

AN AUTOMATIC AND EFFICIENT METHOD OF SNORE EVENTS DETECTION FROM SLEEP AUDIO RECORDINGS

F. Gritti¹, L. Bocchi¹, I. Romagnoli², F. Gigliotti² and C. Manfredi¹

¹ Department of Electronics and Telecommunications, Università degli Studi di Firenze, Firenze, Italy

² Fondazione Don C. Gnocchi, Impruneta, Firenze, Italy

Abstract: Snores are respiratory sounds produced during sleep. They are reported to be a risk factor for various sleep disorders, such as obstructive sleep apnea syndrome (OSA). Diagnosis of OSA relies on the expertise of the clinician that inspects whole night polysomnographic recording. This inspection is time consuming and uncomfortable for the patients. Thus, there is a strong need for a tool to analyze snore sounds automatically. Nocturnal respiratory sounds are composed of two kind of events: “silence” episodes and “sound” episodes that include breathing, snoring and “other” sounds.

In this paper a new method to detect snoring episodes from full night audio recordings is proposed. Signal analysis is performed in three steps: pre-processing, automatic segmentation, extraction of features and classification. With the segmentation step, only the “sound” parts of the audio signal are extracted using a Short-Term Energy and the Otsu thresholding method. The aim of classification step is the detection of snore episodes only, using two Neural Artificial Network applied to four features (length, maximum amplitude, standard deviation and energy).

Data from 24 subject are analyzed using the proposed method; on the dataset, a sensitivity of 86,2% and specificity of 86,3% are obtained.

Keyword: Snore, Obstructive sleep apnea, Neural network, Automatic segmentation

I. INTRODUCTION

Snoring can be defined as a respiratory noise that is generated during sleep when breathing is obstructed by a collapse in the upper air way. Loud and regular snoring is the earliest and most consistent sign of upper airway (UA) dysfunction leading to sleep apnea/hypopnea syndrome [1].

Obstructive sleep apnea (OSA) is the most frequent encountered form of the sleep apnea [1]. In OSA, the upper airways are obstructed during sleep, resulting in the decrease of oxygen flow to the lungs. Patients suffering from OSA often wake up frequently. When there is a full closure of airways, the disease is termed “apnea” while when there is a partial closure, it is known as “hypopnea” [2]. The disease is associated with significant clinical consequences but it is frequently unrecognized and

undiagnosed because simple, low-cost devices for mass screening of the population do not yet exist.

The current “gold standard” method for sleep apnea assessment is Polysomnography (PSG). This technique requires a full night hospital during which the patient is connected to more than ten channels of measurements requiring physical contact with sensors. PSG is thus inconvenient, expensive and unsuited for community screening [3] [4] [5]. Thus, in order to study OSA non-invasively, several researches focused on the analysis of snore sounds from full night audio signal recordings, using signal processing techniques.

Commonly tracheal respiratory sounds are recorded using a microphone placed over the patient’s neck or hung above the patient’s head during the night, leading to long lasting audio signals (6–8 hours). The length of a whole recordings is thus prohibitive for the analysis by listening to and for visual inspection of signal patterns. Hence, automatic methods are needed to speed up the analysis task.

Despite its clinical relevance, a limited number of studies on automatic detection and classification of snore sound has been developed to date [6], [7], [8], [9], [10], [11]. In these works different kind of techniques of analysis are applied, such as: Energy and zero-crossing rate [6][7] [8], Hidden Markov Models (HMMs) and spectral-based features [9], 500Hz sub-band energy distribution [8],[10], normalized autocorrelation coefficient at 1 ms delay and the first predictor coefficient of LPC analysis [6], and frequency range of each formant [11].

However, most often the automatic segmentation step is not included, the snore events being detected manually or with semi-automatic methods.

Hence the motivation of this study was to develop an effective method to detect the snoring episodes, fully automatic and fast enough to allow processing full night recordings in a reasonable amount of time.

A short-term energy measure was implemented for automatic detection of “sound” events and two neural artificial network were applied to four features (length, maximum amplitude, standard deviation and energy), for automatic classification of snore events.

II. METHODS

The aim of the proposed system of analysis is the detection of snoring events from full night audio

recordings. This is achieved by means of the following three steps:

- A. *Pre-processing*: loading of audio signal, band-pass filtering and down sampling;
- B. *Automatic segmentation*: detection of the “sound” parts of the signal;
- C. *Extraction of the features and classification*: identification of snoring events.

The implemented method, named Snore Analyzer, is developed under Matlab 7.11.00 software tool. A flow chart is shown in Figure 1.

Snore Analyzer is provided with a user-friendly interface (Figure 2) that easily allow the user to choose the audio signal to be processed (*Load* bottom) and set the following parameters for subsequent processing: 1) Sampling frequency (44.100 kHz by default); 2) Down sampling frequency (11.025 kHz by default); 3) Starting and ending samples, to select the part of the signal to be processed; 4) Size of analysis window (40 ms by default).

Then the user starts the elaboration of the selected audio signal pushing the *Start* bottom. Through the *Reset* bottom, the user can delete all the items.

The elaboration of whole signal (or a part of it) is fully automatic and the user should not act manually anymore.

The length of each audio signal is about 7-8 hours and the complete analysis of whole signal requires about 30-40 minutes. At the end, the software gives as output a list of extracted “sound” events which are labeled as snore or not-snore.

The next sections (A, B, C) describe each step in detail.

A. Pre-processing

The use of a robust recording system can improve signal acquisition, but noise reduction is required to eliminate interferences. Therefore a pre-processing step is implemented to improve signal to noise ratio.

In this study the audio signal is bandpass filtered by a Butterworth filter of order 5 and a cut-off frequency of 100 – 1000 Hz, to reduce the effects of heart sounds and high-frequency noises [1]. Main frequency components of breathing and snoring sounds are in fact included in this range [12] [13]. After the filtering step, the signal is down sampled (to 11.025 kHz), to reduce the size of the data and hence speed up signal processing.

B. Automatic Segmentation

The audio signal is typically a mixture of two different kind of events: “silence” that do not contain any sound and “sound” that include breathing episodes, snoring episodes and “other” sounds such as oral noise, ambient sounds, patient’s cough, speech and blanket movements, etc.

This step is therefore devoted to identify the “sound” events. Short-Term Energy (STE) is a commonly used

measure for determining the “sound” parts as it increases during “sound” events and decreases during “silence” episodes [14] [15].

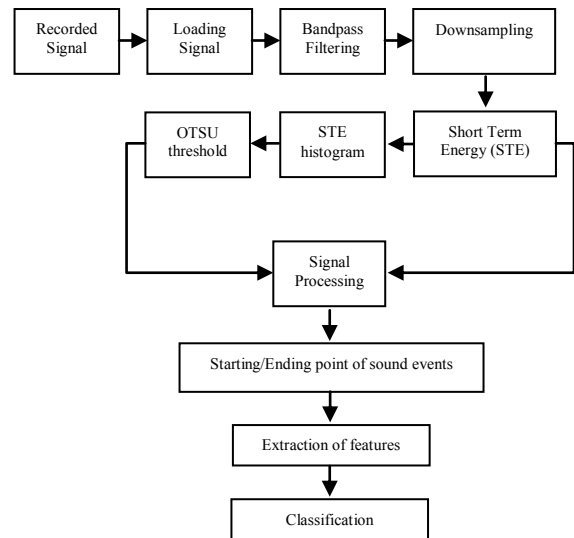


Fig. 1 Flow chart of the analysis system.



Fig. 2 User-friendly interface of the implemented software tool.

In our study, STE is evaluated in signal windows of 40 ms in length with 50% overlap between adjacent windows. In order to determinate boundaries of “sound” events, we computed the histogram of the signal energy and the Otsu method is iteratively applied to obtain two thresholds: the upper one t_u and the lower one t_l [16], [17]. These thresholds are then used to find the starting and ending points of each “sound” event in the audio signal. In particular, when the STE curve overpasses the upper threshold, the first point under the lower threshold (on the left side of the curve with respect to the upper threshold) is detected in order to get the starting point. When the STE curve falls down t_l , the ending point of the event is found (Figure 3).

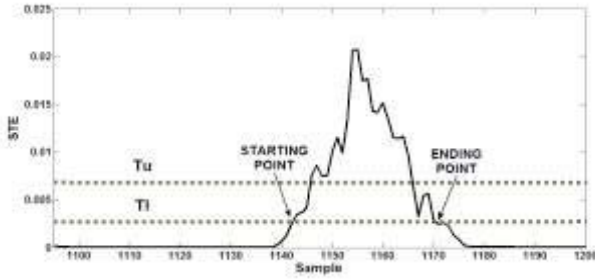


Fig. 3 The starting and ending points of a “sound” event.

At the end of segmentation, the list of extracted “sound” events is saved in a text file to be used in the classification step.

C. Extraction of the features and Classification

Once all the “sound” events from the signal are obtained, they have to be classified as snore or not-snore (i.e. breath and “other” events). In fact for a reliable analysis of OSA, only snore episodes must be detected. This task is carried out in two step: in the first one, a set of four parameters is computed in time domain; in the second one, the events are identified with a classification system.

The first parameter is the length of each “sound” event, calculated as the distance between the starting and the ending point of the event. This feature allows to distinguish between “other” events and breathing/snoring, as the average length (in samples), computed for breathing and snoring sounds, is lower than for “other” sounds, as shown in Table 1.

Table 1 Mean and Standard Deviation of the length of snore/breath and of “other” sounds.

	Length [sample]	
	Mean Value	STD Value
Snore/Breath	$4.7999 \cdot 10^4$	$2.8492 \cdot 10^4$
“Other”	$1.4405 \cdot 10^4$	$1.3602 \cdot 10^4$

The other parameters are: the Standard Deviation (STD), the mean value of Short-Term Energy (STE) and the maximum amplitude of “sound” events, given by the difference between the maximum and the minimum amplitude of the signal.

These features allow to distinguish between snoring episodes and breathing episodes, as the average value of each single feature is higher in the class of snoring events than in the class of breathing events (Table 2) while the behaviour of these parameters is highly variable in “other” sounds.

Hence the following observation can be made: “other” sounds can be found using the length of the events only; snoring and breathing sounds can be distinguished using the STD, the mean value of the STE and the maximum amplitude.

Table 2 Mean and Standard Deviation of STD, mean of STE and Maximum Amplitude.

	STD	
	Mean Value	STD Value
Snore	0.0038	0.0024
Breath	0.0014	0.0005
	STE	
	Mean Value	STD Value
Snore	-5.4120	0.5283
Breath	-6.0974	0.3061
	Maximum Amplitude	
	Mean Value	STD Value
Snore	0.0498	0.0389
Breath	0.0142	0.0061

According to these results, a classifier is designed made up by two artificial neural networks: the first one is used to identify the “other” sounds, while the second one is used to discriminate between snoring and breathing sounds.

The sounds episodes were manually labelled by trained clinicians as snore or not snore to built the training and the testing datasets for the classification system. The training set is made up by 1643 sound signals equally distributed among snoring, breathing and ‘other’ sounds.

The first network is trained with all the events of the training set using only the length of the event as input and the outcome of listening is used as teaching input.

After the training step, the network output is tested and compared with the outcome of listening; the “other” sounds correctly recognized as “other” (true negative) are removed from the training set used in the second network that consists of three inputs, corresponding to the mean value of STE, its STD and the maximum amplitude, respectively.

III. RESULTS

Clinical audio signals (18 patients of different age and sex) are recorded at Fondazione Don Gnocchi, Pozzolatico, Firenze, where the patients slept in single bedroom, separated from television and others predictable sources of noise.

The audio signal are digitized at 16-bit with a sampling frequency $F_s=44.100$ kHz, using a Tascam Us-144 sound card and a unidirectional microphone Shure SM58, positioned at about 30 cm from the mouth of the

patient. The length of single signal is of about 7-8 hours, but, for the analysis, we considered thirty minutes of each recording, selected in the central part of the signal when the patient was sleeping and low environmental noise was detected.

A preliminary evaluation was carried out to assess the performance of the automatic segmentation, evaluated as the percentage of sounds detected over the total number of sounds, Resulting in about 97%.

Concerning the classification step, the first network was tested on 787 “sound” events, different from the original training set. From the analysis of the ROC curve, a “best” threshold was obtained that allows to correctly identify 85.4% of the “other” sounds. These sounds were stored in a list of not-snore events and removed from the test set.

The second network was tested on the remaining sounds and, as for the first network, the best ROC threshold was computed and used to identify snore and not-snore sounds.

The accuracy (number of correct classifications) of the second network was found equal to 86.2%. This result corresponds to a sensitivity (true positive (TP) ratio) equal to 86.2 and a specificity (true negative (TN) ratio) equal to 86.3.

IV. DISCUSSION AND CONCLUSIONS

A full automatic and unsupervised system for snore identification during sleep is proposed.

The proposed automatic segmentation was shown to be a reliable technique for the extraction of sound events as almost all silence events were discarded.

The algorithm for classification correctly identifies the 86.2% of analysed events. However it fails in case of low intensity snores, as such events have low energy and low maximum amplitude. But, as post apnoeic snore events are usually more intense than non-post apnoeic ones, this limitation could be acceptable.

Future work will be devoted to enhancing the procedure, increasing the dataset and defining a reliable method for the identification of post-apnoeic events from the automatically detected snore sounds, e.g. as in [18].

REFERENCES

- [1] M.J. Thorpy, “*The international classification of sleep disorders: diagnostic and coding manual*”, Lawrence KS, ed. Allen Press Inc., USA, 1990, pp. 195–197.
- [2] N.J. Douglas, *Harrison’s Principles of Internal Medicine*, McGraw-Hill, 17th ed. New York, 2008.
- [3] A. Ayatollah. and Z. Moussavi, “Automatic breath and snore sounds classification from tracheal and ambient sounds recordings”, *Medical Engineering & Physics*, vol. 32, pp. 985-990, 2010.
- [4] W.W. Flemons, M.R. Littner, J.A. Rowley, P. Gay, W.M. Anderson, D.W. Hudgel, R.D. McEvoy, and D.I. Loube, “Home diagnosis of sleep apnea: A systematic review of the literature,” *Chest*, vol. 124, pp. 1543–1579, 2003.
- [5] C.A. Kushida, M.R. Littner and T.Morgenthaler, “Practice parameters for the indications for polysomnography and related procedures: An update for 2005,” *Sleep*, vol. 28, no. 4, pp. 499–521, 2005.
- [6] A.S. Karunajeewa, R. Abeyratne and C. Hukins, “Silence–breathing–snore classification from snore-related sounds”, *Physiol. Meas.*, vol. 29, pp. 227–243, 2008.
- [7] U.R. Abeyratne, A. Wakwella and C. Hukins, “Pitch jump probability measures for the analysis of snoring sounds in apnea”, *Physiol. Meas.*, vol. 26, pp. 779–798, 2005.
- [8] A. Ayatollah. and Z. Moussavi, “Automatic and Unsupervised Snore Sound Extraction From Respiratory Sound Signals”, *IEEE Transaction on Biomedical Engineering*, vol. 58, n.5, pp. 1156-1162, May 2011
- [9] W.D. Duckitt, S.K. Tuomi and T.R. Niesler, “Automatic Detection, segmentation and assessment of snoring from ambient acoustic data”, *Physiol. Meas.*, vol. 27, pp. 1047-1056, 2006.
- [10] M. Cavusoglu, M. Kamasaka and O. Erogul, “An efficient method for snore/notsnore classification of sleep sounds”, *Physiol. Meas.*, vol. 28, pp. 1-13, 2007.
- [11] A. Yadollahi and Z. Moussavi, “Formant analysis of breath and snore sounds”, *Proc. IEEE EMBS, Minneapolis, MN, 3-6 Sept. 2009*, pp. 7110 – 7113, 2009.
- [12] J. Fiz, J. Abad, R. Jane, M. Riera, M.A. Mananas and P. Caminal, “Acoustic analysis of snoring in patients with simple snoring and obstructive sleep apnea”, *Eur Respir J*, vol. 9, pp. 146-59, 1996.
- [13] R. Beck, M. Odeh, A. Oliven and N. Gavriely, “The acoustic properties of snores”, *Eur Respir J*, vol. 8(12), pp. 2120-8, 1995.
- [14] J.R. Deller, J.H. Hansen and J.G. Proakis, *Discrete-Time processing of speech signal.*, Mcmillan Publishing Company, New York, 1993, pp. 724-728.
- [15] A. Kulkas, E. Huupponen, “Intelligent methods for identifying respiratory cycle phases from tracheal sound signal during sleep”, *Comput Biol Med*, vol. 39, pp. 1000-1005, 2009.
- [16] N. Otsu, “A threshold selection method from gray-level histograms”, *IEEE Trans. Sys., Man., Cyber*, vol. 9(1), pp.62-66, 1979
- [17] M. Calisti, L. Bocchi, C. Manfredi, I. Romagnoli, F. Gigliotti and G. Donzelli, “Automatic detection of post-apnoeic snore events from home and clinical full night sleep recording, *MAVEBA 09*, Florence, Italy, pp.189-192, 2009
- [18] E. Goldshtein, A. Tarasiuk, and Y. Zigel, “Automatic Detection of Obstructive Sleep Apnea Using Speech Signals” *IEEE Transaction on Biomedical Engineering*, vol. 58, n.5, pp. 1373-1382, May 2011