An ECG-based Algorithm for the Automatic Identification of Autonomic Activations Associated with Cortical Arousal

Mathias Basner, MD, MSc[1]; Barbara Griefahn, MD[2]; Uwe Müller, PhD[1]; Gernot Plath[1]; Alexander Samel, PhD[1,†]

[1]German Aerospace Center (DLR), Institute of Aerospace Medicine, Köln, Germany; [2]Institute for Occupational Physiology at Dortmund University, Dortmund, Germany; †deceased May 19th, 2007

Objectives: EEG arousals are associated with autonomic activations. Visual EEG arousal scoring is time consuming and suffers from low interobserver agreement. We hypothesized that information on changes in heart rate alone suffice to predict the occurrence of cortical arousal.

Methods: Two visual AASM EEG arousal scorings of 56 healthy subject nights (mean age 37.0 ± 12.8 years, 26 male) were obtained. For each of 5 heartbeats following the onset of 3581 consensus EEG arousals and of an equal number of control conditions, differences to a moving median were calculated and used to estimate likelihood ratios (LRs) for 10 categories of heartbeat differences. Comparable to 5 consecutive diagnostic tests, these LRs were used to calculate the probability of heart rate responses being associated with cortical arousals.

Results: EEG and ECG arousal indexes agreed well across a wide range of decision thresholds, resulting in a receiver operating characteristic (ROC) with an area under the curve of 0.91. For the decision threshold chosen for the final analyses, a sensitivity of 68.1% and a specificity of 95.2% were obtained. ECG and EEG arousal indexes were poorly correlated (r = 0.19, P <0.001, ICC = 0.186), which could in part be attributed to 3 outliers. The Bland-Altman plot showed an unbiased estimation of EEG arousal indexes by ECG arousal indexes with a standard deviation of ± 7.9 arousals per hour sleep. In about two-thirds of all cases, ECG arousal scoring was matched by at least one (22.2%) or by both (42.5%) of the visual scorings. Sensitivity of the algorithm increased with increasing duration of EEG arousals. The ECG algorithm was also successfully validated with 30 different nights of 10 subjects (mean age 35.3 ▯ 13.6 years, 5 male).

Conclusions: In its current version, the ECG algorithm cannot replace visual EEG arousal scoring. Sensitivity for detecting <10-s EEG arousals needs to be improved. However, in a nonclinical population, it may be valuable to supplement visual EEG arousal scoring by this automatic, objective, reproducible, cheap, and time-saving method.

Keywords: Sleep; arousals; sympathetic activation; ECG; heart rate; likelihood ratio

Citation: Basner M; Griefahn B; Müller U; Plath G; Samel A. An ECG-based Algorithm for the automatic identification of autonomic activations associated with cortical arousal. *SLEEP* 2007;30(10):1349-1361.

INTRODUCTION

INTRINSIC SLEEP DISORDERS AND EXTERNAL STIMULI CAUSE CHANGES IN SLEEP STRUCTURE. WITH THE DEFINITION OF EEG AROUSALS BY THE AMERICAN Academy of Sleep Medicine (AASM) in 1992,[1] it was acknowledged that the concept of *sleep fragmentation* was not adequately captured by analyses based on the rules of Rechtschaffen and Kales.[2]

Although there is still disagreement as to whether sleep fragmentation independently influences sleep quality,[3,4] the AASM definition of EEG arousals is generally accepted today and is routinely used in clinical and research sleep laboratories. In the past few years, much insight was gained on mechanisms leading to EEG arousals and on effects of sleep fragmentation on daytime function.[5]

Two problems are frequently mentioned in the context of EEG arousal scoring; the analysis according to AASM rules is time consuming, and interobserver agreement is usually very low.[6,7] Therefore, it would be highly desirable to supplement the subjective method of visual EEG arousal analysis with a method that is less time consuming and more objective. Here, an automated algorithm would be the optimal solution, because data analysis time is minimal and there is no interobserver variability.

It has been repeatedly shown that EEG arousals are associated with transient sympathetic activation, such as increases in blood pressure and heart rate.[8-14] These activations can be objectively measured by, for example, ECG, *pulse transit time* (PTT),[15] or *peripheral arterial tonometry* (PAT);[16] therefore, it might be possible to use these signals for the detection of autonomic arousals associated with EEG arousals.

Piston et al.[15] showed a correlation between PTT and EEG frequency shifts in healthy subjects in response to external stimuli. Pillar et al.[17] proposed an autonomic arousal index (AAI) based on increases in pulse rate and decreases in PAT amplitude. Although they demonstrated good agreement between EEG arousal indexes and AAI, the applicability of the AAI is restricted as the PAT signal is not routinely recorded in sleep laboratories, which may partly be due to additional costs for the PAT device and sensors.

We developed an algorithm for the automatic identification of autonomic activations associated with cortical arousal, which is solely based on changes in heart rate, i.e., on the ECG signal. Potential advantages are that ECG data acquisition is cheap and uncomplicated (3 electrodes suffice). The ECG signal is highly reliable and less prone to artifacts than many other electrophysiological signals. It is routinely sampled during polysomnography and is implemented in most ambulatory screening systems. In the latter, information on heart rate changes could be extracted from the pulse wave signal of pulse oximeters, as well. Hence,

Disclosure Statement

This was not an industry supported study. The authors have reported no financial conflicts of interest.

Submitted for publication December, 2006

Accepted for publication June, 2007

Address correspondence to: Mathias Basner, MD, MSc, German Aerospace Center (DLR), Institute of Aerospace Medicine, 51170 Cologne, Germany; Tel: +49-2203-601-3058; Fax: +49-2203-696372; E-mail: mathias.basner@dlr.de

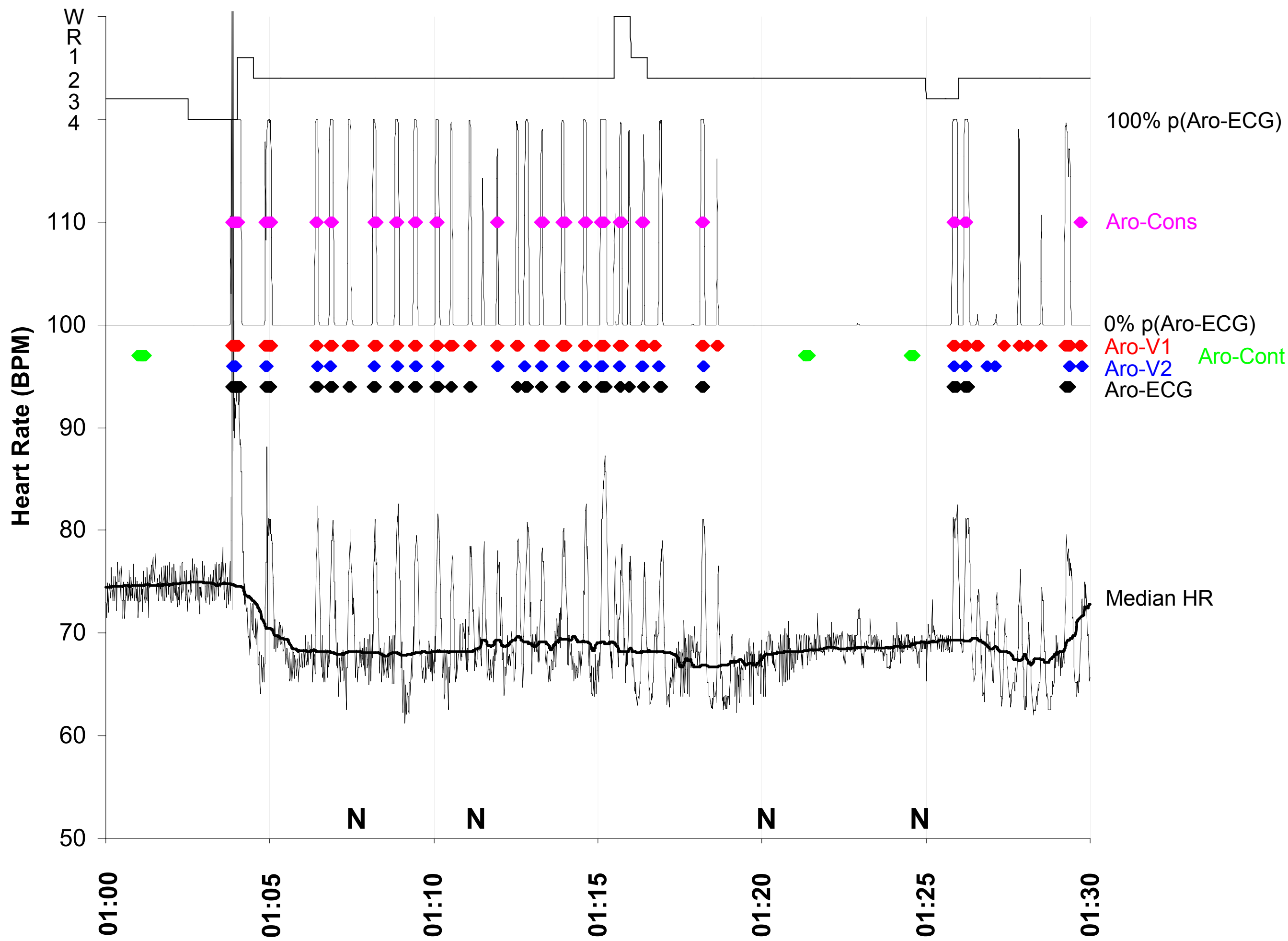

Figure 1—A period of 30 minutes of 1 subject night is shown. Beat-to-beat heart rate (thin black line) is shown in the lower part of the figure together with the median heart rate (Median HR, thick black line). A hypnogram is shown in the uppermost part of the figure. The probability of being the first heartbeat after the onset of an EEG arousal (P[Aro-ECG]) is shown below the hypnogram. The following events are indicated as colored diamonds: Consensus arousals (Aro-Cons, purple), visually scored arousals by human scorer 1 (Aro-V1, red), visually scored arousals by human scorer 2 (Aro-V2, blue), arousals scored by the ECG algorithm (Aro-ECG, black), and control arousals (Aro-Cont, green). N indicates the occurrence of an aircraft noise event. There is a body movement at 1:04 going along with a movement artifact in the ECG and a sleep stage shift from S4 to S1. Although subjects were screened for OSA in order to be excluded from the study, this period may represent several repetitive arousals caused by increased upper airway resistance.

large scale epidemiologic or experimental field studies could be performed with high explanatory power but low methodological expense, compared with polysomnographic studies.

METHODS

Subjects and Protocol

Subjects participated in a polysomnographic study on the effects of aircraft noise on sleep conducted between 1999 and 2004 at the Institute of Aerospace Medicine at the German Aerospace Center (DLR).[18] Subjects were selected in a multilevel selection process, which included a medical history, a physical examination, blood and urine samples, and audiometry. A one-night screening of peripheral hemoglobin oxygen saturation and pulse rate was performed in most of the subjects. Subjects were required to have normal hearing thresholds for age and no history of loud snoring. The study protocol was approved by the ethics commission of the Medical Board of the district North Rhine. Subjects signed an informed consent form before the start of study. The study consisted of laboratory and field phases.

Field Phase for Algorithm Development

Field phase subjects were living in the vicinity of Airport Cologne/Bonn, a freight hub with high densities of nocturnal air traffic. They were polysomnographically investigated for 9 consecutive nights. Sound pressure level (SPL) and noise events were recorded inside and outside the bedroom. People could adhere to their usual bed times, which had to include the period between midnight and 06:00. Design and methods applied in the field phase are described in detail in Basner et al.[18,19] Fifty-six subjects (mean [SD] = 37.0 [12.8] years, 26 male) contributed one night each to the data set that was used for algorithm development. Nights with the highest traffic density (mean [SD] = 7.9 [2.9] aircraft noise events per hour sleep period time [SPT] on average) were chosen. SPT was defined as the time between the first occurrence of stage 2 and the last occurrence of any sleep stage but *wake*.

Laboratory Phase for Algorithm Validation

Laboratory phase subjects were investigated in groups of 8 for 13 consecutive nights in the underground sleep facility of the Institute of Aerospace Medicine. Here, between 4 and 128 aircraft noise events per night with maximum SPLs between 45 and 80 dBA were realistically played back via loudspeakers in the 8 separate bedrooms. Lights were turned off at 23:00 and on again at 07:00. Design and methods applied in the laboratory study are described in detail in Basner et al.[18,20] Ten subjects (mean [SD] = 35.3 [13.6] years, 5 male) contributed 3 nights each for the validation of the ECG algorithm.

Polysomnography

Sleep was polysomnographically recorded with an EEG-amplifier developed at the German Aerospace Center and using the standard setup (EEG C3-A2, C4-A1, left and right EOG and submental EMG). The following additional channels were recorded: nasal airflow (thermistor), electrocardiogram (ECG), chest wall motion (piezo electrodes), body position, actimetry, and SPL and light intensity in the bedroom. Sleep stages were classified by 2 experienced scorers following standard criteria,[2] using 30-second epochs. The interobserver agreement was 88.1% on average (mean total Cohen's[21] kappa 0.812). Body movements accounting for more than half of the epoch that would have otherwise been scored as *movement time* were classified as *wake*, because it was assumed that these kinds of movements do not occur without respective cortical activation.

Electrocardiogram Analysis

The ECG was derived from the chest wall (derivation Einthoven II) and sampled at 1000 Hz. R-waves were automatically detected with a software developed in a LabVIEW environment.[22] The time of each heartbeat was stored in a separate line of an ASCII file together with the heart rate (beats per minute [bpm]) derived from the interval to the preceding beat. Printouts of heartbeat against time were generated automatically for 30-min intervals (see Figure 1) and inspected visually. The beginning and the end of periods with signal loss (e.g., due to temporary disconnections or loose electrodes) were recorded. This procedure took less than 1 to 2 minutes per night. Periods with ECG signal loss were later excluded from the analysis. Of the 56 nights of the field study, there were 23 nights (41%) with ≥1 period of signal loss, covering 1.7% of SPT. Of the 30 nights of the laboratory study, there were 2 nights (6.7%) with at least 1 period of signal loss, covering 0.1% of SPT.

Similar to a moving average, a moving median (see "Median HR" line in Figure 1) was calculated with a time window of ± 90 seconds for every heartbeat in SPT. The moving median was aimed at following trends in basic heart rate throughout the night (e.g., notice the sudden decrease in heart rate between 01:04 and 01:05 in Figure 1). For beats at the beginning or end of SPT, beats before sleep onset and after the final awakening were included in the calculation of the moving median, while periods of ECG signal loss were excluded.

EEG Arousal Analysis

EEG arousals of the 56 field and the 30 laboratory nights were scored according to AASM criteria[1]; i.e., any EEG frequency shift for ≥3 seconds was scored as an arousal in NREM sleep. During REM sleep, an increase in EMG was needed as well. Two visual AASM arousal scorings were obtained for each night (Aro-V1 and Aro-V2). In order to increase transferability, the scorers were deliberately chosen from 2 different centers, DLR (V1) and a collaborating clinical center at Marburg University (V2). An arousal was called consensus arousal (Aro-Cons) if the start of Aro-V1 differed from the start of Aro-V2 by ± 5 s or less. If an arousal met these criteria, start time and duration of Aro-Cons were always obtained from scorer V1, as this scorer marked the start of the arousal event in 87.8% of consensus arousals earlier or at the same time as scorer V2.

For the development of the ECG algorithm, only consensus arousals were used. For each Aro-Cons, a control arousal (Aro-Cont) was randomly placed by a computer program in parts of the night without indication of cortical arousal, i.e. control arousals were not allowed to overlap with Aro-V1, Aro-V2, R&K wake epochs, or times of ECG signal loss (including a safety margin of 60 s). Each Aro-Cont had the same length as its Aro-Cons counterpart. We additionally required that the average heart rate of 3 beats prior to the onset of Aro-Cons and Aro-Cont should differ by less than one bpm. With this procedure, the prevalence of both Aro-Cons and Aro-Cont was artificially set to 0.5, and it was possible to calculate the sensitivity of the ECG algorithm based on Aro-Cons and its specificity based on Aro-Cont. The different types of arousals are shown in Figure 1. For Aro-V1 and Aro-V2, arousal indexes were calculated as the number of arousals per hour of TST minus periods of ECG-signal loss.

Likelihood Ratio Estimation

For each Aro-Cons and Aro-Cont, and for every beat from 2 beats before (b2) to 38 beats after (a38) arousal onset, the difference in heart rate of the beat under investigation to the median heart rate was calculated and stored. Based on the pooled differences of both Aro-Cons and Aro-Cont conditions, likelihood ratios (LRs) were calculated for 10 intervals of heart rate differences. A computer algorithm assured that the number of data points was similar for each of the 10 intervals. If Aro-Cons contributed more data points to an interval than Aro-Cont, LRs greater than 1 resulted. They indicated that heart beat differences falling in such an interval were more likely to be associated with an EEG arousal than with a control condition. Likewise, LRs less than 1 indicated that heart beat differences were less likely to be associated with an EEG arousal than with a control condition.

Receiver operating characteristics (ROC) were generated for each of the 40 categories (b2 to a38), and areas under the curve (AUC) were calculated. They show how accurately each specific heart beat (b2 to a38) would be able to differentiate between arousal and control conditions if applied as a single "diagnostic test." Based on the 40 ROC plots and the associated LRs it was decided to use 5 consecutive beats (a1 to a5) only for the calculation of the probability of each heartbeat being the first beat after the onset of an EEG arousal.

Probability of EEG Arousals Based on Likelihood Ratios

For each heartbeat in SPT, the probability of this beat being the first beat after the onset of an EEG arousal was calculated

Table 1—Likelihood Ratios

Category	a1 LL	a1 UL	a1 LR	a2 LL	a2 UL	a2 LR	a3 LL	a3 UL	a3 LR	a4 LL	a4 UL	a4 LR	a5 LL	a5 UL	a5 LR
1		-3.4	0.5361		-2.8	0.2216		-2.6	0.1469		-2.7	0.1447		-2.8	0.2302
2	-3.3	-1.6	0.5456	-2.7	-1.1	0.2145	-2.5	-1.0	0.1508	-2.6	-1.0	0.0863	-2.7	-1.1	0.1071
3	-1.5	-0.6	0.4153	-1.0	0.0	0.3117	-0.9	0.1	0.1610	-0.9	0.2	0.1212	-1.0	0.1	0.1475
4	-0.5	0.2	0.5675	0.1	1.0	0.3672	0.2	1.4	0.2395	0.3	1.5	0.2376	0.2	1.4	0.2053
5	0.3	1.1	0.5622	1.1	2.1	0.5299	1.5	2.7	0.5021	1.6	3.1	0.4836	1.5	3.0	0.3760
6	1.2	2.0	0.8710	2.2	3.5	1.2317	2.8	4.6	1.1652	3.2	5.3	1.2566	3.1	5.3	1.0231
7	2.1	3.3	1.0482	3.6	5.3	1.8880	4.7	7.1	3.1445	5.4	8.2	3.7838	5.4	8.8	3.6601
8	3.4	5.3	1.7931	5.4	7.8	3.6340	7.2	10.1	7.1163	8.3	11.8	11.9273	8.9	12.6	15.3023
9	5.4	8.8	4.0993	7.9	12.1	7.8734	10.2	14.7	20.4848	11.9	16.3	42.8125	12.7	17.4	38.5556
10	8.9		5.7642	12.2		16.3659	14.8		33.0952	16.4		57.7500	17.5		100.0000

Likelihood ratios (LR) for the 1st beat after the onset of an EEG arousal (a1) until the 5th beat after the onset of an EEG arousal (a5). The lower limit (LL) and upper limit (UL) of differences to median heart rate are shown together with the associated LR for each of 10 categories. Category 1 includes all values equal to or below the UL of this category, while category 10 includes all values equal to or above the LL of this category.

Table 2—Example of EEG Arousal Probability Calculation Based on Likelihood Ratios

Beat#	Median HR	Actual HR	Difference	Category	LR	Prior Prob.	Post. Prob.
a1	49.8	54.7	+4.9	8	1.7931	0.0040	0.0071
a2	49.8	54.3	+4.5	7	1.8880	0.0071	0.0134
a3	49.8	54.7	+4.9	7	3.1445	0.0134	0.0410
a4	49.8	55.6	+5.8	7	3.7838	0.0410	0.1392
a5	49.9	64.8	+14.9	9	38.5556	0.1392	0.8618

Calculation of the probability of being the first beat after the onset of an EEG arousal based on differences of 5 consecutive beats (a1 to a5) to median heart rate (example). Notice that the posterior probability of one "test" is the prior probability of the next "test". In this example, the prior probability of 0.004 increased to a probability of 0.8618 after the 5 "tests". HR: heart rate, LR: likelihood ratio, Prior Prob.: Prior Probability, Post. Prob.: Posterior Probability

Table 3—Pearson's Moment Correlation Coefficients and Associated P-Values for 4 Arousal Scorings and a Visual Analogue Scale (VAS) Score with "Alert" and "Tired" as Anchors

	ECG-algorithm	Visual Scoring 1	Visual Scoring 2	Average of Visual Scoring 1 and 2
Visual Scoring 1	0.284*			
	*0.559****	1		
Visual Scoring 2	0.078	0.766***		
	*0.327**	*0.660****	1	
Average of Visual	0.191	0.938***	0.941***	
Scorings 1 and 2	*0.488****	*0.913****	*0.909****	1
VAS alert/sleepy	0.041	0.134	0.129	0.140
	0.070	*0.085*	*0.079*	*0.090*

Above: r including the 3 outliers shown in Figures 6A and 6B; below (in italics): r excluding the 3 outliers; * = P <0.05, ** = P <0.01, *** = P <0.001

based on prior probability (PrP) and LRs. PrP was calculated as the average of Aro-V1 and Aro-V2 (5940 arousals) divided by the total number of investigated valid heart beats during SPT (1,446,174 beats), yielding a ratio of ~0.004. Hence, on average every 250th heartbeat is likely to be the first beat after the onset of an EEG arousal. Arousals during periods with ECG-signal loss were not considered in this calculation. The prior probability (PrP) was transformed into prior odds (PrO) according to equation (1):

(1). $PrO = PrP / (1 - PrP)$

If there is just one category (e.g., a1), comparable to a single diagnostic test, the difference between the heart rate measured at a1 to the median heart rate is associated with a specific LR. The posterior odds (PstO) are then calculated by multiplying the prior odds (PrO) with the LR, which can be transformed in the posterior probability (PstP) according to equation (2):

(2). $PstP = PstO / (1 + PstO)$

If there is more than one category (e.g., a1 to a5), comparable to several sequential diagnostic tests, PstO is calculated by multiplying all LRs and the PrO with each other, exemplified in equation (3) for 5 categories:

(3). $PstO = PrO * LR_{a1} * LR_{a2} * LR_{a3} * LR_{a4} * LR_{a5}$

For this calculation, it is assumed that the tests are conditionally independent, given the arousal status. The calculation of the probability of an EEG arousal based on LRs is exemplified in Table 2 in the results section. The probability of being the first beat after the onset of an EEG arousal is also shown in Figure 1 as P(Aro-ECG).

Table 4—Pair-Wise Comparison of EEG-Arousal Detection Accuracy Using One Method as Gold Standard

Gold Standard	Method A	Sensitivity Method A	Specificity Method A	Method B	Sensitivity Method B	Specificity Method B	rTPF(A,B) [95% CL]	rFPF(A,B) [95% CL]
V2	V1	0.842	0.902	ECG	0.485	0.921	1.74 [1.57; 1.92]	1.25 [1.01; 1.55]
V1	V2	0.733	0.947	ECG	0.473	0.937	1.55 [1.40; 1.72]	0.85 [0.63; 1.15]
ECG	V1	0.745	0.821	V2	0.664	0.848	1.12 [1.04; 1.20]	1.17 [1.06; 1.30]

V1: visual arousal scoring by human scorer 1; V2: visual arousal scoring by human scorer 2; ECG: arousal scoring by ECG algorithm; rTPF: relative true positive fraction; rFPF: relative false positive fraction; CL: confidence limits

Decision Threshold Determination

Although it directly conveys information about the relevance of heart rate increases, the probability of being the first beat after the onset of an EEG arousal alone is not helpful in differentiating between EEG arousals and non-EEG arousals. A decision threshold is needed (for definition and examples, refer to chapter 3 of Hunink et al.[23]). A two-dimensional decision-threshold space was chosen with number of consecutive beats (dimension 1, values ranging between 1 and 15) above a certain probability (dimension 2, values ranging between 0 and 1) of being the first beat after the onset of an EEG arousal. Based on the number of consecutive beats above the EEG arousal probability, 15 ROC plots were generated for decision thresholds in EEG arousal probability between 0 and 1. A decision threshold with a sensitivity of 68.2% and a specificity of 95.1% was chosen for the following analyses. For this threshold, the beginning and the end of all ECG arousals (Aro-ECG) was determined and stored. Similar to the AASM EEG arousal definition, 2 periods with probabilities above the decision threshold had to be separated by ≥10 seconds with probabilities below the decision threshold to be scored as 2 single events; otherwise they were merged into one event including the interposed period. ECG arousals are shown in Figure 1 in black. EEG arousals are per definition not scored during periods of wakefulness, but these periods were not excluded for ECG arousal analyses. Therefore, the ECG arousal index was based on SPT, and not on TST. It was calculated as the number of arousals per hour of SPT minus periods of ECG-signal loss.

Comparison of Arousal Scorings

Indexes of Aro-V1, Aro-V2 and Aro-ECG were compared for differences between groups with the Friedman test. Wilcoxon tests were performed post hoc to test for pair wise differences, and the significance level was Bonferroni corrected (0.05/3 = 0.017). Pearson's moment correlation coefficients were calculated pair wise for the indexes of Aro-ECG, Aro-V1, Aro-V2, the mean of Aro-V1 and Aro-V2, and a visual analogue scale (VAS) score. Using 10-cm analogue scales the participants assessed their actual state between the extremes alert - tired, active - idle, fresh - sleepy, and enterprising - languid. Statistical tests were performed with SPSS 11.5.1.

Usually, a new method is compared to a gold standard. Here, 2 visual EEG arousal scorings were surrogates for the gold standard. As the true EEG arousal score was unknown, the mean of the 2 scorings was the best estimate of the true score. Therefore, scatter plots as well as Bland-Altman-plots[24] were generated comparing the index of Aro-ECG with the mean of Aro-V1 and Aro-V2 indexes. A case 3 intraclass correlation coefficient (ICC) according to Shrout and Fleiss[25] was calculated with SAS (version 9.1). Additionally, the agreement between the 3 scoring methods was visualized in a Venn diagram, and the proportion of specific agreement (PSA) was calculated. Ninety-five percent confidence limits of PSA were calculated with a bootstrap method based on 1000 replications.

Algorithm Validation

The validation of the ECG algorithm was based on data of 10 subjects who participated in the laboratory study on the effects of aircraft noise on sleep and contributed 3 nights each to the validation data set: A noise-free baseline night, a night with 64 identical noise events with maximum SPLs of 45 dBA each, and a night with 64 identical noise events with maximum SPLs of 65 dBA each. The 45 dBA noise event was generated by reducing the SPL of the 65 dBA noise event by 20 dBA, both lasted for about 25 seconds and reached their maximum SPL after about 15 seconds. Five of the 10 subjects participated both in field and laboratory phases, and one of their field phase nights was one of the 56 nights used for algorithm development.

The validation of the ECG algorithm consisted of 2 parts. In part 1, agreement between EEG arousal scorings obtained by human scorer 1, human scorer 2 and the ECG algorithm was calculated based on all arousals scored during SPT. In part 2, the analysis was restricted to the first 20 s after the start of aircraft noise events that were screened for arousal onset. A period of 10 s prior to noise onset had to be free of arousals in order for that noise event to be included in the analysis. Spontaneous arousal probability was estimated in noise-free baseline nights at the same time noise events occurred in 45 dBA nights. In this way, dose-response relationships were established for EEG arousal scorings of human 1, human 2, and the ECG algorithm, with no noise, 45 dBA and 65 dBA conditions representing the dose and arousal probability representing the response.

RESULTS

ECG Analysis

As shown in Figure 2, major findings of studies on the relationship between EEG arousals and cardiac activation were replicated

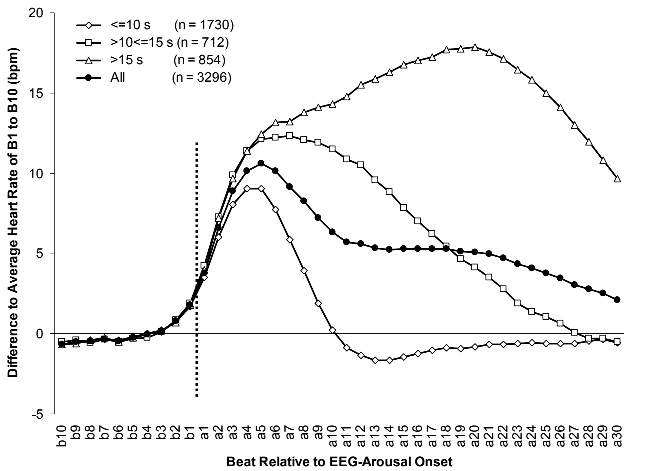

Figure 2—Mean differences in heart rate from 10 beats before (b10) until 30 beats after (a30) EEG arousal onset (vertical line) compared to the average of b10 to b01 depending on the duration of the arousal. 285 EEG arousals were excluded from the analysis because ≥1 of the 40 heart rate values being below 30 bpm or above 120 bpm, indicating an artifact or ectopic heartbeat.

in this investigation.[8,10,13] Descriptively, heart rate started to increase 2 to 3 beats prior to onset of visually scored EEG arousals. Amplitude as well as duration of the increase in heart rate depended on the duration of the EEG arousal. Short EEG arousals were associated with relatively small increases in heart rate, which reached or even fell below the prearousal level rapidly. In contrast, long arousals were associated with prolonged and profound increases in heart rate. In the case of EEG arousals lasting >15 seconds, prearousal levels were still not reached 30 beats after the onset of the arousal. In EEG arousals ≤10 seconds, as well as in the whole group, the maximum average increase in heart rate was reached 5 beats after the onset of the EEG arousal, and heart rate started to return to the prearousal levels afterwards. Hence, it was decided to use only the first 5 beats (a1 to a5) after the onset of an EEG arousal for the calculation of LRs. Of consensus arousals, 49.7% (n = 1779) lasted ≤10 seconds, 20.8% (n = 745) lasted >10 seconds but ≤15 seconds, and 29.5% (n = 1057) lasted >15 seconds.

Figure 3 visualizes cardiac activation for control conditions (Aro-Cont, left in Figure 3) and for consensus EEG arousal (Aro-Cons, right in Figure 3) in a 3D-plot. Here, several histograms with the difference to the median heart rate on the x-axis and the relative frequency of each category on the y-axis are shown from the first heartbeat before EEG arousal onset (-1) to the 21st beat after EEG arousal onset (21). In the control condition, differences to median heart rate were symmetrically distributed around zero from beat -1 to beat 21. In case of consensus arousals, differences to median heart rate were symmetrically distributed around zero before EEG arousal onset. With the onset of the EEG arousal, the distributions swung to the right, i.e., to increased heart rates. Simultaneously, variability increased, i.e., the distributions became wider and more flat. After 5 to 10 heartbeats, the distributions swung back to zero difference.

It is the information embedded in the shift of these distributions that was used for the derivation of an ECG algorithm for the detection of cardiac activations associated with EEG arousals in this publication.

Likelihood Ratios

A total of 7162 heart beat differences (calculated in bpm), 3581 derived from Aro-Cons and 3581 from Aro-Cont, were rounded to one decimal place and sorted in ascending order. The values were assigned to 10 categories with different ranges of heart rate differences. An algorithm assured that each category contained

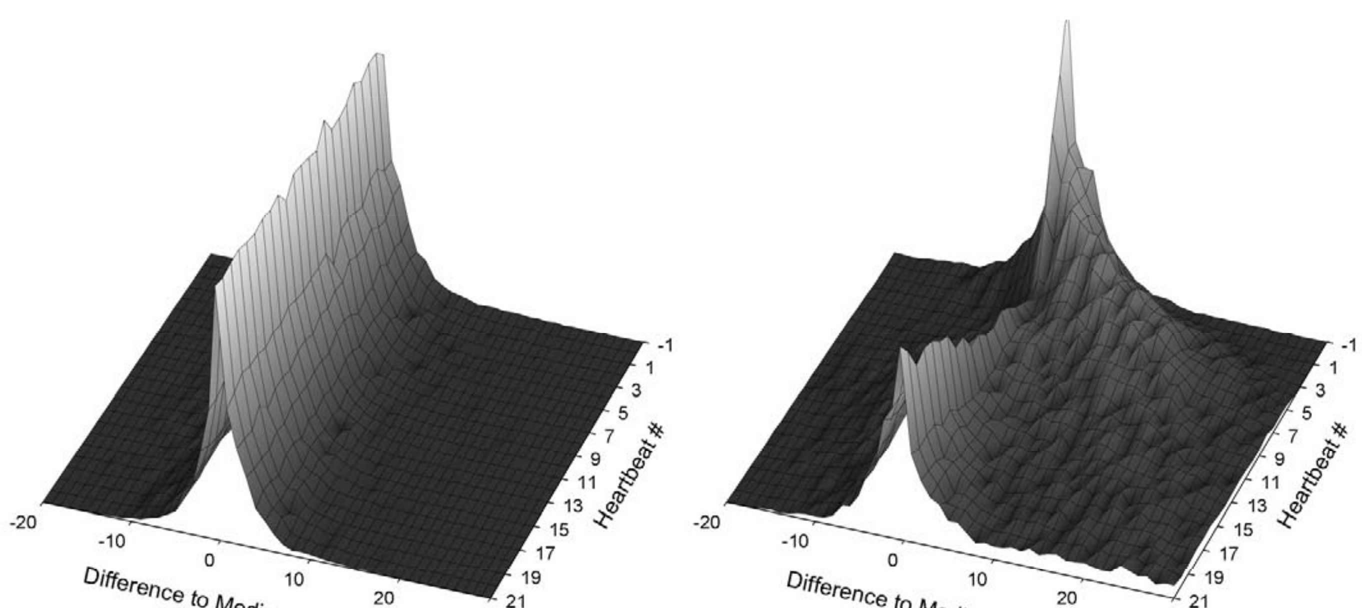

Figure 3—3D-plot of the heart rate responses in control conditions (Aro-Cont, left) and in consensus EEG arousal conditions (Aro-Cons, right) from one beat prior (-1) until 21 beats after event onset (z-axis). The difference to a median heart rate, calculated for a window of ± 90 seconds, is shown on the x-axis. The y-axis (not shown) contains data on the relative frequency of each category.

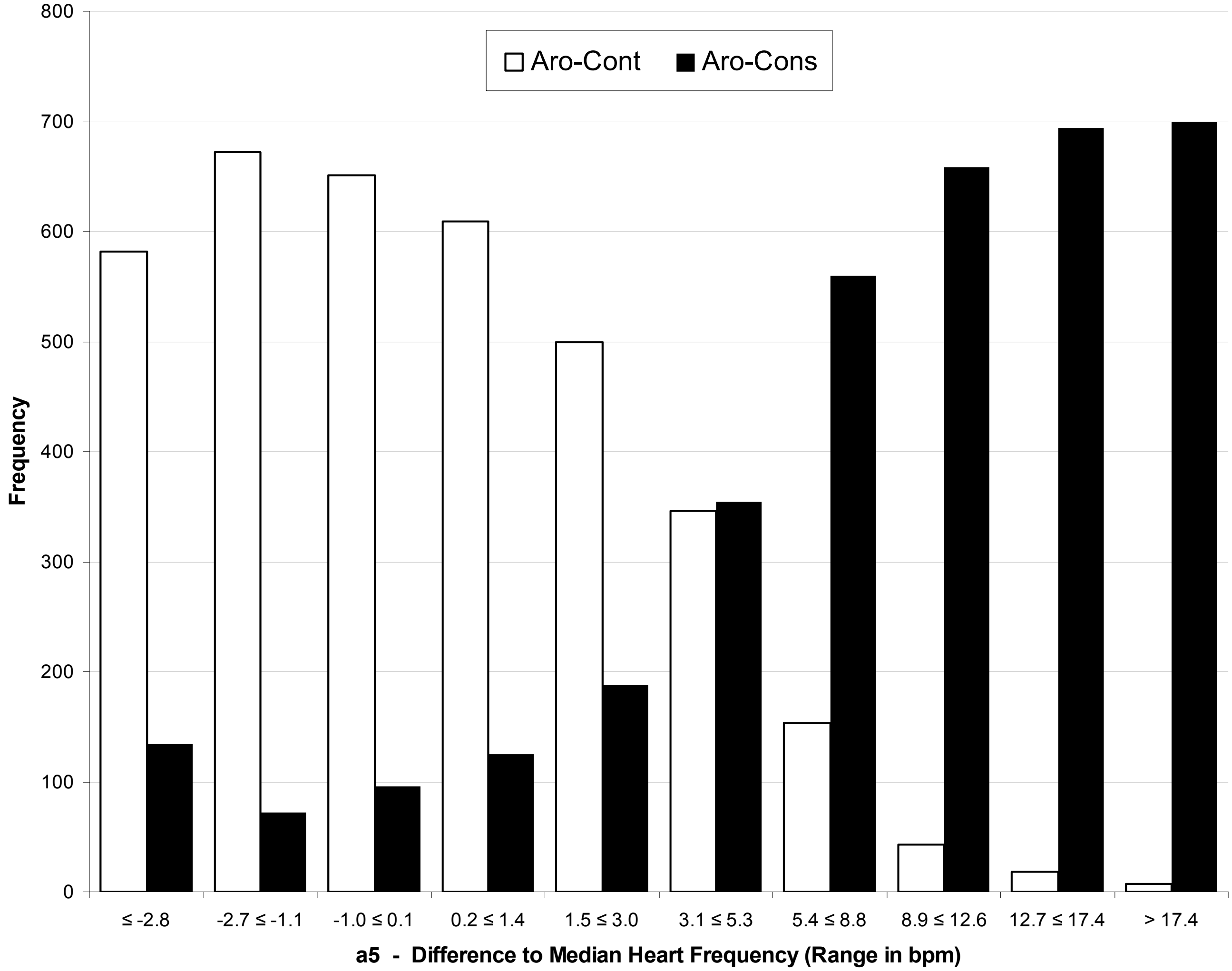

Figure 4—The differences of 3581 heartbeats located 5 beats after EEG arousal onset to median heart rate were calculated both for consensus arousals (Aro-Cons) and control conditions (Aro-Cont) and plotted for 10 categories of heartbeat differences to a moving median. Differences in the control condition were symmetrically distributed around zero, whereas in the EEG arousal condition the prevalence of positive differences increased continuously. As the prevalence of Aro-Cons and Aro-Cont was artificially kept at 0.5, the LR in favor of an EEG arousal can be calculated by dividing the frequency of Aro-Cons values by the frequency of Aro-Cont values in each category. For the last category (10), this procedure lead to a LR of 700/7=100.0, i.e. heart rate differences of more than 17.4 bpm were 100 times more likely associated with an EEG arousal than with a control condition.

approximately 716 values, i.e., one tenth of all 7162 differences. The number of data points per category ranged from 686 to 749. LRs used for the calculation of posterior probabilities are shown in Table 1 for the 1st beat after the onset of an EEG arousal (a1) until the 5th beat after the onset of an EEG arousal (a5). The calculation of LRs is exemplified for the 5th beat after the onset of an EEG arousal (a5) in Figure 4.

Posterior Probabilities

For each heartbeat in SPT, the probability of this beat being the first beat after the onset of an EEG arousal was calculated based on prior probability (PrP) and LRs that depended on the difference of the actual heartbeat (a1) and the following 4 heartbeats (a2 to a5) to the momentary median heart frequency. This procedure is described in detail in the methods section and exemplified for one heartbeat in Table 2. The probability of being the first beat after the onset of an EEG arousal is also shown in Figure 1 as P(Aro-ECG).

Decision Threshold Determination

With a value of 0.912, the AUC, and therefore the accuracy of the criterion, was maximal for the "4 consecutive beats" condition. The empirical ROC plot for this condition is shown in Figure 5. As with every diagnostic test, there are tradeoffs between sensitivity and specificity depending on the choice of the decision threshold. For the results presented below, a decision threshold of 4 consecutive transgressions >35% probability of being the first beat after the onset of an EEG arousal was chosen, going along with a sensitivity of 68.2% and a specificity of 95.1%. Therefore, more than 2 of 3 consensus EEG arousals (Aro-Cons) will be identified correctly (true positives) by the ECG algorithm, whereas less than 5 out of 100 non-EEG arousals (Aro-Cont) will be incorrectly identified as an EEG arousal (false positives). The decision threshold was chosen in favor of specificity as the prevalence of EEG arousals (4 per 1000 heartbeats) was rather low, a situation where a low specificity will decrease the positive predic-

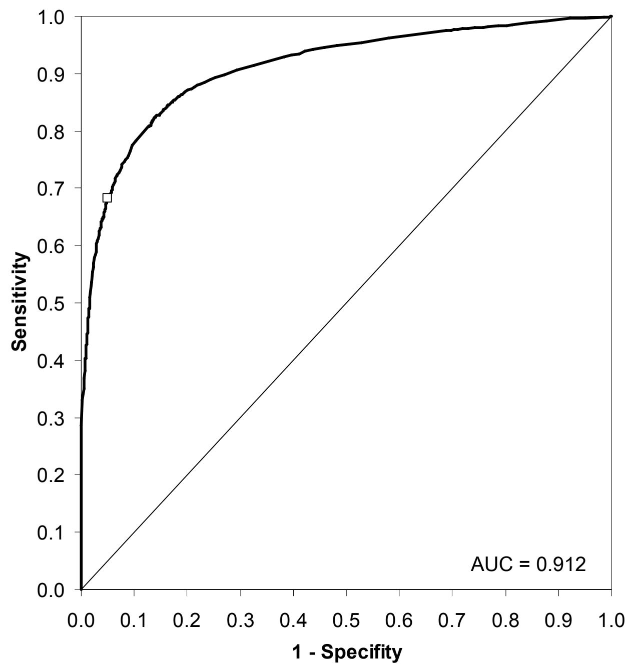

Figure 5—Receiver operating characteristic (ROC) curve for the 4 consecutive heartbeats condition and decision thresholds of EEG arousal probabilities (as estimated by the ECG algorithm) between 0 and 1. The decision threshold used to predict EEG arousals is shown as a white square.

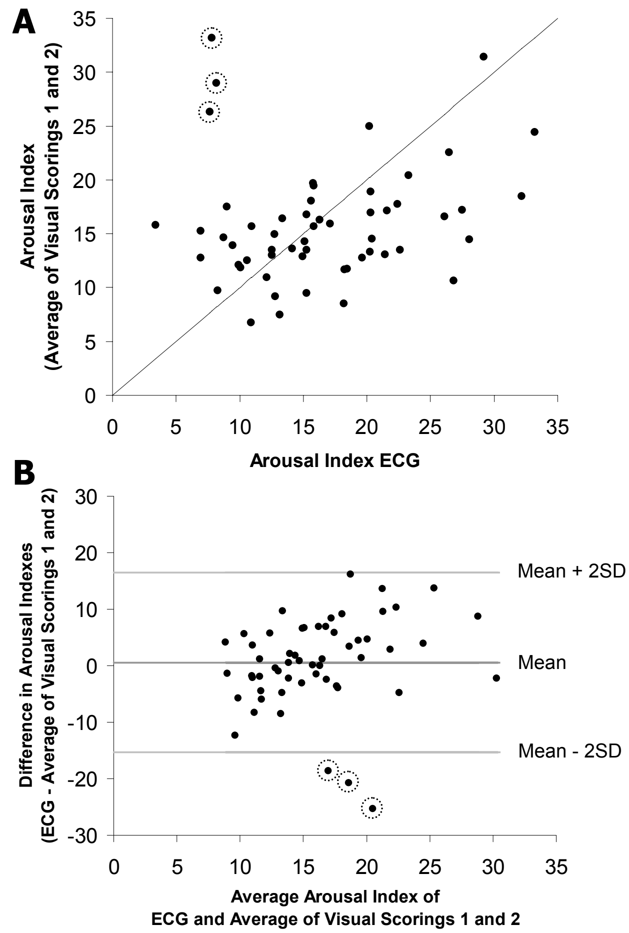

Figure 6—A: Scatterplot of ECG arousal index versus the average of visual arousal indexes. The line represents identical x- and y-values. Three outliers are circled. B: Bland-Altman plot of ECG arousal index versus the average of visual arousal indexes. Lines indicate the average difference as well as ± 2 standard deviations. Three outliers are circled.

tive value of a test result dramatically. The position of the decision threshold in ROC space is shown in Figure 5 as a white square.

Comparison of Arousal Scorings

If periods with ECG-signal loss were included, human 1 scored 7138 EEG arousals and human 2 scored 4883 EEG arousals in total. With the strict rule that the start of 2 arousals had to differ by less than 5 seconds to be called a consensus arousal, 3590 arousals were scored by both, 3548 by human scorer 1 alone and 1293 by human scorer 2 alone. With a more lenient definition, where any overlap in arousal scorings sufficed for an agreement, 3886 arousals were scored by both scorers.

If periods with ECG-signal loss were excluded, there were 3581 consensus EEG arousals (9.5 per hour TST), human 1 scored 7047 EEG arousals (18.9 per hour TST), human 2 scored 4834 EEG arousals (12.9 per hour TST), and the ECG algorithm scored 6641 arousals (16.5 arousals per hour SPT). The Friedman test indicated a significant difference between arousal indexes scored by human scorer 1, human scorer 2 and the ECG algorithm (df=2, χ^2=47.2, P <0.001). Post hoc Wilcoxon Tests showed that all 3 scorings differed significantly from each other with P ≤0.05.

Figure 6A shows a scatterplot of the ECG arousal index and the average of the two visual arousal indexes. Three outliers with low ECG arousal indexes but high visual arousal indexes are circled. Pair wise linear Pearson's correlation coefficients were calculated for ECG arousal index, visual arousal indexes (Aro-V1 and Aro-V2), the average of the 2 visual arousal indexes, and the VAS score. The results are shown in Table 3. All 3 scorings were positively correlated. The correlation decreased in the order visual scoring 1/visual scoring 2 (r = 0.766), ECG algorithm/visual scoring 1 (r = 0.284) and ECG algorithm/visual scoring 2 (r = 0.078). The ICC for the average of 2 visual arousal indexes and the ECG arousal index was 0.186. If the 3 outliers were excluded, correlation coefficients between the ECG arousal index and arousal indexes obtained by visual scorings increased markedly. The ICC (0.449) increased as well. The VAS score was not linearly correlated with any of the arousal measures. Inferences remained unchanged regardless of whether Pearson's or Spearman correlations were calculated.

A Bland-Altman-plot[24] is shown in Figure 6B. The ECG arousal index was practically unbiased, as it overestimated the "true" arousal index by less than 1 arousal per h (+2 SD: +16.5 arousals per h, -2 SD: -15.3 arousals per h). However, the ECG arousal index tended to underestimate low average visual arousal indexes and it tended to overestimate high average visual arousal indexes. The 3 outliers that were already shown in the scatterplott in Figure 6A are again shown and circled in Figure 6B. They were the only data points falling outside the ±2 SD range.

A Venn diagram of 10,123 cases where ≥1 of the 3 methods scored an arousal event is shown in Figure 7A. Here, in contrast to the strict criterion for consensus arousals, agreement was defined as any overlap of arousal periods. Cases with all 3 scorings agreeing, with 2 scorings agreeing, and with no agreement were differentiated and visualized in the Venn diagram. If the

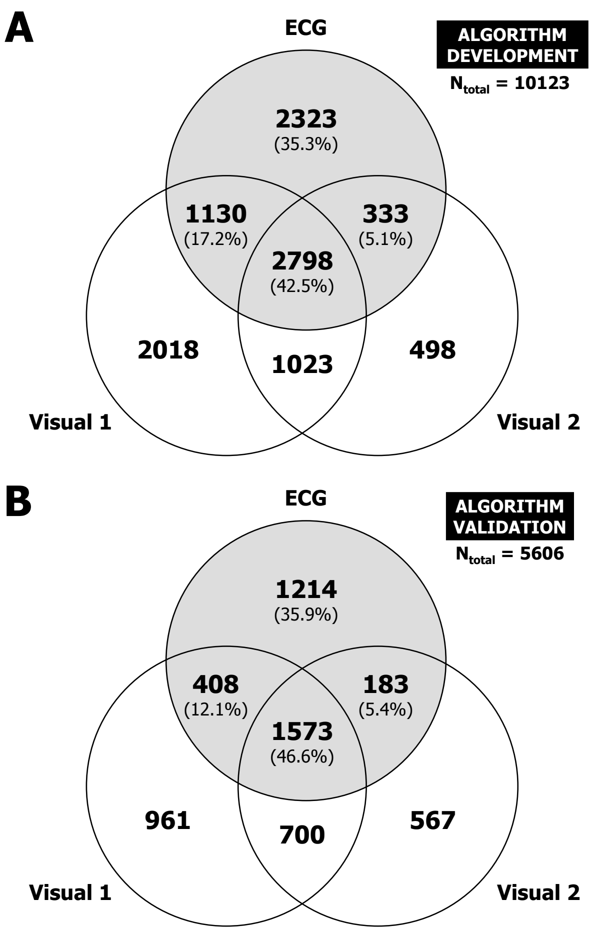

Figure 7—Venn diagrams for arousal scorings of human scorer 1 (Visual 1), human scorer 2 (Visual 2), and the ECG algorithm (ECG) based on nights used for algorithm development (A) and algorithm validation (B). The percentages refer to all arousals scored by the ECG algorithm (gray circle).

arousal period of method 1 overlapped with the arousal period of method 2, and the latter overlapped with the arousal period of method 3, but arousal periods of methods 1 and 3 did not overlap; this was nevertheless considered as one case (with 3 agreements). With this definition, a case with 2 or 3 agreements may contain >1 arousal event scored by the same method every time an arousal event scored by 1 method covers both the end and the beginning of 2 separate arousal events scored by another method. Hence, the reported numbers of arousals scored by each method slightly exceed the numbers shown in the Venn diagram.

In 2798 cases (27.6%) all 3 scorings agreed, in 2486 cases (24.6%) 2 of 3 scorings agreed, and in 4839 cases (47.8%) an arousal event was scored by 1 of the methods with neither of the 2 other methods agreeing.

If only arousal events scored by the ECG algorithm are taken into account, none of the other scorings agreed in 35.3%, one of the other scorings agreed in 22.2% and both other scorings agreed in 42.5%, i.e., in about 2/3 of arousals scored by the ECG algorithm at least 1 of the other 2 scorings agreed. Of the 2323 arousals scored only by the ECG algorithm, 160 (6.9%) occurred during prolonged periods of wakefulness, i.e., during an R&K wake epoch preceded by ≥1 other wake epoch.

Overall, 73.2% of Aro-Cons, i.e., 2798 of 2798 + 1023 arousals, were correctly identified by the ECG algorithm. This number surpasses the expected value of 68.2% by 5%, as the more lenient criteria for consensus arousals were used for this comparison. The proportion of specific agreement between the ECG algorithm and both visual scorings was 59.2% (95% confidence limits 58.3% to 60.2%), and therefore lower than PSA between visual scorings (65.8%, 95% confidence limits 64.8% to 66.7%).

The sensitivity of the ECG algorithm increased with a simultaneous increase in the duration of the EEG arousal visually scored by human scorer 1; 56.2% of EEG arousals ≤10 s, 80.9% of EEG arousals >10 ≤15 s, and 92.5% of EEG arousals >15 s were correctly identified by the ECG algorithm, respectively.

Validation of the ECG Algorithm

The algorithm was validated with 30 laboratory nights of 10 subjects; recordings were analyzed visually by the same 2 scorers and by the ECG algorithm. A Venn diagram of 5606 cases where ≥1one of the 3 methods scored an arousal is shown in Figure 7B. Overall, 69.2% of visual consensus arousals, i.e., 1573 of 1573 + 700 arousals, were correctly identified by the ECG algorithm. There was no relevant change in ECG algorithm performance compared to algorithm development: None of the other methods agreed in 35.9% (was: 35.3%); one of the other methods agreed in 17.5% (was: 22.2%); and both other methods agreed in 46.6% (was: 42.5%). Therefore, compared to algorithm development, cases with 2 agreements slightly decreased in favor of cases with 3 agreements. The proportion of specific agreement between the ECG algorithm and both visual scorings decreased to 55.7% (95% confidence limits 54.4% to 57.0%; was 59.2%), while it increased between visual scorings (68.2%, 95% confidence limits 66.9% to 69.5%; was 65.8%). Of the 1214 arousals scored only by the ECG algorithm, 220 (18.1%) occurred during prolonged periods of wakefulness, and therefore per definition could not have been scored by the other 2 methods. Taking this into account, overall agreement with ≥1 of the other 2 methods was 68.5%.

Results of the event-related analysis are shown in Figure 8. On average, 575 events contributed to each of the 3 x 3 estimated probabilities. Dose-response relationships were observed for all scoring methods, with increasing arousal probability in the order control, 45 dBA and 65 dBA. For the ECG algorithm, arousal probability increased from 8.5% (control) and 14.6% (45 dBA) to 30.7% (65 dBA). Arousal probability in the control condition did not differ between the 3 scorings (Aro-V1: 8.3%, Aro-V2: 8.6%, Aro-ECG: 8.5%; all pair-wise $P > 0.8$ in a random subject effect logistic regression). The ECG algorithm detected significantly less arousals than both human scorers in the 45 dB condition and in the 65 dB condition (all $P < 0.05$). Visual scoring 1 and 2 differed significantly in the 45 dB condition ($P = 0.008$) but not in the 65 dB condition ($P = 0.274$).

Table 4 shows pair-wise comparisons of detection accuracy using 1 of the 3 methods as the gold standard. Compared to the ECG algorithm, human scorer 1 was significantly more sensitive (rTPF = 1.74) and significantly less specific (rFPF = 1.25), while human scorer 2 was significantly more sensitive (rTPF = 1.55) and more specific (rFPF = 0.85), albeit not significantly. Compared to human scorer 2, human scorer 1 was significantly more sensitive (rTPF = 1.12) and significantly less specific (rFPF = 1.17).

An in-depth analysis showed that consensus arousals between visual scorings 1 and 2 that were also detected by the ECG algorithm were significantly longer than those that were not de-

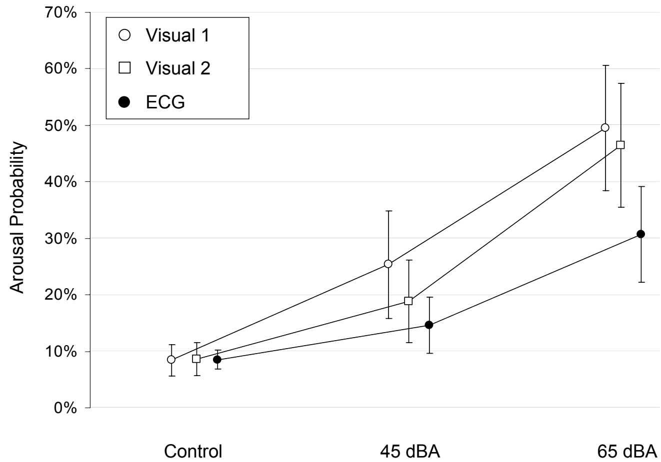

Figure 8—Event-related analysis of arousal probability depending on exposure category (control = no noise, 45 dBA, 65 dBA) and scoring method (Visual 1 = human scorer 1, Visual 2 = human scorer 2, ECG = ECG algorithm). Point estimates and 95% confidence bounds are shown.

tected by the ECG algorithm (average duration no agreement: 10.5 s, agreement: 16.3 s; median duration no agreement: 8.4 s, agreement: 15.3 s; Mann-Whitney-U test: $N_{agreement}$=175, $N_{no\ agreement}$=168, Z=-6.76, P <0.001).

DISCUSSION

This paper describes the development of an ECG-based algorithm for the automatic identification of autonomic activations associated with cortical arousal. There were astonishingly few systematic attempts to develop such algorithms in the past, although there are potentially many advantages (see Introduction). The analysis of heart rate has often been restricted to the maximum increase in heart rate during a specified interval following an event (e.g. an EEG arousal), using the average heart rate prior to the event as baseline. In that way, much of the information embedded in the autonomic response is discarded.

Adachi et al.[26] introduced a so-called pulse rate rise index (PRRI), defined as the number of pulse rate increases ≥ 4-10 bpm per hour of sleep. In their study, the PRRI was compared to the number of EEG arousals associated with apnea, hypopnea, or respiratory effort related increases in esophageal pressure. It tested whether the pulse rate rise index was suitable to distinguish between mild and severe forms of sleep related breathing disorders. Comparisons on the single event level were not presented.

Ayappa et al.[27] tried to determine the frequency and hierarchy of occurrence of oxygen desaturations, EEG arousals, and heart rate changes as immediate consequences of respiratory events. They measured the increase in heart rate in the 10 s following the respiratory event, using the average heart rate calculated during 5 s prior to event termination as baseline. They examined multiple definitions of heart rate increase: Either 4 or 6 bpm increases for either 1 or 2 consecutive beats. They chose 6 bpm increases for 2 consecutive beats as their final criterion. This definition led to a sensitivity of 67.1% and a specificity of 61.1% in terms of EEG arousal detection.

Ayappa et al.[27] also bewail that there is no consensus in the literature on the amplitude necessary to constitute a relevant increase in heart rate.[13] It is impossible to define the relevance of heart rate increases per se. Although Martin et al.[28] suggest that daytime functioning may be impaired by increases in the number of subcortical arousals alone, this has been questioned by Wesensten et al.,[4] because the procedure used by Martin et al. inevitably also induced cortical arousals and changes in sleep structure beside autonomic activations. Recent findings of a carefully designed experiment by Guilleminault et al.[29] support the thesis that EEG arousals are a prerequisite for the detrimental effects of sleep fragmentation on daytime functioning. Therefore, it seems reasonable to determine the relevance of autonomic arousals, depending on whether they are accompanied by cortical arousals.

For our criterion, differences to a moving median were used to determine the size of increases or decreases in heart frequency. The differences of 5 consecutive heart beats were then used to determine the probability of the first of the 5 beats being the first beat following the onset of an EEG arousal using a Bayesian approach, comparable to 5 sequentially applied diagnostic tests. In that way, and in contrast to past approaches, most of the heart rate response information was used for differentiating between EEG arousal and non-EEG arousal conditions.

If the criterion is based on 4 consecutive (first) heartbeats exceeding a certain probability of being associated with an EEG arousal, an ROC curve with an AUC of 0.912 resulted, demonstrating high diagnostic accuracy of the criterion. Arousal indexes generated by the ECG algorithm correlated positively with the average of 2 visual EEG arousal scorings, representing the best estimate of the "true" arousal index. However, the correlation became only significant and explained a reasonable proportion of the variance if 3 outliers were removed from the data set of N = 56. The same holds for the ICC.

According to the Bland-Altman plot, scorings of the ECG algorithm were unbiased with a reasonably small range of ± 2 standard deviations between +16.5 and -15.3 arousals per h sleep. However, the difference between the ECG arousal index and the average of the 2 visual scorings tended to increase with increasing visual arousal indexes. If high numbers of cortical arousals were associated with high numbers of subcortical arousals, which lead to relevant increases in heart rate without cortical activation, this could explain the observed tendency of the ECG algorithm to overestimate EEG arousal frequency in situations where both humans scored a high number of cortical arousals. In general, the Bland-Altman plot is the preferred method for assessing whether an established and a new measurement technique agree. The use of correlation coefficients may be highly misleading in this context.[24]

Both visual scorings agreed with the ECG arousal scorings in 42.5%, whereas ≥1 of the 2 visual scorings agreed in another 22.2%. Cases where the ECG algorithm scored an arousal event exclusively (35.3%) may still indicate relevant autonomic activations not accompanied by cortical arousal, but which may nevertheless play a role in long-term cardiovascular consequences of sleep fragmentation.

The ECG algorithm was also successfully validated with external data. There was no relevant change in agreement between data sets used for algorithm development and validation. Arousal probability determined by the ECG algorithm increased with simultaneously increasing SPL of aircraft noise events, showing its potential usefulness for large-scale field studies on the effects of noise on sleep. However, both visual scorings detected significantly more arousals in the noise exposure conditions than the

ECG algorithm, without detecting more arousals in the control condition. Pair-wise comparisons (Table 4) confirmed that both visual scorings were more sensitive than the ECG-algorithm, but only visual scoring 2 was also more specific. A detailed analysis showed that the higher sensitivity of visual scorings most likely resulted from detections of shorter EEG arousals missed by the ECG algorithm. On the one hand, this indicates that most of the important longer arousals, associated with strong increases in sympathetic activity, seem to be well captured by the ECG algorithm. On the other hand, it shows that efforts on improving the ECG algorithm should be focused on the detection of shorter events in the future. This is also corroborated by the fact that sensitivity for the detection of consensus arousals decreased with increasing arousal duration in the development data set.

Pillar et al.[17] introduced an *autonomic arousal index* (AAI), which is based on pulse rate rises and peripheral arterial tonometry (PAT) amplitude decreases. Similar to our approach, they compared the AAI to visually scored AASM EEG arousal indexes. The reported correlation between the autonomic arousal index and the EEG arousal index was higher than ours (r = 0.82 vs. r = 0.19). At the same time, the variance observed in the Bland-Altman plot was considerably larger. Both phenomena may be explained by the fact that the AAI was based on data of 85 OSA patients and 11 healthy controls, whereas our algorithm is based on healthy subjects only. Some of the sleep apnea subjects showed very high arousal indexes, which may be the reason for the reported high correlation. If the analysis of Pillar et al.[17] was repeated and restricted to the range of arousal indexes observed in our population, i.e. up to 35 arousals per h sleep, the correlation would probably be very similar to ours (see Figure 1 in Pillar et al.[17]). The same may hold for the Bland-Altman plot.

Overall, with 0.88 the AUC of the ROC plot generated by Pillar et al.[17] for the AAI was lower compared to our method (AUC=0.91), indicating a higher accuracy of our method. At the same time, our method is based on the ECG signal only, whereas the method of Pillar et al. was based on heart rate and PAT amplitude, the latter not being routinely recorded in sleep laboratories.

The low interobserver agreement in EEG arousal scorings from 2 human scorers[6,7] was confirmed in this investigation. If agreement is defined as any overlap in arousal scorings, visual scorers agreed in only 49.0% (development data set) and 51.8% (validation data set) of arousal scorings. Hence, only arousals meeting strict criteria were considered as consensus arousals, a surrogate for a "true EEG arousal," and used for the development of the ECG algorithm. The 2 scorers were deliberately chosen from 2 different laboratories. If scorers would have been recruited from a single institution, agreement between scorers might have been higher, but transferability of results would decrease.

Both EEG and ECG arousals were poorly and nonsignificantly correlated with a VAS score with the anchors "alert" and "tired," replicating the findings of many previous studies on EEG arousal frequency and subjective evaluations of fatigue and sleepiness.[17,30,31]

LRs were calculated for 5 consecutive heartbeats and 10 heartbeat intervals each. Alternatively, it would have been possible to calculate both mean and variance of the distribution of heart rate differences (compared to the moving median) for the EEG arousal as well as the non-EEG arousal condition. With this approach, each heartbeat difference could have been translated into a LR on a continuous scale. We decided against this parametric approach for 2 reasons: First, the assumption of normally distributed differences may not hold, which would lead to biased estimates of LRs. Second, the raw data were not corrected for ectopic heartbeats and movement artifacts in order to increase the practicability of the method and to keep expenditure of human labor low. Heart frequency differences derived from ectopic heart beats and movement artifacts are usually very high or very low, i.e. falling in the highest or lowest of the 10 categories. Hence, LRs in the first category were always a little smaller than LRs in the second category (see Table 1). In that way, the small increase in LRs in category 1 compared to category 2 can be regarded as a self-correcting mechanism of the procedure. We did not use more than 10 categories because we wanted the sample size in each of the categories to be high enough to exclude chance as a possible mechanism for explaining the LRs.

Limitations

As information on sleep stages will likely not be available in some of the possible future applications of the algorithm, wake epochs were not excluded from the analysis. Therefore, prolonged periods of wakefulness could lead to biased results: After all, 6.9% (development) and 18.1% (validation) of arousals exclusively scored by the ECG algorithm occurred during prolonged periods of wakefulness. However, it would be easy to exclude periods of wakefulness from the analysis, if sleep stage information is available and this is desired.

Here, the ECG algorithm was trained and validated with SPT data only. If it is to be used without the simultaneous recording of EEG, EMG, and EOG in the future, some method will be needed to reliably determine the beginning and end of SPT. For this purpose, other domains of the ECG signal, e.g., changes in heart rate variability, could be facilitated as well.

We emphasize that this is a first approach for the development of an ECG algorithm for the automatic identification of autonomic activations associated with cortical arousal; it is open for discussion and suggestions. Most probably, the accuracy of the algorithm could be improved by accounting for age, gender, BMI, or other individual variables of the investigated subject.[12] Also, incorporating information on the magnitude of the median heart rate might improve predictions. More than 5 or less than 5 consecutive beats could be used for the calculation of arousal probabilities. The time window of ± 90 seconds for the moving median could be optimized. Dropping the assumption that the tests are conditionally independent given the arousal status may improve algorithm accuracy. Also, the sensitivity of the algorithm increased with the duration of the EEG arousal. Therefore, it might be possible to improve predictions if separate criteria are developed for different lengths of EEG arousals.

As every subject contributed several data points, data were clustered, but this within-subject correlation of data was not taken into account in this first approach. Again, accuracy could possibly be improved if the correlated nature of the data was considered. Those cases where the ECG algorithm underestimated or overestimated "true" arousal frequency relevantly should be inspected more closely in order to learn more about the causes of these biased estimates. So far, the application of the ECG algorithm is restricted to nonclinical populations. Further studies are needed to evaluate the applicability of the ECG-based algorithm in populations different from the one studied in this investigation.

CONCLUSIONS

The ECG algorithm was designed to differentiate between EEG arousal and non-EEG arousal conditions, as defined by the 1992 AASM criteria.[1] This study demonstrated that it is possible to detect visually scored EEG arousals with an algorithm based on the ECG signal alone. However, the accuracy of the ECG algorithm would have to be improved before recommending it as an alternative to visual EEG arousal scoring. It has to be borne in mind that although close association of cortical and autonomic arousals has been repeatedly shown in the past,[8,10,13] they are still distinct events that may occur independently (i.e., cortical arousal without autonomic arousal and vice versa). At least some of the disagreement between visual EEG arousal scoring and ECG algorithm scoring may be attributed to these independently occurring events. Therefore, it is questionable whether it will at all be possible to perfectly identify EEG arousals and hence to *replace* visual EEG arousal scoring with algorithms based on the ECG alone. Regardless, it may still be valuable to supplement visual EEG arousal scoring with this automatic, reproducible, objective, and time-saving method. In screening devices that sample ECG but do not sample EEG, EOG. and EMG, arousal indexes generated by the ECG algorithm could be used as estimates of EEG arousal indexes, thus improving the explanatory power of the screening devices. In terms of studies on the effects of traffic noise on sleep, large-scale field studies could be performed with lower methodological expense but still sufficient explanatory power compared with polysomnographic studies. Finally, apart from its ability to predict EEG arousal indexes, the ECG arousal index itself may be a useful measure of subcortical arousals, which may play a role in nonrestorative sleep[28] and cardiovascular disease.[32] Future versions of the algorithm need to address the low sensitivity for detecting <10-s arousals, and the algorithm still needs to be validated in nonclinical populations.

ACKNOWLEDGMENTS

This paper is dedicated to Alexander Samel, who passed away May 19th, 2007. We miss Alexander as an excellent and honest scientist and as a dear friend. This work was partially supported by the HGF-Virtual Institute "Transportation Noise Effects on Sleep and Performance" (grant #VH-VI-111). We would like to thank Markus Breimhorst and Hans-Peter Bröde of Dortmund University, Wilma Littel, Thomas Penzel, Thomas Ploch and Werner Cassel of Marburg University, and Eva-Maria Elmenhorst and Julia Quehl of DLR Cologne for the productive discussions held on meetings of the HGF-Virtual Institute of Transportation Noise Effects on Sleep and Performance. The data used for the analysis were financed within the DLR/HGF-project "Quiet Air Traffic".

REFERENCES

1. Bonnet M, Carley DW, Carskadon MA et al. EEG arousals: Scoring rules and examples. A preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association. Sleep 1992;2:173-84.
2. Rechtschaffen A, Kales A. 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; 1968
3. Bonnet MH. Differentiating sleep continuity effects from sleep stage effects (Letter to the editor). J Sleep Res 2001:403-6.
4. Wesensten NJ, Balkin TJ, Belenky G. Does sleep fragmentation impact recuperation? A review and reanalysis. J.Sleep Res 1999;4:237-45.
5. Halasz P, Terzano M, Parrino L, Bodizs R. The nature of arousal in sleep. J Sleep Res 2004;1:1-23.
6. Loredo JS, Clausen JL, Ancoli-Israel S, Dimsdale JE. Night-to-night arousal variability and interscorer reliability of arousal measurements. Sleep 1999;7:916-20.
7. Drinnan MJ, Murray A, Griffiths CJ, Gibson GJ. Interobserver variability in recognizing arousal in respiratory sleep disorders. Am J Respir Crit Care Med 1998;2:358-62.
8. Catcheside PG, Chiong SC, Mercer J, Saunders NA, McEvoy RD. Noninvasive cardiovascular markers of acoustically induced arousal from non-rapid-eye-movement sleep. Sleep 2002;7:797-804.
9. Catcheside PG, Chiong SC, Orr RS, Mercer J, Saunders NA, McEvoy RD. Acute cardiovascular responses to arousal from non-REM sleep during normoxia and hypoxia. Sleep 2001;8:895-902.
10. Sforza E, Chapotot F, Lavoie S, Roche F, Pigeau R, Buguet A. Heart rate activation during spontaneous arousals from sleep: effect of sleep deprivation. Clin Neurophysiol 2004;11:2442-51.
11. Sforza E, Nicolas A, Lavigne G, Gosselin A, Petit D, Montplaisir J. EEG and cardiac activation during periodic leg movements in sleep: support for a hierarchy of arousal responses. Neurology 1999;4:786-91.
12. Gosselin N, Michaud M, Carrier J, Lavigne G, Montplaisir J. Age difference in heart rate changes associated with micro-arousals in humans. Clin Neurophysiol 2002;9:1517-21.
13. Sforza E, Jouny C, Ibanez V. Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response. Clin.Neurophysiol. 2000;9:1611-19.
14. Launois SH, Averill N, Abraham JH, Kirby DA, Weiss JW. Cardiovascular responses to nonrespiratory and respiratory arousals in a porcine model. J Appl Physiol 2001;1:114-20.
15. Pitson D, Chhina N, Knijn S, van Herwaaden M, Stradling J. Changes in pulse transit time and pulse rate as markers of arousal from sleep in normal subjects. Clin Sci (Colch.) 1994;2:269-73.
16. Lavie P, Shlitner A, Sheffy J, Schnall RP. Peripheral arterial tonometry: a novel and sensitive non-invasive monitor of brief arousals during sleep. Isr Med Assoc J 2000;3:246-47.
17. Pillar G, Bar A, Shlitner A, Schnall R, Shefy J, Lavie P. Autonomic arousal index: an automated detection based on peripheral arterial tonometry. Sleep 2002;5:543-49.
18. Basner M, Buess H, Elmenhorst D, et al. Effects of nocturnal aircraft noise (Volume 1): Executive summary. Report FB 2004/07/E, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Cologne, Germany, 2004
19. Basner M, Isermann U, Samel A. Aircraft noise effects on sleep: application of the results of a large polysomnographic field study. J Acoust Soc Am 2006;5:2772-84.
20. Basner M, Samel A. Effects of nocturnal aircraft noise on sleep structure. Somnologie 2005;2:84-95.
21. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960:213-20.
22. Samel A, Diedrich A, Drescher J et al. [Long-term monitoring of psychophysiologic values in flight physiology]. Internist (Berl) 1997;8:755-69.
23. Hunink M, Glasziou P, Siegel J, Elstein A, Weinstein M. Decision making in health and medicine: integrating evidence and values. 1st ed. Cambridge, UK: University Press; 2001.
24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;8476:307-10.
25. Shrout PE, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;2:420-8.
26. Adachi H, Mikami A, Kumano-go T et al. Clinical significance of pulse rate rise during sleep as a screening marker for the assessment

of sleep fragmentation in sleep-disordered breathing. Sleep Med 2003;6:537-42.
27. Ayappa I, Rapaport BS, Norman RG, Rapoport DM. Immediate consequences of respiratory events in sleep disordered breathing. Sleep Med 2005;2:123-30.
28. Martin SE, Wraith PK, Deary IJ, Douglas NJ. The effect of nonvisible sleep fragmentation on daytime function. Am J Respir Crit Care Med 1997;5:1596-601.
29. Guilleminault C, Abad VC, Philip P, Stoohs R. The effect of CNS activation versus EEG arousal during sleep on heart rate response and daytime tests. Clin.Neurophysiol. 2006;4:731-39.
30. Pitson DJ, Stradling JR. Autonomic markers of arousal during sleep in patients undergoing investigation for obstructive sleep apnoea, their relationship to EEG arousals, respiratory events and subjective sleepiness. J Sleep Res 1998;1:53-9.
31. Stradling JR, Barbour C, Glennon J, Langford BA, Crosby JH. Prevalence of sleepiness and its relation to autonomic evidence of arousals and increased inspiratory effort in a community based population of men and women. J.Sleep Res 2000;4:381-88.
32. Wolk R, Gami AS, Garcia-Touchard A, Somers VK. Sleep and cardiovascular disease. Curr Probl Cardiol 2005;12:625-62.