This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

A Machine Learning Framework for Automatic and Continuous MMN Detection with Preliminary Results for Coma Outcome Prediction

Narges Armanfard*, Majid Komeili†, James P. Reilly*, John F. Connolly‡

***Abstract*—Mismatch Negativity (MMN) is a component of the event-related potential (ERP) that is elicited through an odd-ball paradigm. The existence of the MMN in a coma patient has a good correlation with coma emergence; however, this component can be difficult to detect. Previously, MMN detection was based on visual inspection of the averaged ERPs by a skilled clinician, a process which is expensive and not always feasible in practice. In this paper we propose a practical machine learning (ML) based approach for detection of the MMN component, thus improving the accuracy of prediction of emergence from coma. Further, the method can operate on an automatic and continuous basis, thus alleviating the need for clinician involvement. The proposed method is capable of MMN detection over intervals as short as two minutes. This finer time resolution enables identification of waxing and waning cycles of a conscious state.**

An auditory odd-ball paradigm was applied to 22 healthy subjects and 2 coma patients. A coma patient is tested by measuring the *similarity* of the patient's ERP responses with the aggregate healthy responses. Because the training process for measuring similarity requires only healthy subjects, the complexity and practicality of training procedure of the proposed method are greatly improved relative to training on coma patients directly.

Since there are only two coma patients involved with this study, the results are reported on a very preliminary basis. Preliminary results indicate we can detect the MMN component with an accuracy of 92.7% on healthy subjects. The method successfully predicted emergence in both coma patients when conventional methods failed. The proposed method for collecting training data using exclusively healthy subjects is a novel approach that may prove useful in future, unrelated studies where machine learning methods are used.

***Index Terms*—Mismatch Negativity detection, Coma outcome prediction, Machine learning, Automatic detection of ERP components.**

I. Introduction

Coma is a state of prolonged unconsciousness that has a variety of etiologies, e.g. traumatic brain injury, stroke, brain tumor, drug or alcohol intoxication [1]. Online assessment of comatose patients is very important because it provides us with the capability to detect short increases in the level of consciousness, thus improving both outcome prediction and the rehabilitation process [2]. A passive rehabilitation regimen could be feasible when suitable markers indicate higher levels of consciousness.

*Dept. Elec. and Comp. Eng., McMaster University, Canada. E-mail: armanfn@mcmaster.ca

†Dept. Elec. and Comp. Eng., University of Toronto, Canada.

‡Dept. of Linguistics and Languages, McMaster University, Canada.

Recently, long latency event-related potentials (ERPs) have been introduced as useful predictors of a positive coma outcome [3]. The potential application of longer latency ERPs in clinical practice is reviewed in [2], [4]. Appropriate auditory paradigms elicit long latency ERPs even in the absence of the patient's attention, making them useful in the assessment of altered states of consciousness. One of the most common paradigms used in the literature examines "pre-attentive" processing - that is, neural processing that is below the level of conscious awareness in the individual yet reflects selective processing of a stimulus by virtue of its deviance from an established sequence of stimulation. This paradigm is known as the *oddball* paradigm and typically consists of two types of auditory stimuli [4]: standard tones and deviant tones, where repetitive standard tones are interspersed with slightly deviant stimuli. This demonstrably useful paradigm elicits two different long latency ERP components: the N1 and the Mismatch Negativity (MMN). The presence of N1 and MMN (elicited at respectively about 100 and 150 millisecond post-stimulus) provides evidence of basic brain function at a level reflecting cortical function. The N1 is an obligatory sensory response evoked by each tone (i.e. both standard and deviant) and highlights the encoding of acoustic input in the auditory cortex.

The MMN is an automatic response to auditory stimuli that deviate from the ongoing context of identical auditory stimuli. It reflects automatic sensory memory processes [5], [6]. Although the MMN is often referred to as a "pre-attentive" response, the evidence from sleep and anesthesia research indicates that a state of consciousness is required for the response to occur as indicated by most of the work on MMN absence during NREM sleep in both oddball and acoustic pattern deviations [7], [8], [9], [10]. Some question whether sleep is the best model for the study of consciousness. Regardless, the sleep studies noting MMN absence in NREM sleep are supported by anesthesia work that has also reported the absence of the MMN during sedation [11], [12]. As Tavakoli has noted in [13] and also Dykstra [14], the evidence suggests strongly that MMN elicitation is dependent on the individual being in a "conscious state" even if the individual is not conscious of the deviant stimuli. Thus, it should be noted that we are making a distinction between consciousness and conscious awareness in this report. Occurrence of the MMN does not require nor necessarily reflect conscious awareness but only a state of consciousness.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

Clinical studies on coma patients demonstrate that the presence of the MMN has a high correlation with coma awakening [15], [16]. The reported results show that more than 90% of patients who were considered as non-awake showed no MMN (i.e. a high specificity) and more than 90% of patients in whom MMN was detected returned to consciousness (i.e. a high positive predictive value). But only about 30% of patients who had regained consciousness showed MMN (i.e. a low sensitivity). This implies that if the patient shows MMN, they will emerge with high probability; however the patient may still emerge even though they do not show an MMN (i.e. perhaps because it is present but not discernible, due to the difficulty in detecting it). Because previous work [15], [16] has demonstrated such high positive predictive values, the guiding principle of this paper is that if the MMN is present, then it is highly likely the patient will emerge.

A major difficulty with current methods is that assessment is typically performed based on the average of ERP signals over a long recording time (typically on the order of 30 min or longer [4], [15]), in order to reduce the effect of background EEG noise. A long averaging interval in the presence of latency jitter in the component timing across trials reduces the detectability of the MMN by visual means, since the individual components become "smeared" together in the averaged signal. We postulate that the use of traditional averaging techniques over long intervals reduces the detectability of the MMN and could be one of the reasons for the low sensitivity of the MMN reported in clinical studies.

In this paper, we alleviate the above difficulties by proposing a machine learning based framework that can accurately detect the presence of the MMN, thus offering the potential to improve the sensitivity of prediction of coma emergence beyond the current 30% figure and so improve the accuracy of the test for emergence.

The proposed method is shown to be capable of detecting the MMN over intervals as short as two minutes. We show that this short window length reduces the smearing effect and reveals evidence of distinct waxing and waning cycles of a conscious state (N.B. not necessarily conscious awareness, however) in coma patients. Experimental verification of these waxing and waning cycles is a novel result of this study.

To assess the prognosis for emergence of coma patients, we employ a novel "similarity" criterion (to be described), whereby ERP responses of a coma patient under test are compared to those of healthy subjects. Since healthy subjects exhibit MMN with high probability [17], [18][1] and if the similarity of a coma patient to healthy subjects is high, then it is highly likely the patient exhibits the MMN. Since the presence of the MMN indicates emergence with high probability, high similarity indicates a positive prognosis for emergence. Furthermore, there may be possibilities that low similarity values will indicate non-emergence, although this assertion has not been validated in this paper and must be verified with more data.

A further advantage of the proposed method is that the training data is provided solely by healthy subjects, which are far easier to recruit than coma patients.

One of the major drawbacks of all previous studies using long–latency ERPs is that they all require manual, visual inspection by a skilled clinician [15], [19], [20] while, in practical situations, such an expert is not always available. A further advantage of the proposed method is that it operates on an automatic and continuous basis and therefore alleviates this need.

Performance of the proposed framework is demonstrated on 22 healthy subjects and two coma patients. MMN detection accuracy on the healthy subjects is 92.7%. The two coma patients both exhibited intervals of high similarity value, suggesting the proposed method predicts emergence. Both patients did in fact emerge. It is interesting to note that no distinctive MMN components were obtained by previous averaging methods in either patient. Thus, our method gave the correct prediction results whereas traditional methods would have failed in this situation.

This paper is based on very preliminary results. Much more extensive data collection and testing must be undertaken before the method can be considered for clinical purposes. However, a novel and potentially useful method for determining coma prognosis is presented, which on the basis of limited data, shows considerable promise in dealing with a difficult problem.

II. Methods

A. Passive oddball paradigm and EEG recording

In this study, the N1 and MMN components were elicited using a modification of a classic auditory oddball paradigm, as described in part in [21]. Stimuli consisted of standard tones (85%) and deviant tones (15%). These stimuli were randomly presented; however, each deviant was preceded by at least two standard tones. We used a duration deviant that is one of the most robust types of "deviant" features, both for evoking the MMN but also for producing one of the most stable MMN waveforms over time [22].

All healthy subjects and coma patients were exposed to the passive oddball paradigm comprised of standard and deviant tones of 800 Hz with durations of 75 ms and 30 ms, respectively. Tone rise/fall time was 5 ms and a stimulus onset asynchrony (SOA) between tones of 610 ms; thus, the onset of one tone (either standard or deviant) occurred 610 ms after the onset of the preceding tone (either standard or deviant). We use the term *epoch* to denote the time interval corresponding to a single stimulus trial. In the analysis phase, only the first 300 msec. of data within an epoch are considered, since all useful information regarding the N1 and MMN components is contained within this 300 msec. interval. The tones were presented in one block of 1880 stimuli (with a total duration of approximately 20 minutes) comprised of 280 deviants and 1600 standards. Note that the numbers of deviants and standards are changed to fit the 85 - 15 percent values.

[1]The study in [17] cites an 82% MMN occurence rate in healthy adults, whereas the recent study [18], which uses duration deviants (as does the present study), obtains a 100% occurrence rate in healthy adults.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

The EEG was recorded with a 32-channel BioSemi headcap[2] with a sampling rate of 512 Hz. The recording incorporated a 0.01-100Hz analog bandpass filtering operation that was digitally filtered offline to a bandwidth between 0.1-30Hz.

B. Data Pre-Processing

The relevant components, corresponding to both standard and deviant stimuli, are contained within an interval 0 to 300 ms after the stimulus onset. To better highlight N1 and MMN components in pathological recordings and to remove the effect of eye blink and muscle artifacts, the extracted 300-ms epochs are filtered by a band-pass FIR filter from 2 Hz to 30 Hz with a filter order of 40 [15], [23], [24] . Then the epochs in which the variance of the Vertical-EOG channel exceeds 500 μv^2 or the signal peak to peak (on any electrode) exceeds 100 μv are excluded [25]. The first 2-min window of the raw data is also excluded.

For feature extraction, a relatively large number M of candidate features are extracted from the signal at hand. Candidate features are parameters in various forms which are extracted from the signal which, on an *a priori* basis, are speculated to be discriminative between the classes (Section II-C explains how we are dealing with a two-class classification problem in the machine learning phase of this work). The specific candidate feature set in our case consists of various statistical quantities at each channel. These quantities are kurtosis, skewness, variance, maximum, minimum and power in eight different frequency bands: Alpha-band (8Hz to 13Hz), Beta1-band (13Hz to 20 Hz), Delta-band (1Hz to 4Hz), Lower-band (1Hz to 8Hz), Total-band (1Hz to 30 Hz), Beta-band (13Hz to 30 Hz), Beta2-band (20Hz to 30Hz) and Theta-band (4Hz to 8Hz). In addition, the wavelet decomposition vector with wavelet 'rbio6.8' [3] at level 3 is also considered (i.e. 62 features per channel so that half are approximation coefficients and the other half are level 3 detail coefficients). This wavelet has proven useful for previous EEG studies [26]. Consequently, all together, each channel is represented with 75 features. This set of candidate features is selected based on the fact they have proven to be discriminative in previous studies; e.g. [27], [28], [29], [30].

C. The LFS Method for Feature Selection and Similarity Measurement

For the machine learning aspect of this work we consider two classes of data, where the first class Y_1 corresponds to the presence of N1 only (i.e. response to standard tones) and the second class Y_2 corresponds to the presence of the MMN and N1 components (i.e. response to the deviant tones). Our methodology is based on designing a machine learning process which discriminates class Y_2 from class Y_1.

Not all of the candidate features are discriminative (i.e. between classes Y_1 and Y_2). Irrelevant features may degrade the accuracy and efficiency of the similarity measurement in the test phase [31]. Therefore, the candidate feature set must be reduced to contain only the most relevant features. This task is performed using a feature selection process, which is common practice in machine learning problems with a relatively small number of training samples [32], [33], [34], [35], [36].

Many feature selection approaches are presented in the literature [37], [32], [33], [38]. Almost all of these methods select a *global* common feature subset for all regions of the sample space. In contrast, the recently–developed localized feature selection (LFS) method [39], [40], assigns a unique and possibly distinct feature set to every training sample. This allows the underlying machine learning model to adapt to nonstationarities, disjoint class clusters, and nonlinear decision boundaries in the sample space. The LFS method has been shown to be immune to the overfitting problem. In addition the training process is convex, and the method naturally provides a *similarity* measure, which is an indication how close a query datum $\mathbf{x}^q$ is to the respective classes. The method is also highly suitable when the number of candidate features far exceeds the number of available training samples. It is for these reasons that the LFS method is preferred for the task at hand.

We give a brief summary of the LFS method. The reader is referred to the references for full details. The algorithm accepts training samples $\mathbf{x}^{(i)}, i = 1, \ldots, N_{tr}$ (i.e. samples of candidate features) as input. N_{tr} is the number of training samples. The samples consist of two classes, Y_1 and Y_2. Each training sample has an associated label $y^{(i)} \in [Y_1, Y_2]$.

Each training sample is treated as a representative point for its surrounding region. Within each region, an optimum feature set is selected so that, locally within the region, same–class samples cluster as closely as possible around the training sample, whereas opposite–class samples are far removed from the training sample. To determine the selected features associated with each training sample, we consider M-dimensional indicator vectors $\mathbf{f}^{(i)}, i = 1, \ldots, N_{tr}$, where M is the number of candidate features. These vectors have the property that

$$\mathbf{f}^{(i)}(m) = \begin{cases} 1, & \text{if the mth candidate feature is selected} \\ & \quad \text{by the ith training sample,} \\ 0, & \text{otherwise.} \end{cases}$$

where $m = 1, \ldots M$. We define two quantities U_1 and U_2 as follows:

$$U_1 = \sum_{\substack{j \in \mathbb{Y}_s \\ j \neq i}} ||\mathbf{x}_p^{(i)} - \mathbf{x}_p^{(j)}||_{\mathcal{L}},$$

and

$$U_2 = \sum_{\substack{j \in \mathbb{Y}_o \\ j \neq i}} ||\mathbf{x}_p^{(i)} - \mathbf{x}_p^{(j)}||_{\mathcal{L}},$$

where $\mathbb{Y}_s$ ($\mathbb{Y}_o$) is the set of training samples having the same (opposite) class as $\mathbf{x}^{(i)}$, and $||\cdot||_{\mathcal{L}}$ is a norm which induces locality; i.e., samples far from $\mathbf{x}_p^{(i)}$ are given less weight in the sum. The subscript p indicates that the

[2] http://www.biosemi.com/headcap.htm

[3] available in MATLAB

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

respective quantity has been projected into the coordinate space induced by $\mathbf{f}^{(i)}$; i.e., elements of $\mathbf{x}^{(i)}$ are set to zero if the corresponding element in $\mathbf{f}^{(i)}$ is zero. U_1 is a measure of the combined intra-class distances to $\mathbf{x}_p^{(i)}$ (in the local coordinate system determined by $\mathbf{f}^{(i)}$), whereas U_2 gives a measure of the combined distance of local opposite–class samples to $\mathbf{x}_p^{(i)}$ in the same coordinate system.

The value of $\mathbf{f}$, which determines the set of selected features, is obtained as the solution to an optimization procedure that simultaneously minimizes U_1 and maximizes U_2; i.e.

$$\begin{aligned} \mathbf{f}_{opt}^{(i)} \quad = \quad & \arg\min_{\mathbf{f}^{(i)}} U_1 \\ & \text{and} \\ & \arg\max_{\mathbf{f}^{(i)}} U_2. \end{aligned} \tag{1}$$

Thus, the end result is that features are selected so that local clustering of intra–class samples is as tight as possible, whereas inter–class samples are as far removed as possible from $\mathbf{x}_p^{(i)}$. In [39], this objective is re-formulated into an readily–solved linear program, whereas in [40] the problem is recast as a convex program.

The LFS method precludes classification by conventional means, since the selected features vary over the training sample space. Classification in the LFS case is accomplished by assigning a hypersphere $\mathcal{Q}^{(i)}$ which is centered on each $\mathbf{x}_p^{(i)}$. Each hypersphere is given a class label which is the same as the corresponding $\mathbf{x}^{(i)}$. The local coordinate system of each sphere corresponds to the features selected by the LFS algorithm. The radius $r^{(i)}(\gamma)$ is determined such that the "impurity level" within the hyper-sphere $\mathcal{Q}^{(i)}$ is not greater than the user-defined parameter γ. The "impurity" level is the ratio of the number of inter-class samples within $\mathcal{Q}^{(i)}$ to the number of intra-class samples within $\mathcal{Q}^{(i)}$. In all our experiments, γ is fixed to its default value 0.2. For more details see [39].

The *similarity* $S_{Y_\ell}(\mathbf{x}^q)$ of a query datum $\mathbf{x}^q$ to class $Y_\ell \in \{Y_1, Y_2\}$ is measured based on how many hyper-spheres with class label Y_ℓ contain $\mathbf{x}^q$. The similarity measure is instrumental in subsequent discussion.

To this end, we define a set of binary variables $s^{(i)}(\mathbf{x}^q)$ as

$$s^{(i)}(\mathbf{x}^q) = \begin{cases} 1 & \text{if } \mathbf{x}_p^q \text{ is inside } \mathcal{Q}^{(i)} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The *similarity* $S_{Y_\ell}(\mathbf{x}^q)$ of $\mathbf{x}^q$ to the class Y_ℓ is computed through an aggregation process as follows:

$$S_{Y_\ell}(\mathbf{x}^q) = \frac{\sum_{i \in \mathbb{Y}_\ell} s^{(i)}(\mathbf{x}^q)}{\eta_\ell} \tag{3}$$

where $\mathbb{Y}_\ell$ indicates the set of all regions whose class labels are Y_ℓ. η_ℓ is the cardinality of $\mathbb{Y}_\ell$. Thus, the similarity of $\mathbf{x}^q$ to class Y_ℓ is the ratio of the number hyperspheres with class Y_ℓ which contain $\mathbf{x}^q$, to the total number of training samples having class Y_ℓ.

Once the training process is complete, the class label y^q of a query datum $\mathbf{x}^q$ is determined as the class corresponding to the largest similarity value.

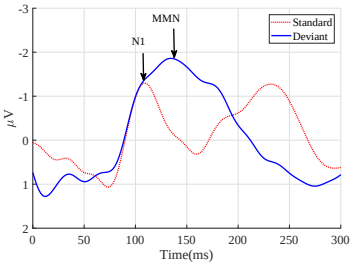


Fig. 1: Average of de-artifacted epochs corresponding to standard (red) and deviant (blue) stimuli of a typical healthy subject at channel Fz. The obligatory N1 component is elicited for both standard and deviant stimuli while the MMN occurs only for the deviant.

D. Methodology for Determining Prognosis

We can now discuss the overall methodology for determining whether a particular coma patient is likely to emerge. The process involves the training phase, the evaluation phase, and the testing phase.

1) The Training Phase: We use the LFS method to train a machine learning model to discriminate between classes Y_1 and Y_2 in healthy subjects. The data are pre-processed in the manner described in Sect. II-B. To generate reliable and stable training points, for each training subject, the waveforms over all available deartifacted epochs associated with standard and deviant tones are averaged together separately. Therefore, each of N subjects provides two training samples; one for standard tones and the other for deviant tones. Thus altogether we have $N_{tr} = 2N$ training samples. Two such training sample waveforms (at channel Fz) for a typical healthy training subject are shown in Fig. 1 where, as expected, it is seen the standard and deviant stimuli elicit the N1 component and the MMN and N1 components, respectively.

Each of the N training subjects gives us two averaged waveforms from each of the 32 channels, corresponding to the two types of tones. We extract 75 candidate features from each of these averaged waveforms, as described in Sect. II-B. We concatenate the features from the channels together for each subject, to form two sets of M–dimensional candidate feature vectors $\mathbf{x}_{std}^t(i)$ and $\mathbf{x}_{dev}^t(i), i = 1, \ldots, N$, where $M = 75 \times 32 = 2400$. The superscript t specifies *training* data. Since it is known that each vector $\mathbf{x}_{std}^t(i)$ is from class Y_1, these data each have label Y_1. Similarly all $\mathbf{x}_{dev}^t(i)$ vectors have label Y_2.

The training vectors $\mathbf{x}_{std}^t(i)$ and $\mathbf{x}_{dev}^t(i)$ and their associated class labels $y^{(i)}$ are fed into the LFS algorithm where they assume the role of the training samples $\mathbf{x}^{(i)}, i = 1, \ldots, N_{tr}$. The LFS procedure then identifies an optimal subset of selected features for each input training vector, and also specifies the radius $r^{(i)}$ of the hypersphere $\mathcal{Q}^{(i)}$, associated with each training sample, according to the discussion in Sect. II-C. Once this process is complete, the LFS method can then calculate the similarity of a query datum to each class and therefore classify the query datum sample.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

2) The Evaluation Phase: Once the machine learning model has been trained, we must assess its accuracy on healthy subjects. To do this the epochs from a healthy test subject are now averaged over what we refer to as a *P1 interval*, which implies an interval extending 2 minutes into the past that is updated in 1 minute segments. Performance is evaluated in each P1 interval using a Leave-One subject-Out (LOO) cross validation strategy. One round of cross-validation involves partitioning the N total available subjects into two sets: $N-1$ training subjects and one test subject. The training phase explained in section II-D1 is applied to the $N-1$ training subjects and then the test phase explained in Section II-D3 is applied to the one remaining test subject. Here, since ground truth (i.e., the labels) of the test subject ERPs are known, validation is performed by comparing the test result to the ground truth value.

In this paper the total number N of available healthy subjects is 22; hence the number of available training samples at each LOO round is 42 (i.e. $2 \times (N-1)$) (21 samples for each of classes Y_1 and Y_2). In each LOO round, in each P1 interval of the test subject, there is one query datum for each class [4]. The LOO procedure is performed 22 times, so that over all rounds each subject is left out once.

In each P1 interval in each LOO round where the ith subject is left out, we extract two query candidate feature vectors $\mathbf{x}^q_{\text{std}}(i,t)$ and $\mathbf{x}^q_{\text{dev}}(i,t)$ from the respective waveforms, corresponding to standard and deviant responses and where t is the P1 interval index, in exactly in the same manner as described above for the training phase. The only difference is that in this case the data are averaged only over a P1 interval instead of over all available data. These data pairs, over the entire available range for both i and t, are then fed into the LFS algorithm for classification, and the output test values are compared to the known class values and the results tallied.

Three criteria are used for performance evaluation: these are true positive rate (TPR), true negative rate (TNR) and Accuracy, which in the general case are defined as follows:

$$\text{TPR} = \frac{1}{N}\sum_{k=1}^{N}\frac{TP^{(k)}}{P^{(k)}},\ \text{TNR} = \frac{1}{N}\sum_{k=1}^{N}\frac{TN^{(k)}}{N^{(k)}}, \quad (4)$$

$$\text{Accuracy} = \frac{1}{N}\sum_{k=1}^{N}\frac{TP^{(k)}+TN^{(k)}}{P^{(k)}+N^{(k)}}, \quad (5)$$

where, at round k of the LOO process, $TP^{(k)}$ ($TN^{(k)}$) are the number of query data with class Y_2 (Y_1) that are correctly predicted as class Y_2 (Y_1), and $P^{(k)}$ ($N^{(k)}$) are the total number of query points with class label Y_2 (Y_1). High accuracy with equally distributed TPR and TNR is desired. The results of this evaluation phase are reported in Sect. III.

[4]Within each P1 interval, the average rate of rejection (due to the artifact removal explained in Section II-B) for standard and deviant epochs are respectively 14.45% and 14.20% for healthy subjects, and 3.81% and 3.6% for coma patients.

3) The Test Phase: Here, our objective is to determine whether a previously unseen coma patient is likely to emerge. Here, all the 22 available healthy subjects are used as training subjects. In each P1 interval of the coma patient, we pre-process the patient data, average the waveforms over the respective P1 interval, and extract candidate features as described in Sect. II-B. These features form two "query" M-dimensional vectors $\mathbf{x}^q_{\text{std}}(t), \mathbf{x}^q_{\text{dev}}(t)$, corresponding to standard and deviant stimuli respectively as before.

Within each P1 interval, using the selected features and hyperspheres already established in the training procedure using healthy subjects, we evaluate the similarity value $S_{Y_\ell}(\mathbf{x}^q)$ of the test samples $\mathbf{x}^q_{\text{std}}(t), \mathbf{x}^q_{\text{dev}}(t)$ to their respective classes, according to eq. (3). If the similarity to both classes is sufficiently high, then the patient's responses are close to those of healthy responses, from which we can conclude that the MMN exists within that P1 interval with high probability and so the patient is likely to emerge.

E. Remarks

To apply the proposed method in a clinical setting, we would evaluate similarity of a patient's deviant and standard responses to class Y_2 and Y_1, respectively. If it exceeds a specified threshold value with sufficient density over a specified number of P1 intervals, then we would declare emergence. Thus the methodology we describe in this paper is a positive indicator for emergence. More data is required before we can specify a threshold value and an appropriate number of P1 intervals.

At this point we can speculate that our method could also be used as a negative indicator for emergence. That is, if the similarity is beneath a threshold for a specified interval and also no evidence of high similarity exists, then we would declare non-emergence. However this aspect of the method has significant ethical implications, since the ramifications of a false negative can be severe. This is a topic for further work that must be executed carefully.

An alternative machine learning approach to determining emergence may be to collect training data directly from coma patients, where the requisite labels would correspond to whether or not the patient eventually emerged. The difficulty with this alternative approach is that coma patients exhibit waxing and waning cycles of a conscious state, as demonstrated in Figure 3 in Sect. III. Even in patients who do emerge, it may be seen that the similarity level, and hence level of brain function, varies with the observation interval. Since the level of brain function is highly salient to emergence, some intervals will be indicative of emergence and others not. It is difficult to determine beforehand which intervals are indicative. So the training data samples with this approach are not consistent in representing the patient's state of emergence. This inconsistency in the training samples reduces the ability of a machine learning model to classify properly. Furthermore, it is difficult to conceive how detection of the MMN can be incorporated into this alternative approach. As we have seen, the presence of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

MMN is a highly salient indicator of emergence, so by ignoring it we are not taking advantage of this valuable information.

In contrast, the proposed method trains using only healthy subjects, who exhibit the MMN on deviant tones with high probability. Therefore with the proposed method, the training data does not exhibit the waxing and waning phenomenon and is therefore consistent across time. The method is also based on detection of the highly salient MMN component, and therefore takes advantage of this available information.

III. Results

In Section III-A, performance of the proposed methodology is demonstrated on healthy subjects. Section III-B provides preliminary evidence of the effectiveness of the proposed framework for assessment of comatose patients.

A. Results on Healthy Subjects

Here we report results for testing the LFS method's accuracy in discriminating class Y_1 from Y_2 using the LOO method described in Sect. II-D2. The results are summarized in Table I.

In these experiments, the parameter α used by the LFS method, that sets an upper bound on the number of selected features, ranges from 1 to 10. The TPR, TNR and Accuracy figures are averaged over all available P1 intervals for each value of α. The maximum Accuracy value with respect to α, along with the corresponding TPR and TNR values for the proposed methodology are presented in the second column of Table I where the default parameter values are used for the LFS method. The high prediction accuracy (92.7%) demonstrates the effectiveness of the proposed methodology to identify the presence of the MMN component in healthy subjects. Also, for the interested reader, classification performance of the proposed strategy using seven state-of-the-art global feature selection algorithms is also reported in Table I (columns 3-9) where a linear SVM is used as a classifier with parameters set to their default values and the number of selected features also ranges from 1 to 10. The results show that the localized approach is a better fit to this problem.

The high performance on healthy subjects indicates that the MMN can be detected with high accuracy in healthy subjects in deviant epochs. This suggests that the MMN can also be detected with similar accuracy on coma patients, provided the MMN waveform characteristics of the coma patient are similar to those of healthy subjects[5]. Thus on deviant and standard epochs, if the coma patient's response has a sufficiently high similarity respectively to classes Y_2 and Y_1, then it is highly likely that the MMN exists. Then, according to [15], [16], the patient will emerge. Therefore the validity of the proposed method is verified.

Furthermore, it is noted from Sect. II-D1 that determination of similarity for any query datum requires only healthy subjects for training. This justifies our proposed training methodology. The ability to train only on healthy subjects is a major advantage of the proposed method, since they are much more easily recruited than coma patients.

B. Preliminary evidance for coma prognosis

In this section the proposed methodology is applied to our 2 available comatose patients.

Patient 1 is a 29-year old male who was involved in a motor vehicle collision. In total, 17 sessions of data were acquired. He scored 4 on the Glasgow Coma Scale [6] at the initiation of data collection. The patient emerged, gradually improved and was sent to the rehab unit.

Patient 2 is a 21-year old male who was involved in a collision between an all-terrain vehicle and a tree. He has a Glasgow Coma Scale rating of 7 at the time of recording. In total, 7 sessions of data were acquired. He was transferred to stepdown 15 days post-injury. He slowly emerged over the next couple days, showing more and more alertness, tracking, and command following. He was then discharged to the rehab unit 2 months after his collision.

For both patients, recordings were conducted in sessions, where each session was approximately 20-min in duration, with an inter-session interval of about 2 hours. Each session consists of 17 P1 intervals.

The layout of the remaining part of this sub–section is as follows: First, we analyze the available coma patients using traditional analysis with averaging. These results indicate only indefinite evidence of the MMN at best. Secondly, we perform the analysis using the proposed method. In this case, evidence of the MMN is clearly evident in the both patients, although we observe the level of similarity to wax and wane over time. Finally, we show signals averaged over short 2-min intervals in the last part of this section, where responses are averaged over typical P1 intervals. The detection result (i.e. presence/absence of the MMN component) obtained from visual inspection of these typical P1 intervals coincides with the estimated similarity level corresponding to the same P1 interval.

1) Analysis of Available Coma Patients Using Traditional Approaches: Traditional approaches visually inspect one averaged signal over a limited number of channels (typically channels Fz and Cz) to detect the presence of the N1 and MMN components [48], [15]. This signal is the average of *all* epochs extracted from *all* recording sessions. In this paper we refer to such a signal as the *subject-average* signal. The subject-average signals corresponding to standard and deviant tones for the 2 patients at channels Fz and Cz are shown in Fig. 2.

[5]This assertion has been verified by comparing waveforms from healthy subjects with those from the two available coma patients. In some rarer cases, the latency and/or amplitude of patient MMN responses may fall outside the range expected in most healthy subjects. The proposed method can at least partially accommodate these cases by optimizing the values of the threshold β (defined in Section III-B2) and the LFS parameter γ, for overall best prediction error, when more data becomes available.

[6]The Glasgow Coma Scale (GCS) is the most common clinical indicator that describes level of consciousness based on clinical assessment [43], e.g. asymmetry in pupillary responsiveness, dilatation and constriction, verbal and motor responses [44], [45], [46], [47]. The GCS ranges between 3 (deep unconsciousness) to 15 (best response).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics

TABLE I: LOO Accuracy (averaged over all available P1 intervals) of the proposed method, along with the corresponding TPR and TNR (in percent), using both local and global feature selection algorithms. Standard deviations (in percent) are presented in parentheses.

	LFS	**FDA** [38]	**mRMR** [37]	**Logo** [32]	**FMS** [33]	**DEFS** [34]	**KCSM** [41]	**SIMBA** [42]
Accuracy	**92.7(8.1)**	86.8(10.0)	86.9(9.3)	87.1(8.9)	86.8(10.0)	73.3(11.5)	86.3(9.8)	83.3(11.5)
TPR	92.9	83.6	79.4	80.4	83.6	96.9	82.1	76.3
TNR	92.4	90.0	94.4	93.9	90.0	49.0	90.6	90.3

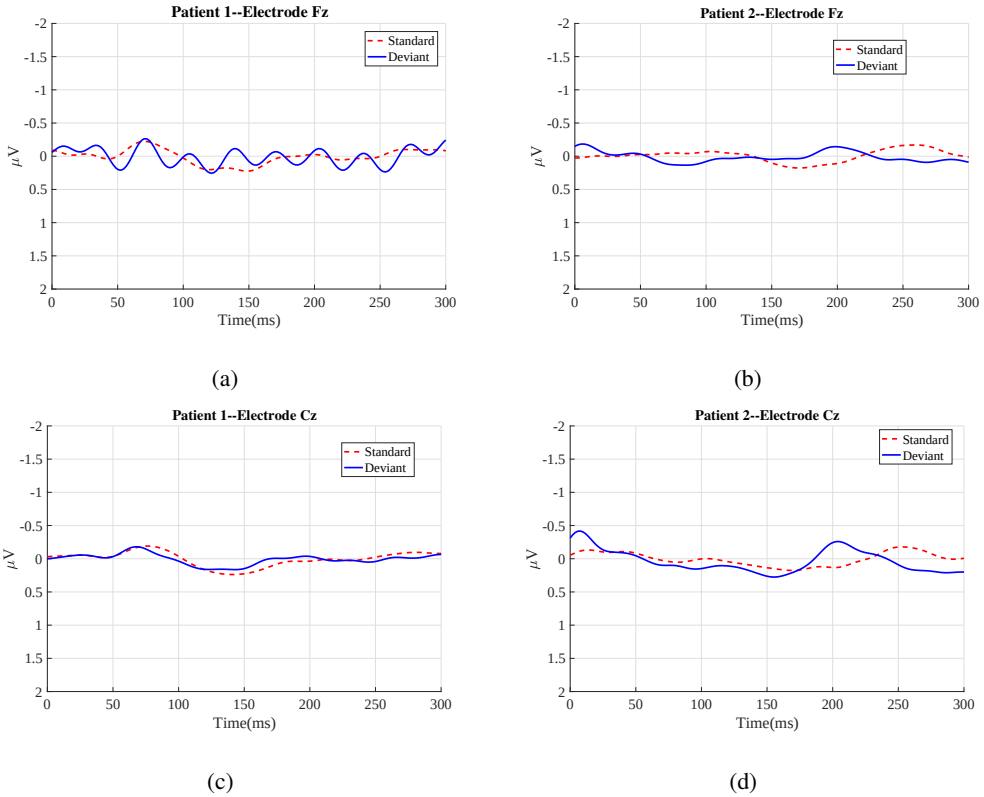


Fig. 2: Subject-average signals corresponding to standard and deviant tones for patients 1 and 2 for the sites positioned at electrodes Fz and Cz.

It is seen from these figures that in all cases, the N1 and MMN components are not clearly discernible from either response, or if they are discernible, they are both very weak. Yet, both patients emerged, suggesting these components may indeed be present in their ERP responses, although obscured. We believe this outcome is a manifestation of the poor sensitivity of the conventional MMN test to determine prognosis based on the subject-average signals.

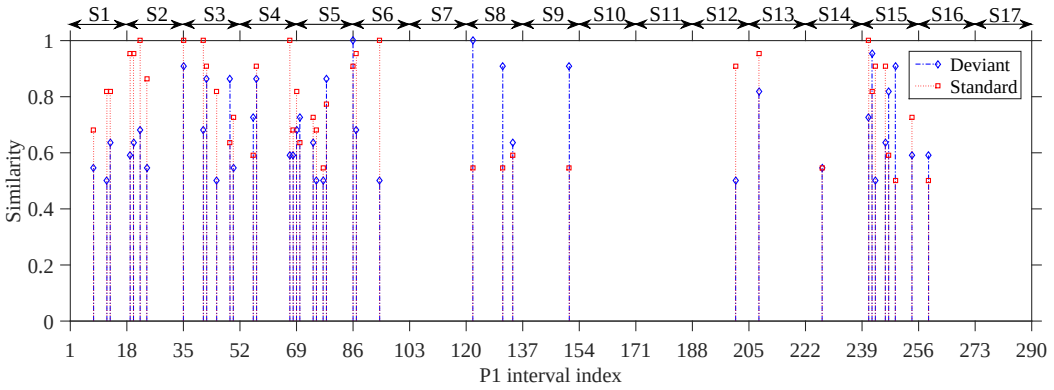
In this paper, we refer to the average of epochs extracted from a single session as a *session–average* signal. It is interesting to note that in addition to the subject–averaged signals above, for both patients, there also existed multiple session–average signals where the appropriate components were not evident, or were very weak, in the responses to both standard and deviant tones. Therefore, since there is no definitive MMN component that can be detected by visual inspection, the session–average signals also indicate that there is no evidence of an appropriate brain response to the deviant stimuli; whereas the detection results of the proposed method shown in Fig. 3 (discussed in the following section) indicate that there are many active P1 intervals that have high similarity to both standard and deviant responses of a healthy brain.

2) Analysis of Available Coma Patients by Proposed Method: We hypothesize that discernibility of the N1 and MMN components in the subject-average signals in Fig. 2 is obscured, or smeared out, by the averaging procedure. To verify this idea, we used patients 1 and 2 as test subjects and assessed their similarities.

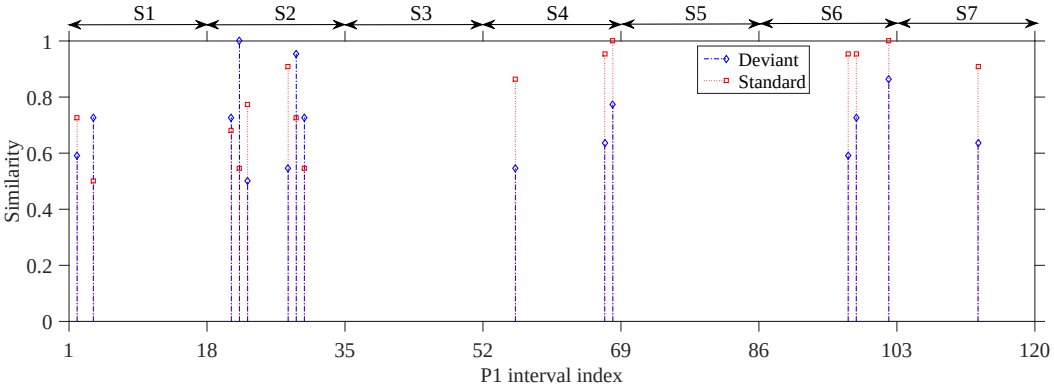
Fig. 3 shows similarities of the patients vs. the P1 interval index, but only for the "active" P1 intervals. Active intervals are those for which both standard and deviant similarities, as defined by (3), exceed a threshold β, which in our experiments is set to 0.5. Hence, active intervals are used as a predictor for the eventual emergence. In each sub–figure of Fig. 3, the first 17 ticks correspond to the 17 P1 intervals of Session 1 and 18-34 correspond to the intervals of Session 2, etc. Sessions are concatenated even though they are not contiguous in time.

In contrast to the session–average and subject–average results, the results shown in Fig. 3a demonstrate that patient

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics



(a) Patient 1–emerged



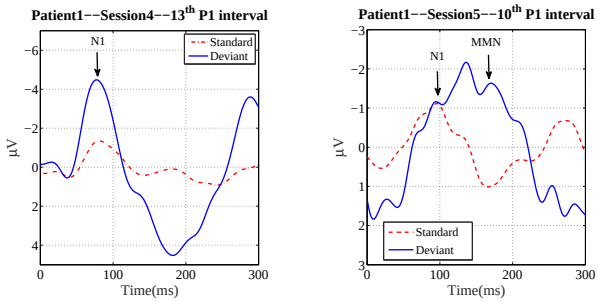
(b) Patient 2–emerged

Fig. 3: Similarity for active P1 intervals of patients 1 and 2 vs. the P1 interval index where β is set to 0.5. The red and blue graphs respectively correspond to $S_{Y_1}(x^q_{std})$ and $S_{Y_2}(x^q_{dev})$. The active P1 intervals are used as a predictor of eventual emergence.

1 has many P1 intervals in which significant responses to both standard and deviant stimuli occurred– i.e. where the brain activity was clearly similar to that of healthy subjects. However, due to the significantly shortened averaging window, this figure also verifies the existence of the waxing and waning cycles of the patient's level of cortical functionality, where in some sessions there were a large number of high–similarity P1 intervals (waxing) while in other sessions there was a decreased number of such intervals (waning). Patient 2 from Fig. 3b also demonstrates similar waxing and waning behavior.

3) Visual examples: In this section we show some typical active and non-active P1 intervals from different sessions of all patients.

We refer to the average of the de-artifacted clean epochs extracted from a *single* P1 interval as a *sub-session* average signal. Sub-session average signals corresponding to standard and deviant tones over two typical *single* P1 intervals of patient 1 are shown in Fig. 4. Visual inspection of these sub-session average signals demonstrates that there is no MMN at the 13th P1 interval of session 4 (Fig. 4a) either in channel Fz or, as verified by the authors, in any other electrode location. However, at the 10th P1 interval of Session 5 (Fig. 4b) an MMN component does exist (as indicated by the arrow). An N1 clearly appears in



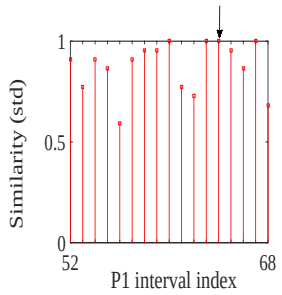
(a) (b)

Fig. 4: Sub-session average signals corresponding to standard and deviant stimuli, at channel Fz, of patient 1 at P1 intervals a) 13 of Session 4, and b) 10 of Session 5.

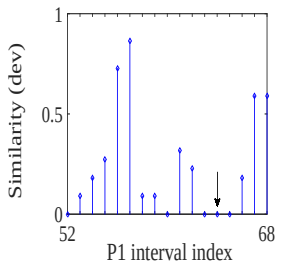
both of the corresponding P1 intervals. Note that none of these components are evident from Fig. 2 nor the session averaged signals. This can be considered a verification of our idea that waxing and waning of the level of functionality exists, and that traditional averaging can obscure this behaviour.

The similarity of patient 1 brain function to those of the healthy brains over all P1 intervals of Sessions 4 and 5 are shown in Fig. 5. Figs. 5a and 5c show that, as expected,

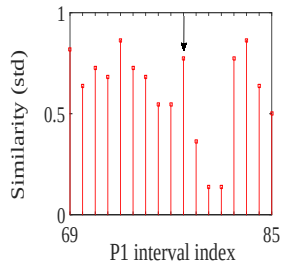
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2018.2877738, IEEE Journal of Biomedical and Health Informatics



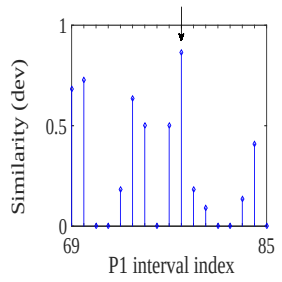
(a) Standard–Patient 1, Session 4



(b) Deviant–Patient 1, Session 4

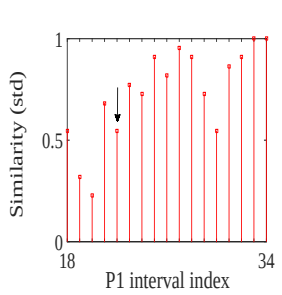


(c) Standard–Patient 1, Session 5

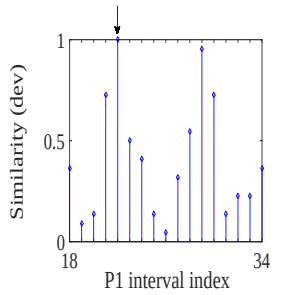


(d) Deviant–Patient 1, Session 5

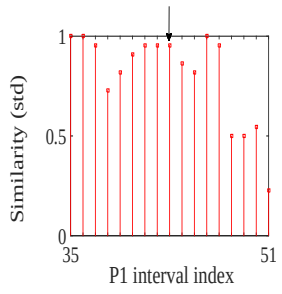
Fig. 5: Similarities corresponding to the P1 intervals of Sessions 4 and 5 of patient 1. (a) and (c) correspond to standard stimuli, (b) and (d) correpond to deviant stimuli. P1 intervals of interest are indicated by arrows.



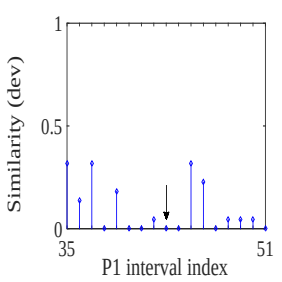
(a) Standard–Patient 2, Session 2



(b) Deviant–Patient 2, Session 2

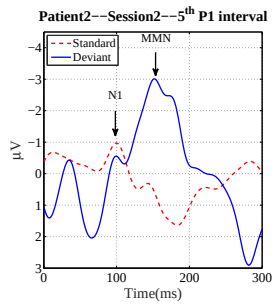


(c) Standard–Patient 2, Session 3

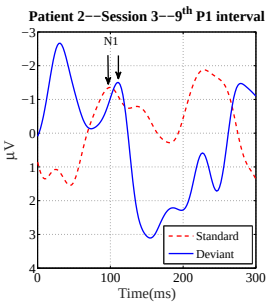


(d) Deviant– Patient 2, Session 3

Fig. 7: Similarities corresponding to the P1 intervals of Sessions 2 and 3 of patient 2. (a) and (c) correpond to standard stimuli, (b) and (d) correspond to deviant stimuli. P1 intervals of interest are indicated by arrows.



(a)



(b)

Fig. 6: Sub-session average signals corresponding to standard and deviant stimuli, at channel Fz, of patient 2 at P1 interval a) 5 of Session 2 and b) 9 of Session 3.

patient's brain response to the standard stimuli has a good similarity to those of the healthy subjects– this verifies the existence of the N1 component in the brain response to auditory stimuli (i.e. existence of the low level brain function). Fig. 5b demonstrates that, over P1 interval 13 of Session 4, there is no similarity between the patient's response to deviant stimuli and those of healthy subjects – this indicates that the patient did not have an MMN component within this P1 interval (waning). Furthermore, Fig. 5d demonstrates that P1 interval 10 of Session 5 shows a high similarity to the corresponding ERPs of healthy subjects – i.e. the average ERP corresponding to the deviant stimuli within this P1 interval shows an MMN component, suggesting the patient is likely to emerge from coma (which in fact he did). These results are consistent with the visual inspection results shown in Fig. 4.

The sub-session average for patient 2 corresponding to the standard and deviant stimuli over two typical P1 intervals of Sessions 2 and 3 are shown in Fig. 6, where visual inspection indicates that the N1 component appears in both these windows, but the MMN is manifest only in the 5th P1 interval of Session 2.

As in the case of patient 1, the similarity of the brain function of patient 2 over all P1 intervals of Sessions 2 and 3 to healthy brain function is shown in Fig. 7. As expected, all P1 intervals of Sessions 2 and 3 have relatively high similarity to the standard responses of the healthy subjects which indicates the presence of the component N1 over all P1 intervals. The high and low similarities respectively shown in Figs. 7b and 7d demonstrate that the patient's functionality was in his waxing phase in Session 2 and in his waning phase in Session 3. In addition, the 5th interval of Session 2 has a high similarity to both standard and deviant responses of the healthy subjects which indicates the presence of MMN component – this is consistent with the visual inspection result (see Fig. 6a). Furthermore, the 9th interval of Session 3 has both a high and low similarity to healthy subjects by virtue of exhibiting a similar response to standard stimuli (a clear N1) and a dissimilar response to deviant stimuli (no apparent MMN) – this result is also consistent with the ground truth obtained by visual inspection of the corresponding interval (see Fig. 6b) which again demonstrates the performance of the proposed method for automatic and continuous assessment of ERPs for the purpose of coma outcome prediction.

Note that both patients 1 and 2 demonstrate intervals where a MMN is present, indicating a higher level of func-

tionality over these periods. Since these patients emerged, there is hypothetical evidence to suggest that even short periods of higher functionality are predictive of a positive prognosis.

IV. Conclusions

In this paper we propose a machine learning based methodology for automatic and continuous assessment of ERPs for identifying the presence of the MMN component. The method consists of two phases: learning and testing. In the former phase, discriminative sub-spaces are trained using healthy training subjects. The trained subspaces are then used in the test phase for continuous assessment of a test subject. The performance of the proposed method is demonstrated on 22 healthy and 2 coma patients. The proposed method reduces load, cost and impracticality of requiring frequent expert assessment over the course of possibly many days.

Prognosis is evaluated using the novel approach of comparing the *similarity* of the coma patient's ERP responses to those of healthy subjects. High similarity values indicate emergence. The similarity measure can be established by training using only healthy subjects which greatly alleviates the training process.

The proposed method was evaluated using a Leave-One-subject-Out (LOO) procedure on the healthy training set and achieved an accuracy of 92.7%. Also, our method correctly predicted emergence for our 2 coma patients, whereas conventional methods using long averaging windows failed to identify MMN activity and thus were of low predictive value in identifying whether the patient would emerge or not.

The performance of the proposed machine learning method allows us to significantly reduce the analysis interval over which ERPs are averaged. The reduced window length reduces the smearing effect when ERP components are averaged together in the presence of latency jitter, and furthermore exposes the presence of waxing and waning cycles of the conscious state of coma patients. This fact allows us to hypothesize that the presence of even short waxing intervals is a salient indicator for emergence.

It is important to note that the results obtained using only 2 coma patients are preliminary and must be confirmed with a larger sample of data. Nevertheless, we consider our initial results to be very promising. We submit that the proposed method incorporates several novel advances in the study of disorders of consciousness, such as the fundamental idea of an automated machine learning paradigm for coma prognosis, the use of healthy subjects for training, and the verification of the existence of waxing and waning cycles of the state of consciousness.

Acknowledgment

The authors would like to thank Richard Mah, Dr. Alison Fox-Robichaud and Dr. Cindy Hamielec for their help in clinical experiments. The authors acknowledge the financial contributions of the Natural Science and Engineering Research Council of Canada (NSERC), and MITACS.

References

[1] G. B. Young, A. H. Ropper, and C. F. Bolton, "Coma and impaired consciousness: a clinical perspective," 1998.

[2] N. M. Kane, S. R. Butler, and T. Simpson, "Coma outcome prediction using event-related potentials: P3 and mismatch negativity," *Audiology and Neurotology*, vol. 5, no. 3-4, pp. 186–191, 2000.

[3] H. L. Lew, J. H. Poole, A. Castaneda, R. M. Salerno, and M. Gray, "Prognostic value of evoked and event-related potentials in moderate to severe brain injury," *The Journal of head trauma rehabilitation*, vol. 21, no. 4, pp. 350–360, 2006.

[4] C. C. Duncan, R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Näätänen, J. Polich, I. Reinvang, and C. Van Petten, "Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1883–1908, 2009.

[5] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, "The mismatch negativity (mmn) in basic research of central auditory processing: a review," *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2544–2590, 2007.

[6] R. Näätänen, A. W. Gaillard, and S. Mäntysalo, "Early selective-attention effect on evoked potential reinterpreted," *Acta psychologica*, vol. 42, no. 4, pp. 313–329, 1978.

[7] D. H. Loewy, K. B. Campbell, D. R. de Lugt, M. Elton, and A. Kok, "The mismatch negativity during natural sleep: intensity deviants," *Clinical Neurophysiology*, vol. 111, no. 5, pp. 863–872, 2000.

[8] M. Macdonald, P. Jamshidi, and K. Campbell, "Infrequent increases in stimulus intensity may interrupt central executive functioning during rapid eye movement sleep," *Neuroreport*, vol. 19, no. 3, pp. 309–313, 2008.

[9] L. D. Sculthorpe, D. R. Ouellet, and K. B. Campbell, "Mmn elicitation during natural sleep to violations of an auditory pattern," *Brain research*, vol. 1290, pp. 52–62, 2009.

[10] M. Strauss, J. D. Sitt, J.-R. King, M. Elbaz, L. Azizi, M. Buiatti, L. Naccache, V. van Wassenhove, and S. Dehaene, "Disruption of hierarchical predictive coding during sleep," *Proceedings of the National Academy of Sciences*, vol. 112, no. 11, pp. E1353–E1362, 2015.

[11] S. Blain-Moraes, R. Boshra, H. K. Ma, R. Mah, K. Ruiter, M. Avidan, J. F. Connolly, and G. A. Mashour, "Normal brain response to propofol in advance of recovery from unresponsive wakefulness syndrome," *Frontiers in human neuroscience*, vol. 10, 2016.

[12] W. Heinke, R. Kenntner, T. C. Gunter, D. Sammler, D. Olthoff, and S. Koelsch, "Sequential effects of increasing propofol sedation on frontal and temporal cortices as indexed by auditory event-related potentials," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 100, no. 3, pp. 617–625, 2004.

[13] P. Tavakoli, "How the detection of change of an acoustic stimulus can result in the interruption of the central executive controlling the involuntary capture of attention," Ph.D. dissertation, University of Ottawa, 2017.

[14] A. R. Dykstra, P. A. Cariani, and A. Gutschalk, "A roadmap for the study of conscious audition and its neural basis," *Phil. Trans. R. Soc. B*, vol. 372, no. 1714, p. 20160103, 2017.

[15] D. Morlet and C. Fischer, "MMN and novelty P3 in coma and other altered states of consciousness: A review," *Brain topography*, vol. 27, no. 4, pp. 467–479, 2014.

[16] C. Fischer, D. Morlet, P. Bouchet, J. Luaute, C. Jourdan, and F. Salord, "Mismatch negativity and late auditory evoked potentials in comatose patients," *Clinical neurophysiology*, vol. 110, no. 9, pp. 1601–1610, 1999.

[17] D. V. M. Bishop and M. J. Hardiman, "Measurement of mismatch negativity in individuals: A study using single-trial analysis," *Psychophysiology*, vol. 47, no. 4, pp. 697–705, 2010.

[18] N. M. T. Lapinskaya, "EEG Assessment of Disordered Consciousness: A Framework and a Case Study," Master's thesis, McMaster University, 2018.

[19] L. Naccache, J. Sitt, J.-R. King, B. Rohaut, F. Faugeras, S. Chennu, M. Strauss, M. Valente, D. Engemann, F. Raimondo *et al.*, "Reply: Replicability and impact of statistics in the detection of neural responses of consciousness," *Brain*, vol. 139, no. 6, pp. e31–e31, 2016.

[20] D. Gabriel, E. Muzard, J. Henriques, C. Mignot, L. Pazart, N. André-Obadia, J.-P. Ortega, and T. Moulin, "Replicability and impact of statistics in the detection of neural responses of consciousness," *Brain*, p. aww065, 2016.

[21] C. Fischer, F. Dailler, and D. Morlet, "Novelty P3 elicited by the subjects own name in comatose patients," *Clinical neurophysiology*, vol. 119, no. 10, pp. 2224–2230, 2008.

[22] C. Escera, E. Yago, M. D. Polo, and C. Grau, "The individual replicability of mismatch negativity at short and long inter-stimulus intervals," *Clinical Neurophysiology*, vol. 111, no. 3, pp. 546–551, 2000.

[23] N. Armanfard, M. Komeili, J. P. Reilly, and L. Pino, "Vigilance lapse identification using sparse EEG electrode arrays," in *Electrical and Computer Engineering (CCECE), 2016 IEEE 29th Canadian Conference on*. IEEE, 2016.

[24] N. Armanfard, M. Komeili, J. P. Reilly, R. Mah, and J. F. Connolly, "Automatic and continuous assessment of erps for mismatch negativity detection," in *2016 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2016.

[25] M. Ravan, G. Hasey, J. P. Reilly, D. MacCrimmon, and A. Khodayari-Rostamabad, "A machine learning approach using auditory odd-ball responses to investigate the effect of clozapine therapy," *Clinical Neurophysiology*, vol. 126, no. 4, pp. 721–730, 2015.

[26] F. Cong, T. Ristaniemi, and H. Lyytinen, *Advanced Signal Processing on Brain Event-related Potentials: Filtering ERPs in Time, Frequency and Space Domains Sequentially and Simultaneously*. World Scientific, 2015, vol. 13.

[27] A. Khodayari-Rostamabad, J. P. Reilly, G. M. Hasey, H. de Bruin, and D. J. MacCrimmon, "A machine learning approach using EEG data to predict response to ssri treatment for major depressive disorder," *Clinical Neurophysiology*, vol. 124, no. 10, pp. 1975–1985, 2013.

[28] M. Ravan, D. MacCrimmon, G. Hasey, J. P. Reilly, and A. Khodayari-Rostamabad, "A machine learning approach using P300 responses to investigate effect of clozapine therapy," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 5911–5914.

[29] D. W. Orme-Johnson and C. T. Haynes, "EEG phase coherence, pure consciousness, creativity, and TM–Sidhi experiences," *International Journal of Neuroscience*, vol. 13, no. 4, pp. 211–217, 1981.

[30] L. Torsvall *et al.*, "Sleepiness on the job: continuously measured EEG changes in train drivers," *Electroencephalography and clinical Neurophysiology*, vol. 66, no. 6, pp. 502–511, 1987.

[31] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 4–37, 2000.

[32] Y. Sun, S. Todorovic, and S. Goodison, "Local-learning-based feature selection for high-dimensional data analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1610–1626, 2010.

[33] Q. Cheng, H. Zhou, and J. Cheng, "The Fisher-Markov selector: fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 6, pp. 1217–1233, 2011.

[34] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Feature subset selection using differential evolution and a statistical repair mechanism," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11 515–11 526, 2011.

[35] Z. Zhu, Y.-S. Ong, and M. Dash, "Markov blanket-embedded genetic algorithm for gene selection," *Pattern Recognition*, vol. 40, no. 11, pp. 3236–3248, 2007.

[36] S. Sun, Q. Peng, and A. Shakoor, "A kernel-based multivariate feature selection method for microarray data classification," *PloS one*, vol. 9, no. 7, p. e102541, 2014.

[37] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2001.

[39] N. Armanfard, J. Reilly, and M. Komeili, "Local feature selection for data classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 6, pp. 1217–1227, 2016.

[40] ——, "Feature selection and classification based on local modeling of the sample space," *IEEE Trans. Neural Networks and Learning Systems*, 2017.

[41] L. Wang, "Feature selection with kernel class separability," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 9, pp. 1534–1546, 2008.

[42] R. Gilad-Bachrach, A. Navot, and N. Tishby, "Margin based feature selection-theory and algorithms," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 43.

[43] C. Jones, "Glasgow coma scale." *AJN The American Journal of Nursing*, vol. 79, no. 9, pp. 1551–1557, 1979.

[44] A. Marmarou, J. Lu, I. Butcher, G. S. McHugh, G. D. Murray, E. W. Steyerberg, N. A. Mushkudiani, S. Choi, and A. I. Maas, "Prognostic value of the Glasgow Coma Scale and pupil reactivity in traumatic brain injury assessed pre-hospital and on enrollment: an impact analysis," *Journal of neurotrauma*, vol. 24, no. 2, pp. 270–280, 2007.

[45] N. A. Mushkudiani, C. W. Hukkelhoven, A. V. Hernández, G. D. Murray, S. C. Choi, A. I. Maas, and E. W. Steyerberg, "A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes," *Journal of clinical epidemiology*, vol. 61, no. 4, pp. 331–343, 2008.

[46] Y. Lee, T. Phan, D. Jolley, H. Castley, D. Ingram, and D. Reutens, "Accuracy of clinical signs, SEP, and EEG in predicting outcome of hypoxic coma a meta-analysis," *Neurology*, vol. 74, no. 7, pp. 572–580, 2010.

[47] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness: a practical scale," *The Lancet*, vol. 304, no. 7872, pp. 81–84, 1974.

[48] I. Holeckova, C. Fischer, M.-H. Giard, C. Delpuech, and D. Morlet, "Brain responses to a subject's own name uttered by a familiar voice," *Brain research*, vol. 1082, no. 1, pp. 142–152, 2006.