CHARACTERIZATION OF AROUSALS IN POLYSOMNOGRAPHY USING THE STATISTICAL SIGNIFICANCE OF POWER CHANGE

Junjie V. Liu and H. Klar Yaggi

Section of Pulmonary, Critical Care and Sleep Medicine, Department of Internal Medicine,
Yale University School of Medicine, New Haven, Connecticut, USA
junjie.liu @ yale.edu

Abstract — Arousals are neural events during sleep represented as abrupt increases of high-frequency electroencephalogram (EEG) signals in polysomnography (PSG). In clinical practice, a human scorer uses visual detection to demarcate the starts and ends of arousals, whereas other properties of arousals are hardly ever studied. Here we characterized arousals by the statistical significance of arousal-associated changes in the EEG signal power. To evaluate the test-retest reliability, we used a database of 1026 men who completed two PSGs separated by several years. Ten-second segments of EEG signals that either contained or were without arousals were extracted. For each segment, the power of EEG signal filtered in delta, theta, alpha, beta or gamma frequency band was computed. Then for each PSG, statistical significance of the difference in power between the arousal-containing and the arousal-absent group of EEG segments was computed. The statistical significance showed good test-retest reliability (intraclass correlation coefficient ICC > 0.40). In comparison, the numeric value of the difference in power showed generally poor test-retest reliability (ICC < 0.10). The statistical significance had higher test-retest reliability in theta band (4-8 Hz) than in other frequency bands, and higher reliability in Stage 2 sleep than in rapid eye movement (REM) sleep. Furthermore, the statistical significance in theta band was not influenced by the incidence rate of arousals. Thus, statistical significance of power change in the theta band is a robust metric of arousal-associated EEG signal changes, which may become useful in studying diseases associated with abnormal arousals.

I. Introduction

Polysomnography (PSG) is the cornerstone of diagnostics for a wide range of sleep disorders. A key component of PSG is electroencephalogram (EEG), which is used to determine the sleep stages and denote important neural events such as the arousals. Arousals during sleep are transient increases of neural activities in pathways that promote wakefulness, particularly the ascending reticular activating system that terminates diffusely in the cerebral cortex [1]. In PSG, arousal-associated neural activities in the cerebral cortex are represented as transient increases of the high-frequency components of EEG signals [2]. Arousals are clinically important because many sleep disorders are accompanied with repetitive arousals, which lead to sleep fragmentation and excessive daytime sleepiness [3]. In obstructive sleep apnea (OSA), a very common disorder that impairs millions of Americans, arousals are critical to the termination of apneas and hypopneas, at least in part by resuming the neural inputs that control upper airway dilator muscles to the levels that resolve airway collapses [4][5]. Hence, an accurate characterization of arousals in PSG may improve our understanding and treatment of OSA and other sleep disorders [6].

In routine clinical practice, arousals are identified from the PSG by a human scorer's visual examination of EEG signals, and only the beginnings and ends of arousals are denoted [2]. Other properties of arousals have only been studied by a few research groups. The arousal-associated increase of EEG power in high frequencies was described for groups of subjects [7], but has not been used as a measure of arousal intensity for each individual subject. Indeed, numeric changes in EEG power might not be reliable measures of arousals for individual subjects, in part because of the very large variations in the baseline EEG power values across individual normal subjects [8][9]. Alternatively, a measure of arousal intensity has been proposed using two analysis steps: first, a large training database of arousals was rated with an arousal intensity score between 0 and 9 based on the subjective opinion of a single scorer, then for each testing arousal segment, the best match from the training database was found based on 33 features from the wavelet analysis, and the arousal intensity score was assigned accordingly [10]. This arousal intensity measure had good test-retest reliability across two PSGs on different days [11], but is a subjective measure, cumbersome to employ, and does not clearly reflect the increase of high-frequency EEG signals.

In this study, we propose a new measure of the arousal-associated EEG power changes that not only has good test-retest reliability but is objective and easy to compute. Within each PSG, we computed the statistical significance of the EEG power differences between the group of time segments containing arousals, and the group of time segments without arousals (i.e. baseline). This measure essentially normalizes the arousal-associated EEG power changes by the intrinsic variations of EEG power across the PSG (i.e. the entire night of sleep). The influence of large variations in the baseline EEG power values across subjects is minimized, consequently, good test-retest reliability is achieved even

across PSG studies that are separated by years. Using this statistical significance measure, we characterized the arousal-associated EEG power changes in medium (theta) and low (delta) frequencies as reliable, individualized traits independent from the power changes in high frequencies.

II. METHODS

PSG data were obtained from the National Sleep Research Resource (NSRR) on the Internet (http://sleepdata.org) [12]. We focused on the osteoporotic fractures in men (MrOS) study because it contained two PSGs, suitable for our study of the test-retest reliability. In MrOS study, in-home PSG was conducted on community-dwelling elderly men during the initial visit in year 2003-2005 and the return visit in year 2009-2012, and 1026 men completed PSG in both visits [13][14]. At the return visit, the age of this cohort was 81.0±4.4 years and the body mass index (BMI) was 26.9±3.8 kg/m2, similar to that reported for the initial visit [15]. We then excluded 46 subjects who used CPAP or BiPAP during PSG (all at the return visit), and 6 subjects whose PSG scoring data were missing or erroneous, resulting in 974 subjects for our analysis. The PSG data were read and scored by certified researchers during previous studies, and the scoring results were also available at the NSRR. The scoring result of each PSG included the timings of sleep stages, arousals (following the ASDA definition [2]), and respiratory events (sleep apnea or hypopnea). We limited analysis to the rapid eye movement (REM) sleep stage and the Stage 2 sleep, which is the most dominant non-REM sleep stage. The arousal index (ArI) was calculated as the number of scored arousal events divided by the total duration of sleep epochs during each sleep stage. Similarly, the apnea-hypopnea index (AHI) was the number of scored apnea and hypopnea events divided by the total duration.

In each sleep stage of each PSG study, we demarcated 10-second-long non-overlapping time segments that either contained arousal events or did not. Note that the segments not containing arousals might contain other events, such as respiratory apneas and hypopneas, periodic leg movements, and "unsure" events (defined by NSRR as hypopneas with a >50% decrease in flow). Each 10-second segment was demarcated with the following rules: i) the segment was during a sleep stage that lasted 90 seconds (three epochs) or longer, and started after 15 seconds after the start of the stage and ended before 15 seconds before the end of the stage; ii) each arousal-containing segment contained at least three seconds of a time period scored as arousal (the minimum duration of an ASDA-defined arousal event [2]). Taking together, this demarcation procedure resulted in two groups of 10-second segments ("arousal" and "control" groups) for each sleep stage.

EEG signal in each PSG was derived by subtracting the sum of A1 and A2 channels from the sum of C3 and C4 channels. The power of EEG signal within each 10-second segment was computed for five conventional EEG frequency bands (delta: 1-4 Hz; theta: 4-8 Hz; alpha: 8-13 Hz; beta: 13-25 Hz; gamma: 25-45 Hz), using the *periodogram* function in Matlab (Mathworks, MA, USA). This computation yielded two groups of power values ("arousal" and "control" groups) for each frequency band in each sleep stage.

Next, we tested the null hypothesis that the EEG power was the same between the group of arousal-containing time segments and the group of control segments. The null hypothesis was tested only if there were at least 10 segments in each group, and the statistical significance (z-score) were computed using the Wilcoxon two-sided signed-rank test with normal approximation. The null hypothesis was rejected (arousal-associated power change was significant) if $|z| > 2.576$ (corresponding to $P < 0.01$). For comparison, we also computed the numeric ratio and the numeric difference between the median power values of the two groups.

Test-retest reliability between the two PSG studies per subject was evaluated using the intraclass correlation coefficient (ICC), specifically the subtype of one-way random single measures [16]. ICC between the first and the second PSG across all subjects was computed for the statistical significance (z-score) as well as the numeric ratio and difference of the arousal-associated EEG power change. ICC below 0.40 was interpreted as poor test-retest reliability, whereas ICC above 0.60 was interpreted as good [17].

We also studied the relationships between the statistical significance values (z-scores) of arousal-associated power change during REM sleep and during Stage 2 sleep, using the Spearman's rank correlation test. Similarly, the relationships between the z-scores and other PSG parameters (ArI, AHI) were studied.

III. RESULTS

Consistent with the definition of arousals, EEG power increased significantly in high-frequency (alpha, beta, gamma) bands during arousals compared to control time segments (Table 1). The statistical significance of this increase in high-frequency power was higher during Stage 2 sleep (Wilcoxon's z-score was roughly 11±4, as mean ± standard deviation across PSG sessions), compared to during REM sleep (z-score roughly 6±2), though in both sleep stages, the increase was significant ($z > 2.576$) in more than 94% of the subjects. Meanwhile, EEG power in the theta band increased significantly in roughly half of the subjects (z-score: 2.8±4.9 in Stage 2, 2.8±2.6 in REM), whereas EEG power in the delta band was essentially unchanged. In comparison, the arousal-

Power Metric	Delta	Theta	Alpha	Beta	Gamma
Z-score, Stage 2	-0.8 ± 3.9	2.8 ± 4.9	9.8 ± 4.2	12.1 ± 4.0	11.1 ± 3.9
Z-score, REM	0.3 ± 2.6	2.8 ± 2.6	5.8 ± 2.1	6.7 ± 2.1	6.1 ± 2.3
Ratio, Stage 2	1.0 ± 1.0	1.2 ± 0.8	1.8 ± 0.9	2.2 ± 1.5	4.0 ± 7.7
Ratio, REM	1.3 ± 2.1	1.4 ± 0.8	2.0 ± 0.8	2.2 ± 1.4	3.8 ± 8.0
Diff., Stage 2	1.8 ± 51.4	0.9 ± 4.6	1.4 ± 2.0	0.4 ± 0.7	0.2 ± 0.5
Diff., REM	-0.7 ± 75.2	0.9 ± 4.8	1.2 ± 2.3	0.4 ± 0.6	0.2 ± 0.5

Table 1. Arousal-associated EEG power changes as measured by the statistical significance (Z-score), and the numeric ratio and numeric difference between power values during and without arousals. Mean ± standard deviation across PSG sessions for the two sleep stages (Stage 2, REM) in five frequency bands were shown.

Power Metric	Delta	Theta	Alpha	Beta	Gamma
Z-score, Stage 2	0.45	0.66	0.52	0.44	0.35
Z-score, REM	0.33	0.45	0.22	0.19	0.32
Ratio, Stage 2	0.03	0.05	0.15	0.07	0.10
Ratio, REM	0.00	0.08	0.26	0.12	0.09
Diff., Stage 2	0.00	0.03	0.25	0.03	0.05
Diff., REM	0.00	0.05	0.13	0.06	0.05

Table 2. Test-retest reliability of arousal-associated EEG power changes as measured by the statistical significance (Z-score), numeric ratio and numeric difference. Intraclass correlation coefficient (ICC) between two PSG sessions of the same subject was shown. Note that ICC were much higher for the Z-score than the ratio or the difference.

associated EEG power increases were often poorly reflected by the numeric ratio and the numeric difference of power, which had large standard deviations owing to the wide ranges of numeric values of EEG power across subjects (rows 4-7, Table 1).

The statistical significance of arousal-associated EEG power change was much more consistent between two PSG sessions from the same subject, compared to how consistent the numeric ratio and the numeric difference of power were. Table 2 summarized the test-retest reliability between the two PSG sessions of each subject, which were separated by at least four years in our study. During Stage 2 sleep, test-retest reliability of the z-score of EEG power change was good (ICC: 0.66) for theta band and was fair (ICC: 0.35-0.52) for other bands. During REM sleep, the test-retest reliability of z-score was fair (ICC: 0.45) for theta band and was lower (ICC: 0.19-0.33) for other bands. In comparison, the test-retest reliability of the numeric ratio, and that of the numeric difference of power, were in general very poor (ICC < 0.15), and were slightly better in the alpha band (ICC: 0.13-0.26) then in other bands.

The arousal index (ArI), namely the incidence rate of arousals, was 29.0±15.8 and 25.0±13.3 during Stage 2 and REM sleep, respectively. The test-retest reliability of ArI was almost fair (ICC: 0.37) during Stage 2 sleep, but much lower (ICC: 0.18) during REM sleep. The apnea-hypopnea index (AHI) was 16.4±12.7 and 20.9±12.3 during Stage 2 and REM sleep, respectively. The test-retest reliability of AHI was poor (ICC: 0.25) during Stage 2 sleep, and very poor (ICC: 0.05) during REM sleep. Notably, the test-retest reliabilities of both ArI and AHI were generally lower than the reliabilities of z-score of power change at all frequencies.

For both the EEG power changes and the ArI and AHI, there were strong correlations between Stage 2 sleep and REM sleep within the same PSG session. The Spearman's rank correlation coefficients were 0.36, 0.47, 0.22, 0.19, 0.21 for the z-score of EEG power increase in delta, theta, alpha, beta, gamma bands, and were 0.41 and 0.43 for ArI and AHI, respectively (all with $P < 10\text{-}8$). Figure 1 showed the tight correlation between z-scores of the theta band in Stage 2 and REM sleep, and the weaker correlation between z-scores of the beta band.

Last, we studied the relationships among z-scores in the five bands, ArI and AHI within the same sleep stage of a PSG session. In Stage 2 sleep (Figure 2A), arousal-associated power changes in delta band were dissociated from the changes in beta and gamma bands ($P > 0.10$). ArI was strongly correlated with the z-scores of power changes in high-frequency (alpha, beta, gamma) bands (Spearman's rank correlation coefficient: 0.44, 0.52, 0.39 respectively), and was correlated with AHI (coefficient: 0.38). However, ArI and AHI were not associated with the z-scores of power changes in delta and theta bands ($P > 0.10$). In REM sleep (Figure 2B), the dissociation between low- and high-frequency bands was less distinct, as the z-score of power change in delta band was

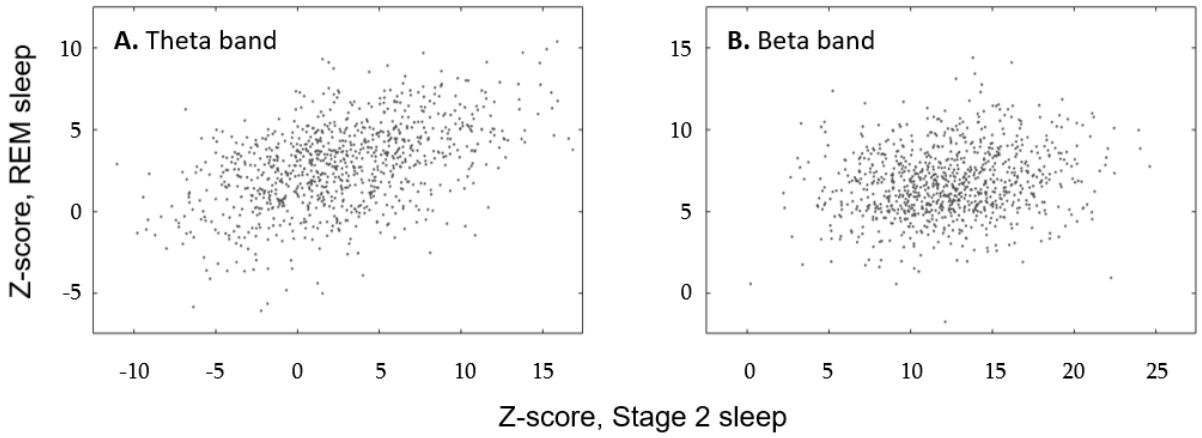

Figure 1. Correlations between the statistical significances (Z-scores) of the EEG power increases during Stage 2 sleep and REM sleep were (A) strong for the theta frequency band, and (B) weaker for the beta band. Each dot represents one PSG session.

correlated with that in beta and gamma bands (coefficient: 0.14, 0.18). ArI remained correlated with power changes in high-frequency (alpha, beta, gamma) bands, though the correlations were weaker than in Stage 2 sleep. ArI and AHI remained dissociated from power changes in delta and theta bands. Thus, in both stages of sleep, arousal-associated EEG power changes in low- (delta) and medium-frequency (theta) bands were not influenced by the incidence rates of arousals or respiratory events.

IV. DISCUSSION

In this paper, we developed a new method to quantitatively characterize the arousal-associated changes in EEG power (signal amplitude). For each EEG frequency band in each sleep stage of each PSG, we computed the statistical significance of EEG power differences between arousal events and the baseline, when there were enough arousal events (in our study, ten or more). This statistical significance metric showed

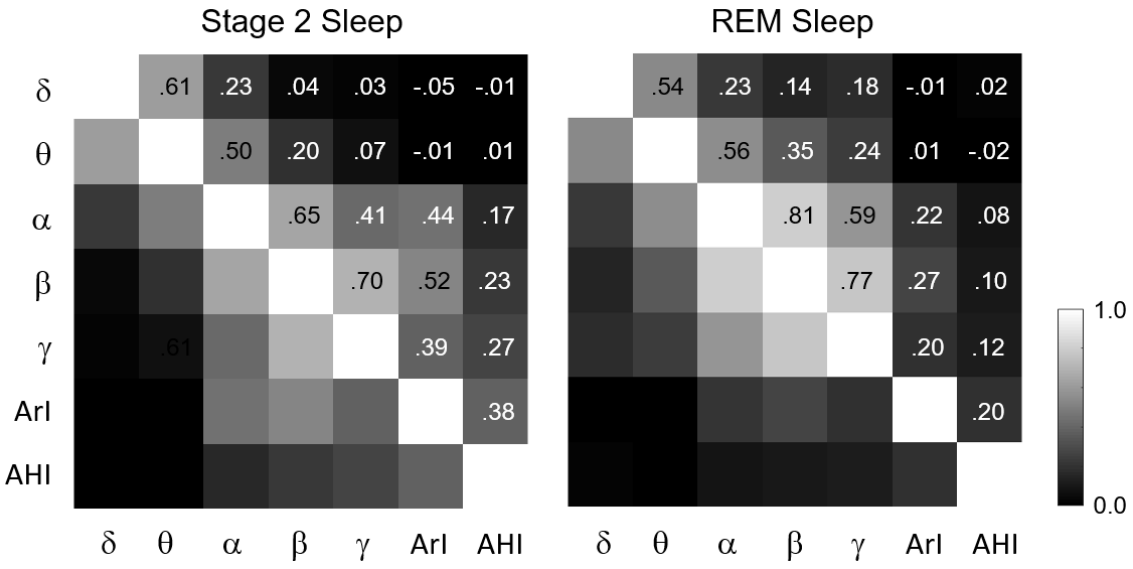

Figure 2. Relationships among the statistical significances (z-score) of EEG power increase in five frequency bands (δ: delta, θ: theta, α: alpha, β: beta, γ: gamma), and arousal index (ArI) and apnea-hypopnea index (AHI), during (A) Stage 2 sleep and (B) REM sleep. The correlation coefficient between each pair of parameters across all PSG sessions was shown in grayscale (black: zero correlation; white: total correlation).

much better test-retest reliability than that of the numeric value changes of EEG power, likely because the influence of variations across human subjects was eliminated. Specifically, the ICC of the statistical significance of EEG power in alpha band (8-13 Hz) during Stage 2 sleep was 0.52, which was remarkable when considering the large number of subjects (near 1000) and the long gap between test and retest (4-9 years). In comparison, the arousal intensity metric based on subjective rating had an ICC of 0.72 in a cohort of 28 subjects between test and retest on two consecutive nights [11]. Thus, the statistical significance of EEG power in alpha band may serve as a better metric of arousal intensity, in that it is easy to compute, not based on any subjective rating, and is directly correspondent to the definition of arousals as increases of EEG power in high frequencies.

Our findings for the EEG power changes in the low and medium frequencies suggest that they are individualized traits that can be used to identify the subtypes of arousals. In non-rapid eye movement (NREM) sleep, which primarily consists of Stage 2 sleep, it has been well established that arousals occur not randomly, but rather in a cyclic alternating pattern (CAP) where short time segments of arousal-like EEG signals are separated by segments of baseline EEG signals [18]. Furthermore, the arousals can be categorized into three subtypes: A1, A2 and A3, where the low- and medium-frequency EEG signals are relatively strong in A1, less in A2, and weak in A3. However, the scoring of these subtypes was quite subjective [18][19]. Here, the statistical significances of EEG power changes in delta and theta bands can serve as quantitative measures to distinguish the subtypes of arousals. In our study, only about half of subjects had significance arousal-associated increase of power in theta waves. Arousals in these subjects might correspond to the subtypes A1 and A2.

Remarkably, the arousal-associated EEG power changes in delta and theta bands were largely uncorrelated to the changes in high-frequency (alpha, beta, gamma) bands, and uncorrelated to the occurrence rate of arousals (ArI). Whereas future studies are needed to elucidate the phenotype representations of such arousal-associated increases of slow EEG signals, they have been associated with protecting sleep stability and consolidation [18]. The independence from ArI suggests that such increases might represent certain stereotypical neuroprotective responses to every arousal event. In contrast, arousal-associated EEG power changes in high-frequency bands were strongly correlated to ArI (second-to-last column, Figure 2A). High ArI is common at the age of our subject cohort and is closely correlated with frequent awakening [20]. Thus, we might speculate that individuals with larger arousal-associated increases of fast EEG signals have higher excitability of the cerebral cortex and are more prone to frequency awakening. Furthermore, in OSA, frequent arousals (high ArI) are associated with frequent respiratory apneas and hypopneas (high AHI). Because of this association, arousal-associated EEG power changes in high-frequency bands were also correlated to AHI, whereas arousal-associated EEG power changes in delta and theta bands were not correlated to AHI (last column, Figure 2A). Thus, the differential arousal-associated changes in different frequency bands may help distinguish the subtypes of OSA, some of which might be associated with the CAP and could be treated differently [21].

During REM sleep, arousal-associated EEG power changes were generally less significant and less reliable in test-retest. These may be contributed by several factors, including the shorter total duration and the generally lower baseline EEG power during REM compared to non-REM sleep, and the ASDA requirement for concurrent electromyogram activity to score an arousal in REM (thus purely EEG-defined arousals may not be scored as arousals) [2]. During REM sleep, EEG power changes in delta and theta bands became somewhat correlated to EEG power changes in high-frequency bands, (Figure 2B) possibly because the baseline EEG power is low. Remarkably, individuals with significant arousal-associated increases of slow EEG signals (delta and theta bands) during Stage 2 sleep also tend to have such increases during REM sleep (Figure 1A), suggesting that such increases are individual traits independent of the sleep stage, and may protect the individual's sleep stability throughout the night. In contrast, the correlation between Stage 2 sleep and REM sleep was much lower for arousal-associated increases of fast EEG signals (Figure 1B), suggesting that such increases are dependent on the sleep stage.

In summary, we developed a new method to measure arousal-associated EEG power changes by comparing short time segments with and without arousals within the same PSG. The statistical significance of arousal-associated EEG power change is a metric with good test-retest reliability and little impact from inter-subject variations. Using this metric, we found that arousal-associated EEG power changes in low- and medium-frequency are implicated for protecting sleep stability, and are largely independent from arousal-associated EEG power changes in high frequency, which are implicated for promoting awakening. Thus, quantitative characterization of arousals is incomplete without taking into account the differential changes of EEG power across the frequency spectrum.

ACKNOWLEDGMENTS

Author HKY was supported by the National Institutes of Health (K24 HL132093). Moreover, we are indebted to the researchers who organized the NSRR project and the

MrOS study, and the National Institutes of Health which funded the NSRR (R24 HL114473) and the MrOS study (R01 HL070837, HL070838, HL070839, HL070841, HL070842, HL070847, HL070848, HL071194).

REFERENCES

[1] Jones BE, "Arousal Systems," *Front. Biosci.*, vol. 8, pp. s438-s451, 2003.

[2] ASDA (American Sleep Disorders Association). "EEG arousals: scoring rules and examples," *Sleep*, vol. 15, no. 2, pp. 173-184, 1992.

[3] Stepanski E. "Sleep Fragmentation and Daytime Sleepiness," *Sleep*, vol 7, no. 1, pp. 18-26, 1984.

[4] Jordan AS, White DP. "Pharyngeal Motor Control and the Pathogenesis of Obstructive Sleep Apnea," *Respir. Physiol. Neurobiol.*, vol. 160, no. 1, pp. 1-7, 2008.

[5] Younes M. "Role of Arousals in the Pathogenesis of Obstructive Sleep Apnea," *Am. J. Respir. Crit. Care Med.*, vol. 169, no. 5, pp. 623-633, 2004.

[6] Amatoury J, Azarbarzin A, Younes M, et al. "Arousal Intensity is a Distinct Pathophysiological Trait in Obstructive Sleep Apnea," *Sleep*, vol. 39, no. 12, pp. 2091-2100, 2016.

[7] De Carli F, Nobili L, Beelke M, et al. "Quantitative Analysis of Sleep EEG Microstructure in the Time-frequency Domain," *Brain Res. Bull.*, vol. 63, no. 5, pp. 399-405, 2004.

[8] Carrier J, Land S, Buysse DJ, et al. "The Effects of Age and Gender on Sleep EEG Power Spectral Density in the Middle Years of Life (Ages 20-60 Years Old)," *Psychophysiology*, vol. 38, no. 2, pp. 232-242, 2001.

[9] De Gennaro L, Ferrara M, Vecchio F, et al. "An Electroencephalographic Fingerprint of Human Sleep," *Neuroimage*, vol. 26, no. 1, pp. 114-122, 2005.

[10] Azarbarzin A, Ostrowski M, Hanly P, Younes M. "Relationship between Arousal Intensity and Heart Rate Response to Arousal," *Sleep*, vol. 37, no. 4, pp. 645-653, 2014.

[11] Azarbarzin A, Ostrowski M, Younes M, et al. "Arousal Responses during Overnight Polysomnography and their Reproducibility in Healthy Young Adults," *Sleep*, vol. 38, no. 8, pp. 1313-132, 2015.

[12] Dean DA, Goldberger AL, Mueller R, et al. "Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource," *Sleep*, vol. 39, no. 5, pp. 1151-1164, 2016.

[13] Blank JB, Cawthon PM, Carrion-Petersen ML, et al. "Overview of Recruitment for the Osteoporotic Fractures in Men Study (MrOS)," *Contemp. Clin. Trials*, vol. 26, no. 5, pp. 557-568, 2005.

[14] Orwoll E, Blank JB, Barrett-Connor E, et al. "Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study – a large observational study of the determinants of fracture in older men," *Contemp. Clin. Trials*, vol. 26, no. 5, pp. 569-585, 2005.

[15] Mehra R, Stone KL, Blackwell T, et al. "Prevalence and Correlates of Sleep-disordered Breathing in Older Men: Osteoporotic Fractures in Men Sleep Study," *J. Am. Geriatr. Soc.*, vol. 55, pp. 1356-1364, 2007.

[16] Shrout PE, Fleiss JL. "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psychol. Bull.*, vol. 86, no. 2, pp. 420-428, 1979.

[17] Cicchetti DV, Sparrow SA. "Developing Criteria for Establishing Interrater Reliability of Specific Items: Applications to Assessment of Adaptive Behavior," *Am. J. Ment. Defic.*, vol. 86, no. 2, pp. 127-137, 1981.

[18] Terzano MG, Parrino L. "Origin and Significance of the Cyclic Alternating Pattern (CAP)," *Sleep Med. Rev.*, vol. 4, no. 1, pp. 101-123, 2000.

[19] Terzano MG, Parrino L, Smerieri A, et al. "Atlas, Rules, and Recording Techniques for the Scoring of Cyclic Alternating Pattern (CAP) in Human Sleep," *Sleep Med.*, vol. 3, no. 2, pp. 187-199, 2002.

[20] Bonnet MH, Arand DL. "EEG Arousal Norms by Age," *J. Clin. Sleep Med.*, vol. 3, no. 3, pp. 271-274, 2007.

[21] Thomas RJ, Terzano MG, Parrino L, Weiss JW. "Obstructive Sleep-disordered Breathing with a Dominant Cyclic Alternating Pattern - a Recognizable Polysomnographic Variant with Practical Clinical Implications," *Sleep*, vol. 27, no. 2, pp. 229-234, 2004.