AUTOMATIC DETECTION OF POST-APNOEIC SNORE EVENTS FROM HOME AND CLINICAL FULL NIGHT SLEEP RECORDINGS

M. Calisti [a], L. Bocchi [a], C. Manfredi [a], I. Romagnoli [b], F. Gigliotti [b], G. Donzelli [c]

[a] Department of Electronics and Telecommunications, Università degli Studi di Firenze, V.S.Marta 3, 50139 Firenze, Italy
[b] Fondazione Don C. Gnocchi, Via Imprunetana 124, 50023, Firenze, Italy
[c] Departments of Paediatrics, Università degli Studi di Firenze,V.le Pieraccini 24, 50139 Firenze Italy

Abstract: **Snoring is the hallmark of the Obstructive Sleep Apnoea Syndrome and several studies explore possible correlations between them. In this work an improved methodology with respect to [4] is proposed, based on a proper energy threshold applied on audio recordings for sound/silence detection, and on a feature vector of 14 elements (13 Mel Frequency Cepstral Coefficient plus the number of zero crossings) for sound classification. This feature vector is obtained from a 62-elements one by applying a genetic algorithm, fitted to obtain the best classification of the training/validation sets.**
The feature vector is analyzed by means of a radial basis neural network to perform snore events identification. Finally, formant frequencies and time analysis are also investigated to split up post-apnoeic snores and normal ones.
Audio data from 26 patients of different age and sex are used to test the methodology: 6 patients (3 male and 3 female) were used to train the nets (1800 snores) and 4 patients to validate the classification (600 snores). On the whole dataset of patients, a sensitivity between 69% and 84% is obtained in the detection of post-apnoeic snores.

Keywords: **Snore, neural network, Mel frequency cepstral coefficients, genetic algorithm, obstructive sleep apnoea.**

I. INTRODUCTION

Obstructive Sleep Apnoea (OSA) is a pathological condition where the upper airways collapse, reducing or cutting the flow to the mouth/nose. The diagnosis of Obstructive Sleep Apnoea Syndrome (OSAS) is commonly made by means of Polysomnographic (PSG) examination. PSG is mainly performed in a clinical environment (sleep laboratories), but could also be performed in home environment. However, PSG examination is bothering for patient, unsuited for mass screening purposes and expensive. Hence, new, simpler and non-invasive methods are investigated to detect OSAS. At present, according to the Italian guidelines, OSA is detected from full-night sleep analysis (uninterrupted recordings lasting from 6 to 10 hours) by means of PSG. Such a huge amount of data implies several technical problems concerning acquisition, storage and processing of data. Hence, efforts are made in the scientific community to define reliable OSA identification techniques from the audio signal only. At present, processing is made over the whole signal that is commonly classified into three classes: snore, breath, silence [1] or five classes: snore, breath, silence, duvet noise, other noise [2]. Other works consider just temporal features [3].

In this work we propose an automatic detection of snore events, that extends the results obtained in [4], followed by an evaluation of the number of Apnoeas or Hypopnoea events (AHI Index) with the methodology proposed in [5]. Our approach allows to split-up snores from other sounds, without predefining other sound classes, thus reducing the total length of the signal to be processed. The method is developed under Matlab2007a®. Full-night audio data (26 patients) come both from clinical and home recordings.

II. METHOD

The flow chart proposed in [4] is revised here, with the aim of performing a faster analysis and a more careful sound/silence segmentation. Firstly, we evaluate the histogram of the audio signal energy to perform an Otsu thresholding [6]. This method has the advantage that it does not require data pre-filtering, as a good energy separation between sound and silence is expected from our recordings [1], even in home environment. This assumption has been verified with a careful setup of the process, both as far as the device and the environmental setup are concerned. Specifically, a unidirectional microphone has been used to perform recordings connected to an external sound card to reduce noise artefacts of the laptop sound card. Patients were separated from bed partner and/or from pets, television and other predictable sources of noise. Moreover, the first 30 minutes of each recording were cut off, to avoid noise due to patient's movements, speaking with the clinician and similar ones. After the selection of sound events, a proper classification is proposed based on features extracted from 60 Mel Frequency Cepstral

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 6th international workshop: December 14-16, 2009*, ISBN 978-88-6453-094-9 (print), 978-88-6453-096-3 (online), © 2009 Firenze University Press.

Coefficients (MFCC) plus short term energy (STE) and the number of zero crossing (NZC), where:

$$\mathrm{STE} = \log\left(\frac{\sum_{i=1}^{n} s(i)^2}{n}\right) + k \qquad (1)$$

$$\mathrm{NZC} = \frac{\sum_{i=1}^{n-1} \left|\mathrm{sign}\big(s(i+1)\big) - \mathrm{sign}\big(s(i)\big)\right|}{2} \qquad (2)$$

Where n=441 is the number of elements in each window, sign() is the sign function, s is the signal and k is a small constant value to avoid log(0). Mean and standard deviation of the MFCCs are obtained as in [4].
As the choice of the number of coefficients is often arbitrary or derived from boundary conditions, we performed a careful search of the most representative MFCCs by means of a genetic algorithm (GA), where each gene of a phenotype represents a MFCC. The population of feature vectors was processed by the neural network until we obtained the best fitting according to proper classification. Furthermore, after low-pass filtering the audio signal (2 kHz cut-off), the number of zero crossings has been used as a selection feature for snore/non-snore events.
Several methods are proposed in literature to separate OSA events from non-OSA ones. Here we adapted the one proposed in [5] with the aim of identifying snore episodes after apnoea ones. This allows obtaining an AHI index related to apnoeic events only. A detailed flow chart of the process for the best feature vector selection is reported in Fig.1.
Short term energy, number of zero crossing and MFCCs extraction from the signal are performed according to [4]. Mean (m) and standard deviation (std) of the MFCCs for all frames of an event are also evaluated.
To detect starting and ending points of the event, the Otsu methodology [6] was iteratively applied to obtain two thresholds, the upper one and the lower one. The histogram was settled up to 2000 levels. After a first upper threshold detection t_u, a second Otsu thresholding was performed from level zero to level t_u, to obtain a lower threshold t_l. When the STE of the signal overpasses t_u, a starting point is detected, when the STE of the signal falls down t_l, the ending point of the event is found. As in [4] this procedure allows to obtain two sets representing the starting and the ending points of the events. Filtering only these events instead of the whole signal greatly speeds up the signal processing.
Once we have obtained all the events from the recording, we listened and classified the various frames as snoring or non-snoring frames to prepare a training set. We classified about 600 events from 6 patients (3 male and 3 female) without regarding the prevalence of the pathology, for a total of about 1800 snoring frames and 1500 non-snoring frames.
At present, most of the approaches try to classify snoring and "other events", e.g. mainly breath. However different noise events are included in "other events" that are difficult to classify. Hence a different approach is presented here, where we train the net with feature vectors representing only snore.

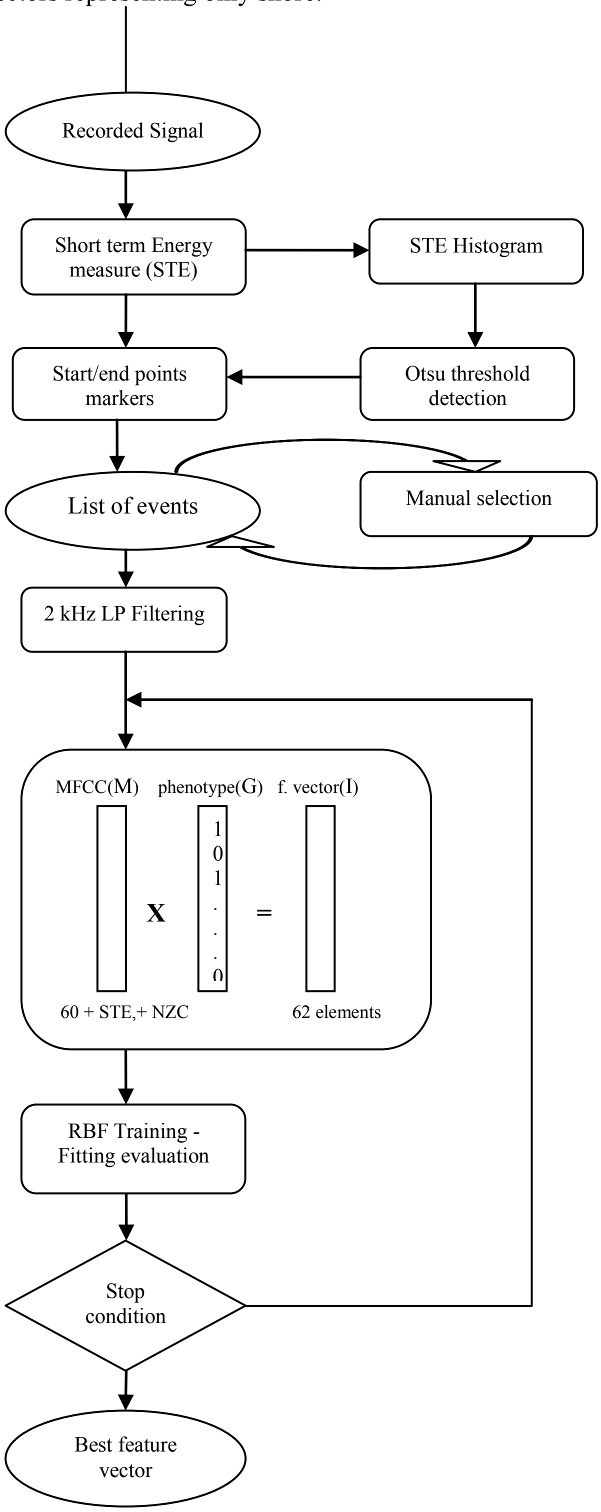

Figure 1. Flow chart of the sound detection and snore identification

Hence, if a unclassified feature vector (named here *void vector*) is presented to the net, according to the similarity of this new feature vector to the ones presented to the net during the training step, snore frames can be separated from non-snore ones.
In our opinion, this approach is more general as it is not proper to assume that cough, bed noise, breath and similar sounds belong to the same class, as was proposed in our previous work [4], though quite good results were obtained.
The improvement proposed here is based on a Radial Basis Neural Network (RBNN) that provides, as output of the hidden layer, a vector D representing the distance between the input vector (our feature vector of 62 elements) and the input weight matrix. For the radial basis neuron, the output is equal to 1 when the distance between the weight vector and the input vector is 0. Hence, the maximum value of D points out if the feature vector represents a snore or not.
All the 1800 frames representing snores are used as training set. Then, presenting several input vectors to the net, and taking into account the maximum of the output vector, 600 frames not already presented to the net are used to evaluate the network. The number of correct classifications over the validation set of 600 frames represents the value of our fitting function:

$$fit = -\frac{TP+TN}{600} \tag{3}$$

Where TP=true positive (a snore correctly recognized as a snore), TN=true negative (a non-snore correctly recognized as a non-snore). The minus sign is due to the fact that the most common genetic algorithm tools aim to minimize the fitting function.
To perform the GA, the input vector I to the net was obtained as the product between the vector M of 62 elements (30 mean values and 30 std of the MFCCs plus STE and NZC) and a binary feature vector called phenotype G that represent the elements of the vector M that will belong to the input vector I or not. The various individuals of the population for the GA are different Gs with different combinations. The stop condition was set at 30 minutes of elaboration. The whole process is shown in Fig.1.
After GA optimization, the resulting best feature vector was used to train an Optimized Radial Basis Neural Network (ORBN) and to test the net on the whole database of patients. Here “optimized net” means a net trained with the optimized input set.
The snore event recognized as snore is than processed to identify the post-apnoeic snore event, according to [5]. Also, a temporal feature is taken into account, based on the assumption that at least 10s of silence should exist before the snore to satisfy the apnoea definition (air flow absence lasting 10s at least [7]).

III. RESULTS

Experiments were carried on under the same conditions as in [4], and with the same equipment. Mainly three blocks of the chain in Fig.1 affect our results: the sound detection from the whole recording; the snores recognition from the sound; the OSA-snores recognition from the snores.
The first step, mainly related to the reduction in time of the whole recording, gives good results. As an example, results for 4 subjects are shown in Table 1.

Table1.Examples of reduction in time of the recordings.

Patient	Time of whole recording (min)	Time of whole events (min)
Subject 1	572	37
Subject 2	446	47
Subject 3	592	70
Subject 4	476	31

The accuracy of this step, evaluated as the number of sounds detected over the total number of sounds, is about 96,65% (ranging from 93% to 99%). This accuracy was computed by listening to about 1 hour of recording for 10 patients.
The second step was validated by listening to 50 events extracted from 6 different patients. As in [4], an event is classified as snore if there is at least one frame recognized as snore in the whole event. The sensitivity, measured as TP/(TP+FN) varied from 87,1% to 97,82% with a good improvement with respect to [4]. Here FN (false negative), represents a snore wrongly recognized as non-snore.
The best phenotype was obtained running five times the GA, with the stop condition of 30min running, but in all cases the problem was optimized after 10 generations. From the five best phenotypes obtained, only the elements common to all of them were used, thus discarding 6 elements. Thus, the best phenotype is composed by 14 elements from the 62 of the original one, as shown in Table 2.
Notice that the OSA evaluation was carried on offline after the automatic snore extraction. Only the snores that follow a silence longer than 10s were analyzed.
Finally, we extended what suggested in [5], considering as apnoeic snores only the snores occurring after an apnoea event. In this way, we notice a little increasing of the post apneic snore formant frequencies.
Taking into account the difference on formant frequencies and the temporal consideration regarding a 10s silence before sound, we obtained a sensitivity varying from 85% to 87%.

Table 2. Best phenotype obtained from GA

Element of I	Element of M	Meaning
1	M(1)	Mean of 1° MFCC
2	M(4)	Std of 2° MFCC
3	M(5)	Mean of 3° MFCC
4	M(7)	Mean of 4° MFCC
5	M(14)	Std of 7° MFCC
6	M(15)	Mean of 8° MFCC
7	M(23)	Mean of 12° MFCC
8	M(24)	Std of 12° MFCC
9	M(30)	Std of 15° MFCC
10	M(37)	Mean of 19° MFCC
11	M(42)	Std of 21° MFCC
12	M(54)	Std of 27 ° MFCC
13	M(56)	Std of 28° MFCC
14	M(61)	Number of Zero Crossing

IV. Discussion

The proposed sound/silence detection algorithm mainly fails with low intensity snores, as such events have not enough energy to be classified as sound signals by the Otsu methodology. However, as post apnoeic snore events are more intense than non-post apnoeic ones, this limitation could be acceptable. Moreover, the Otsu thresholding fails if very few snore events are present in the recording. Specifically in 2 cases out of the 26 analyzed, manual thresholding was required, as the patient snored few times as compared to the length of the whole recording. In this case, thresholds were not coherent with the sound. This happened for one laboratory recording where some devices added a continuous noise during the night and for one home recording, where the patient snored few times over the whole recording (about 6 minutes out of 7 hours of recording). However, as Table 1 points out, the reduction in time could be relevant. Hence further analysis is required to overcome these limitations and possibly define a time threshold that points out if the recording is acceptable or not.

The sensitivity of the ORBN was really good, achieving in some case the 98% of recognition. The large variety of different kind of snores does not allow for a perfect recognition, but these first results seem quite good also as compared to existing literature.

At the end of the whole chain, the post apnoeic snore recognition varies from 69% to 84%, using the approach in [5]. Actually, sound detection and sound classification are hold on in automatic way, while the post apnoeic snore is analyzed offline, with a methodology not yet implemented in the algorithm.

V. Conclusion

We provide a full automatic highly sensitive system for snore identification during sleep that takes into account aspects of the problem not considered in other approaches. The search of the most meaningful features that identify the snore from other sounds could be further explored to provide a link between snoring arousal and other sound features.

A post apnoeic classification provides a first attempt to validate the system from data recordings for syndrome evaluation. However, we point out two weaknesses: first, the non automatic performance of the post apnoeic identification step and second, the used approach that does not perfectly fit our needs, but that was chosen for its easy applicability.

Finally, larger testing is needed to further validate our approach and compare its capability against the traditional home polysomnography approach.

References

1. A. S. Karunajeewa, U. R. Abeyratne, C. Hukins, “Silence-breathing-snore classification from snore-related sounds”, Physiol. Meas. 29, p.227–243, 2008.
2. W. D. Duckitt, S. K. Tuomi, T. R. Niesler, “Automatic detection, segmentation and assessment of snoring from ambient acoustic data”, Physiol. Meas. 27, p.1047–1056, 2006.
3. Y.-L. Hsu, M.-C. Chen, C.-M. Cheng, C.-H. Wu, “Development of a portable device for home monitoring of snoring”, IEEE Int. Conf. Systems, Man and Cybernetics, vol.3, p.2420-24, 2005.
4. M. Calisti, L. Bocchi, C. Manfredi, I. Romagnoli, F. Gigliotti, G. Donzelli, “Automatic detection of snore episodes from fill night sound recordings: home and clinical application”, AVFA09, Madrid, Spain, 2009.
5. A.K. Ng, T.S. Koh, E. Baey, T.H. Lee, U.R. Abeyratne, K. Puvanendran, “Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnoea?”, Sleep Medicine, 2007.
6. N. Otsu, “A threshold selection method from gray–level histograms”, IEEE Trans. Sys., Man., Cyber. Vol. 9(1), p.62–66, 1979.
7. M. Herzog, A. Schmidt, T. Bremert, B.Herzog, W. Hosemann, H.Kraftan, “Analysed snoring sounds correlate to obstructive sleep disorder breathing”, Eur. Arch. Otorhinolaryngol. Vol. 265, p.105–113, 2008.
8. J. Sola-Soler, R. Jaine, J. A. Fiz, J. Morera, “Automatic classification of subjects with and without Sleep Apnoea through snoring analysis”, Proc. 29th Int. Conf. IEEE EMBS, Lyon, France, 2007.