 10.3217/978-3-99161-014-4-056

TO REPEAT OR NOT TO REPEAT? ERP-BASED ASSESSMENT OF THE LEVEL OF CONSCIOUSNESS - A CASE STUDY

S. Halder[1], A. Matran-Fernandez[1], R. Nawaz,[1], M. Lopes da Silva[2], T. Bertoni[3], J.-P. Noel[3], J. Jöhr[2], A. Serino[3], K. Diserens[2], R. Scherer[1], S. Perdikis[1]

[1]School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom
[2]Department of Clinical Neurosciences, Acute Neurorehabilitation Unit, Division of Neurology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland
[3]MySpace Lab, Department of Clinical Neurosciences, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland

E-mail: serafeim.perdikis@essex.ac.uk

ABSTRACT: Determination of the wakefulness and consciousness state in patients with disorders of consciousness (DOC) is vital for clinical decision-making. Typically, behavioral indicators and motor responses are employed. Recent advancements in neuroimaging have enabled motor independent assessment of DOC patients.

We present a single-case analysis of a 24-year-old female, selected from a sample of n=77 patients, diagnosed with a DOC. We investigated the single-trial classification of stimuli within the peri-personal space (PPS) using event-related potential (ERP) features. Data from two sessions, conducted ten days apart, were analysed.

We observed significant differences in classification accuracies between sessions (high in session one, low in session two), which did not correspond to the patient's recovery from UWS to MCS. ERP analyses confirmed the difference between sessions, supporting the observed changes in classification accuracies.

Our study underscores the importance of longitudinal assessments to accurately diagnose DOC patients. In future research we aim to expand our analyses to the full dataset.

INTRODUCTION

Reliably determining the state of wakefulness and consciousness of patients with disorders of consciousness (DOC) is crucial for clinical decision-making, providing appropriate care and ensuring patient rights. Usually, behavioral indicators and motor performance in response to specific instructions are used to determine this. The Glasgow Coma Scale [1], for example, assesses a person's level of consciousness based on their ability to open their eyes and perform verbal and motor responses. The Coma Recovery Scale-Revised (CRS-R) [2] is a more comprehensive assessment tool that covers multiple domains and allows for a more detailed assessment and differentiation between states of consciousness such as coma, vegetative state (VS), minimally conscious state (MCS), and locked-in syndrome.

In recent times, researchers have been exploring functional neuroimaging technologies and brain-computer interface-based approaches with the aim to detect unique cognitive patterns when assessing the consciousness state of patients who cannot exhibit motor behavior due to brain injuries [3–5]. One of the several brain networks that have been targeted for this purpose is the cortical network that encodes the Peri-Personal Space (PPS). The PPS is the space surrounding the body that defines the immediate physical domain and is relevant to the interaction between self and others or self and the environment [6]. It is assumed that the related cortical network is linked to bodily self-consciousness and therefore hypothesised to be altered in patients with DOC. Indeed, a physiological index of PPS was identified in evoked electroencephalogram (EEG) responses to tactile, auditory, or audio-tactile stimulation at distances within and outside the PPS [7]. Seventeen patients with DOC participated in the study. The results suggest that the extracted multisensory evoked responses degrade in patients with DOC and correlate with the Lempel-Ziv complexity, a metric used to predict global states of consciousness in continuous EEG signals, but not with CRS-R scores [7]. Although these results seem to be in line with neuroscientific findings, they are not yet conclusive and more data is needed to make more precise statements. Among other things, because the Lempel-Ziv complexity as a measure of conscious experience has been called into question [8].

To minimise errors in diagnosis, it is recommended to repeat the CRS-R at least five times within a time period of a few weeks [9]. Given the non-stationarity and inherent variability of EEG signals, the question arises of how many repetitions are required before reliable clinical decision about the state of consciousness of a patient with DOC can be made when using EEG. To answer this question, the experiment in [7] was repeated in a larger cohort of patients with DOC and the EEG and CRS-R assessments were repeated several times per patient. The study is still ongoing, but in this paper we present initial

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-056

results and a case study that highlights and emphasizes the need for repeated measurements for making informed decisions.

MATERIALS AND METHODS

Patients: A dataset of 84 patients (23 female, median age 53 years, range=18–84) with a disorder of consciousness (CRS-R at assessment median=15, range=0–23) was recorded at the University Hospital of Lausanne (CHUV), Switzerland. Seven patients were excluded from the analysis in this paper due to incomplete data. For the remaining n=77 patients a total of 202 sessions (median=2 sessions, range=1–7 sessions) were recorded.

The patient (Patient A) selected for detailed analysis in this paper was 24 years old at the time of the experiment, female and admitted with a traumatic brain injury. Session 1 was performed two days and Session 2 twelve days post-admittance to the acute care unit. The patient was diagnosed with unresponsive wakefulness syndrome (UWS; CRS-R 7) in Session 1 and minimally conscious state minus (MCS-; CRS-R 11) in Session 2. Approximately three months post-injury the patient emerged from the MCS.

Experiment: Three different stimuli were administered: (1) auditory close (AC; distance 5 cm from extended arm; 65.2 dB SPL; 50 ms of white noise via speaker), (2) auditory far (AF; distance 75 cm; 64.1 dB SPL; 50 ms of white noise via speaker) and (3) tactile (T; two FES electrodes attached to dorsal part of arm near elbow; 50 ms of continuous, sub-threshold stimulation at 35 Hz). Furthermore, auditory and tactile were combined (tactile + auditory close (TAC); tactile + auditory far (TAF)). One experimental block consisted of 50 presentations for each of the five stimuli (250 total). Sessions were planned as a set of three blocks. However, this was not always possible due to constraints of the clinical environment. Further details can be found in the description of the original study [7].

EEG recording: EEG data were recorded with 16 channels positioned at Fz, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, and CP4 (all referenced to the right earlobe) at a sampling rate of either 500 Hz or 512 Hz (g.USBamp or g.Nautilus respectively, both by g.tec medical engineering GmbH, Graz, Austria.).

Preprocessing: Recordings were read in GDF format using Python MNE [10]. Continuous data were high-passed filtered at 1 Hz. On epoched data (-1 to 2 s) bad channels were marked using RANSAC [11] and bad trials were marked using AutoReject [12, 13]. Next a 20 Hz low-pass filter was applied to the continuous data. The continuous data was then epoched (-1 to 2 s referred to stimulus onset), bad channels were interpolated, and trials rejected. All the 512 Hz epochs were resampled to 500 Hz. No baseline was applied. This resulted in a median number of 1358 trials per patient (range=319–4810 trials).

Classification: For classification, epochs from 0 to 1 s after stimulus presentation were used. These epochs were resampled to 10 Hz. We trained a shrinkage linear discriminant analysis (LDA) classifier on two classes (AC vs. AF and TAC vs. TAF) and five classes (All vs. All) using scikit-learn [14]. Performance was assessed via mean accuracy using stratified 10-fold cross-validation independently on each session. We used the median accuracy across sessions to select the patient for discussion in this paper.

ERP Analysis: We compared the ERPs between Session 1 and Session 2 by computing the median response across channels CP1, CPz and CP2 in the time window from -0.25 s to 1.0 s around stimulus presentation using the pre-processed data. Furthermore we performed a time-frequency decomposition using eight Morlet Wavelets in a range from 2–18 Hz using 1–9 cycles per frequency band and used this to compute the inter-trial coherence (ITC) for Session 1 and Session 2 separately using all trials. A baseline from -0.25 s to 0.0 s was applied both for ERP and ITC visualisation.

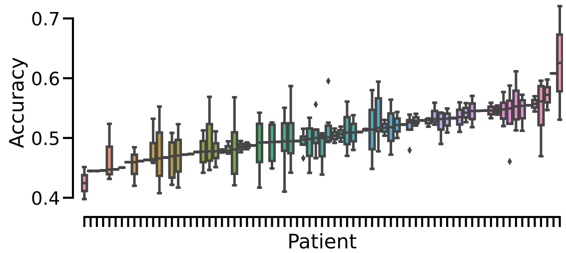

Figure 1: Single-trial accuracy at distinguishing close and far in auditory only mode (i.e,. AC vs. AF). Each boxplot represents the data of one patient across all sessions of that patient. Patients on the x-axis were sorted by median accuracy. A high variance of accuracy was expected as we assumed that some patients in the sample would not be conscious. For this paper we decided to investigate the patient with the highest median accuracy (on the far right; referred to as Patient A in this paper).

RESULTS

Patient selection: Binary classification accuracies (AC vs. AF) of single-trial ERPs ranged from a median of 42% to 62% (median of whole dataset 50%; see Figure 1). For this paper we chose to investigate the patient with the highest median accuracy (referred to as Patient A in this paper).

Patient A accuracies: Accuracies dropped for all classification approaches from Session 1 to Session 2 (see confusion matrices for two-class AC vs. AF in Figure 2, two-class TAC vs. TAF in Figure 3, and five-class All vs. All in Figure 4). Binary single-trial classification accuracies for Session 1 are close to 70% for both the auditory (Session 1 median=0.69, SD=0.11, range=0.53–0.87; Session 2 median=0.54, SD=0.11, range=0.38–0.76) and tactile stimulus modalities (Session 1 median=0.66, SD=0.07; range=0.62-0.83; Session 2 median=0.58, SD=0.09, range=0.38-0.7). Five-class classification accuracy drops in Session 2 in par-

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-056

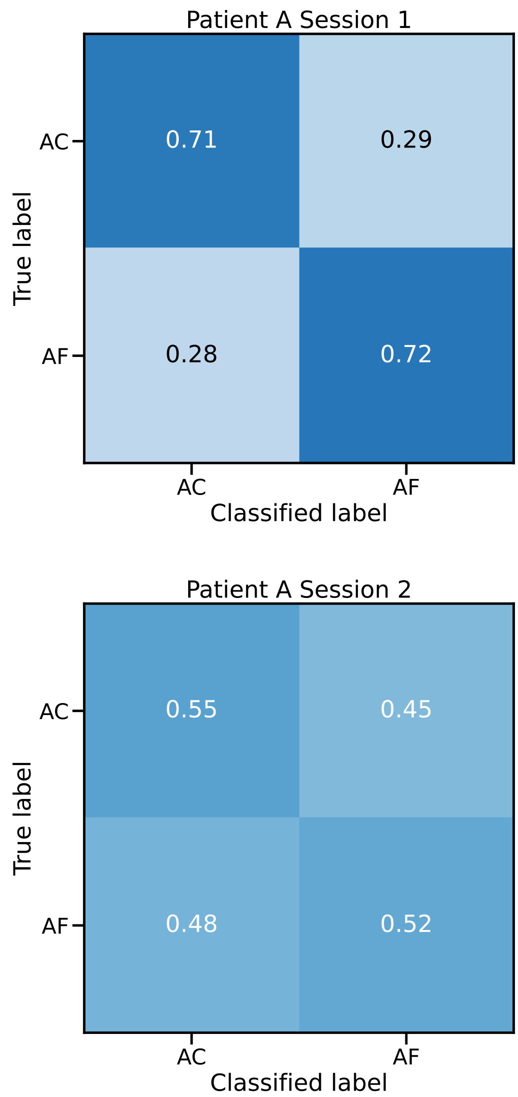

Figure 2: Two-class (AC vs. AF) normalised confusion matrices for Patient A. The accuracy drops considerably from session 1 (top) to session 2 (bottom).

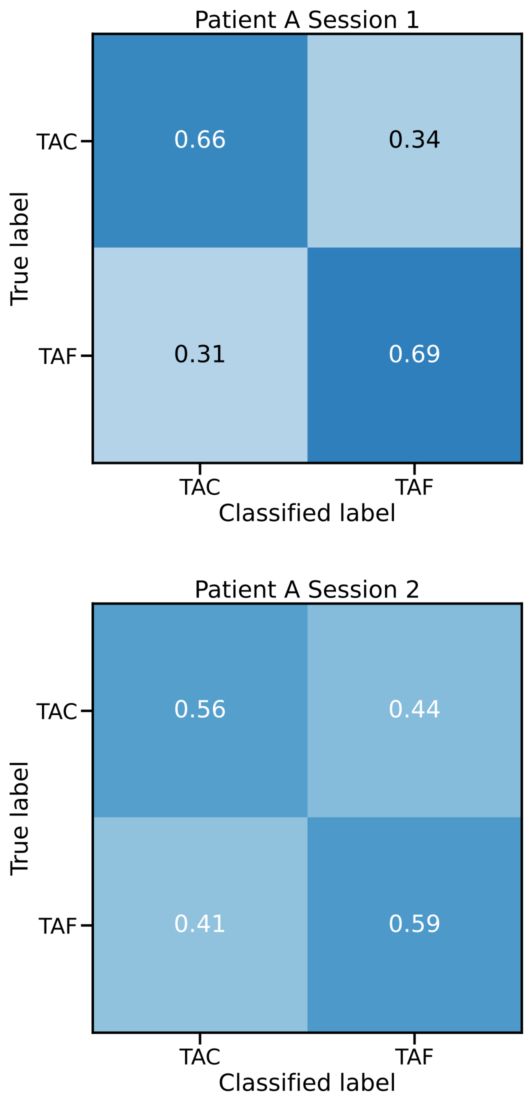

Figure 3: Two-class (TAC vs. TAF) normalised confusion matrices for Patient A. The accuracy drops considerably from session 1 (top) to session 2 (bottom).

ticular due to incorrect classification of pure tactile (T) and tactile-auditory-close (TAC) trials (Session 1 median=0.54, SD=0.06; range=0.46–0.66; Session 2 median=0.35, SD=0.06, range=0.27–0.47). The difference in accuracy between Session 1 and 2 is statistically significant according to t-tests for independent samples with Bonferroni correction (see Figure 5; AC vs. AF Session 1 vs. Session 2 $t_{18} = 3.73, p = .004$; TAC vs. TAF Session 1 vs. Session 2 $t_{18} = 2.72, p = .04$; All vs. All Session 1 vs. Session 2 $t_{18} = 6.99, p < .0001$).

Patient A ERPs: Investigation of the event-related potentials aligns with the classification results. See Figure 6 for responses to tactile-auditory close and far stimuli shown separately for Session 1 and Session 2. Phase-locked responses to the tactile stimulus are visible in Session 1 between 100 and 300 ms post stimulus. Differences in the response to close and far stimuli were clearly visible in Session 1 between 300 and 600 ms. No phase-locked responses were observed in Session 2. The time-domain responses align with inter-trial coherence (see Figure 7). In Session 1 strong phase-locking was observed between 0 and 500 ms post-stimulus with the strongest response around 7 Hz. The visualisation of the ITC has no coherent pattern in Session 2.

DISCUSSION

The case study presented in this paper offers insights into the delineation of PPS in patients with disorders of consciousness using ERP features and single-trial classification. To limit the analysis, as a starting point we selected the patient with the highest median classification accuracy in our database (Patient A), who also exhibited high variability in classification performance across two sessions which took place 10 days apart.

The results demonstrate our approach is capable to dif-

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-056

Figure 4: Five-class confusion matrices for Patient A. The accuracy drops considerably from session 1 (top) to session 2 (bottom).

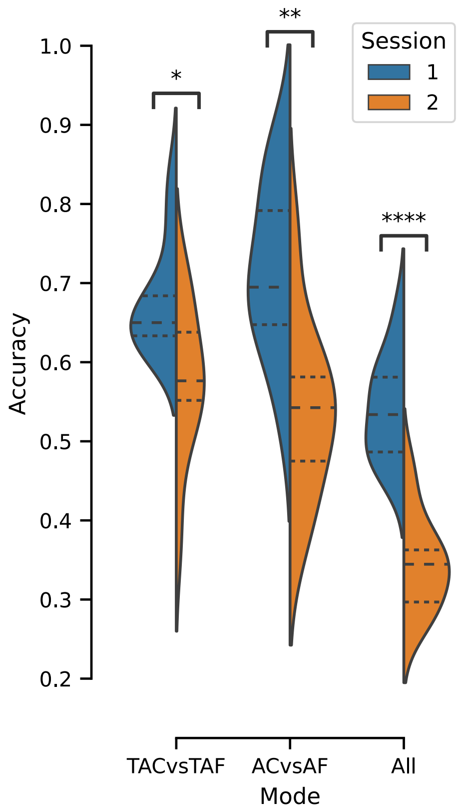

Figure 5: Distribution of accuracy in the 10-fold cross-validation performed on the data of each session. Colours indicate session: blue Session 1 and orange Session 2. The first and second violins represent the binary classification results (TAC vs. TAF; AC vs. AF, resp.) and the third five-class classification results (All vs. All). Stars indicate significance according to t-test for independent samples with Bonferroni correction: *: $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$. The central dashed line was placed at the median of each violin, the finer dashed lines at the first quartiles.

ferentiate stimuli presented within the PPS across various conditions (T, AC, AF, TAC, TAF), with notably high classification accuracies in the first session (particularly as these results are obtained on single-trial ERPs). These findings underscore the sensitivity of PPS delineation as a potential tool for assessing consciousness in DOC patients independently of motor output [7, 15].

After suffering a car accident, Participant A's recordings were performed 2 and 12 days post-injury (Session 1 and Session 2, resp.), with the patient being deemed unresponsive (CRS-R=7, UWS) in the first session and minimally conscious (CRS-R=11, MCS-) in the second one. She was discharged from acute care 21 days post-injury and emerged from minimally conscious state (CRS-R=21) 2.5 months post-discharge.

The variation in classification accuracy between the first and second sessions, particularly with the observed recovery from DOC as indicated by the final available CRS-R score, may suggest an alteration in sensory processing or awareness levels as the patient regained consciousness. This is further supported by the presence of clear ERP peaks for combined tactile and auditory conditions in the first session, which seem to diminish alongside improved consciousness levels. However, the reduction in classification accuracy during the second session poses questions about the dynamics of PPS and its neural correlates as patients recover, such as the representation of PPS and how it might be centered around the body, offering insights into the neural mechanisms that could be involved in PPS processing [16]. Thus, the lower accuracy might occur as a consequence of a reorganization of sensory processing networks or changes in the salience of peripersonal stimuli as the patient's cognitive state evolves. These observations are critical for developing a nuanced understanding of consciousness and its manifestations in DOC patients, providing a foundation for future investigations into the mechanisms underlying consciousness recovery.

On the other hand, there are other plausible explanations

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-056

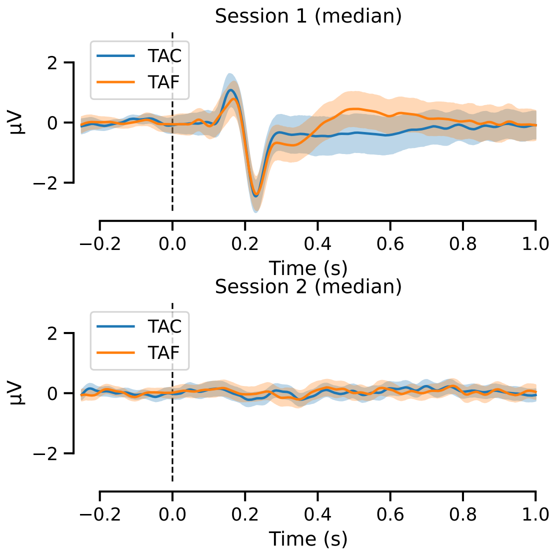

Figure 6: Median ERPs (across channels CP1, CPz and CP2) from Session 1 (top) following combined tactile-auditory stimulation and Session 2 (bottom) of Patient A. Epochs were baselined to the average of the 250 ms before stimulus presentation. The blue lines indicate the response to the close stimulus (tactile-auditory close; TAC), the orange the response to the far stimulus (tactile-auditory far; TAF). The vertical dashed line indicates the timepoint of stimulus presentation. The shaded area indicates the parametric confidence interval (95%).

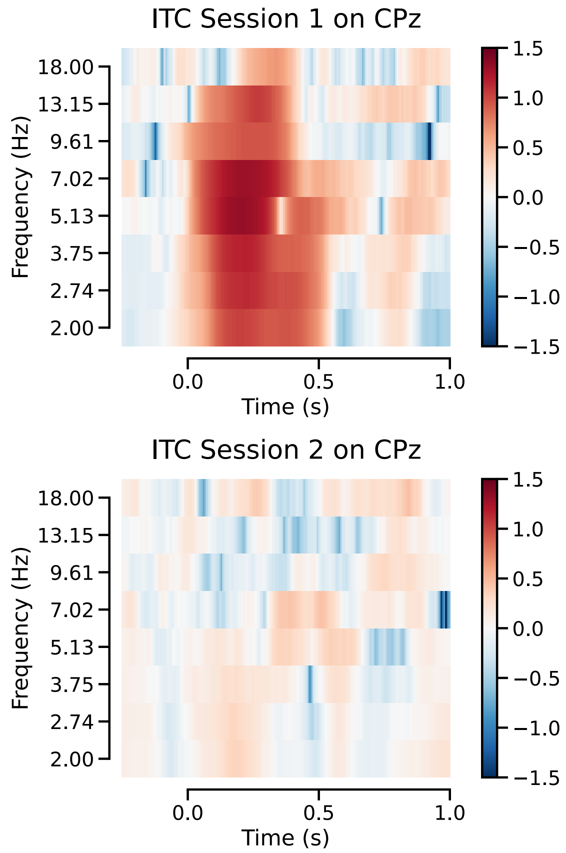

Figure 7: Inter-trial coherence computed for all trials of Session 1 (top) and 2 (bottom) of Patient A. Stimulus presentation occurred at 0.0 s. ITC was calculated for 8 frequency bands in the range of 2-18 Hz. Epochs were baselined to the average of the 250 ms before stimulus presentation. Colour indicates the strength (red stronger, blue weaker) of the ITC.

for the observed low performance in the second session, which is particularly surprising as the patient was discharged shortly after and was associated with a higher CRS-R than Session 1. Firstly, it is conceivable that the patient may have been in a state of sleep during the second session, akin to the absence of responsiveness we would expect to observe if the patient was unconscious [17, 18]. Alternatively, technical issues with the recording equipment or environmental factors, the likelihood of which increases due to the harsh experimental conditions at bedside in an acute unit, could have influenced the quality of the recordings and therefore the final performance of our classifiers. Finally, the discrepancy between the results and the CRS-R scores for each of the sessions could be due to the inherent limitations of the CRS-R themselves [19, 20], particularly as clinical underestimation of conscious awareness may occur (which might have happened in Session 1 of Patient A). However, it is important to note that we cannot definitively conclude which of the four scenarios is applicable in this particular case and further exploration of the larger database is needed. The seeming contradiction of CRS-R and ERP classification results may be resolved by including further measures of consciousness in the analysis, such as measures of EEG signal diversity, such as the Lempel-Ziv Complexity [21, 22] as suggested in [7]. This measure can be computed from the spontaneous EEG enabling us to include this in a future analysis.

A key takeaway from our study is the recognition that relying solely on data from the second session would have led to an erroneous diagnosis for Patient A. This is particularly evident in light of the patient's reemergence to a state of minimum consciousness shortly after the recording of this session. Consequently, our findings align with the perspective advocated by Wannez and colleagues [9] regarding the necessity of conducting multiple recording sessions over time, encompassing various times of the day to account for circadian rhythms. This approach is essential for accurately assessing DOC and avoiding potentially misleading interpretations based on discrete observations. In essence, our study underscores the importance of adopting a longitudinal approach to clinical assessments in this domain.

CONCLUSIONS

In conclusion, our study adds to the body of literature advocating in favour of conducting multiple recording sessions over time when assessing DOC to minimise erroneous diagnoses of patients. However, it is important to acknowledge that we are basing this stance on results from a single patient selected from a larger dataset. In the future, we plan to expand our analyses to the other patients, with the aim of providing a more comprehensive understanding of the factors influencing variability

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-056

in performance and the implications for clinical practice. By broadening our scope and methods (e.g., by including the Lempel-Ziv Complexity measure), we can contribute to enhancing the accuracy and reliability of diagnostic assessments in this challenging area of healthcare.

REFERENCES

[1] Teasdale G, Jennett B. Assessment of coma and impaired consciousness. a practical scale. The Lancet. 1974;304.

[2] Giacino JT, Kalmar K, Whyte J. The JFK coma recovery scale-revised: Measurement characteristics and diagnostic utility. Archives of Physical Medicine and Rehabilitation. 2004;85.

[3] Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. Detecting awareness in the vegetative state. Science. 2006;313.

[4] Horki P *et al.* Detection of mental imagery and attempted movements in patients with disorders of consciousness using EEG. Frontiers in Human Neuroscience. 2014;8.

[5] Real RG *et al.* Information processing in patients in vegetative and minimally conscious states. Clinical Neurophysiology. 2016;127.

[6] Rizzolatti G, Fadiga L, Fogassi L, Gallese V. The space around us. Science. 1997;277(5323):190–191.

[7] Noel JP *et al.* Peri-personal space encoding in patients with disorders of consciousness and cognitive-motor dissociation. NeuroImage: Clinical. 2019;24:101940.

[8] Orłowski P, Bola M. Sensory modality defines the relation between EEG Lempel–Ziv diversity and meaningfulness of a stimulus. Scientific Reports. 2023;13.

[9] Wannez S, Heine L, Thonnard M, Gosseries O, Laureys S, Collaborators CSG. The repetition of behavioral assessments in diagnosis of disorders of consciousness. Annals of Neurology. 2017;81(6):883–889.

[10] Gramfort A *et al.* MNE software for processing MEG and EEG data. NeuroImage. 2014;86:446–460.

[11] Fischler MA, Bolles RC. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM. 1981;24(6):381–395.

[12] Jas M, Engemann D, Raimondo F, Bekhti Y, Gramfort A. Automated rejection and repair of bad trials in MEG/EEG. In: 2016 International Workshop on Pattern Recognition in NeuroImaging (PRNI). 2016, 1–4.

[13] Jas M, Engemann DA, Bekhti Y, Raimondo F, Gramfort A. Autoreject: Automated artifact rejection for MEG and EEG data. NeuroImage. 2017;159:417–429.

[14] Pedregosa F *et al.* Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830.

[15] Patané I, Cardinali L, Salemme R, Pavani F, Farnè A, Brozzoli C. Action planning modulates peripersonal space. Journal of Cognitive Neuroscience. 2019;31(8):1141–1154.

[16] Serino A *et al.* Body part-centered and full body-centered peripersonal space representations. Scientific reports. 2015;5(1):18603.

[17] Chennu S, Bekinschtein TA. Arousal modulates auditory attention and awareness: Insights from sleep, sedation, and disorders of consciousness. Frontiers in psychology. 2012;3:17334.

[18] Cote KA. Probing awareness during sleep with the auditory odd-ball paradigm. International Journal of Psychophysiology. 2002;46(3):227–241.

[19] Cruse D *et al.* Bedside detection of awareness in the vegetative state: A cohort study. The Lancet. 2011;378(9809):2088–2094.

[20] Jöhr J, Halimi F, Pasquier J, Pincherle A, Schiff N, Diserens K. Recovery in cognitive motor dissociation after severe brain injury: A cohort study. PLoS One. 2020;15(2):e0228474.

[21] Lempel A, Ziv J. On the complexity of finite sequences. IEEE Transactions on information theory. 1976;22(1):75–81.

[22] Schartner M *et al.* Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. PloS one. 2015;10(8):e0133532.

Published by
Verlag der Technischen Universität Graz