

Automatic differentiation of obstructive and central hypopneas with esophageal pressure measurement during sleep

C. Morgenstern, *Student Member IEEE*, M. Schwaibold, W. Randerath, A. Bolz
and R. Jané, *Member IEEE*

Abstract— The differentiation between obstructive and central respiratory events is one of the most recurrent tasks in the diagnosis of sleep disordered breathing. Esophageal pressure measurement is the gold-standard method to assess respiratory effort and identify these events. But as its invasiveness discourages its use in clinical routine, non-invasive systems have been proposed for differentiation. However, their adoption has been slow due to their limited clinical validation, as the creation of manual, gold-standard validation sets by human experts is a cumbersome procedure. In this study, a new system is proposed for an objective automatic, gold-standard differentiation between obstructive and central hypopneas with the esophageal pressure signal. First, an overall of 356 hypopneas of 16 patients were manually scored by a human expert to create a gold-standard validation set. Then, features were extracted from each hypopnea to train and test classifiers (Discriminant Analysis, Support Vector Machines and adaboost classifiers) to differentiate between central and obstructive hypopneas with the gold-standard esophageal pressure signal. The automatic differentiation system achieved promising results, with a sensitivity of 0.88, a specificity of 0.93 and an accuracy of 0.90. Hence, this system seems promising for an automatic, gold-standard differentiation between obstructive and central hypopneas.

I. INTRODUCTION

ONE OF THE MOST IMPORTANT ASPECTS OF THE DIAGNOSIS of Sleep Disordered Breathing (SDB) in regards of the appropriate choice of treatment, is the correct identification of respiratory events. Particularly, the correct differentiation between central and obstructive apneas/hypopneas is one of the most recurrent tasks due to the prevalence of the corresponding syndromes (OSAHS/CSAHS) [1]. Currently, esophageal pressure (Pes) measurement is considered the gold-standard technique for measurement of respiratory effort and the identification of obstructive and central events [2, 3]. Still, the complexity and invasiveness of esophageal pressure manometry and its impact on sleep [4] limits its usage in clinical routine. So, researchers have been recently trying to develop non-invasive systems for the

differentiation between central and obstructive apneas/hypopneas [5-7]. However, the clinical adoption of these techniques has been slow, mostly due to their limited clinical validation. The bottleneck when creating a gold-standard validation set is usually found in the manual identification of the mentioned events by a human expert, as it is a cumbersome procedure that may suffer of interscorer differences and subjective interpretation. Hence, the development of an objective and efficient method for automatic invasive assessment of central and obstructive events is desirable.

In this work a new automatic classifier is proposed based upon artificial intelligence (AI) techniques to automatically differentiate between central and obstructive hypopneas with the gold-standard Pes signal. Our system focused on the differentiation of hypopneas, because pressure swings during a hypopnea are more subtle than during other events, therefore being considered one of the most challenging tasks [2, 3]. In a first step, hypopneas were manually scored by a human expert to create a gold-standard validation set. Then, a specific set of features was extracted from the Pes-signal of each hypopnea in order to train and test the AI classifiers. Finally the performance of the different classifiers is evaluated and compared.

II. METHODOLOGY

A. Subjects

Sixteen male subjects had full nocturnal polysomnography (NPSG) with an 18-channel recorder (Somnolab V2.01 Weinmann GmbH, Hamburg, Germany) at the sleep laboratories of *Klinikum Bethanien* hospital in Solingen, Germany, according to a completely new protocol specifically designed for this purpose and approved by the hospital's Ethics Committee. Mean values \pm SD of the studied population for Apnea-Hypopnea Index (AHI) were 13.5 ± 11.3 events/h (range 1.8 – 46.9 events/h), body-max-index (BMI) 28.6 ± 4.02 kg/m² (range 24 – 40 kg/m²) and age 53.6 ± 15.8 years (range 21 – 78 years).

The respiratory signal was obtained with a nasal cannula device/pressure transducer system and a sampling frequency of 32Hz according to the AASM indications [3]. Esophageal manometry (UniTip catheter by UNISENSOR AG, Attikon, Switzerland and ISOPRE-P amplifier, Standard instruments GmbH, Karlsruhe, Germany) was systematically recorded for each patient with a sampling frequency of 16Hz. Other recorded physiological signals included arterial oxygen saturation (SpO₂), body position, pulse and plethysmograms

Manuscript received April 22nd, 2009. This work was supported in part by the Grant 2008FIC00369 of the Generalitat de Catalunya, Spain, and by the Spanish *Ministerio de Ciencia e Innovación* under grant TEC2007-68076-C02-01. C. Morgenstern and R. Jané are with Dept. ESAII, Universitat Politècnica de Catalunya (UPC), Institut de Bioenginyeria de Catalunya (IBEC) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Pau Gargallo 5, 08028, Barcelona, Spain (e-mail: cmorgenstern@ibec.pcb.upc.es, Raimon.Jane@upc.edu). M. Schwaibold is with MCC-Med GmbH & Co. KG, Karlsruhe, Germany W. Randerath is with Klinikum Bethanien, Solingen, Germany A. Bolz is with Institute of Biomedical Engineering, University of Karlsruhe, Karlsruhe, Germany

(inductive belts (ProTech, Services Inc, Mukilteo, WA, USA) around the rib cage), each with a sampling frequency of 32Hz. Also 2 EEG (C3/C4), 2 EOG (right/left), 1 submental EMG, 1 leg-EMG and 1 ECG channels were recorded with a sample frequency of 256Hz each.

Sleep stages, apneas, hypopneas and other respiratory events were scored applying standard criteria [2, 3]. According to these criteria [2], a hypopnea is identified by a clear decrease (>50%) from baseline in the amplitude of a valid measure of breathing during sleep and lasts for at least 10 seconds. The baseline is defined as the mean amplitude of stable breathing and oxygenation in the two minutes preceding onset of the event. For the scoring of a central apnea/hypopnea, a clear reduction in esophageal pressure swings from baseline (as defined before) is required. According to this guideline, there is no relative or absolute reduction in esophageal pressure inside the event's interval that can be used to distinguish a central from an obstructive, event, which increases the difficulty of the automatic differentiation.

A human expert reviewed the full-night recordings of our 16 patients and manually scored hypopneas with the mentioned criteria, primarily using the airflow and Pes signals. A total of 181 obstructive hypopneas and 175 central hypopneas were scored, resulting in an overall of 356 hypopneas.

B. Pre-processing and detection of inspirations

The time markers of the manually scored hypopneas were imported for our automatic processing, as the automatic detection of apneas/hypopneas in the flow signal has already been proficiently solved by others [7]. The specific purpose of this study is to propose a new method for the differentiation of obstructive/central hypopneas.

The Pes signal presented noise and physiological disturbances (like swallowing or coughing artifacts) that had to be reduced. After interpolating the Pes signal to 32 Hz, a 15-point Moving Average (low pass) filter with a cut-off frequency of 1Hz at -3dB was applied to the Pes signal, as the main frequency content of the Pes signal is located below 1Hz [8]. Flow pre-processing stages were performed as detailed in [9].

In order to differentiate a hypopnea, the Pes signal of the hypopnea (hPes) and the 2 min prior to the hypopnea's start (Pes2min) were extracted and processed separately. In case hPes or Pes2min presented a significant baseline oscillation, linear trends were automatically removed from the signals (detrending). The robust detection of the maxima and minima of the pressure swings in the filtered Pes signal is an essential part of our analysis. So, criteria, like a minimal separation time between two inspirations (1 second), were defined to assure that only the Pes maximum of each inspiration was correctly identified, while avoiding the detection of small oscillations or artifacts. Minima were detected by locating the local minimum between two contiguous maxima. In case the hPes or Pes2min signals did not start with a maximum, the manual markers were automatically shifted (maximal 1 second) until a maximum was found at the beginning of each signal (fig. 1 a). After the

detection of the local extrema, in case a maximum's amplitude surpassed the mean amplitude of the signal's maxima plus twice the standard deviation, the corresponding maximum/minimum pair was excluded. This was especially necessary to exclude positive Pes peaks caused by swallowing and coughing (fig. 1 b). Finally, the number of maxima and minima that had been detected in the Pes2min and hPes signals, respectively, was compared. In case the number of maxima and minima was not identical, the hypopnea was classified as an artifact. A total of 5 hypopneas (1.4%) were discarded as artifacts and were not further processed.

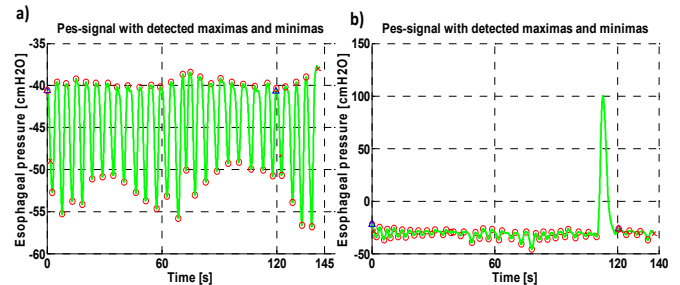


Fig.1 Two examples of the Pes signals of two hypopneas (a, b) with detected maxima and minima. The crosses (red) on the right indicate the original hypopnea (hPes) markers while the triangles (blue) indicate the new markers after relocation. The markers at the left indicate the beginning of the original/moved Pes2min segment, respectively. The swallowing artifact in hypopnea (b) was discarded by the pre-processing.

C. Feature extraction

The accurateness of the classification will primarily depend on how well the extracted features characterize the central and obstructive hypopneas. As no relative or absolute reduction in esophageal pressure during a hypopnea's interval can be used to distinguish between central and obstructive hypopneas, we had to compare the relative changes in amplitude in hPes in respect to Pes2min. In order to minimize the effects of possible baseline drifts during the observed segments of the Pes signal, we worked with the amplitude difference of the corresponding maximum and minimum of an inspiration (instead of its absolute amplitude values). The overall number of amplitude difference values of the hPes signal was used as the first feature, (Table 1).

As seen in (fig. 2), we computed the median of the amplitude differences of the Pes2min signal (see fig. 2, blue

TABLE I
EXTRACTED FEATURES

Feature		Feature	
1	Number of amplitude difference values of hPes	6	Median of Pes2min
2	Correlation index of hPes	7	Median of the superior group of values of Pes2min
3	Hypopnea location index	8	Median of the inferior group of values of Pes2min
4	Median of the 1 st derivative of the maxima in hPes	9	Median of hPes
5	Median of the 1 st derivative of the minima in hPes	10, 11	Median of hPes \pm standard deviation of hPes

solid line on the left) that divided the Pes2min amplitude difference values in two groups, one located above this median and another located below. For each of these two groups, their respective median was computed again (see dashed blue lines on the left). The standard deviations of the Pes2min signal around the median are represented as magenta dash-dot lines on the left. For the hPes signal the median was computed (see red solid line on the right) \pm the standard deviation of hPes (dashed red lines on the right). All these parameters were used as features (6-11) for the classifiers (Table 1).

With these visual references, we defined a location index, see feature (3) (Table 1), that defined four amplitude difference intervals where the median of hPes could be located in respect to Pes2min. We numerated these intervals from 1 - 4 (fig. 2): e.g. if the median of hPes was located above the median of the inferior group of Pes2min and below the median of Pes2min, the location index was assigned the value 3 (fig. 2 b).

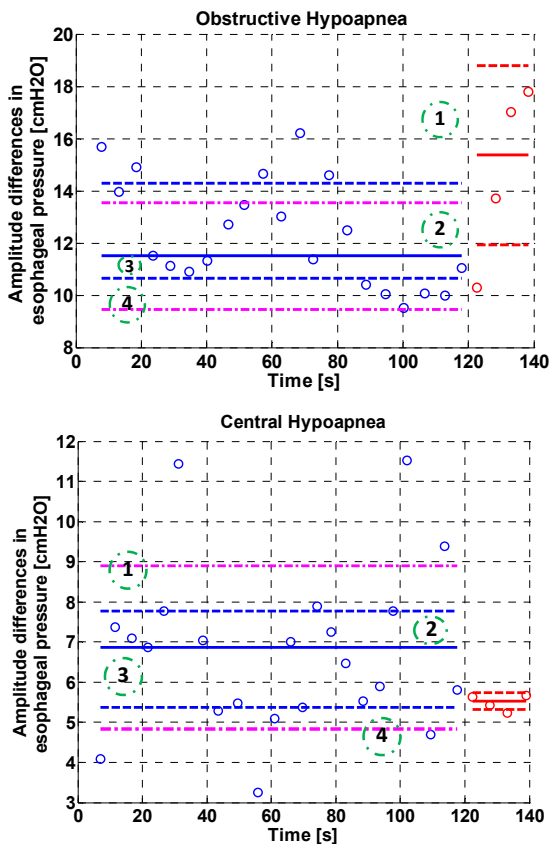


Fig.2 Examples of the distribution of amplitude differences for an obstructive (left) and a central hypopnea (right). The amplitude difference values for the Pes2min signal are located on the left (blue circles), while the difference values for the hPes segment are on the right (red circles). The values of the location index are indicated in the dash-dot (green) circles. Note that the hPes difference values for the obstructive apnea increment sequentially while those for the central apnea do not substantially vary.

Furthermore, we also calculated Pearson's correlation coefficient of the hPes amplitude differences, as we observed that an obstructive hypopnea usually presented a higher linear correlation of its amplitude differences (fig 2 a)

in comparison to a central hypopnea (fig. 2 b), see feature (2) (Table 1).

Finally, we also usually observed a divergent behavior of the envelope of the hPes maxima and minima for obstructive hypopneas (fig. 1a). So, we computed the median of the first derivatives of the amplitudes of the maxima and minima, respectively, see features (4, 5) (Table1).

D. Training and testing of the classifiers

Discriminant Analysis (DA), Support Vector Machines (SVMs) [10] and boosting algorithms like adaboost [11] belong to the family of supervised machine learning methods. These classification methods require a training phase where the classifier learns to solve a bi-class problem. The training set usually consists of a $(n \times k)$ matrix with a number k of n -dimensional elements and of a label vector of length k containing the class label $\{+1; -1\}$ for each element. In our case, the labels corresponded to the manual, gold-standard scorings for central $\{-1\}$ and obstructive $\{+1\}$ hypopneas.

Training, validation and test sets were created with a cross-validation algorithm (MATLAB v.7.5, The Mathworks Inc., Natick, MA, USA) in order to randomly assign 35% of the hypopneas to the training set, 15% to the validation set and 50% to the test set. Thus, the training set consisted of 123 hypopneas, the validation set of 53 hypopneas and the test set of 175 hypopneas (with an approximately 50/50% within group distribution of obstructive and central hypopneas). The DA analysis was performed with 5 different functions: linear DA (LDA), diagonal linear DA (DLDA), quadratic DA (QDA), diagonal quadratic DA (DQDA) and Mahalanobis DA (MDA). In order to find the optimal parameters for the SVM classification (SVM-light v. 6.01, University of Dortmund, Dortmund, Germany), we evaluated SVMs on the validation set with a polynomial kernel of degrees (4, 8, 16, 24) and with a Gaussian kernel with sigma values (0.1, 0.2...0.9). The adaboost implementation (GML adaboost Matlab Toolbox, Moscow, Russia) used here uses classification and regression trees (CARTs) as weak classifiers. A CART with only 1 split would be equivalent to a classical stump learner as described in [11]. This toolbox delivers three different variations of the adaboost algorithm: standard adaboost [11], "gentle" adaboost [12] and "modest" adaboost [13], each with different generalization capabilities. Again, several parameters had to be optimized with the validation set, such as the number of CART splits (1, 3, 6, 12, 24, 36, 64, 128) and maximum iterations of the adaboost algorithm (100 and 300 cycles). Sequential forward selection was used for each classifier in order to select the most relevant subset of the described features, respectively.

III. RESULTS

The best classification results for DA were achieved with the quadratic DA (QDA) function and a feature vector comprising the features (1, 3, 11). The best classification results for the SVM classifier were achieved with a Gaussian

kernel, a sigma of 0.2 and a feature vector with the features (1, 3, 5). Of all tested adaboost classifiers, the best results were achieved with the standard adaboost algorithm, 24 CART splits and after 100 cycles of maximum iteration (Table 2). The feature vector for the adaboost classifier comprised all eleven features.

TABLE 2
CLASSIFICATION RESULTS
SENSITIVITY (SE), SPECIFICITY (SP), POSITIVE PREDICTIVE VALUE (PPV),
NEGATIVE PREDICTIVE VALUE (NPV), ACCURACY (ACC)

	QDA	SVM	adaboost
SE	0.88	0.90	0.88
SP	0.90	0.84	0.93
PPV	0.90	0.85	0.93
NPV	0.88	0.89	0.88
ACC	0.89	0.87	0.90

As seen in the ROC curves (fig. 3), the adaboost classifier seems to deliver a better overall performance than the SVM and the QDA classifiers.

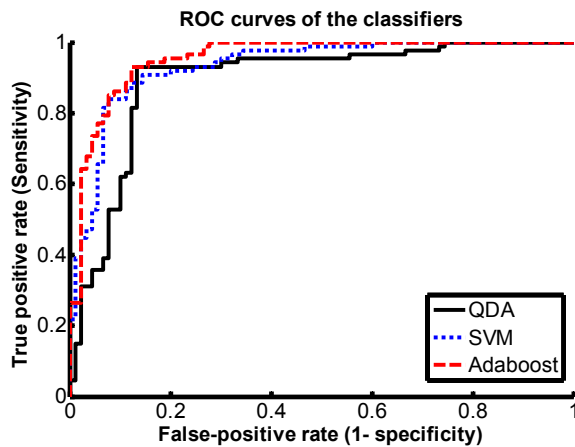


Fig.2 Receiver Operator Characteristic (ROC) curves for the QDA, the best SVM and best adaboost classifiers for the differentiation between central and obstructive hypopneas. The curves were obtained by varying the threshold value of the classifier's output in the range between -1 and $+1$.

IV. DISSCUSION AND CONCLUSIONS

An overall of 357 manual, gold-standard hypopnea scorings were used in this study, which represents twice the size for a gold-standard annotation set than in other studies [5, 7]. As our classifiers were trained and tested with these gold-standard scorings, the results of our automatic classification should also be acceptable from a human scorer's perspective. The information contained in the selected features seems to be adequate for the differentiation between obstructive and central hypopneas. As only the Pes-signal was used for the differentiation, this system should also be applicable for the differentiation of other events (e.g. apneas), as the same differentiation criteria apply [2].

Our best adaboost classifier showed the best overall classification accuracy (0.90), good results for obstructive (0.88) and the best overall classification results for central hypopneas (0.93), although it was also the classifier with the

highest computational complexity, the largest feature vector and processing time. The SVM classifier showed the best overall results when identifying obstructive hypopneas (0.90) but also the poorest results when identifying central hypopneas (0.84). The QDA classifier showed a remarkable performance given the fact that it is the classifier with the lowest computational complexity. Thus, from a computational cost/performance point of view, we would choose the QDA algorithm for the differentiation of the hypopnea set presented here.

The scores reported in this study represent the upper limit of human interscorer agreement [14]. Given that currently the only accepted gold-standard annotations are manual scorings by human experts, these results should also correspond to the overall limit to be reached by an automatic classifier. Hence, the classification techniques presented in this study seem to be promising for the automatic, invasive differentiation of obstructive and central hypopneas and an objective and time-efficient creation of gold-standard validation sets. This should help to validate more efficiently and objectively non-invasive differentiation systems that have already been or are to be developed.

REFERENCES

- [1] Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993;328:1230-1235.
- [2] The Report of an American Academy of Sleep Medicine Task Force, Sleep-Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research, *SLEEP*, Vol. 22, No. 5, 1999
- [3] Iber, C., Anoni-Israel S., Chesson A. L., Qua S.F., The AASM Manual for the Scoring of Sleep and Associated Events, American Academy of Sleep Medicine, Westchester, IL, 2007
- [4] Chervin R., Aldrich M., Effects of Esophageal Pressure Monitoring on Sleep Architecture, *Am J Respir Crit Care Med*;156:881-885, 1997
- [5] Argod J, Pépin JL, Lévy P., Differentiating Obstructive and Central Sleep Respiratory Events through Pulse Transit Time, *Am J Respirat Crit Care Med* 158:1778-1783, 1998
- [6] Luo YM, Tang J, Steier J, Zhong NS, Moxham J, Polkey MI., Distinguishing Obstructive From Central Sleep Apnea Events: Diaphragm Electromyogram and Esophageal Pressure Compared., *Chest*. 135 (5), 1133-1141, 2009
- [7] O. Fontenla-Romero, B. Guijarro-Berdiñas, A. Alonso-Betanzos, V. Moret-Bonillo, A new method for sleep apnea classification using wavelets and feedforward neural networks, *Artificial Intelligence in Medicine*, 34:1, pp. 65-76, 2005
- [8] P Ask, PA Oberq, L Tibbling, Frequency content of esophageal peristaltic pressure, *Am J Phys Endocr Metab*, 236: E296-E300, 1979
- [9] Morgenstern C., Schwaibold M., Randerath W., Bolz A., Jané R., Assessment of changes in upper airway obstruction by automatic identification of inspiratory flow limitation during sleep, *IEEE Trans. on Biomedical Engineering*, DOI: 10.1109, Vol. 56 (9), 2009
- [10] Vapnik, V. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995
- [11] Freund Y, Schapire R. Game theory, on-line prediction and boosting. In *Proc. 9th ann. Conf. on Comp. Learning Theory*, pp. 325-332, 1996
- [12] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 38(2):337-374, 2000.
- [13] Vezhnevets A. Vezhnevets V., *Modest AdaBoost- teaching AdaBoost to generalize better*. Graphicon 2005
- [14] Steltner H, Staats R, Timmer J, Vogel M, Guttman J, Matthys H, Virchow JC, Diagnosis of Sleep Apnea by Automatic Analysis of Nasal Pressure and Forced Oscillation Impedance, *Am J Respir Crit Care Med* Vol 165. pp 940-944, 2002