Medical Engineering & Physics 34 (2012) 1213–1220

Contents lists available at SciVerse ScienceDirect

Medical Engineering & Physics

journal homepage: www.elsevier.com/locate/medengphy

Multiclass classification of subjects with sleep apnoea–hypopnoea syndrome through snoring analysis

Jordi Solà-Soler [a,b,*], José Antonio Fiz [c,b], José Morera [c], Raimon Jané [a,b]

[a] Department of ESAII, Universitat Politècnica de Catalunya, Barcelona, Spain
[b] Institut de Bioenginyeria de Catalunya (IBEC), CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain
[c] Pulmonology Service, Hospital Universitari Germans Trias i Pujol, Badalona, Spain

ARTICLE INFO

Article history:
Received 25 August 2011
Received in revised form
13 December 2011
Accepted 14 December 2011

Keywords:
Snoring
Sleep apnoea
Bayes classifier
Kernel density estimation

ABSTRACT

The gold standard for diagnosing sleep apnoea–hypopnoea syndrome (SAHS) is polysomnography (PSG), an expensive, labour-intensive and time-consuming procedure. Accordingly, it would be very useful to have a screening method to allow early assessment of the severity of a subject, prior to his/her referral for PSG. Several differences have been reported between simple snorers and SAHS patients in the acoustic characteristics of snoring and its variability. In this paper, snores are fully characterised in the time domain, by their sound intensity and pitch, and in the frequency domain, by their formant frequencies and several shape and energy ratio measurements. We show that accurate multiclass classification of snoring subjects, with three levels of SAHS, can be achieved on the basis of acoustic analysis of snoring alone, without any requiring information on the duration or the number of apnoeas. Several classification methods are examined. The best of the approaches assessed is a Bayes model using a kernel density estimation method, although good results can also be obtained by a suitable combination of two binary logistic regression models. Multiclass snore-based classification allows early stratification of subjects according to their severity. This could be the basis of a single channel, snore-based screening procedure for SAHS.

1. Introduction

Sleep apnoea–hypopnoea syndrome (SAHS) is a common disorder the first symptom of which is usually heavy snoring. The impact of snoring ranges from no sleep disruption to continuously disrupted sleep [1]. The prevalence of SAHS is 3.2 times higher in snorers than in non-snorers [2]. Accordingly, a snoring analysis system could help to provide further indication of level of risk. The gold standard for diagnosing SAHS is polysomnography (PSG). This is a very expensive, labour-intensive and time-consuming procedure. It would be desirable to have a *screening* procedure that helped respiratory physicians to rapidly determine the severity of a patient, in order to establish priority amongst candidates waiting for PSG. An ideal screening procedure should neither consider anyone with SAHS as healthy, nor send any healthy individual to the hospital for PSG.

Recently, some authors have investigated the possibility of identifying SAHS through the analysis of nocturnal oximetry [3] or oronasal airflow pressure [4]. Acoustic analysis of snoring reveals information relating to the site and degree of obstruction of the upper airway [5]. Several studies have found statistically significant differences in the acoustic characteristics of snoring between patients with SAHS and simple snorers [6–10]. Most of these studies have classified snoring individuals into two classes by means of an apnoea–hypopnoea index (AHI) threshold. However, no further information about the severity of the subject is provided. A recent publication of our group has described multiclass analysis of snoring subjects with SAHS [11]. Other authors have used a Bayes classifier with Gaussian density estimation to characterise individuals according to features of snoring and apnoea [12], but in general these variables do not follow a normal distribution.

Our approach is based on a single channel, namely the sound signal, and in particular we exclusively use the acoustic information extracted from snores, without knowing the number or the duration of apnoeas. Good classification rates of snoring subjects with SAHS can be achieved with this tight constraint if (1) a deep analysis of snoring episodes is carried out, something that necessarily includes a wide range of snoring features and their variability, as we have shown in previous articles [13]; and (2) an automatic algorithm is used for the selection of the best set of features, for a given performance measure.

In a preliminary study we analysed a Bayes classifier with a kernel density estimation method, using a range of snoring features. In this paper, we analyse the performance of this classifier when it

* Corresponding author at: Institut de Bioenginyeria de Catalunya (IBEC), Baldiri Reixac, 4, Torre I, 9th Floor, 08028 Barcelona, Spain. Tel.: +34 934137358.
E-mail addresses: jordi.sola@upc.edu (J. Solà-Soler), jafiz@msn.com (J.A. Fiz), josepmorera.germanstrias@gencat.cat (J. Morera), raimon.jane@upc.edu (R. Jané).

1350-4533/$ – see front matter © 2011 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2011.12.008

Table 1
Polysomnographic characteristics of the patients on the database.

		C_1	C_2	C_3	p_{12}	p_{13}	p_{23}
Subjects, N = 36 (M = 25, F = 11)	M/F	10/3	7/4	8/4	–	–	–
Age (yr)	m	45	48	52	0.706	0.096	0.460
	s	11	12	10			
BMI (kg/m^2)	m	27.1	28.9	32.9	0.339	0.012	0.085
	s	4.4	4.1	6.0			
AHI (h^{-1})	m	1.8	8.7	44.1	<0.001	<0.001	<0.001
	s	1.5	2.5	20.2			
Number of snores, T = 65625	m	1226	2243	2084	0.068	0.044	0.580
	s	1289	1177	937			
	T	15942	24672	25011			

M, Males; F, Females; m, mean; s, standard deviation; T, total number of snores; BMI, Body mass index; AHI, apnoea hypopnoea index; C_x, Class x; $x = 1{:}3$, C_1 = AHI < 5, $C_2 = 5 \leq$ AHI < 30, C_3 = AHI $\geq$ 30; p_{xy} = statistical significance of the Mann–Whitney U-test between classes x and y.

takes into account variability of the snoring features. We also introduce a novel classification procedure based on the combination of two binary logistic regression models. We compare these two methods to the Bayesian approach with Gaussian density estimation, and to a recently developed kernel density estimation method based on diffusion processes [14].

2. Methods

2.1. Signal acquisition

A prototype of a single-channel device (Snoryzer Uno, Sibel S.A., Barcelona, Spain) was used to record respiratory sound signals during sleep. Snoring sounds were noninvasively recorded using a unidirectional electret condenser microphone coupled to the skin surface through a conic air cavity designed following the guidelines of Kraman et al. [15]. Snoring sound signals were acquired while full-night polysomnography was performed. The microphone was placed over the trachea at the level of the cricoid cartilage using an elastic strap. The sound signal was amplified and filtered using a second-order analogue Butterworth band-pass filter between 70 and 2000 Hz and then digitised with a sampling frequency of 5000 Hz and a 12 bit A/D converter.

2.2. Subject database

The respiratory sound signal database contained data from 36 subjects (11 females and 25 males with an age range of 23–69 years and AHI range of 0–90.8 h^{-1}) who were enrolled from the Sleep Disorders laboratory of the University Hospital Germans Trias i Pujol in Badalona, Spain. All were free of any upper airway infection and other diseases throughout this study. None had undergone treatment for snoring or were taking any medication at the time of data collection. The study was approved by the hospital's research ethics committee and informed consent was obtained from all patients. The characteristics of the 36-subject database including the number of snores per group (T = 65625 snores) are reported in Table 1. All the snores detected during the night were used for subject classification.

Snoring episodes were identified by a previously trained and validated automatic detector and analyser [11], developed by our research group (DLL Snore Analyser v.5.5). The snoring detector was designed to identify snoring episodes in simple snorers and in SAHS patients, and to reject respiratory sounds from regular inspiration and exhalation, coughs, speech and other artefacts. The detector has a segmentation subsystem and a classification subsystem [11]. The segmentation subsystem establishes the time boundaries of possible snoring episodes. The classification subsystem is based on a 2-layer feed-forward multilayer neural network. The output layer has 2 neurons that show presence or absence of snore. The input pattern consists on 22 temporal and spectral features of the sound segment. This pattern allows the distinction between the snoring sound and the remaining respiratory sounds. Therefore, the detector identifies each snoring episode and defines automatically the time boundaries of the event. The snoring detector was validated by means of a database of 948 annotated episodes by a medical doctor [16]. This manual annotation included non-apnoeic snores, inspiration and exhalation sounds, voice, cough and noise artefacts.

Most published studies on classification of snoring subjects with SAHS focus on a two class approach, i.e., using only one AHI threshold at a time [6–10,17]. Recently, we have investigated the analysis of acoustic snore frequency characteristics with three levels of SAHS severity, but without carrying out multiclass classification of patients [11]. The American Academy of Sleep Medicine proposes grading the severity of SAHS into four levels, according to the thresholds AHI = 5, 15 and 30 h^{-1} [18]. In this paper, we opted to explore the classification of snoring subjects into three classes of SAHS severity given by thresholds AHI = 5 and 30 h^{-1}. In this way, each class had a comparable size, given the subjects available on our database. The classes were C_1 (no SAHS, AHI < 5 h^{-1}), C_2 (mild to moderate SAHS, 5 h^{-1} $\leq$ AHI < 30 h^{-1}) and C_3 (severe SAHS, AHI $\geq$ 30 h^{-1}). All patients were snorers (the total number of snores detected per individual ranged from 117 to 4214).

2.3. Snore characterization

Several techniques in time and frequency domains have been developed in our previous work [13] for the analysis and characterization of snores (Table 2). In the time domain, snores are characterised by their mean and maximum sound intensity (I_{mean}, I_{max}) [11] and by the period of the sound vibrations or pitch. The pitch of a snore is described by its mean value (P_m), standard deviation (P_s) and interquartile range (P_{iqr}); the pitch density (P_{dens}), defined as the fraction of time with pitch over the total duration of a snore; and the number of intervals with pitch within a snore episode (P_{ints}).

The frequency content of a snore is assessed by its power spectral density (PSD). The shape of the PSD is characterised by a set of parameters [11]: the mean, median, peak and maximum frequencies (F_{mean}, F_{med}, F_{peak}, and F_{max}, respectively); the standard deviation of the frequency (*StdDev*); and the symmetry and flatness coefficients (*CSymm*, *CFlatn*). The power distribution of the PSD is measured by energy ratios in three frequency bands of interest: B = (0, 500)Hz, (100, 500)Hz and (0, 800)Hz. For each band B, energy ratios are calculated with respect to the total energy (RW_B) and the energy outside that band ($Rout_B$).

The oral and nasal cavities produce resonances within a snoring sound. These, also called formants, correspond to the peaks of the spectral envelope and can be estimated using an autoregressive (AR) model. The formants of snores measured at the trachea were

Table 2
Snore characterization [11,13].

Snore intensity parameters	
I_{mean}	Mean sound intensity
I_{max}	Maximum sound intensity
Snore pitch parameters	
P_m	Pitch mean value
P_s	Pitch standard deviation
P_{iqr}	Pitch interquartile range
P_{dens}	Pitch density
P_{dens}	Pitch intervals
Snore frequency (PSD) Parameters	
F_{mean}	Mean frequency
F_{med}	Median frequency
F_{peak}	Peak frequency
F_{max}	Maximum frequency
CSymm	Symmetry coefficient
CFlatn	Flatness coefficient
RW_B	Ratio of energy in band B to total energy
$Rout_B$	Ratio of energy in band B to energy outside B
Spectral envelope snore parameters	
F_i	Frequency of the *i*th formant
L_i	Attenuation of the *i*th formant
M_i	Amplitude of the *i*th formant

B is the frequency band: *B* = (0, 500)Hz, (100, 500)Hz or (0, 800)Hz. *i* = 1:5.

found to lie in five frequency bands B_1–B_5. Each formant can be characterised by its frequency F_i, its amplitude with respect to the maximum M_i and its attenuation L_i, $i = 1:5$.

In our previous work [13], we have found that the variability of snoring features over the night is significantly higher in SAHS patients with AHI ≥ 10 than in snorers with AHI < 10. These differences are much more significant when the variability is measured on a snore-by-snore basis as follows: for a given parameter *P* of snore *s*(*j*), $P_{s(j)}$, the first difference $dP_{s(j)} = P_{s(j)} - P_{s(j-1)}$ is calculated. The time series *dP* oscillates around zero and its amplitude is assessed in terms of the standard deviation (*SdP*) and the interquartile range (*IQdP*). The difference *dP* is also computed over the time interval *dT* ($dtP \equiv dP/dT$) between the beginning of consecutive snores. For several snore parameters, the amplitude of its first difference, *dP*, was found to be correlated with the AHI [13].

In this paper, for multiclass classification, the independent variables X_k of the classification models are selected from amongst all the snore parameters derived from the sound intensity, the pitch, the PSD, and the AR Spectral Envelope. For every snore parameter *P*, six independent variables X_k are obtained from the measures described in Table 3. All these variables are available for selection by the corresponding optimum selection algorithms described in the following sections.

2.4. Logistic regression classifier

A binary logistic regression model was previously used [13] for the classification of snoring subjects into two classes with a threshold AHI = 10 h^{-1}. This approach has the advantage over other classification methods, such as discriminant analysis, that it does not assume that the data are normally distributed and that the model can be adjusted for a desired sensitivity or specificity. For a given AHI threshold (*TH*), a dichotomous variable Y_{TH} is defined that takes a value of 0 in snorers with AHI < *TH* and $Y_{TH} = 1$ in snorers with AHI ≥ *TH*. The probability that $Y_{TH} = 1$ is calculated by the logistic model

$$p_j = p(Y_{TH} = 1 \left| x_{j1}, \ldots, x_{jk} \right.) = \frac{1}{1 + \exp\{\beta_o + \beta_1 x_{j1} + \cdots + \beta_K x_{jK}\}} \quad (1)$$

where the model parameters β_k, $k = 0{:}K$, are estimated by the maximum likelihood method from the available observations ($x_{j1}, \ldots, x_{jK}$), $j = 1{:}N_{obs}$, of the independent variables X_k, $k = 1{:}K$.

In this paper, we propose the use of two binary logistic regression classifiers simultaneously with thresholds TH = 5 h^{-1} and TH = 30 h^{-1}. This will allow the classification of subjects into three classes, depending on the values of the dependent variable of the two models, Y_5 and Y_{30} (Table 4). The optimum independent variables of each model are automatically selected from all the available snoring parameters (see Section 2.3, Tables 2 and 3) by the forward stepwise selection (FSS) algorithm [19] (Fig. 1a).

2.5. Naïve Bayes classifier

The Bayes' rule provides a direct way of classifying individuals into multiple classes. The variables X_k are assumed to be independent (Naïve Bayes assumption) so that their joint probability density function (PDF) can be factored into the product of the individual PDF's. In order to avoid managing products of extremely low numbers, the logarithm of the PDF's can be used instead of the PDF's themselves. Thus, a given observation ($x_{j1}, \ldots, x_{jK}$) would be classified into the class, $C_{classif}$ given by:

$$\begin{aligned} C_{classif} &= \underset{i=1:3}{\arg\max} \left\{ P(C_i) \cdot p_{X_1,\ldots,X_K}(x_{j1}, \ldots, x_{jK} \left| C_i \right.) \right\} \approx \\ &\approx \underset{i=1:3}{\arg\max} \left\{ \hat{P}(C_i) \cdot \prod_{k=1}^{K} \hat{p}_{X_k}(x_{jK} \left| C_i \right.) \right\} = \\ &= \underset{i=1:3}{\arg\max} \left\{ \log(\hat{P}(C_i)) + \sum_{k=1}^{K} \log(\hat{p}_{X_k}(x_{jk} \left| C_i \right.)) \right\} \end{aligned} \quad (2)$$

The Naïve approximation has been reported to give good results in most scenarios [20]. The probability $P(C_i)$ of pertaining to each class is estimated by the number of subjects in that class over the total number of individuals. Previous studies have been based on the Gaussian assumption for the conditional PDF $p_{Xk}(X_k|C_i)$ of the independent variables [12]. However, in general the available data do not follow a Gaussian distribution, so it is preferable to use kernel-based PDF estimation [20]. Given a set of *N* observations $(x_{1k}, \ldots, x_{Nk})^T$ of the variable X_k, the estimated PDF is given by

$$\hat{p}_{X_k}(x) = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^{N} \exp\left\{ -\left(\frac{x - x_{jk}}{2h} \right)^2 \right\} \quad (3)$$

where *h* is the kernel bandwidth [14]. The performance of the method is dependent on the selection of the bandwidth *h*. In this paper, we use two approaches to estimate *h*: the empirical approach provided by kernel density estimation (kernel ksdensity) method (*ksdensity* function, available in MATLAB R2009b), and the bandwidth estimation procedure recently proposed by Botev et al. [14], based on linear diffusion processes (kernel diffusion). Both kernel-based approaches will also be compared to the conventional Gaussian method for estimating the conditional densities.

The optimum independent variables are automatically selected by a sequential floating forward selection (SFFS) algorithm described in Fig. 1b [20]. The maximum number of variables allowed is 10. At each step, variables are chosen to enter or leave the model based on the optimisation of a performance measure *J*(·).

Table 3
Independent variables derived from each snore parameter.

Name	Description
P	Mean value of the parameter
SP	Standard deviation of the parameter
SdP	Standard deviation of the first difference of the parameter
IQdP	Interquartile range of the first difference of the parameter
SdtP	Standard deviation of the first difference of the parameter over time
IQdtP	Interquartile range of the first difference of the parameter over time

P stands for any of the snore parameters defined in Table 2. The same symbol *P* is used for the parameter and its mean value.

Table 4
Logistic regression classification rule.

Value of the dependent variable	Y_5	0	0	1	1
	Y_{30}	0	1	0	1
Classified class		C_1	C_3[a]	C_2	C_3

C_1 = AHI < 5, C_2 = 5 ≤ AHI < 30, C_3 = AHI ≥ 30.
[a] In case of inconsistency, the subject is classified in the most severe group because we aim to maximise the sensitivity.

Consider the confusion matrix C in which the original class is in the rows and the classified class in the columns:

$$C = \begin{pmatrix} C1_{C1} & C2_{C1} & C3_{C1} \\ C1_{C2} & C2_{C2} & C3_{C2} \\ C1_{C3} & C2_{C3} & C3_{C3} \end{pmatrix} \qquad (4)$$

where Ci_{Cj} is the number of subjects of class C_j classified into class C_i. We define three performance measures of interest: the negative predictive value (NPV) of the healthy subjects (C_1, AHI < 5 h^{-1}) given by

$$\mathrm{NPV}_{C1} \equiv \frac{C1_{C1}}{C1_{C1} + C2_{C1} + C3_{C1}} \qquad (5)$$

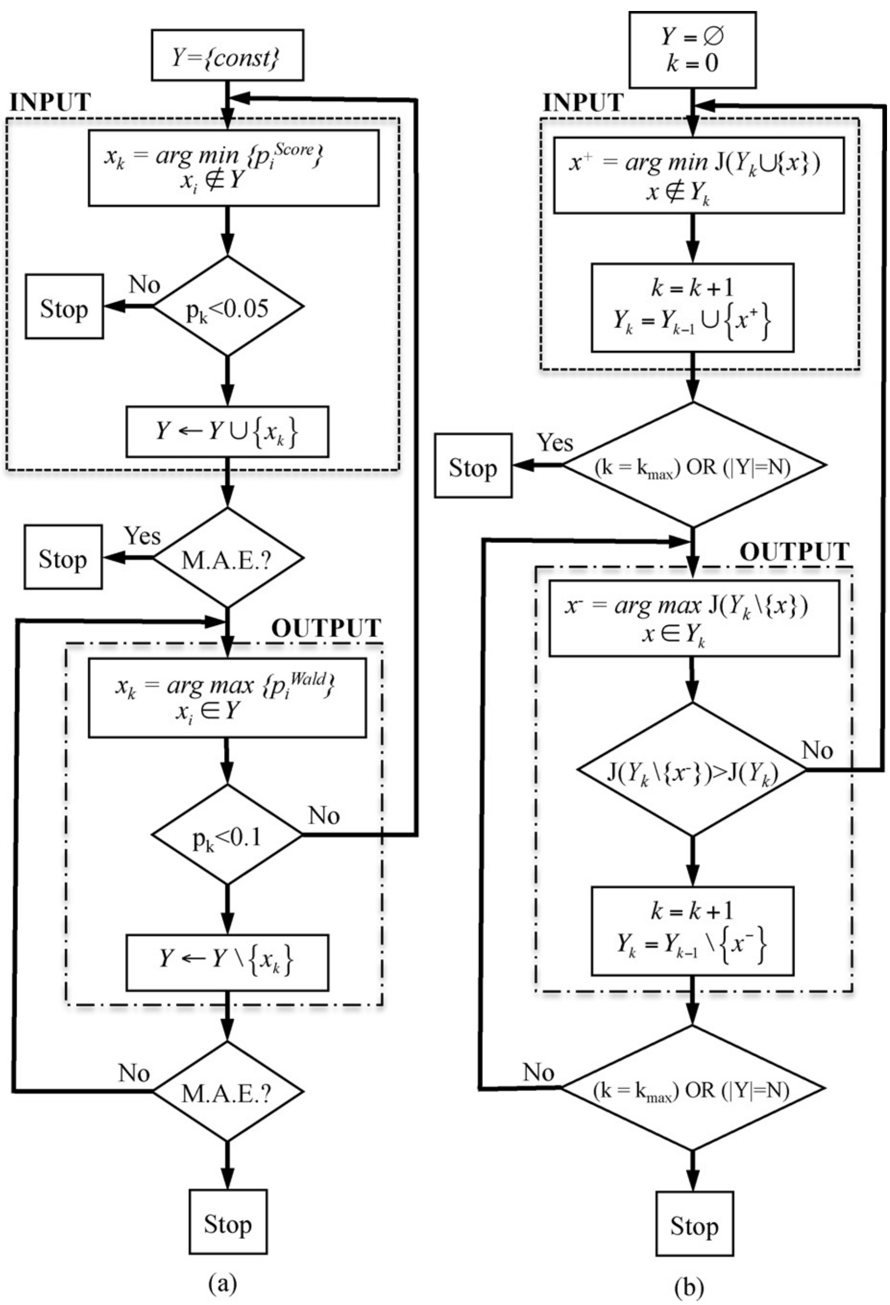

Fig. 1. (a) Forward stepwise selection algorithm used in the logistic regression classifier. Variables enter/leave the model as a function of the statistical significance p of the Score/Wald statistic, respectively. M.A.E. stands for model already evaluated. (b) Sequential floating forward selection algorithm used in the Bayes Classifier. After the end of the algorithm, the optimum model is Y_{opt} = arg max{$J(Y_k)$}.

the total false negatives (TFN), i.e., the number subjects classified into a class with too low a severity

$$TFN \equiv C1_{C2} + C1_{C3} + C2_{C3} \tag{6}$$

and the accuracy (ACC), defined as

$$ACC \equiv \frac{tr(C)}{sum(C)} \tag{7}$$

where $sum(C)$ is the sum of all the elements and $tr(C)$ is the trace of the confusion matrix C. The SFSS algorithm involves a single (main) performance measure $J(\cdot)$. If, in a given iteration, the maximum or minimum of $J(\cdot)$ is reached for several variable sets, a secondary measure $J2(\cdot)$ is used to select the best set. We have studied the behaviour of the SFSS algorithm with the six possible combinations of $J(\cdot)$ and $J2(\cdot)$ using the three performance measures (5)–(7). For each Bayes' classifier, we report the outcomes with the combination of $J(\cdot)$ and $J2(\cdot)$ that yielded the best classification rates.

2.6. Classification performance

For both logistic regression and Bayes' classifiers, the optimum set of independent variables is selected by the suitable forward selection algorithm (Fig. 1). In each case, all the observations available are used for the estimation. After that, the optimum models are validated by means of the leave-one-out cross-validation procedure.

3. Results

Table 5 shows the optimum independent variables automatically selected by FSS (logistic regression) and the SFSS (Bayes) algorithms for each classifier studied. We can see that the selected variables are mainly variability measures of the snoring features.

The evolution of the performance functions of the Bayes' classifiers is shown in Fig. 2 for the Gaussian-based, kernel *diffusion*-based and kernel density (*ksdensity*) estimation methods. The Bayes' classifier with *ksdensity* estimation achieves the best performance (100%) in the three measures of interest.

The classification performance of each individual logistic regression model under cross-validation is shown in the ROC curve of Fig. 3. We can see that the area under the curve (AUC) is above 0.9 in all cases.

Fig. 4 shows the data histograms and the PDFs estimated by the Gaussian and the two kernel methods for several independent variables of the optimum Bayes model with kernel density estimation (Table 5). Here we can clearly see that the PDFs of the snore features do not seem to follow a Gaussian distribution in any case. The kernel diffusion-based approach provides an estimated PDF that usually fits better the data than the Gaussian PDF, but in some cases its behaviour is not as expected at the edges, especially at the origin. The kernel *ksdensity* PDF estimation provides the best fit to the data.

Table 6 shows the confusion matrix of the optimum classification models. The Bayes model with Gaussian PDF estimation has the worst performance (lowest ACC and 1-TFN) amongst the three Bayes classifiers. The logistic regression method exhibits zero TFN and a high accuracy, but a slightly lower NPV_{C1}. The best results are again obtained with the Bayes classifier with kernel *ksdensity* PDF estimation.

The confusion matrix for the cross-validation of all the models is shown in Table 7. The validation rates of the kernel diffusion approach (ACC = 58.3%, 1-TFN = 80.7%, NPV_{C1} = 61.5%) are much lower than the Gaussian PDF ones (ACC = 75%, 1-TFN = 88.9%, NPV_{C1} = 76.9%). As in the optimum classification models, the highest classification rates are obtained by the kernel *ksdensity*

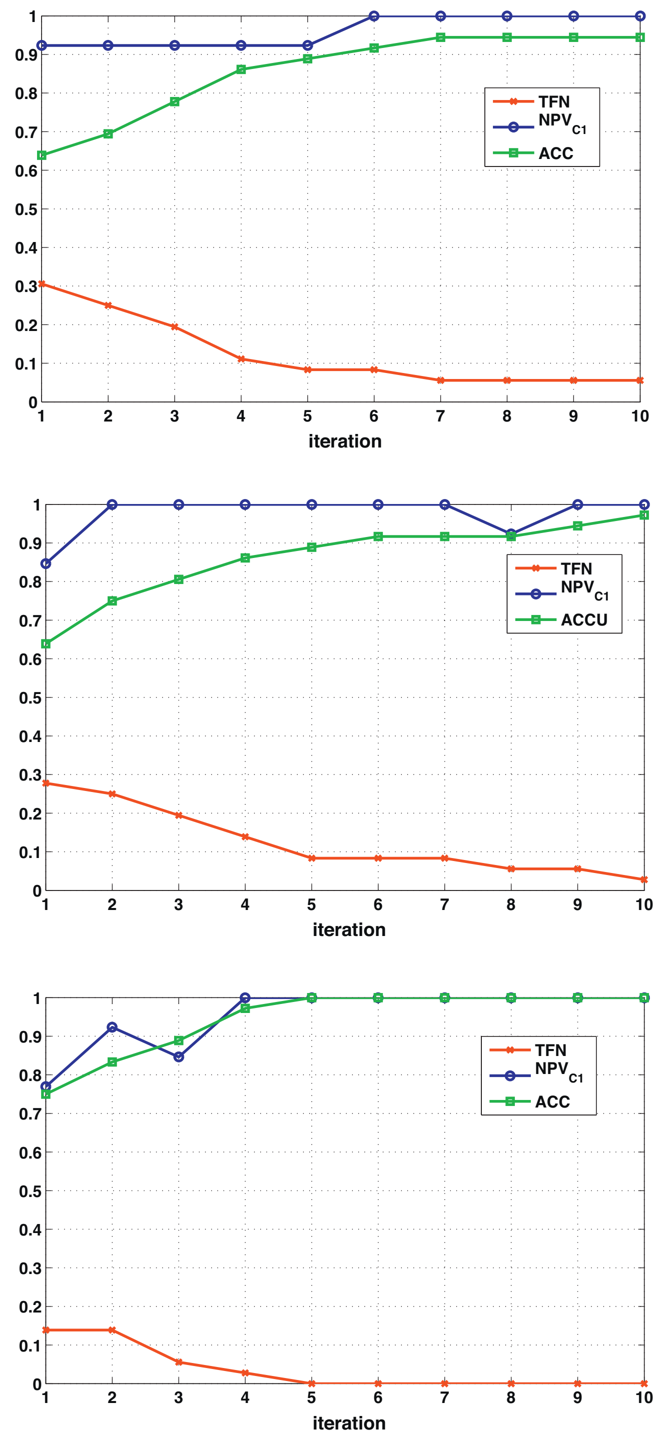

Fig. 2. Evolution of the optimum performance measures used in the sequential floating forward selection algorithm of the Bayes classifier, with PDF estimation by (a) Gaussian, (b) diffusion-based Kernel and (c) ksdensity Kernel approaches.

approach (ACC = 83.3%, 1-TFN = 100%, NPV_{C1} = 84.6%). It is worth noting that with this model, no moderate (C_2) or severe (C_3) SAHS patient would be classified as mild (C_1), that is TFN = 0% and this is not the case with any of the other models. Moreover, with the *ksdensity* kernel, 100% of severe patients (C_3) are classified as a C_3. In summary, the proposed Bayes classifier with kernel *ksdensity* PDF estimation method provides the highest classification rates, both in the optimum model as well as in the cross-validation.

Table 5
Optimum variables for each classification model.

Classifier	Average features	Variability features
LR (5)	–	$IQdtP_m$, $IQdtP_s$, $IQdtP_{dens}$, SdF_{mean}, $IQdRout_{500}$, SF_3
LR (30)	P_{iqr}	SP_m, SP_{dens}, $IQdtI_{mean}$, $IQdF_{med}$, SdCSymm, $IQdF_{max}$, SdF_3, SdL_3
B_{Gauss}	F_2	$IQdtP_m$, $IQdP_{iqr}$, SdCFlatn, $\mathrm{SdRout}_{100–500}$, $IQdtF_{peak}$, $IQdRout_{500}$, SL_1, SdL_2, SM_1
B_{Kdiff}	F_{mean}, I_{max}	$IQdtP_s$, SdF_{mean}, $IQdRout_{100–500}$, SdF_{peak}, $IQdF_{max}$, SRW_{500}, SF_3, SL_1
B_{Kdens}	*CFlatn*, F_{peak}, I_{mean}	$IQdP_s$, SCSymm, IQdCFlatn, SdF_{mean}, $SdRW_{500}$, SF_2, SdM_3

LR (*X*), Logistic regression with AHI threshold $X = 5, 30\ \mathrm{h}^{-1}$; *B*, Bayes; *Gauss*, Gaussian PDF; *Kdiff*, *diffusion*-based Kernel density function; *Kdens*, *Ksdensity*-based Kernel density function. The snore features are defined in Section 2.3 and in Table 2.

Table 6
Confusion matrix of the optimum classification models with variability features.

(a) Logistic Regression, ACC=94.4%, 1-TFN=100%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*84.6*	0	15.4
C_2	0	100	0
C_3	0	0	100

(b) Bayes Gaussian PDF, J=ACC, J2=NPV_{C1}, ACC=94.4%, 1-TFN = 94.4%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*100*	0	0
C_2	0	100	0
C_3	0	16.7	83.3

(c) Bayes Kernel_{diff}, J=ACC, J2=TFN, ACC=97.2%, 1-TFN = 97.3%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*100*	0	0
C_2	0	100	0
C_3	0	8.3	91.7

(d) Bayes Kernel_{kdens}, J=ACC, J2=TFN, ACC=100%, 1-TFN = 100%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*100*	0	0
C_2	0	100	0
C_3	0	0	100

C_1, AHI < 5; $C_2 = 5 \leq \mathrm{AHI} < 30$; $C_3 = \mathrm{AHI} \geq 30\ \mathrm{h}^{-1}$; ACC, Accuracy; TFN, Total false negatives; Kernel_{diff}, Diffusion-based kernel density function; Kernel_{kdens}, *Ksdensity*-based kernel density function.

4. Discussion

Our results show that high classification rates of snoring subjects with different levels of SAHS can be achieved based exclusively on the acoustic information extracted from snore signals, without any requiring information on the duration or the number of apnoeas. Each classification model contains the optimum independent variables (Table 5) automatically selected by the appropriate selection algorithm (Fig. 1). It is worth noting that, in all cases, most independent variables were describing variability in the characteristics of a patient's snores, and mainly snore-to-snore variability. This fact confirms our tenet that multiclass classification of snoring subjects with SAHS based only on the information extracted from snores is possible, provided that the variability in snoring characteristics is included in the models. We obtained much higher classification rates than in our previous work, in which variability in snoring characteristics was not considered.

It is natural to use binary logistic regression for the classification of subjects into two classes. In our previous studies using this technique, we obtained good classification rates with thresholds $\mathrm{AHI} = 5\ \mathrm{h}^{-1}$, $10\ \mathrm{h}^{-1}$ or $15\ \mathrm{h}^{-1}$, separately [7,11]. In a recent study, Karunajeewa et al. [17] have used logistic regression with snoring features for the classification of subjects into two classes using a threshold of $\mathrm{AHI} = 10\ \mathrm{h}^{-1}$. The ROC curves obtained in the present study (Fig. 3) are comparable to the ones obtained in their study. However, those authors did not set out to achieve a multiclass classification. We have shown that two binary classifiers can be easily combined to use logistic regression for a multiclass classification. The result is a classifier with a high accuracy and very low cross–validation error in terms of TFN (Table 7).

The Bayes method with a Gaussian PDF estimation was found to give the worst classification performance (lowest ACC, NPV_{C1} and 1-TFN), amongst the Bayes classifiers. This is probably due to the fact that snoring features do not follow a Gaussian distribution, as

Table 7
Confusion matrix of the cross-validation of the classification models with variability features.

(a) Logistic Regression, ACC=72.2%, 1-TFN = 91.7%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*53.8*	30.8	15.4
C_2	9.1	81.8	9.1
C_3	0	16.7	83.3

(b) Bayes Gaussian PDF, J=ACC, J2= NPV_{C1}, ACC=75.0%, 1-TFN = 88.9%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*76.9*	7.7	15.4
C_2	9.1	72.7	18.2
C_3	8.3	16.7	75.0

(c) Bayes Kernel_{diff}, J=ACC, J2=TFN, ACC=58.3%, 1-TFN = 80.7%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*61.5*	38.5	0
C_2	9.1	63.6	27.3
C_3	0	50.0	50.0

(d) Bayes Kernel_{kdens}, J=ACC, J2=TFN, ACC=83.3%, 1-TFN = 100%

Original Class (%)	Classified Class (%) C_1	C_2	C_3
C_1	*84.6*	7.7	7.7
C_2	0	63.6	36.4
C_3	0	0	100

C_1, AHI < 5; C_2, $5 \leq \mathrm{AHI} < 30$; C_3, $\mathrm{AHI} \geq 30\ \mathrm{h}^{-1}$; ACC, Accuracy; TFN, Total False Negatives; Kernel_{diff}, Diffusion-based kernel density function; Kernel_{kdens}, *Ksdensity*-based kernel density function. In each Bayes case, we only show the combination of performance measures $J(\cdot)$ and $J2(\cdot)$, which is what led to the best classification rates.

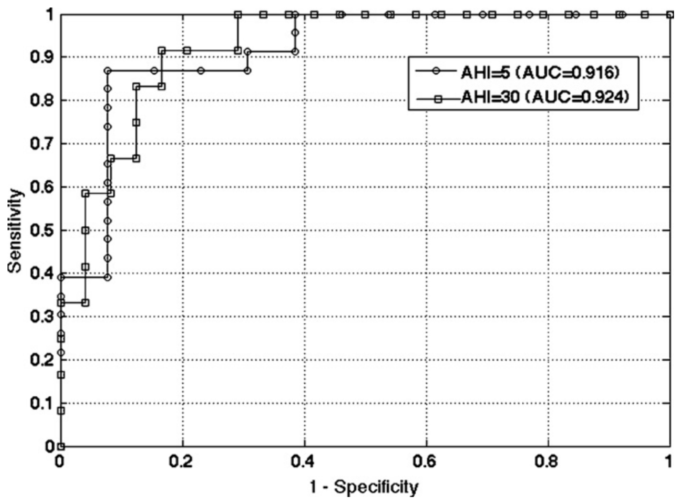

Fig. 3. Receiver operating curves for the cross-validation of the logistic regression models with thresholds AHI = 5 and 30 h^{-1}. AUC stands for area under the curve.

the data histograms and the PDF estimates clearly illustrate (Fig. 4). The classification performance is greatly improved if a kernel density estimation is used, which is not constrained by the Gaussian PDF assumption (Tables 6 and 7). Despite the optimum bandwidth selection implemented with the diffusion-based kernel method, the best results are obtained by the kernel classifier based on the *ksdensity* function. We suspect that this behaviour is related to the PDF estimated by the diffusion method being a poor fit to the actual data distribution, especially near the origin (see Fig. 4b and d).

The optimum snore-based Bayes classification, the one obtained with the *ksdensity* kernel density estimation, has two main advantages. On the one hand, it provides the highest percentage of healthy subjects (C_1) correctly classified (NPV_{C1} = 84.6%). On the other hand, the number of SAHS patients (C_2 or C_3) that are mistaken for healthy subjects is null (TFN = 0%), even after cross-validation of the model (Tables 6d and 7d). We must bear in mind that, in an early screening process, every healthy individual that is mistakenly sent to a second stage of diagnosis (in this case, PSG) means a substantial unnecessary cost to a public health care system. Nevertheless, the alternative, which occurs in all the other classification approaches explored, namely that some subjects would be erroneously diagnosed as not having SAHS, is also not desirable in a screening process.

In conclusion, in this paper we have shown that multiclass classification of snoring subjects with different levels of SAHS can be achieved with high accuracy, based exclusively on the acoustic information extracted from snore signals, without needing to know the duration or the number of apnoeas. The best approach is a Bayes kernel density estimation classifier, although good results can also be obtained by a suitable combination of two binary logistic regression models. The proposed method is a promising tool for screening SAHS as it provides a way of prioritizing patients for a second stage of diagnosis depending on their severity.

The present study has some limitations. The results obtained need to be validated using a database with a larger number of subjects. In addition, as stated in Section 2, the signals were recorded from subjects who were enrolled from the Sleep Disorders

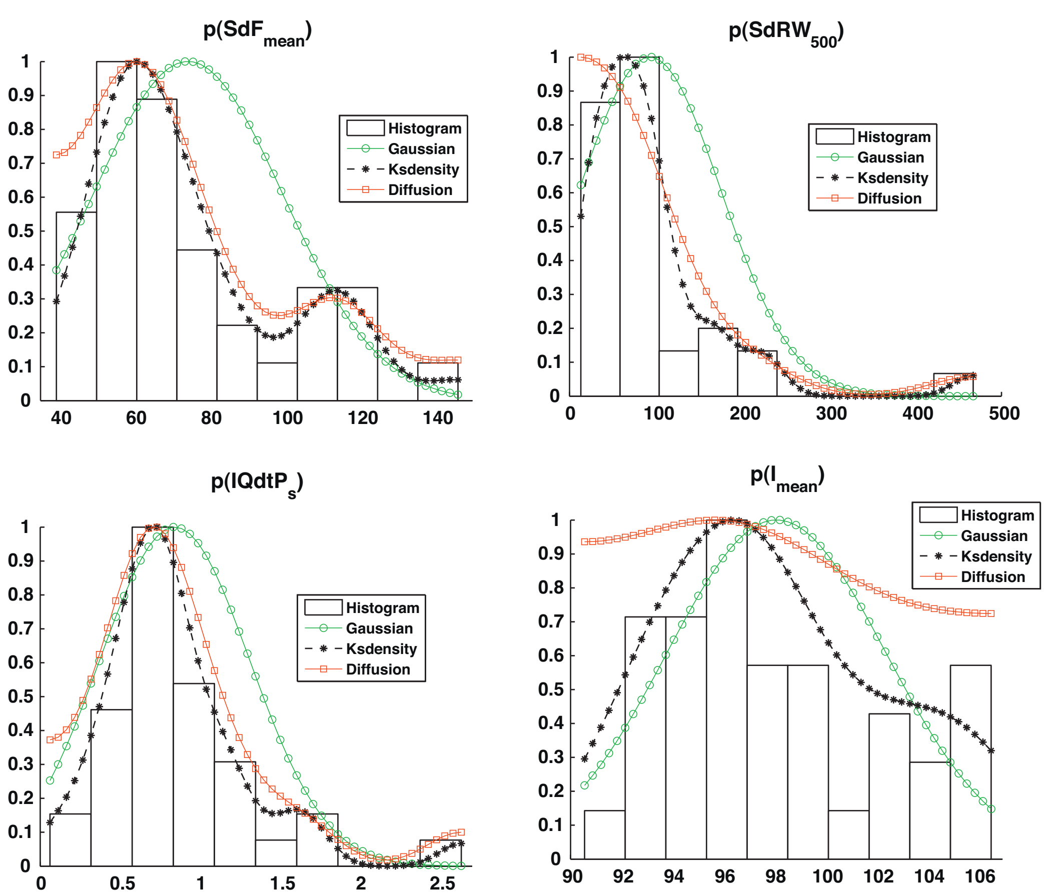

Fig. 4. Data histograms and PDFs estimated by means of the three proposed methods, for four of the independent variables shown in Table 4. The data histograms do not seem to follow a Gaussian distribution in any case. The ksdensity Kernel density estimation appears to provide the best fit to the data.

laboratory of the hospital. There are some subjects that, despite being habitual snorers, do not suffer from any other symptom such as sleepiness. Those subjects are usually derived to the Otorhinolaryngology service, and were not included in the present study. Nevertheless, in this article 13 subjects of the database (from a total number of 36 subjects) have an AHI < 5 h^{-1}, and therefore they have been diagnosed as non-SAHS.

Conflicts of interest

None.

Acknowledgements

This work was supported in part by the Ministerio de Ciencia e Innovación from Spanish Government under grant TEC2010-21703-C03-01. The authors would like to thank Pulmonology Service of the *Hospital Germans Trias i Pujol* for providing access to the snoring sound signals acquired in their Sleep Disorders Laboratory. The authors would also like to thank SIBEL S.A., for its collaboration with the medical instrumentation and for developing the Snoryzer Uno System.

References

[1] Eckert DJ, Malhotra A. Pathophysiology of adult obstructive sleep apnea. Proc Am Thorac Soc 2008;5(2):144–53.
[2] Wilson K, Stoohs RA, Mulrooney TF, Johnson LJ, Guilleminault C, Huang Z. The snoring spectrum, acoustic assessment of snoring sound intensity in 1,139 individuals undergoing polysomnography. Chest 1999;115(3):762–70.
[3] Marcos JV, Hornero R, Alvarez D, del Campo F, Zamarrón C. Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry. Med Eng Phys 2009;31(8):971–8.
[4] Caseiro P, Fonseca-Pinto R, Andrade A. Screening of obstructive sleep apnea using Hilbert–Huang decomposition of oronasal airway pressure recordings. Med Eng Phys 2010;32(6):561–8.
[5] Pevernagie D, Aarts RM, De Meyer M. The acoustics of snoring. Sleep Med Rev 2010;14(2):131–44.
[6] Perez-Padilla JR, Slawinski E, Difrancesco LM, Feige RR, Remmers JE, Whitelaw WA. Characteristics of the snoring noise in patients with and without occlusive sleep apnea. Am Rev Respir Dis 1993;147(3):635–44.
[7] Fiz JA, Abad J, Jané R, Riera M, Mañanas MA, Caminal P, et al. Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea. Eur Respir J 1996;9(11): 2365–70.
[8] Herzog M, Schmidt A, Bremert T, Herzog B, Hosemann W, Kaftan H. Analysed snoring sounds correlate to obstructive sleep disordered breathing. Eur Arch Otorhinolaryngol 2007;265(1):105–13.
[9] Cavusoglu M, Ciloglu T, Serinagaoglu Y, Kamasak M, Erogul O, Akcam T. Investigation of sequential properties of snoring episodes for obstructive sleep apnoea identification. Physiol Meas 2008;29(8):879–98.
[10] Ng AK, Koh TS, Abeyratne UR, Puvanendran K. Investigation of obstructive sleep apnea using nonlinear mode interactions in nonstationary snore signals. Ann Biomed Eng 2009;37(9):1796–806.
[11] Fiz JA, Jané R, Solà-Soler J, Abad J, García MA, Morera J. Continuous analysis and monitoring of snores and their relationship to the apnea–hypopnea index. Laryngoscope 2010;120(4):854–62.
[12] Ben-Israel N, Tarasiuk A, Zigel Y. Nocturnal sound analysis for the diagnosis of obstructive sleep apnea. In: Conf Proc IEEE Eng Med Biol Soc. 2010. p. 6146–9. PMID: 18003405 [PubMed – indexed for MEDLINE].
[13] Solà-Soler J, Jané R, Fiz JA, Morera J. Automatic classification of subjects with and without Sleep Apnea through snoring analysis. In: Conf Proc IEEE Eng Med Biol Soc. 2007. p. 6094–7. PMID: 18003405 [PubMed – indexed for MEDLINE].
[14] Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. Ann Stat 2010;38(5):2916–57.
[15] Kraman SS, Wodicka GR, Oh Y, Pasterkamp H. Measurement of respiratory acoustic signals. Effect of microphone air cavity width, shape, and venting. Chest 1995;108(4):1004–8.
[16] Jané R, Fiz JA, Solà-Soler J, Blanch S, Artis P, Morera J. Automatic snoring signal analysis in sleep studies. In: Conf Proc IEEE Eng Med Biol Soc. 2003. p. 366–9.
[17] Karunajeewa AS, Abeyratne UR, Hukins C. Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome. Physiol Meas 2011;32(1): 83–97.
[18] Flemons WW, Littner MR, Rowley JA, Gay P, Anderson WM, Hudgel DW, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. Chest 2003;124(4):1543–79.
[19] SPSS. SPSS 16.0 Algorithms, http://support.spss.com/ProductsExt/SPSS/Documentation/Manuals/16.0/SPSS16.0Algorithms.pdf.
[20] Jain AK, Duin RPW, Jianchang M. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 2000;22(1):4–37.