Recurrence analysis of the EEG during sleep accurately identifies subjects with mental health symptoms

David E. McCarty a, Naresh M. Punjabi b, Paul Y. Kim a, Clifton Frilot II c, Andrew A. Marino a,∗

a Division of Sleep Medicine, Department of Neurology, LSU Health Sciences Center, Shreveport, LA, USA
b Department of Pulmonary & Critical Care Medicine, Johns Hopkins Medicine, Baltimore, MD, USA
c School of Allied Health Professions, LSU Health Sciences Center, Shreveport, LA, USA

ABSTRACT

Analysis of brain recurrence (ABR) is a novel computational method that uses two variables for sleep depth and two for sleep fragmentation to quantify temporal changes in non-random brain electrical activity. We postulated that ABR of the sleep-staged EEG could identify an EEG signature specific for the presence of mental health symptoms. Using the Mental Health Inventory Questionnaire (MHI-5) as ground truth, psychological distress was assessed in a study cohort obtained from the Sleep Heart Health Study. Subjects with MHI-5 < 50 (N = 34) were matched for sex, BMI, age, and race with 34 subjects who had MHI-5 scores > 50. Sixteen ABR markers derived from the EEG were analyzed using linear discriminant analysis to identify marker combinations that reliably classified individual subjects. A biomarker function computed from 12 of the markers accurately classified the subjects based on their MHI-5 scores (AUROC = 82%). Use of additional markers did not improve classification accuracy. Subgroup analysis (20 highest and 20 lowest MHI-5 scores) improved classification accuracy (AUROC = 89%). Biomarker values for individual subjects were significantly correlated with MHI-5 score (r = 0.36, 0.54 for N = 68, 40, respectively). ABR of EEGs obtained during sleep successfully classified subjects with regard to the severity of mental health symptoms, indicating that mood systems were reflected in brain electrical activity.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The relation between psychological distress and the pattern of the electroencephalogram (EEG) recorded from distressed subjects has been studied since the discovery of the EEG (Lemere, 1936). In major depressive disorder (MDD), for example, many attempts were made to identify visual features, spectral characteristics, or other linear properties of the signal that would allow identification of risk, confirm diagnosis, permit monitoring of the effect of treatment, and/or predict treatment response (Olbrich and Arns, 2013). Changes in absolute or relative alpha power were probably the most frequently identified variables associated with MDD, but not with sufficient consistency to warrant clinical application (Knot and Lapierre, 1987; Pozzi et al., 1995; Grin-Yatsenko et al., 2009; Jaworska et al., 2012).

Various methods based on analysis of the nonlinear dynamical complexity in the EEG were proposed for studying mental disorders (Bystritsky et al., 2012). Within the limitations of this perspective (Rapp, 1994), various approaches were developed to distinguish between the presence and absence of MDD (Olbrich and Arns, 2013) and to predict treatment efficacy (Arns et al., 2014). Similar observations were reported for other mental disorders including schizophrenia (Paulus and Braff, 2003) and autism (Rosl et al., 2011).

Analysis of brain recurrence (ABR) is a computational method designed to detect and quantify deterministic temporal patterns in the EEG (non-random brain activity) not detectable by conventional EEG methods such as pattern-recognition or spectral analysis (Carrubba et al., 2012a). ABR was used to study a range of problems in basic and clinical neuroscience (Frilot et al., 2014). Patients with multiple sclerosis were identified using ABR (Carrubba et al., 2010; Carrubba et al., 2012b), and it was used to create a novel paradigm in which the concepts of sleep depth and variability could be quantified (Carrubba et al., 2012a; Wang et al., 2013). Employing markers based on these variables, patients with mild or moderate obstructive sleep apnea were distinguished using the sleep-staged EEG from a single derivation (Wang et al., 2013), illustrating the concept that a complex physiologic disorder leaves an objectively discernible and specific footprint on brain electrical activity.

We became interested in whether the sleep-acquired EEG could similarly be used to classify subjects with psychological distress. Our ultimate goal was to develop objective analytical methods to help in the diagnosis and classification of subjects with neurocognitive

http://dx.doi.org/10.1016/j.pscychresns.2014.10.004
disorders. In the present study we tested the hypothesis that sleep depth and fragmentation markers extracted from the staged, sleep-acquired EEG could be employed to accurately assign subjects into classes with higher or lower levels of distress, using scores from the Mental Health Inventory questionnaire (MHI-5) as ground truth. If the subjects could be correctly classified, we planned to interpret the result as an indication that psychological distress was objectively associated with a specific type of algorithmically-determinable change in the sleep EEG.

2. Methods

2.1. Patients

Fig. 1 shows the basic stages of the analysis. The study cohort was chosen from the 6441 participants in the Sleep Heart Health Study (SHHS), a multi-center study sponsored by the National Heart, Lung and Blood Institute and conducted in ten U.S. communities to determine the cardiovascular and other consequences of sleep-disordered breathing (Quan et al., 1997). All SHHS participants underwent overnight polysomnography (PSG) between 1995 and 1998. The present investigation took advantage of the EEG data in the PSG, the scores from the MHI-5, and relevant covariate information collected during the baseline SHHS examinations.

The cohort studied was chosen from SHHS participants for whom mental health status, age, gender, body mass index (BMI), and race were ascertained within 1 year of polysomnography. Participants with sleep apnea, type 2 diabetes, stroke, myocardial infarction, angina, heart failure, coronary angioplasty, or coronary artery bypass graft surgery were excluded. We arbitrarily regarded a scaled MHI-5 score greater than 50. When there was more than one possible choice, the subject was chosen randomly. The two sub-cohorts were well matched on all pertinent characteristics except for MHI-5 score (the variable used to define the sub-cohorts) (Table 1). The low RDI scores indicated that the subjects did not have sleep apnea. We arbitrarily chose subcohorts of 34 subjects, which was about half of the subjects available in the less-than-50 group. All research-related procedures were approved by the institutional review boards for human research at the institutions where the data was collected.

![Fig. 1. Experimental design.](image)

Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>MHI-5 &lt; 50</th>
<th>MHI-5 &gt; 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Race (C/non-C)</td>
<td>29/5</td>
<td>29/5</td>
</tr>
<tr>
<td>Age (years)</td>
<td>58.6 ± 2.3</td>
<td>58.7 ± 2.3</td>
</tr>
<tr>
<td>Male/Female</td>
<td>12/22</td>
<td>12/22</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.6 ± 0.6</td>
<td>25.9 ± 0.6</td>
</tr>
<tr>
<td>RDI</td>
<td>1.6 ± 0.2</td>
<td>1.3 ± 0.2</td>
</tr>
<tr>
<td>MHI-5</td>
<td>40.5 ± 1.4</td>
<td>78.4 ± 2.1</td>
</tr>
</tbody>
</table>

2.2. Measure of mental health

The MHI-5 screening instrument asked “How much of the time during the last month have you: (1) been a very nervous person; (2) felt calm and peaceful; (3) felt downhearted and blue; (4) been a happy person; and (5) felt so down in the dumps that nothing could cheer you up?” Each answer was scored 1–6 (subject range 5–30), with higher scores indicating better mental health. For analysis, the total score was linearly transformed into a variable with a range from 0 to 100. In a population-based sample, the overall accuracy of the MHI-5 in identifying mood disorders is 88%, using a cut-off of 60 points or less (Rumph et al., 2001). In a population of patients with HIV, the best cut-off score for major depression (84%) was 52 (Holmes, 1998).

2.3. Polysomnograms

PSGs were recorded using the Compumedics P Series system (Abbotts Ford, Victoria, Australia) (Quan et al., 1997) and were obtained from the SHHS database (National Heart Lung & Blood Institute, 2012). Details regarding the recording procedures were described elsewhere (Redline et al., 1998). Each PSG was about eight hours in duration and had been divided by the original SHHS investigators into 30-second epochs and classified into one of five mutually exclusive stages, four stages of sleep (REM, N1, N2, N3) or the stage of wake after sleep onset (WASO). The PSGs contained EEGs recorded from C3–M2 and C4–M1, sampled at 125 Hz, and were provided as MAT files. For analysis, the EEGs were interpolated to 500 Hz (our laboratory standard sampling frequency for the EEG) using a standard algorithm (Matlab, Mathworks, Natick, MA, USA), filtered using an FFT digital filter to pass 0.5–35 Hz, and evaluated by means of custom codes in a standard numerical computing environment (Matlab). We chose 35 Hz as the cut-off because we had previously determined that it produced the best balance between capturing essentially all the energy in the EEG while providing good protection against common artifacts that appear in the sleep EEG above 35 Hz. The EEGs from both derivations were analyzed, but only the results from C3–M2 are presented here because the C4–M1 results were essentially identical.

2.4. Analysis of brain recurrence

Analysis of brain recurrence (ABR) is a nonlinear technique for extracting information from the EEG. The basic idea is that although brain electrical activity appears irregular, it actually exhibits recurrent patterns that can be detected and quantified, thereby permitting evaluation of the relation between the recurrences and behavioral or clinical observations. ABR is based on the conjecture that brain function is mediated by electrical activity in localized neuronal networks and their inter-network electrical synchronization (Carabba et al., 2012a). In this perspective, an EEG from any derivation is regarded as a delocalized measure of the amount of law-governed activity in the EEG (as opposed to random) behavior contained in the brain. The amount of such activity is quantified using ABR. The basic signal-processing techniques and their applicability to model-independent analyses of nonstationary signals like EEGs were previously described (Zbilut and Webber, 2006). Briefly, 5-component vectors (points in a five-dimensional mathematical hyperspace) were formed that consisted of the EEG amplitude at t and four earlier times identified by four successive lags of five points (10 ms). The sequence in hyperspace of all such vectors obtainable from one second of the EEG (480 vectors, given our choices of sampling rate, vector dimension, and delay time) was interpreted as a path that was determined by law-governed (non-random) activity in the brain. The amount of such activity was quantified using the variables percent recurrence (r), defined as the percent of the 480 vectors in the path that were near other vectors (and hence were recurrent), and percent determinism (d), defined as the percent of the recurrent points that were adjacent to at least one other recurrent point. Detailed analysis of these variables provides a theoretical rationale for why they quantify the amount of law-governed activity in the EEG (Zbilut and Webber, 2006; Frilot...
et al., 2014). The Euclidean norm was used for measuring distance, and vectors were identified as near if they were within 15% of the distance between the two vectors that were furthest apart (threshold radius). These choices as well as those for the delay time (10 ms), embedding dimension (five), and threshold radius (15%) were identified on the basis of previous experience and found to be useful for analyzing the vigilant and sleep EEG (Carrubba et al., 2012a; Friloi et al., 2014).

Both r and d were computed for each second of the EEG, resulting in approximately 60 s × 60 mm × 8 h = 28,800 values for a typical eight-hour overnight EEG. The continuous series of those values, r(t) and d(t), were interpreted as independent measures of sleep depth whereas lower values were associated with wake or light sleep and higher values with deeper sleep (Carrubba et al., 2012a). For statistical evaluations, in each subject the variables were averaged epoch-by-epoch and then stage-by-stage over the entire overnight sleep study, resulting in 8 markers for sleep depth (Fig. 1).

Markers for fragmentation in sleep depth were created by generalizing the conventional definition of EEG arousals (American Academy of Sleep Medicine, 2007). For both r(t) and d(t), the ratio of the mean over 3 s (one value per second) to the mean of the preceding 10 s (ten values) was determined, and the process was repeated using successive steps of 3 s, resulting in a time series of approximately 9000 ratios for an overnight EEG. Whenever the ratio increased by more than 100% for r(t) or 50% for d(t) the change was counted as an arousal, and the hourly rate of arousals, termed the generalized arousal index (GAI) was determined for each stage.

The procedure was performed using r(t) and d(t) which, after averaging, resulted in 8 GAI markers for sleep fragmentation (4 from r(t) and 4 from d(t)). N1 epochs occurred only rarely (<3% of the ~57,000 epochs in the study), consequently they were combined with N2 for discriminant analysis and descriptive statistics.

2.5. Discriminant analysis

Fisher’s linear discriminant analysis was used to determine the coefficients of biomarker functions that combined the sleep markers in the way that best separated the subjects into groups with MHI-5 scores that were below or above 50 (Theodoridis and Koutrombas, 2008). Classification accuracy was assessed using area under the receiver operating characteristics curve (AUROC) (Matlab) (Theodoridis and Koutrombas, 2008).

The reliability of the biomarker function was assessed by means of a 10-fold cross validation process in which a biomarker function was determined using 90% of the subject’s data and then used to classify the remaining 10% (validation process). For both r(t) and d(t), the ratio of the mean over 3 s (one value per second) to the mean of the preceding 10 s (ten values) was determined, and the process was repeated using successive steps of 3 s, resulting in a time series of approximately 9000 ratios for an overnight EEG. Whenever the ratio increased by more than 100% for r(t) or 50% for d(t) the change was counted as an arousal, and the hourly rate of arousals, termed the generalized arousal index (GAI) was determined for each stage.

The procedure was performed using r(t) and d(t) which, after averaging, resulted in 8 GAI markers for sleep fragmentation (4 from r(t) and 4 from d(t)). N1 epochs occurred only rarely (<3% of the ~57,000 epochs in the study), consequently they were combined with N2 for discriminant analysis and descriptive statistics.

2.6. Discriminant analysis

Fisher’s linear discriminant analysis was used to determine the coefficients of biomarker functions that combined the sleep markers in the way that best separated the subjects into groups with MHI-5 scores that were below or above 50 (Theodoridis and Koutrombas, 2008). Classification accuracy was assessed using area under the receiver operating characteristics curve (AUROC) (Matlab) (Theodoridis and Koutrombas, 2008).

The reliability of the biomarker function was assessed by means of a 10-fold cross validation process in which a biomarker function was determined using 90% of the subject’s data and then used to classify the remaining 10% (validation process). For both r(t) and d(t), the ratio of the mean over 3 s (one value per second) to the mean of the preceding 10 s (ten values) was determined, and the process was repeated using successive steps of 3 s, resulting in a time series of approximately 9000 ratios for an overnight EEG. Whenever the ratio increased by more than 100% for r(t) or 50% for d(t) the change was counted as an arousal, and the hourly rate of arousals, termed the generalized arousal index (GAI) was determined for each stage.

The procedure was performed using r(t) and d(t) which, after averaging, resulted in 8 GAI markers for sleep fragmentation (4 from r(t) and 4 from d(t)). N1 epochs occurred only rarely (<3% of the ~57,000 epochs in the study), consequently they were combined with N2 for discriminant analysis and descriptive statistics.

3. Results

3.1. Cohort characteristics

Percent recurrence (r) and percent determinism (d) (independent measures of sleep depth) (Wang et al., 2013) exhibited ultradian rhythms of 2–5 cycles (Fig. 2a). As expected, sleep-stage-specific mean values were typically lowest during WASO and progressively greater during the NREM stages, with REM between N1 and N2. The rates of generalized arousals (measures of sleep fragmentation) also depended on sleep stage (Fig. 2b). Arousal markers were most likely during WASO, least likely during N3, and had intermediate probabilities during the other sleep stages.

3.2. Subject-level characteristics

The hypothesis that the stage-specific markers could be combined statistically to classify individual subjects was tested using discriminant analysis of 16 ABR markers extracted from the C3 EEG and AUROC to assess classification accuracy. Initially all possible combinations of 2 markers were considered, but all resulting AUROC values were <70%. When all possible combinations of 3–12 markers were systematically evaluated, progressively greater AUROC values were obtained up to 82%, which occurred for one of the 1820 possible combinations of 12 markers. All combinations of 13–16 markers yielded an AUROC ≤82%, and were therefore not considered further.

The best AUROC results for combinations of 4, 6, 8, 10, and 12 markers are shown in Table 2. For the cases of 4 and 6 markers, only markers for sleep fragmentation were identified by the discriminant analysis as necessary for optimal subject classification. For cases where more than 6 markers were considered, the discriminant algorithm also incorporated sleep-depth markers for optimal results (Table 2). When discriminant analyses were performed using only one stage of sleep, the respective AUROC results using only WASO, N1/N2, N3, or REM were 66%, 58%, 64%, and 58%, indicating that multi-stage markers were needed for accurate classification.

To evaluate the possibility that the AUROC values were due to chance rather than subject classification based on MHI-5 scores, surrogate sampling distributions were created by randomizing the higher/lower classification of each subject and calculating the resulting AUROC. A total of 104 randomizations were performed, and in each case (Table 2) AUROC was significant (P < 0.05).

The likelihood that the results (Table 2) would generalize to other samples of subjects was considered by means of cross-validation analyses. The resulting AUROC values were 70–72%, and the correlation coefficients for the association between the value of the biomarker function and the MHI-5 score was 0.23–0.36, depending on the number of markers used in the analyses.

To assess the effect of the sub-cohort distributions of MHI-5 scores on the results for accuracy and biomarker correlation, the sub-cohorts were restricted to subjects with the 20 highest and the 20 lowest MHI-5 scores. Both AUROC and the correlation coefficient increased notably in the partial sub-cohorts (Fig. 3).

3.3. Sub-cohort comparisons

Mean normalized (with respect to WASO) stage-specific sleep depth and general arousal indices did not differ between the sub-cohorts (Figs. 4 and 5).
4. Discussion

When applied to the sleep-acquired, staged EEG, ABR yielded delocalized variables that tracked traditional stage-related concepts of sleep depth and fragmentation. We postulated that ABR-derived markers could be combined statistically to identify a signature specific for the presence of mental health symptoms. The assumption was tested by ascertaining whether subjects could be reliably classified according to symptom severity, as assessed using MHI-5 scores.

Using 12 ABR markers, the specific identity of which was determined from an a priori group of 16 markers by means of discriminant analysis, we classified 68 subjects into either MHI-5 < 50 or MHI-5 > 50 groups, with an AUROC of 82% (Table 2), corresponding to a sensitivity and specificity of 79% and 77%, respectively (Fig. 3a). The significance of the AUROC (P < 0.05) was established by means of a standard surrogate analysis. The biomarker function had a dose-dependent relationship with symptom severity (Figs. 3a, r = 0.36, P < 0.05). Moreover, when subjects with mid-range MHI-5 scores were removed, subjects with intact mental health could be differentiated from those with more severe affective symptoms with even more precision (Fig. 3b, AUROC = 89%, r = 0.54, P < 0.05). We interpret the overall results to mean that symptoms of psychological distress could be objectively detected in the EEG using sleep-stage-specific recurrence markers from each subject.

Mean normalized (with respect to WASO) stage-specific sleep depth was marginally greater in all sleep stages in the sub-cohort with lower MHI-5 scores, but none of the individual comparisons were pair-wise significant (P > 0.05) (Fig. 4). Similarly, mean stage-specific general arousal indices differed insignificantly between the sub-cohorts (Fig. 5). Both results were expected because no individual marker was sufficiently impacted by the presence of mental disorder to permit a statistically determinable impact on the group mean. Discriminant analysis, in contrast, focused on

### Table 2

<table>
<thead>
<tr>
<th>N</th>
<th>Highest AUROC (%)</th>
<th>GAL in WASO</th>
<th>GAL in N1/REM</th>
<th>GAL in N1/N2</th>
<th>GAL in W2S</th>
<th>r in N1/N2</th>
<th>d in N3</th>
<th>d in WASO</th>
<th>d in REM</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 (1820)</td>
<td>74</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6 (8008)</td>
<td>77</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8 (12870)</td>
<td>79</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10 (8008)</td>
<td>81</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>12 (1820)</td>
<td>82</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Fig. 3. AUROC analyses and biomarker-versus-MHI-5 scatterplots. (a) N = 34 in each sub-cohort. (b) Restricted sub-cohorts chosen by including only the subjects with the 20 highest MHI-5 values in the MHI-5 > 50 sub-cohort and the subjects with the 20 lowest MHI-5 values in the MHI-5 < 50 sub-cohort. Dotted line indicates thresholds for classification into higher or lower groups, 0.44 and 0.53 for (a) and (b), respectively.
assessing an impact at the individual level and achieved this objective by integrating many relatively small impacts of mental disorder on different markers.

There are many reports of a relationship between nonlinear dynamics and health. In general, increased order (thought to impair the body’s adaptability) is associated with disease (Goldberger et al., 2002). Abnormal dynamics have previously been reported in connection with mental disorders, including depression (Gottschalk et al., 1995; Linkenkaer-Hansen et al., 2005; Carlino et al., 2012). Our method extends these ideas to the level of the individual subject.

Limitations of the study include the retrospective design, the particular choices of delay time and other ABR parameters, and the relative crudity of the neurobehavioral assessment instrument. Archived retrospective data inherently contains some unintended heterogeneity in the study cohort, which limits the applicability of the results. To fully understand what ABR measures (and how it is captured by differing choices of ABR parameters), the neural activity to which it is related must be understood. We did not address that issue. Our aim was to show that ABR reliably classified subjects based on the MHI-5, the implication of which would be that ABR actually captured meaningful neural activity. In addition, the non-specificity of the mental disorder reflected in the MHI-5 is well recognized. Our intention was to classify subjects as either healthy or with any kind of mental disorder. A more specific assessment tool, for depression as an example, would probably allow refinement of the biomarker approach, resulting in better classification precision.

Negative affective symptoms accompany numerous pathologies, including sleep deprivation (Mustahsan et al., 2013), substance abuse (Gleason et al., 2013), the psychological burden of caring for a sick child (Gallagher and Hannigan, 2013), chronic pain (Iliffe et al., 2009), heart disease (Garfield et al., 2014), and primary mental illnesses. We did not address the underlying mechanism for such negative symptoms, but instead focused on whether they could be reliably identified by distinguishing those with poor self-reported mental health from those without such problems, using only one sleep-acquired EEG signal. Our results were a proof-of-concept, suggesting that ABR of the sleep-acquired, staged EEG is a practical and incrementally valuable data source for identifying the presence of impaired mental health. Diagnostic and prognostic information about mental illness may be hiding in plain sight.

5. Conclusion

Based on algorithmic analyses of single-derivation staged, sleep-acquired EEGs, subjects in a population-based sample of adults could be correctly classified at a respectable accuracy level regarding the presence or absence of negative neurobehavioral symptoms, using the MHI-5 score as ground truth.

References


Hoboken, pp. 2979–2986.


