Refinement and validation of an ECG based algorithm for detecting awakenings

S. McGuire[1], U. Müller[2], G. Plath[2], M. Basner[1]

[1] Division of Sleep and Chronobiology, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA, smcgu@upenn.edu

[2] Division of Flight Physiology, Institute of Aerospace Medicine, German Aerospace Center (DLR), Cologne, Germany

ABSTRACT

The most sensitive method for measuring the impact of noise on sleep is electroencephalogram based polysomnography. However, this approach is somewhat invasive and has a high methodological expense as trained individuals are needed for both instrumentation and data analysis. An alternative, less intrusive and expensive approach that has been proposed is to use a single channel electrocardiogram to measure heart rate. An algorithm was previously developed which automatically identifies increases in heart rate associated with cortical arousals, greater than or equal to 3 seconds in duration. However, the EEG awakening (i.e., an activation greater than or equal to 15 seconds in duration) is currently the most agreed upon indicator of noise-induced sleep disturbance. Therefore, refinements have been made to the original algorithm in order to identify EEG awakenings. The data used for refining and validating the algorithm were gathered in field studies which examined the effects of aircraft noise on sleep. The changes made include the re-estimation of model parameters, the use of actigraphy data, and the addition of a sleep onset and offset algorithm. Awakenings detected with the algorithm were compared to awakenings identified visually using polysomnography data for 106 subject nights and agreement between the two approaches was high relative to conventional standards. Therefore, due to the high sensitivity in detecting awakenings this ECG based approach could be a useful method for examining noise-induced sleep disturbance in larger subject samples with lower methodological expense compared to polysomnography and more reliable and meaningful results compared to actigraphy.

INTRODUCTION

As nighttime noise exposure continues to change we need a methodological approach that will allow the examination of the impact of noise on sleep, which is sensitive in identifying awakenings but also low in cost so that larger and more representative samples in exposed communities can be examined. The most sensitive method for evaluating sleep is to use EEG-based polysomnography which is the simultaneous measurement of the electroencephalogram (EEG), the electrooculogram (EOG), and the electromyogram (EMG). While this approach has been implemented in a few field studies (Basner, Samel, & Isermann 2006; Flindell et al. 2000; Ollerhead et al. 1992), it is expensive to implement as trained individuals are needed to apply and remove the electrodes each night and trained individuals are needed to visually score the data. There is also high intra- and inter-rater variability in sleep stage and arousal scoring (Bonnet et al. 2007).

A more cost-effective approach that has been used in previous studies on aircraft noise and sleep is actigraphy (Ollerhead et al. 1992; Passchier-Vermeer et al. 2002). Actigraphs are devices typically worn on the wrist and measure acceleration. Awakenings during the night are identified based on the amount of movement. While this approach is inexpensive and noninvasive, different algorithms are used to identify awakenings and analysis is conventionally done using 60 second epochs (e.g. Cole et al. 1992) which may result in an underestimation of the number of awakenings. In the study by Ollerhead et al. (1992), it was found that less than one percent of aircraft events induced an actigraphy defined arousal while over 4 percent induced an EEG awakening or arousal. In addition, comparison studies between awakenings identified using actigraphy and polysomnography have found high sensitivity in identifying sleep epochs during the night but a low specificity (below 0.40) in identifying wake epochs (Jean-Louis et al. 2001).

When a cortical arousal occurs during the night, an autonomic arousal also often occurs, which includes an increase in blood pressure and heart rate. Basner et al. (2007) developed an algorithm for automatically identifying changes in heart rate that are associated with a cortical arousal, 3 seconds or longer in duration. Basner et al. (2008) compared the probability of awakening to an aircraft noise event calculated based on arousals identified with the ECG algorithm to the probability of awakening calculated based on EEG scored awakenings. They found that once the probability of spontaneous reactions was accounted for the two approaches provided similar results. There are several advantages to using ECG to identify awakenings; it is non-invasive and individuals can be taught how to apply the electrodes themselves, therefore like actigraphy the devices can be used by individuals unattended. Also with an automatic scoring algorithm, data can be scored the same within and across studies.

In noise effects research, the indicator of noise induced sleep disturbance that is most agreed upon is awakenings, cortical activations 15 seconds in duration or longer. Basner, Brink, and Elmenhorst (2012) stated that arousals, less than 15 seconds in duration, occur often during the night and are therefore not a specific indicator of noise induced sleep disturbance. Normal adults can have over 80 arousals per night (Bonnet & Arand 2007). In addition it is awakenings that are being used in developed sleep models (Basner, Samel, & Isermann 2006) and is the definition of awakenings used in proposed night-time noise protection guidelines (Schreckenberg, Thomann, & Basner 2009). Therefore, refinements have been made to the original algorithm by Basner et al. in order to only predict cortical arousals 15 seconds or greater in duration. The modifications made include re-estimation of model parameters, the addition of an algorithm to identify the start and end of the sleep period, and the use of actigraphy data for both predicting the sleep period and to improve arousal predictions. In order to validate the algorithm comparisons have been made to arousals identified visually using polysomnography data. An overview of the algorithm and the agreement between arousals identified using the ECG-based algorithm and cortical arousals are discussed.

ALGORITHM

The algorithm identifies arousals based on both heart rate and movement independently. The data used for refining the algorithm is from the STRAIN field study, which was conducted by the Institute of Aerospace Medicine at the German Aerospace Center (DLR) around Cologne-Bonn Airport (Basner, Samel, & Isermann 2006). Heart rate was obtained from the ECG measurements which were sampled at 1024 Hz. The time of each R-wave was detected and stored in a file along with the inter-beat interval. The signal used to identify movement is an accelerometer signal; the device was worn on the chest, and was sampled at 8 Hz**.**

Arousals are identified in the algorithm by using matrices of likelihood ratios. The likelihood ratios are an indicator of whether the difference in heart rate or the amount of movement is associated with an arousal. There is a separate matrix used for both types of measurements. For heart rate, the difference in beat to beat heart rate to a 3 minute moving median heart rate was calculated. The heart rate difference during 1066 visually scored cortical arousals from 56 different subject nights was extracted. The visually scored arousals were between 15 and 60 seconds in duration and occurred after different sleep stages. The heart rate difference during 1066 segments of the night that did not contain arousals was also extracted. Likelihood ratios were calculated using the 2132 samples, and for the model the likelihood ratios for 1 to 34 beats after the start of an arousal is used. For each beat there are 10 likelihood ratios.

To estimate movement the derivative of the accelerometer signal is calculated using an FIR filter. Then the square root of the sum of the squared derivative for every one second is calculated. An example of the generated movement signal is shown in Figure 2a. Segments of the movement signal during and not during cortical arousals were also extracted and used to calculate the likelihood ratios, and for the model the likelihood ratios for 1 to 34 seconds after the start of an arousal is used. For each second there are 10 likelihood ratios. The matrices used for movement and heart rate are shown in Figure 1. The values in the matrices change with time and increase with the amplitude of the movement signal and with larger differences in heart rate.

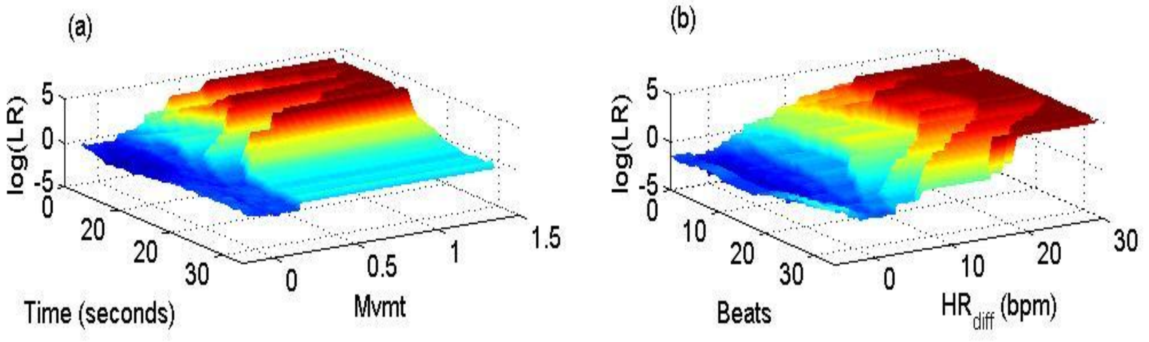

Figure 1: Likelihood ratio matrices used in the program for identifying arousals based on (a) movement and (b) heart rate.

To identify an arousal based on the difference in heart rate, likelihood ratios are multiplied for every beat from 1 beat to 34 beats after the time point that is under investigation. If the result obtained from multiplying the likelihood ratios is above a preset threshold then an arousal is scored. The beginning and end of each arousal is corrected if necessary by examining the ratio of the heart rate difference before and after each time point of interest. This ratio provides an indication of where heart rate starts increasing and where it stops decreasing. Arousals based on movement are identified in a similar manner and arousals using both approaches are combined. If the duration between identified arousals is less than 10 seconds they are also combined. An example of the signals used to identify arousals and a comparison between arousals scored with the algorithm and those identified using EEG are shown in Figure 2a. The signals used to identify sleep onset and the end of the sleep period are shown in Figure 2b. The average ratio of the beat to beat heart rate in each 30 second segment to a 20 minute median heart rate is calculated and the percent of each 30 second segment that contains movement is calculated. For sleep onset, 3 minutes are examined around each 30 second epoch to determine if the heart rate has decreased and is within 5.0% of the median and if there is no movement, for sleep offset the algorithm follows a similar procedure starting from the end of the signals.

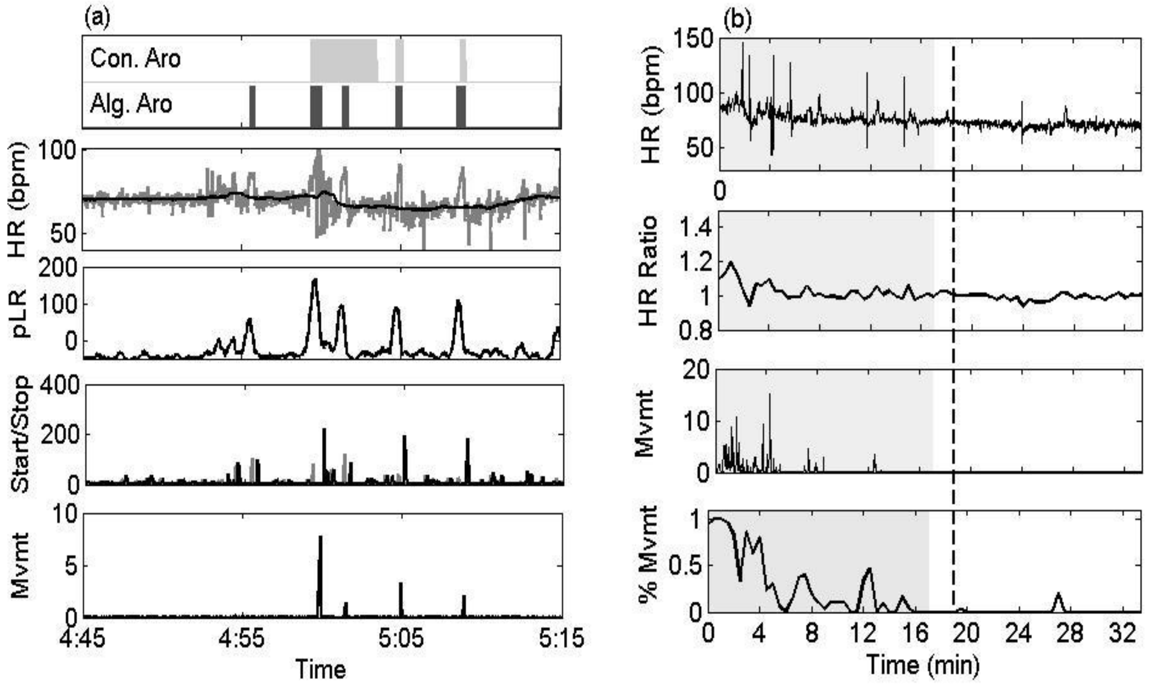

Figure 2: (a) An example of the signals used in the ECG-based algorithm to identify arousals; heart rate and the median heart rate (HR), the log of the signal obtained from multiplying likelihood ratios (pLR), start and stop which is the ratio of heart rate before and after each time point of interest, and the signal used to identify movement (Mvmt). Alg. Aro is arousals identified with the algorithm, and Con. Aro is EEG identified arousals. (b) Heart rate, heart rate ratio, movement signal, and the percent of each 30 second segment containing movement. The black dotted line indicates the visually scored sleep onset time, and the gray region is the period of the night that is identified as occurring before sleep onset by the algorithm.

VALIDATION METHOD

To validate the changes made to the algorithm arousals scored visually using polysomnography were compared to arousals identified with the ECG-based algorithm. The data used for the validation are from year one of the NORAH sleep study, which was conducted in 2011 by the Institute of Aerospace Medicine at the German Aerospace Center (DLR) around Frankfurt Airport. There were 49 subjects in the study. The subjects ranged in age from 18 to 76, 25 were female. Sleep was measured using EEG-based polysomnography. A total of 106 subject nights of data was used in the analysis.

Two independent visual scores of sleep stages and cortical arousals were obtained. Sleep stages were scored for each 30 second epoch according to the rules of Rechtschaffen and Kales (1968). Arousals were scored according to the AASM criteria (Bonnet et al. 1992). During NREM sleep if there was a change in EEG frequency for 3 seconds or greater than an arousal was scored, during REM sleep a change in EMG activity also had to occur.

RESULTS

The agreement between arousals, 15 seconds or greater in duration, scored visually based on polysomnography data and those calculated using the ECG-based algorithm were evaluated by calculating sensitivity, specificity, and Cohen's kappa values. The sensitivity values indicate how well the algorithm identified the EEG arousals, while specificity indicates how well the algorithm was at identifying time periods when EEG-arousals were not occurring. Kappa is a chance corrected measurement of agreement (Landis & Koch 1977).

The arousals identified using the ECG-based algorithm was compared to both sets of visually scored arousals and to a set of consensus arousals. For a consensus arousal to be scored the 2 visually scored arousals had to be equal to or greater than 15 seconds in duration and they had to overlap. There was no requirement for the duration of the overlap. The start of the consensus arousal was defined as the earlier of the start times of the two scored arousals. The end of the consensus arousal was defined as the later of the end times of the two scored arousals.

In order to calculate specificity a set of control arousals was defined. The control arousals were defined to be of the same duration as each visually scored or consensus arousal up to a duration of 300 seconds. If a visually scored arousal was greater than 300 seconds in duration, then a control arousal of 300 seconds was defined. Control arousals were randomly assigned to parts of the night in which the sleep stage was not scored as wake and there was not a visually scored arousal. The mean results for sensitivity and specificity are in Table 1. The specificity values are the mean values based on 100 calculations, for each calculation the position of the control arousals was randomly varied.

Table 1: Sensitivity values and specificity values (based on 100 calculations) comparing the agreement between visually scored arousals that were greater than or equal to 15 seconds in duration and arousals identified using the ECG-based algorithm. The values in parentheses are the standard deviations for the 100 calculations.

	Sensitivity	Specificity
Visual Scoring 1 (Gold) vs. ECG-Based Algorithm	0.790	0.979 (0.003)
Visual Scoring 2 (Gold) vs. ECG-Based Algorithm	0.820	0.977 (0.003)
Consensus Visual Scoring (Gold) vs. ECG-Based Algorithm	0.893	0.976 (0.004)
Visual Scoring 1 (Gold) vs. Visual Scoring 2	0.773	0.997 (0.001)
Visual Scoring 2 (Gold) vs. Visual Scoring 1	0.801	0.992 (0.002)

Cohen's kappa was also calculated for each comparison between visually scored arousals and those identified using the ECG-based algorithm. Arousals identified with the algorithm when only using movement or only using heart rate were also compared to the consensus arousals. Kappa values were calculated for each subject. The mean results for the kappa calculations are shown in Figure 3. Paired t-tests were completed to determine if the difference between kappa values were statistically significant, the results are indicated in Figure 3.

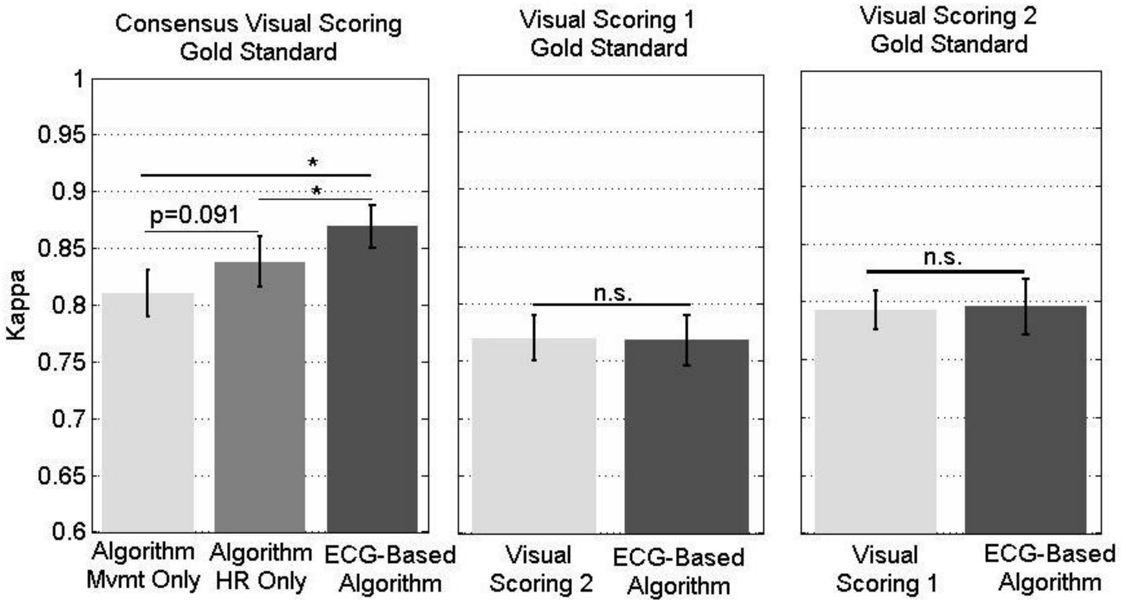

Figure 3: Kappa values calculated for comparisons between visually scored arousals greater than or equal to 15 seconds in duration and arousals identified using the ECG-based algorithm. Data expressed as mean ± standard error of the mean, *p< 0.001, n.s. not significant.

DISCUSSION

The ECG-based algorithm was refined in order to identify arousals during the night that are 15 seconds or greater in duration. Arousals identified with the algorithm were compared to two sets of arousals that were visually scored using polysomnography data and to a set of consensus arousals. The agreement between the two approaches, EEG arousals and those identified using ECG and actigraphy combined, calculated using Cohen's kappa was near or exceeded 0.80 for all comparisons. Landis and Koch (1977) define a kappa value above 0.8 as indicating near perfect agreement between approaches. In addition there was not a statistically significant difference between kappa values for comparisons between arousals identified based on the algorithm and EEG scored arousals and the kappa values for comparisons between the two sets of visually scored arousals. Therefore, a high level of agreement was achieved between the two approaches, polysomnography and ECG plus actigraphy. On an individual basis, for 38 out of 49 subjects a kappa value over 0.8 was obtained, for 9 a kappa value between 0.61 and 0.80 was obtained which indicates substantial agreement, and for 2 the kappa value was below 0.61. We did identify these two subjects as having apnea events. Improving the algorithm to identify these events as well as using actigraphy to obtain a measure of respiration, will be explored in the future.

Arousals were also identified using either heart rate alone or movement alone and compared to the consensus arousals. High agreement was obtained using heart rate alone; however a statistically significant improvement was made when both ECG and actigraphy are used. The agreement between arousals identified based on movement only and the consensus arousals was also high but was lower than ECG alone and ECG plus actigraphy. Also the approach used to identify movement in this analysis involves examining a continuous accelerometer signal, the signal was measured from the chest not wrist, respiration rate was detected in the signals and activity was evaluated every 1 second therefore this may be a more sensitive measurement of movement then provided by conventional actigraphy scoring algorithms and used in past actigraphy studies on the effects of noise on sleep.

The use of ECG and a continuous measurement of movement were found to provide a sensitive measure of arousals during the night. This approach is noninvasive, low in cost, and arousals can be identified automatically, quickly, and consistently between studies. In addition to detecting awakenings, repetitive autonomic arousals may contribute to more longer term health problems, and therefore the collection of heart rate data alone could be useful for understanding whether noise-induced sleep disturbance impacts cardiovascular health.

ACKNOWLEDGEMENTS

The authors would like to thank the FAA/NASA/Transport Canada/DOD/EPA PARTNER Center of Excellence and the Environment & Community Center/Forum Airport & Region, Kelsterbach, Germany, that carries out the NORAH study for providing funding for this research. Opinions, findings, conclusions and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of PARTNER or NORAH sponsoring organizations.

REFERENCES

Basner M, Brink M, Elmenhorst E-M (2012). Critical appraisal of methods for the assessment of noise effects on sleep. Noise & Health 14(61): 321-329.

Basner M, Griefahn B, Müller U et al. (2007). An ECG-based algorithm for the automatic identification of autonomic activations associated with cortical arousal. Sleep 30(10):1349-1361.

Basner M, Müller U, Elmenhorst E-M et al. (2008). Aircraft noise effects on sleep: a systematic comparison of EEG awakenings and automatically detected cardiac activations. Physiological Measurement 29(9):1089-1103.

Basner M, Samel A, Isermann U (2006). Aircraft noise effects on sleep: application of the results of a large polysomnographic field study. Journal of the Acoustical Society of America 119(5): 2772-2784.

Bonnet MH, Arand DL (2007). EEG arousal norms by age. Journal of Clinical Sleep Medicine 3(3): 271-274.

Bonnet M, Carley D, Carskadon M et al. (1992). EEG arousals: Scoring rules and examples. A preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association. Sleep 15(2): 173-184.

Bonnet MH, Doghramji K, RoehrsT et al. (2007). The scoring of arousal in sleep: reliability, validity, and alternatives. Journal of Clinical Sleep Medicine 3(2): 133-145.

Cole RJ, Kripke DF, Gruen W et al. (1992). Automatic sleep/wake identification from wrist activity. Sleep 15(5): 461-469.

Flindell I, Bullmore AJ, Robertson KA et al. (2000). Aircraft noise and sleep, 1999 UK trial methodology study. ISVR Consultancy Services, ISVR, University of Southampton, UK.

Jean-Louis G, Kripke DF, Mason WJ et al. (2001). Sleep estimation from wrist movement quantified by different actigraphic modalities. Journal of Neuroscience Methods 105(2): 185-191.

Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics 33(1):159-174.

Ollerhead JB, Jones CJ, Cadoux RE et al. (1992). Report of a field study of aircraft noise and sleep disturbance. London, United Kingdom, Department of Transport.

Passchier-Vermeer W, Vos H, Steenbekkers JHM et al. (2002). Sleep disturbance and aircraft noise exposure - exposure effect relationships. Netherlands, TNO, Report No.: Report 2002.027.

Rechtschaffen A, Kales A (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Public Health Service, U.S. Government, Printing Office, Washington, DC.

Schreckenberg D, Thomann G, Basner M (2009). FFI and FNI-two effect based aircraft noise indices at Frankfurt Airport. In Proceedings of Euronoise 2009, Edinburgh, Scotland.