relationships of the powers of different EEG frequencies in a single understandable index. The scale ranges from 0 - 2.5 with values of zero only occurring during sleep and a value of 2.5 only occurring during wakefulness. A value of 1.25 predicts an equal probability of being awake or asleep. The current investigation is undertaken to obtain normal values of non-REM ORP in good sleepers in their home environments and to establish the repeatability of this measure from night to night.

Methods: Good sleepers (GS) were recruited and defined as non-shift workers sleeping 7–9 hours per night who considered their sleep of good quality and didn't have daytime symptoms attributed to their sleep. Acceptable participants couldn't be habitual nappers, have chronic health or sleep diagnoses, use psychotropic or recreational drugs, or have reached the menopause. ESS, ISI and FOSQ 10 had to be in the normal range. OSA was excluded based on an HSAT measuring EEG (Prodigy). A second EEG study was performed one week later.

Results: Thirty-three people completed the study (9 (20–35) females, 8 (20–35) males, 8 (35–50) males and 8 (35–50) females. The mean ORP for non-REM sleep during both nights of the study was 0.54 ± 0.17 (66 nights). Excluding three outlier nights, results in a tighter distribution of 0.51 ± 0.13. Distribution fitting suggests a normal distribution (Kolmogorov-Smirnov test, p=0.254). No difference was noted in non-REM ORP between first and second nights (0.55 ± 0.17 vs 0.53 ± 0.16). The study was powered to find a difference of 0.08 units between first and second nights with an α error (0.05) and β error (0.2).

Conclusion: GS normal values for non-REM ORP is ≤0.76 units. Repeatability of ORP suggests no systematic second night effect. There is considerable variability in non-REM ORP between nights.

Support (If Any): Financial and in-kind support was provided by Cerebra Health Inc.

0318

COMPARISON ON FEATURE EXTRACTION METHODOLOGIES FOR SLEEPINESS DETECTION USING ELECTROENCEPHALOGRAPHY

Park J[1], Choi H[2], Yoon J[1], Yoon S[2], Yun C[1]

[1]Seoul National University Bundang Hospital, Seong Nam, KOREA, REPUBLIC OF, [2]Department of Electrical and Computer Engineering, Seoul National University, Seoul, KOREA, REPUBLIC OF

Introduction: Sleepiness from chronic sleep deprivation and sleep disorders, is a major risk factor for traffic accidents and occupational injuries, and a main determinant of on-the-job performance. The aim of this study was to compare feature extraction methodologies for the detection of sleepiness using electroencephalography (EEG).

Methods: Eight healthy subjects (age 20 - 30 yeas old; 4 male) free from sleep disorders and major medical conditions were included in this study. The subjects underwent 10 hours of EEG monitoring during daytime in two conditions, sleep sufficient and acutely sleep deprived conditions (<4 hours of night sleep). During EEG recording, the objective sleepiness was measured using psychomotor vigilance task at every 2-hour. Four feature extraction methods were used for the analysis of EEG; band pass (BP), power spectrum density (PSD), multitaper spectrum density (MPSD), wavelet packet decomposition (WPD). In order to compare the usefulness of the four techniques, we used Support Vector Machine (SVM) and Random Forest (RF) to measure supervised classification performance. To evaluate the generalization ability of each classifier, a 10-fold cross validation has been applied. Additionally, the four feature extraction methods were compared with visualization using the t-Distributed Stochastic Neighbor Embedding (t-SNE).

Results: In comparing the classification performance of the four feature extraction methods, MPSD showed the best classification performance with the highest AUC value in both SVM and RF. WPD and BP were quite similar in performance. PSD scored the lowest AUC value in both SVM and RF. In analysis with visualization using t-SNE, BP, WPD and MPSD showed markedly separated values between sleepy and awake status, but PSD did not.

Conclusion: The findings suggest that MPSD might be useful feature extraction methodology for the detection of sleepiness using EEG.

*JP and HC contribute equally to this work.

Support (If Any):

0319

A SIMPLIFIED BIPOLAR FRONTAL MONTAGE FOR RECORDING AND STAGING SLEEP.

Punjabi NM[1], Kaplan PW[1], Margolick J[1], Aurora RN[1]

[1]Johns Hopkins University, Baltimore, MD

Introduction: The montage recommended by the American Academy of Sleep Medicine (AASM) for recording sleep consists of referential derivations, with input electrodes over the frontal, central, and occipital regions. Monitors that only use frontal electrodes are becoming increasingly used due to their sophistication, simplicity of use, and lower susceptibility for contamination from pulse and sweat artifact. The overall objective of the present study was to compare the visually scored data derived from a bipolar frontal montage and the conventional montage as recommended by the AASM.

Methods: Concurrent recordings of sleep were obtained using two Nox-A1 recording systems (Nox Medical) with independent electrode derivations. Both monitors were synchronized for simultaneous data collection. The first recorder was configured with the conventional AASM montage (F4-M1, C4, M1, and O2-M1). The second recorder was configured using a bipolar frontal montage (F4-E3, F4-F7, F3-F8, F3-E2, and Fpz-E3E4). Both recordings were scored for wake, NREM (1, 2, and 3) and REM sleep using the current AASM guidelines. Overall percent agreement for all sleep stages and percent agreement stratified by stage were determined at the epoch level.

Results: A total of 3461 recorded epochs derived from each monitor were scored in a blinded fashion. The distribution of sleep stages based on the conventional vs. bipolar montage was as follows: wake (7.9% vs. 7.6%), N1 (6.1% vs. 7.1%), N2 (41.1% vs. 40.8%), N3 (23.8% vs. 23.6%), and REM sleep (21.0% vs 20.8%). Overall percent agreement across all stages was 88.7% (95% CI: 0.88–0.90). Percent agreement by stage (using the conventional montage as the reference) was as follows: wake (90.5%; 95% CI: 0.87–0.94), N1 (77.8%; 95% CI: 0.72–0.83), N2 (88.8%; 95% CI: 0.87–0.90), N3 (87.7%; 95% CI: 0.85–0.90), and REM sleep (92.0%; 95% CI: 0.90–0.94).

Conclusion: A frontal bipolar montage for recording and staging can sleep provide a high level of agreement for all sleep stages when compared to the conventional montage currently recommended by the AASM.

Support (If Any):

0320

A CROSS-VALIDATION APPROACH TO INTER-SCORER RELIABILITY ASSESSMENT

Fernandez C[1,2], Rusk S[1,2], Glattard N[1,2], Hensen B[2], Shokoueinejad M[3], Creado S[2], Hungerford J[4]

[1]Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, [2]EnsoData Research, EnsoData, Inc.,

Madison, WI, [3]Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, [4]Department of Pediatrics Sleep Medicine, University of Arkansas for Medical Sciences, Little Rock, AR

Introduction: Inter-scorer variability is a challenge in sleep medicine. The inter-scorer reliability (ISR) program aids clinicians in assessment of inter-scorer variability and accreditation. Leave-one-out-cross-validation (LOOCV) is a powerful technique borrowed from machine learning for evaluating how well a statistical analysis will generalize to an independent dataset. In the present study, we adapt the LOOCV approach to ISR analysis, proposing a novel application of the methodology, whereby we introduce the concept of overfitting in the sleep scoring context and characterize its impact on ISR assessment reproducibility.
Methods: A cohort (N=72) was selected using stratified sampling with proportionate allocation to control for sleep apnea severity, medical conditions, medications, and demographic factors. The cohort was scored by four independent sleep technologists (RPSGT). ISR was assessed using epoch-by-epoch agreement for sleep stages, respiratory, arousal, and movement events under two LOOCV settings. First, average agreement was calculated with each clinician as the designated reference scorer (DRS) compared to each of the three “held out” clinicians. Second, average agreement was calculated by constructing a DRS based on events that a 2/3 majority of clinicians agree with compared to the fourth “held out” clinician.
Results: Across all event types, 42–60% of all event-epochs were marked with the presence of an event by only one of four clinicians. All four clinicians agreed on 6–14% of all event-epochs evaluated. No statistically significant differences were observed in the percentage of event-epochs marked by 2/3 majority versus the percentage of event-epochs marked by clinicians individually. Comparatively, the observed agreement estimates were greater for all event types in the 2/3 majority setting than the individual setting.
Conclusion: Cross-validation presents an opportunity to improve the generalization of agreement estimates in ISR assessments. This work demonstrates that consensus-based DRS’s may be constructed and used for ISR assessments. Given the substantial percentage of epochs that were marked by a single clinician, utilizing a consensus-based reference can serve to regularize overfit scenarios where inter-scorer variability would be amplified by artifacts in an individual’s scoring. Therefore, cross-validation approaches may enable measurement of scoring agreement with greater reproducibility.
Support (If Any):

0321
WEARABLE DEVICE ECG AND G-SENSOR-BASED SLEEP STAGE EVALUATION USING PSG-BASED LEARNING SIGNAL

lin Y[1], Wang P[2,3], Lin C[2], Sadrawi M[4], Lin C[5], Hsieh Y[1], Kuo C[1], Chien J[5], Haraikawa K[5], Abbod MF[6], Shieh J[4]
[1]Healthcare and Beauty RD Center,Cal-Comp Inc., New Taipei City, TAIWAN, [2]Thoracic Division, Department of Internal Medicine, Far Eastern Memorial Hospital., New Taipei City, TAIWAN, [3]Oriental Institute of Technology, New Taipei City, TAIWAN, [4]Department of Mechanical Engineering and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, TAIWAN, [5]Healthcare and Beauty RD Center,Kinpo Electronics Inc., New Taipei City, TAIWAN, [6]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, UNITED KINGDOM

Introduction: Sleep is essential for the human life. Insufficient sleep has negative effect on the cardiac system. Furthermore, a study revealed the correlation between sleep and diabetes. Meanwhile, sleep is also related memory loss. Autonomic nervous system (ANS) is an advanced part of the sleep quality measurement. Heart rate variability (HRV); collected through electrocardiography (ECG); is one of the substantial options for ANS. In order to have a better understanding of sleep behavior, this study aims to classify sleep stages by utilizing features extracted from ECG and G-Sensor signals utilizing wearable device that is based on PSG as the reference.
Methods: The dataset was collected from 24 patients during whole sleep time. The PSG machine and single lead BC1 ECG with the 6-axis G-Sensor device (Bio Clothing One, XYZ life BC1, Kinpo Inc., Taipei, Taiwan) are simultaneously utilized for every patient to collect the data. Finally, the ECG and G-Sensor data from the BC1 ECG device are analyzed to evaluate the sleep stage with a reference generated by the PSG. Features extracted from the raw ECG and G-Sensor signals each 30 seconds with a 5-minute sliding window are used in the evaluation. In addition, R-R detection evaluation is performed for the HRV evaluation. Hence, a combination of RR-interval, ECG-derived respiration (EDR) and R-R amplitude differences from ECG with addition to features from the 6-axis G-Sensor are used as the ANN inputs. The output is the sleep stage level.
Results: The evaluation of HRV from ECG with G-Sensor signal provides acceptable results. Data from 18 and 6 patients are used for training and testing of the ANN respectively. Results show the best achieved accuracy using 10-fold cross validation ANN are 76.68% and 75.38% for training and testing respectively.
Conclusion: Sleep evaluation utilizing wearable device ECG and G-Sensor signals can be applied to train ANN with PSG signal as reference. However, this investigation needs further advanced evaluations for comparison to sleep study results utilizing EEG signal as the input evaluation signal.
Support (If Any): Kinpo Electronics, Inc.

0322
FIELD-BASED SLEEP MEASUREMENT: CONCORDANCE BETWEEN COMMERCIAL ACTIVITY MONITORS AND AN ACTIGRAPH

Kubala AG, Santos EC, Barone Gibbs B, Buysse DJ, Patel SR, Hall MH, Kline CE
University of Pittsburgh, Pittsburgh, PA

Introduction: Many commercial activity monitors (CAM) now measure sleep-wake patterns. However, it is unknown whether these CAMs provide sleep data that are similar to research-grade actigraphy. The purpose of this study was to examine the concordance between multiple CAMs and a validated research-grade actigraph.
Methods: On the same wrist, 28 healthy adults (50% female, 25.0 ± 4.2 y) wore an Actiwatch Spectrum (AW) and alternated wearing 6 CAMs for one 24-h period (Samsung Gear Fit2, Fitbit Alta, Polar A360, Jawbone Up3, Xiaomi Mi Band 2, Mistfit Shine 2). A daily sleep diary was also completed. Sample sizes for comparisons ranged from 18–27 due to CAM device error. Unedited AW (default setting) and CAM outputs were used for comparisons. Comparison variables included total sleep time (TST) and wake after sleep onset (WASO). Direct comparisons between single-night AW and CAM data were made via paired t-tests, mean absolute percent error (MAPE) calculations and intra-class correlations (ICC).
Results: On average, Fitbit, Jawbone, Misfit, and Xiaomi Mi significantly overestimated TST relative to AW (54.4–76.7 min, $p \leq .007$), while Polar underestimated TST (67.3 min, $p \leq .002$). Samsung TST did not