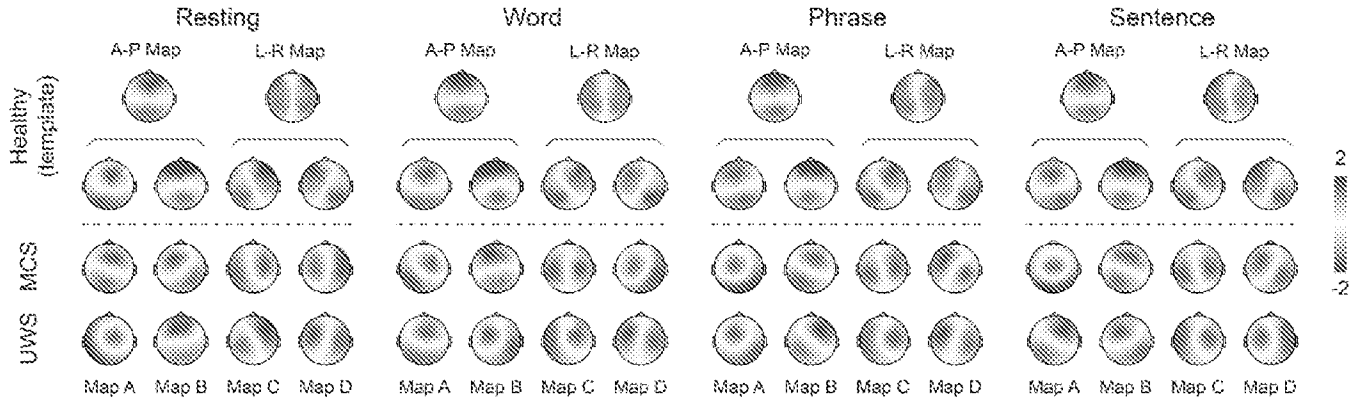


FIG. 9

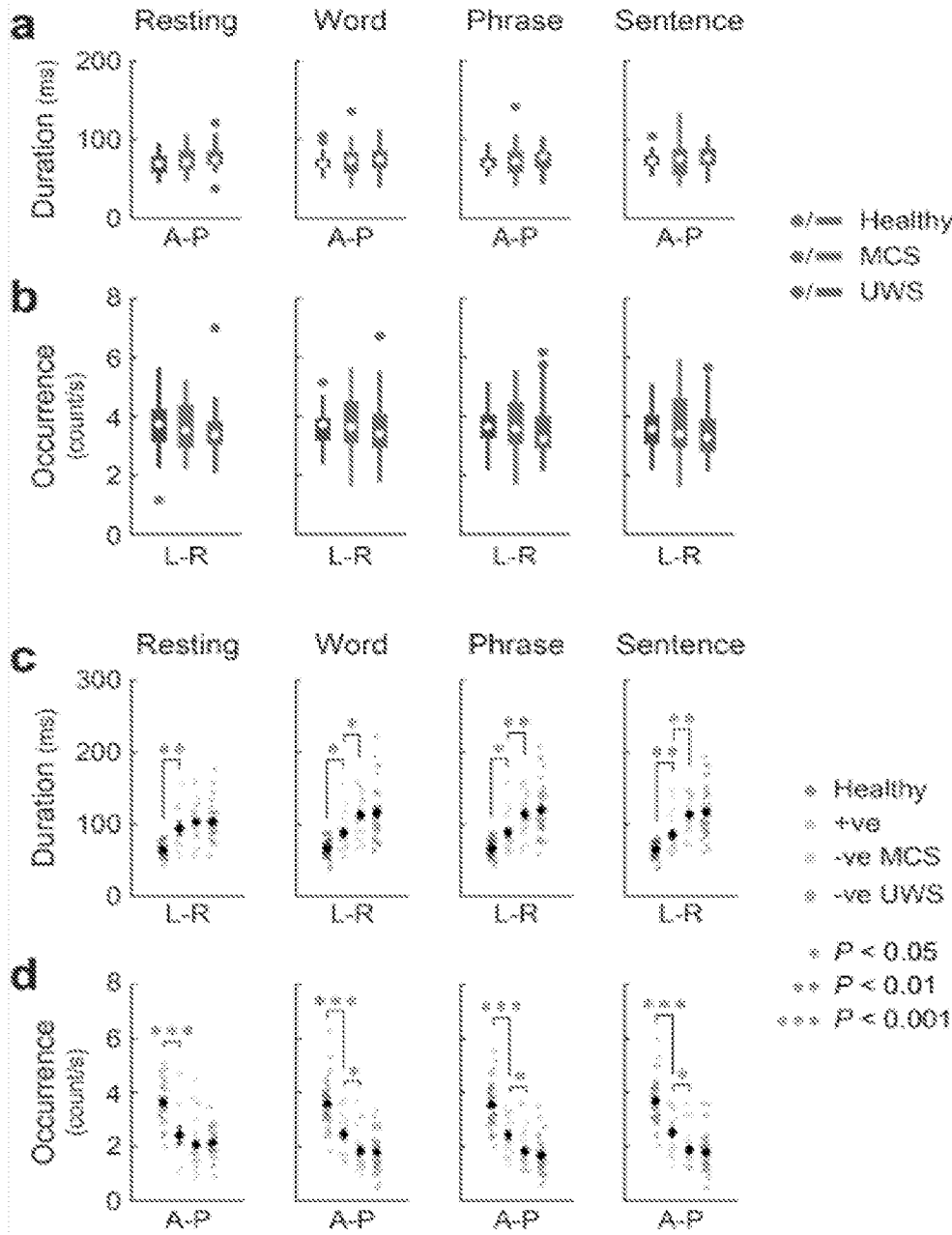


FIG. 10

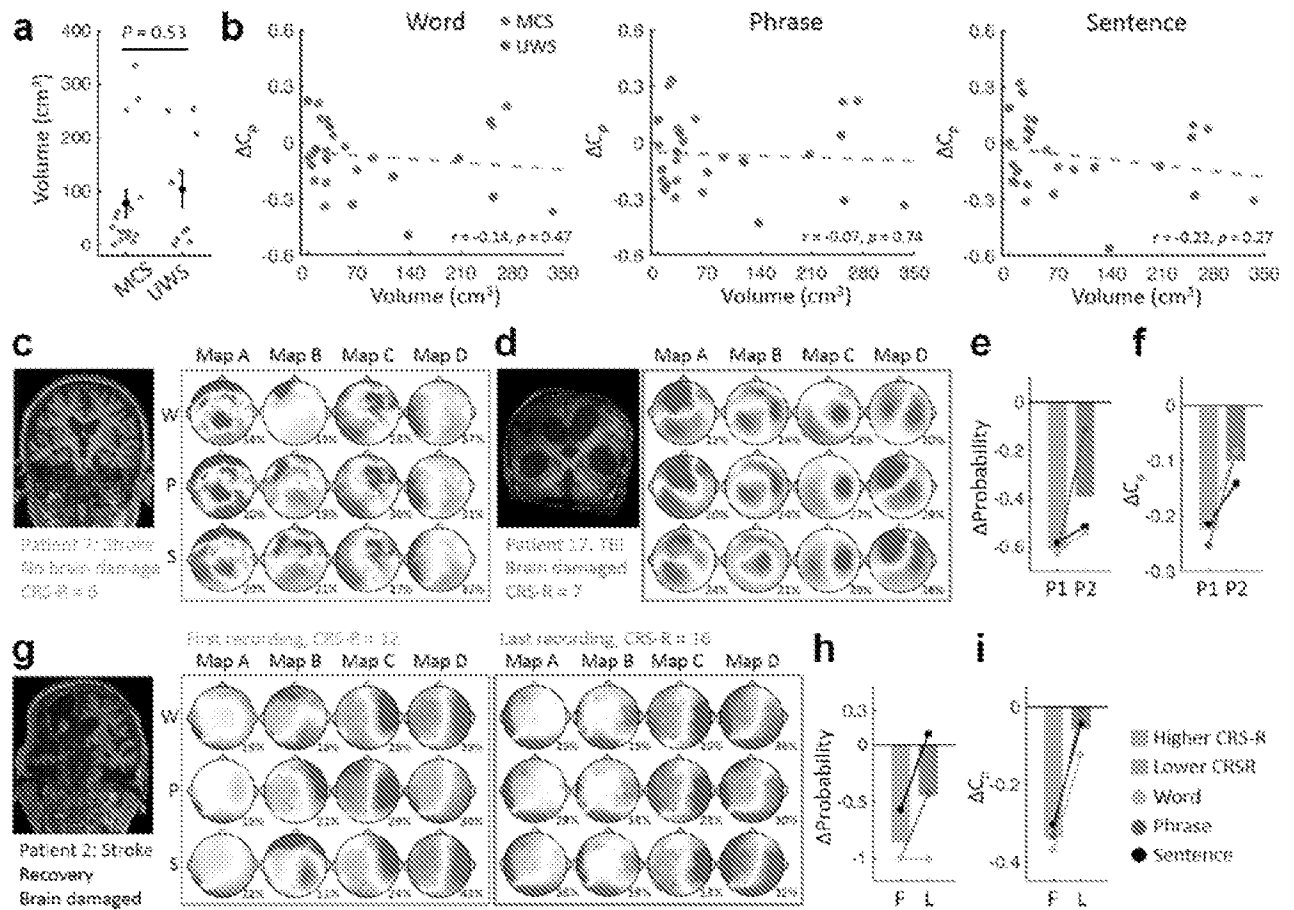


FIG. 11

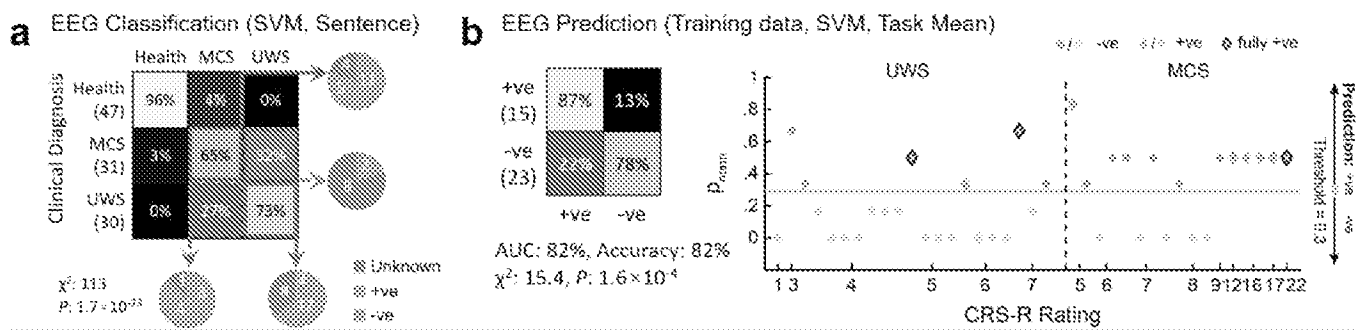
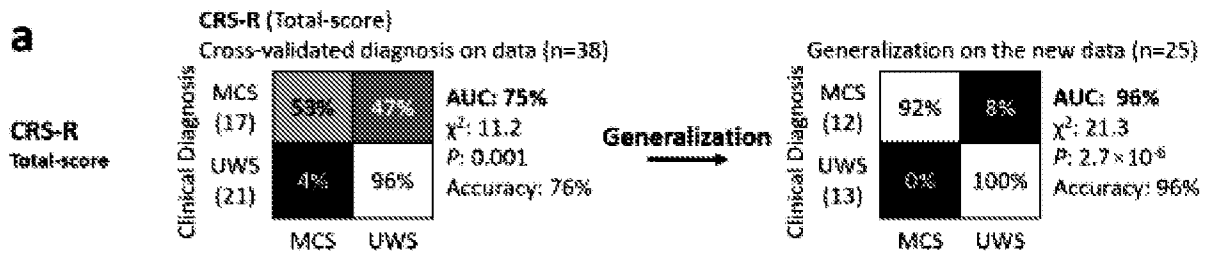
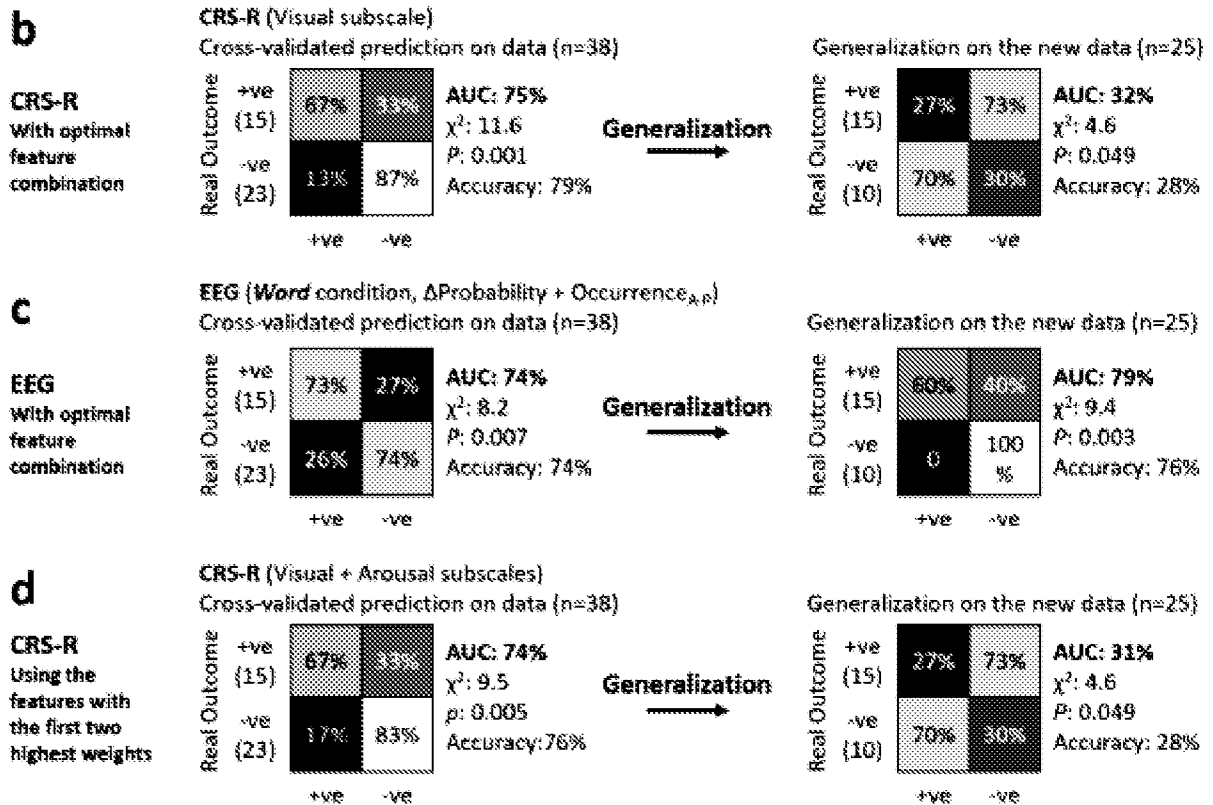


FIG. 12

## Diagnosis



## Outcome prediction with feature selection



## Outcome prediction without feature selection

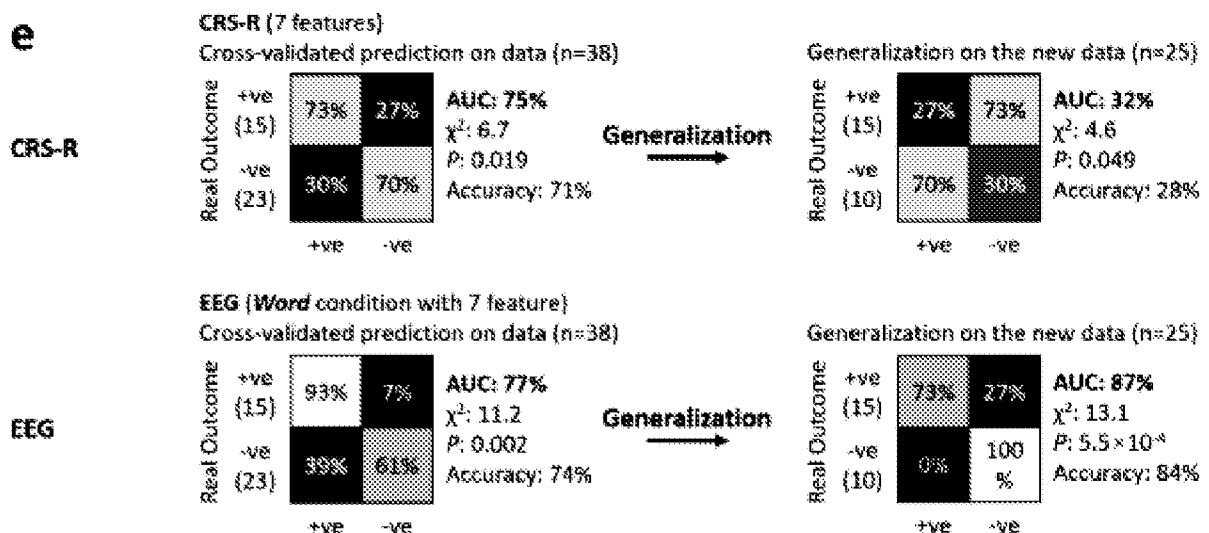


FIG. 13

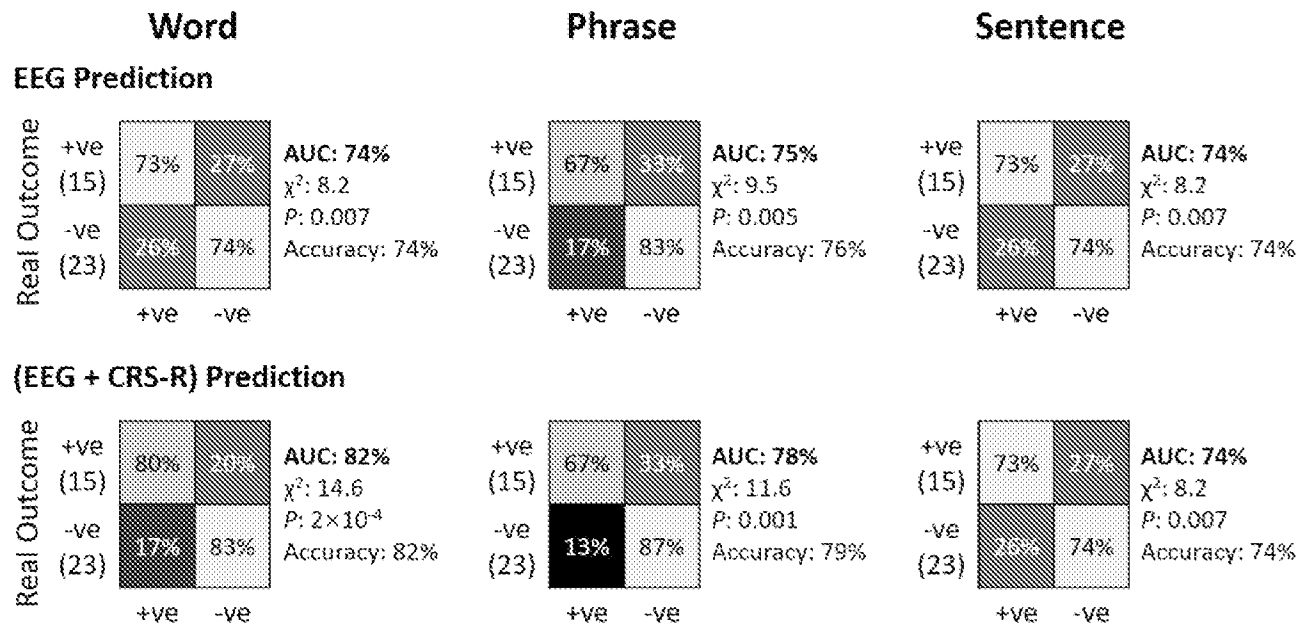


FIG. 14

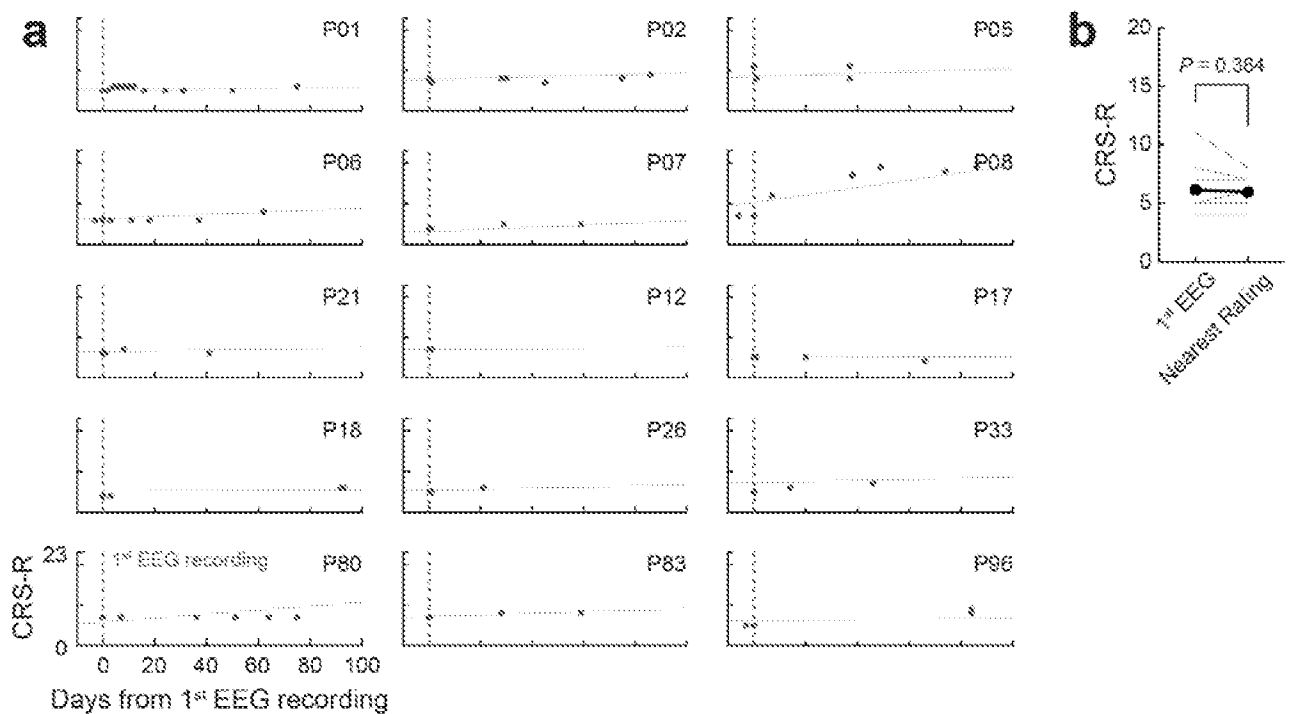


FIG. 15

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/092186

**A. CLASSIFICATION OF SUBJECT MATTER**

A61B 5/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

A61B 5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS,CNTXT,VEN,WOTXT,EPTXT,USTXT,CNKI: consciousness, disorder+, MCS, UWS, coma, EEG, assess+, analy+, evaluat+

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                               | Relevant to claim No. |
|-----------|----------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | US 2020015696 A1 (UNIV MCMASTER) 16 January 2020 (2020-01-16)<br>paragraphs [0080]-[0272], [0340]-[0360], figures 1-5, 11, 12    | 11                    |
| Y         | US 2020015696 A1 (UNIV MCMASTER) 16 January 2020 (2020-01-16)<br>paragraphs [0080]-[0272], [0340]-[0360], figures 1-5, 11, 12    | 11-21                 |
| Y         | US 2020077918 A1 (ICM INST DU CERVEAU ET DE LA MOELLE EPINIERE et al.) 12<br>March 2020 (2020-03-12)<br>paragraphs [0062]-[0065] | 11-21                 |
| A         | CN 101155548 A (Mentis Cura EHF) 02 April 2008 (2008-04-02)<br>the whole document                                                | 11-21                 |
| A         | US 2019246927 A1 (CERENION OY) 15 August 2019 (2019-08-15)<br>the whole document                                                 | 11-21                 |
| A         | CN 104921723 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 23 September 2015<br>(2015-09-23)<br>the whole document                    | 11-21                 |
| A         | CN 109645994 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 19 April 2019<br>(2019-04-19)<br>the whole document                        | 11-21                 |



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

20 February 2021

Date of mailing of the international search report

03 March 2021

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC  
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing  
100088  
China

Authorized officer

SHEN, Yanyan

Facsimile No. (86-10)62019451

Telephone No. 62085634

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2020/092186**

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: **1-10**  
because they relate to subject matter not required to be searched by this Authority, namely:  
[1] The subject matter of claims 1-10 relate to diagnostic methods excluded by PCT Rule 39(iv).
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2020/092186**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| US                                        | 2020015696 | A1 | 16 January 2020                      | WO                      | 2020014780 | A1 | 23 January 2020                      |
| US                                        | 2020077918 | A1 | 12 March 2020                        | None                    |            |    |                                      |
| CN                                        | 101155548  | A  | 02 April 2008                        | NO                      | 20074590   | B  | 03 December 2007                     |
|                                           |            |    |                                      | US                      | 2009220429 | A1 | 03 September 2009                    |
|                                           |            |    |                                      | RU                      | 2007135627 | A  | 10 April 2009                        |
|                                           |            |    |                                      | BR                      | PI0608448  | A2 | 29 December 2009                     |
|                                           |            |    |                                      | EP                      | 1861003    | A1 | 05 December 2007                     |
|                                           |            |    |                                      | WO                      | 2006094797 | A1 | 14 September 2006                    |
|                                           |            |    |                                      | JP                      | 2008531158 | A  | 14 August 2008                       |
|                                           |            |    |                                      | NO                      | 20074590   | L  | 03 December 2007                     |
|                                           |            |    |                                      | CA                      | 2603913    | A1 | 14 September 2006                    |
|                                           |            |    |                                      | NO                      | 20074590   | A  | 03 December 2007                     |
| US                                        | 2019246927 | A1 | 15 August 2019                       | CN                      | 111712193  | A  | 25 September 2020                    |
|                                           |            |    |                                      | WO                      | 2019158810 | A1 | 22 August 2019                       |
| CN                                        | 104921723  | A  | 23 September 2015                    | CN                      | 104921723  | B  | 01 December 2017                     |
| CN                                        | 109645994  | A  | 19 April 2019                        | CN                      | 109645994  | B  | 28 April 2020                        |