35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

HMM-based Snorer Group Recognition for Sleep Apnea Diagnosis*

Dulip L. Herath, Udantha R. Abeyratne, *Senior Member, IEEE,* and Craig Hukins

Abstract— **This paper presents an Hidden Markov Models (HMM)-based snorer group recognition approach for Obstructive Sleep Apenea diagnosis. It models the spatio-temporal characteristics of different snorer groups belonging to different genders and AHI severity levels. The current experiment includes selecting snore data from subjects, identifying snorer groups based on gender and AHI values (AHI < 15 and AHI > 15), detecting snore episodes, MFCC computation, training and testing HMMs. A set of multi-level classification rules is employed for incremental diagnosis of OSA. The proposed method, with a relatively small data set, produces results nearly comparable to any existing methods with single feature class. It classifies snore episodes with 62.0% (male), 67.0% (female) and recognizes snorer group with 78.5% accuracy. The approach makes its diagnosis decision at 85.7% (sensitivity), 71.4% (specificity) for males and 85.7% (sensitivity and specificity) for females.**

I. Introduction

OBSTRUCTIVE Sleep Apnea (OSA) is a breathing disorder that is caused by complete or partial obstruction of the upper airway (UA) while sleeping. Clinically, its severity level is described by Apena-Hypopnea Index (AHI) where apnea and hypopnea refer to complete and partial obstructions respectively. This index is defined as the average number of such UA obstructions that occur repetitively per hour.

If untreated, OSA patients are likely to be subjected to a number of medical conditions viz., cardiovascular morbidity, systemic and pulmonary hypertension, ischemic heart disease and stroke. OSA also increases the risk associated with these conditions. Further, OSA can lead to consequences such as daytime drowsiness, mood changes, deteriorated cognitive abilities, headaches [1]-[3]. Studies say that, 24% of middle-aged men and 9% of women suffer from OSA condition (with AHI>5) and symptomatic OSA is prevails at rates of 4% and 2% respectively [4].

Polysomnography (PSG) is the established diagnosis method for OSA [5]. However, the inherent limitations of PSG such as cost, labor and patient inconvenience make it impractical for large-scale population screening. Therefore, developing a simple, fast, non-invasive and portable solution is highly desirable [6], [7]. The acoustics of snoring is a powerful alternative information source as snoring is the most prevalent early OSA symptom [8],[9] and studies show that snore sounds (SS) contain vital information for OSA diagnosis [10]-[12].

The uniqueness of the articulatory system is determined by the anatomical structure and physiological function. This uniqueness is also manifested in the acoustic properties of the speech signals [13]-[15]. Moreover, speech and SSs share a number of common features with respect to the articulators, vocal tract dimensions, and air stream mechanisms [16], [17]. Therefore, logic dictates that the acoustic properties of speech and SSs are similarly affected by the anatomy and physiology of the articulatory system and the UA. Hence, it is plausible to assume that the acoustic variations in SS signals are correlated with anatomical and physiological variations of the UA.

In Automatic Speaker Recognition (ASR), speakers or speaker groups are discriminated by effectively exploiting the idea that every speaker has an anatomically and physiologically unique articulatory system which is manifested in the speech sounds of that speaker acoustically[18]-[20].

The severity of OSA and the acoustic properties of SS signals of a subject are both affected by the anatomical and physiological characteristics of his/her UA. Therefore, assuming the subjects with different OSA severity levels have unique acoustic characteristics in their SS signals, it should be possible to discriminate subjects into different OSA severity groups by modeling the acoustic behavior of their SS signals, by taking an approach similar to ASR.

Close examination of snore episodes, especially pre and post apnea event snore episodes, reveals that there are significant spatio-temporal patterns in SSs that are useful for OSA diagnosis. However, in the literature, it is hardly reported any SS-based OSA diagnosis approach that is capable of capturing both spatial and temporal dynamics of simultaneously. Therefore, developing a diagnostic method based on spatio-temporal dynamics of SS is highly beneficial. Hidden Markov Models (HMMs) [21] with Mel Frequency Cepstral Coefficients (MFCCs) and their differentials as acoustic features [22] are well-known for capturing spatio-temporal dynamics of speech signals.

This paper proposes a novel approach based on HMMs for snore episode classification and OSA diagnosis by modeling the spatio-temporal behavior of snorer groups. The current approach can be extended to a number of useful applications such as modeling intra-snorer snore evolution, apnea and hypopnea event prediction/detection, and bridging respiratory and neurology domains. A set of multi-level rules have been proposed to enable incremental diagnosis.

*This research is partly supported by the Australian Research Council under an ARC Discovery Grant (DP120100141) to Abeyratne. The work is also supported by The University of Queensland via a PhD Scholarship to Herath.

D. L. Herath and Udantha R. Abeyratne (e-mail: udantha@itee.uq.edu.au) are with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia

C. Hukins is with the Department of Respiratory and Sleep Medicine, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia.

II. METHOD

A. Data Acquisition Setup and Subject Data Set

Data of the current work (see Table I) are drawn from a repository of CD-quality high-fidelity SS recordings of the patients who underwent PSG diagnostic studies at the Respiratory and Sleep Disorders Unit of The Princess Alexandra Hospital, Brisbane, Australia. The recording environment has a non-contact recording setup that captures SS in parallel to the PSG studies.

There are two identical low-noise microphones (Model NT3, RODE®, Sydney, Australia) and a low-end professional quality preamplifier and A/D converter unit (Model Mobile-Pre USB, M-Audio®, CA, USA). The two microphones are kept at a nominal distance of 50cm. The sampling rate and size are 44.1 kHz and 16-bits respectively.

TABLE I
DEMOGRAPHIC DETAILS OF THE SUBJECTS

	Gender	No of Subjects	Age Mean±Std	BMI Mean±Std	AHI Mean±Std
0≤ AHI <15	M	7	51.4±15.7	33.2±6.9	6.3±3.2
	F	7	50.7±9.5	33.3±9.2	8.0±2.3
AHI > 15	M	7	51.7±15.7	34.7±6.3	29.3±8.3
	F	7	58.1±12.6	33.3±13.3	33.4±9.5

B. Front-end Processing and Feature Extraction

Snore Episode Segmentation

The snore episodes were extracted from the audio signals of overnight recordings by employing the automatic snore segmentation algorithm described in [16]. This algorithm returns the time-domain boundaries of the audio segments that satisfy the objective definition of snore given in [16] based on the existence of signal periodicity. These episodes are further scrutinized by a human listener to eliminate the episodes mistakenly recognized as snore episodes by the algorithm, e.g. speech, door sounds, and duvet noises.

Pre-emphasis, Framing and Windowing

Prior to extracting features, a first order pre-emphasis filter (with a standard constant 0.97) is applied on each snore episode to enhance the high frequency components. Then, each episode is converted into a sequence of frames of length N samples per frame with $r\%$ overlap with neighboring frames. Finally, spectral distortions due to short-time framing is minimized by applying a Hamming window (with standard values for $\alpha = 0.54$ and $\beta = 0.46$).

Mel Frequency Cepstral Coefficient Features

MFCCs are widely used in speech processing tasks including ASR. MFCCs are a high uncorrelated feature set that is capable of representing speech signals at a frequency resolution which is close to that of human ear in an intensity independent manner. In the current work, frames obtained from snore episodes are represented in terms of MFCCs and their 1st and 2nd differentials to ensure effective capturing of spatio-temporal signal dynamics. The log-normalized energy, which is given by the 0th coefficient of MFCCs, is highly correlated with UA anatomy and physiology. Therefore, log-normalized energy is also used to in the feature set to bring snorer specific information [22].

C. HMM-based Snorer Group Models

A set of snorer groups Λ is defined based on gender and OSA severity levels, where $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_P\}$, and $|\Lambda| = P$, consists of snorers belonging to different gender groups and AHI-value ranges.

Snore is a phenomenon that is progressing from left to right in the time domain while rapidly changing in the spatial domain. The probability distributions associated with each HMM; *initial state, state transition and emission probability distributions;* collectively describe the acoustic behavior of each snorer group that it represents. [21], [22].

Each snorer group is modeled as standard left-to-right HMM with 5 states. The states, 2, 3, and 4, are defined as emitting states whereas the states 1 and 5 are non-emitting initial and final states respectively, see Figure I.

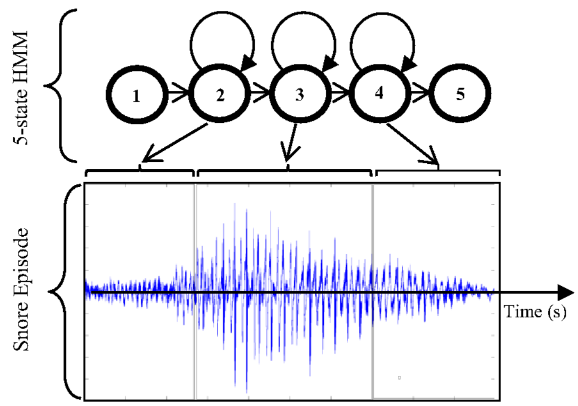

Figure 1: 5-state HMM is used to model onset, body and end of snore episode of a snorer

These emission states are correspondent with different UA configurations that are dynamically changing as the SS signal is progressing. Therefore, the observation feature vectors belonging to different stages of a snore episode belong to different emissions states of the respective HMM.

The probability of emitting the observation vector o_t from state i is governed by the output probability density function $b_i(o_t)$$\mathrm{b_i(o_t)}$given in (1) which is defined as a multivariate Gaussian Mixture Model (GMM).

$$b_i(o_t) = \sum_{j=1}^{M} m_{ij} b_{ij}(o_t)\ \mathrm{b_i(o_t)} = \sum_{\mathrm{j=1}}^{\mathrm{M}} \mathrm{m_{ij} b_{ij}(o_t)} \quad (1)$$

, where m_{ij} is the j^{th}mixture weight of the i^{th} state of HMM and $b_{ij}(\boldsymbol{o_t})$ is defined as in (2)

$$b_{ij}(\boldsymbol{o_t}) = \frac{1}{(2\pi)^{D/2} |\Sigma_{ij}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{o_t} - \boldsymbol{\mu_{ij}})^T (\Sigma_{ij})^{-1} (\boldsymbol{o_t} - \boldsymbol{\mu_{ij}})\right\} \quad (2)$$

, where μ_{ij} and $\sum_{ij}$ are the mean and the covariance matrix of the j^{th} Gaussian distribution of the i^{th} state, D is the dimension of the feature vector.

The HMM parameters are estimated by using Baum-Welsh re-estimation procedure on the Viterbi alignments

[23] of the snore episodes of subjects given in Table I.

The likelihood of a snore episode $O = o_1 o_2 \ldots\ldots o_T$, with T observation vectors from the model λ is expressed in terms of the log-likelihood (LL) computed according to (3).

$$LL = \log P(\mathrm{O}|\lambda) = \log P(o_1 o_2 \ldots\ldots o_T|\lambda) = \sum_{t=1}^{T} \log P(o_t|\lambda) \quad (3)$$

A Viterbi decoder returns $\hat{\lambda}$ the best snorer model that gives the highest log-likelihood score by computing $P(o_t|\lambda)$ in terms of HMM probability distributions according to (4).

$$\hat{\lambda} = \arg\max_{\lambda \in \Lambda} \log P(\mathrm{O}|\lambda) \quad (4)$$

D. Classification Rules

Rule 1: Snore Episode Classification

Let $\Omega_X = \{O_1, O_2, \ldots\ldots, O_R\}$ be a set of all the snore episodes that belong to a new subject X where O_i is the i^{th} snore episode of the subject and $|\Omega_X| = R$.

$$\forall\, O \in \Omega_X, \boldsymbol{If}\ \hat{\lambda}_j = \arg\max_{\lambda \in \Lambda} \log P(\mathrm{O}|\lambda)\ \boldsymbol{then}\ O \xrightarrow{is\ attributed\ to} j$$

The snore episodes belonging to one gender group are compared only with the models belong to the same gender group.

Rule 2: Snorer Group Recognition

Applying Rule 1 repeatedly on Ω_X leads to a mutually-exclusive and exhaustive partition of Ω_X with number of blocks $\leq P$ where block B_j is the subset of Ω_X that contains the snore episodes attributed to snorer j by Rule 1.

$$\boldsymbol{If}\ \hat{B}_k = \arg\max_{\forall_i B_i \in Partition\ of\ \Omega_X} size(B_i)\ \boldsymbol{then}\ X \xrightarrow{is\ classified\ as} k$$

Rule3: Snore Episode Probability Distribution

This rule describes how X's overnight snore episodes are distributed across the snorer groups by assigning a probability measure to each block in the partition of Ω_X .

$$Pr(X = k) = \begin{cases} \frac{size(B_k)}{|\Omega_X|}, & B_k \in Partition\ of\ \Omega_X \\ 0, & B_k \notin Partition\ of\ \Omega_X \end{cases}, \ k \in \{1,2,..,P\}$$

Rule 4: Diagnostic Rule

This rule is based on the probability distribution defined under Rule 3. It defines a decision rule to classify new subjects to clinically important groups: *OSA* or *non-OSA*. These groups are defined on the basis of OSA severity level described in terms of AHI values.

Let *threshold* be a value on the AHI scale, then the decision rule can be defines as follows:

$$\boldsymbol{If}\ \frac{\sum_{k=1}^{P} \boldsymbol{Pr}(X=k\ and\ k<threshold)}{\sum_{k=1}^{P} \boldsymbol{Pr}(X=k\ and\ k>threshold)} > 1$$
$$\boldsymbol{then}\quad X \in non-OSA\ \boldsymbol{else}\ X \in OSA$$

III. Results

A. Snorer Groups and Data Set

The four snorer groups defined for the current experiments by setting *threshold* =15, i.e. $\Lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$, are given in Table II. Each group contains equal number of snorers.

TABLE II
FOUR CATEGORIES OF SNORER GROUPS

Snorer Group	Gender	AHI Range	Number of Subjects	Number of Episodes
λ_1	M	0 ≤ AHI < 15	7	11148
λ_2	M	AHI > 15	7	16570
λ_3	F	0 ≤ AHI < 15	7	6570
λ_4	F	AHI > 15	7	12688

B. Framing and Feature Extraction

The frame size N was set to $N = 2646$ ($44.1\ kHz \times 60\ msec$) to ensure the periodicity of frames as [34] states that SS signals show a pseudo-periodicity around 60 *milliseconds*. The overlapping percentage r is set to 50% to guarantee effective capturing of spectral dynamics. Then, as described in Section II-B, each frame is represented as a 39-dimension feature vector by choosing first 12 MFCCs, log-energy and their differentials as features.

C. Snorer Group Model Training and Testing

For each snorer group defined, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, an HMM snorer model was created and trained as described in Section II-B. During the training, the number of Gaussian mixtures of the emission states of HMMs was incremented in the order of 1, 2, 4, 6, and 8. Male and Female models are separately trained and tested with 14-fold cross validation with leave-one-out policy.

The snore episode classification accuracy, as shown in Table III, is strictly increasing as the number of Gaussian mixtures increases in both gender groups up to 6. However, the snorer group recognition (Rule 2) and diagnosis results (Rule 3 and 4) given in Table IV and V respectively show that increasing the number of Gaussians further will lead to deterioration of performance in terms of sensitivity and specificity mainly due to overtraining.

TABLE III
SNORE EPISODE CLASSIFICATION ACCURACY (%) –RULE 1

Gender	Gaussian Mixture Components				
	1	**2**	**4**	**6**	**8**
Male	53.9	56.8	58.6	**61.3**	59.4
Female	55.6	58.2	63.8	**66.1**	61.2

TABLE IV
SNORER GROUP RECOGNITION ACCURACY (%) – RULE 2

Gender	Gaussian Mixture Components				
	1	**2**	**4**	**6**	**8**
Male	64.2	71.4	64.2	**78.5**	78.5
Female	50.0	50.0	64.2	**78.5**	78.5

In this experimental setup, the best performance is reported with 5-state HMMs with 6 Gaussian mixtures for each snorer group, snore episodes can be discriminated to two groups ($0 \leq$ AHI < 15 and AHI >15) with 61.3% (male) 66.1% (female) accuracy. Further, snorer group recognition accuracy is reported 78.5% for both genders.

Finally, the best (sen)sitivity and (spe)cificity are reported as 85.7% (sen), 71.4% (spe) and 85.7% (sen), 85.7% (spe) for male and female groups respectively, see Table V. Even though the subject data set used in this experiment is relatively small, HMM with a single feature class MFCCs has given promising results which are comparable to other single feature class SS-based OSA diagnosis approaches.

TABLE V
SENSITIVITY AND SPECIFICITY OF OSA DIAGNOSIS (%) - RULE 3 AND 4

	Snorer Models			
Mixture Components	Male		Female	
	Sensitivity	Specificity	Sensitivity	Specificity
1	42.8	85.7	42.8	57.1
2	57.1	85.7	42.8	57.1
4	57.1	71.4	57.1	71.4
6	**85.7**	**71.4**	**85.7**	**85.7**
8	85.7	71.4	85.7	71.4

IV. CONCLUSION

In this paper, an approach based on snorer group recognition was proposed for OSA diagnosis. The methodology is based on modeling intra-snore episode signal dynamics using HMMs with MFCC-based features. The acoustic models associated with HMMs are defined in terms of multivariate GMMs. These models were trained and tested on a data set of 28 subjects (14 male, 14 female).

From the experimental outcomes, it can be concluded that:

- HMMs are capable of classifying snore episodes effectively by taking the spatio-temporal behavior of snore sounds.
- The acoustic behavior of SSs of subjects with less severe OSA (normal or mild, i.e. AHI < 15) is significantly different from that of subjects with high severe OSA (moderate or high, i.e. AHI >15).
- Female snore episodes can be classified with relatively higher accuracy than male snore episodes.
- Snore episode (probability) distribution across different snorer groups of a subject can be used to diagnose his or her OSA severity.
- Number of Gaussian components in the acoustic model has a significant impact on the snore episode classification, snorer group recognition and diagnosis.

ACKNOWLEDGMENT

Authors would like to acknowledge Mr. Brett Duce, Laboratory Manger, Princess Alexandra Hospital, Brisbane, Australia for his help with clinical data acquisition.

REFERENCES

[1] K. M. Hla, T. B. Young, T. Bidwell, M. Palta, J. B. Skatrud, and J. Dempsey, "Sleep apnea and hypertension: A population-based study," *Ann. Intern, Med.*, vol. 120, no.5, pp. 382-388, 1994.

[2] M. H. Kryger, T. Roth, and W.C. Dement, *Principles and Practice of Sleep Medicine,* 3rd ed. Philadelphia, PA: W.B. Saunders, pp. 997-1040, 2000.

[3] L. J. Findley, J. T. Barth, D. C. Powers, S. C. Wilhoit, D. G. Boyd, and P. M. Suratt, "Cognitive impairment in patients with obstructive sleep apnea and associated hypoxemia," *Chest*, vol. 90, no. 5, pp. 686-690, 1986.

[4] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *N. Engl. J. Med.,* vol.328, no. 17, pp. 1230-1235, 1993.

[5] K. E. Bloch, "Polysomnography: A systematic review," *Technol. Health Care,* vol. 5, pp. 285-305, 1997.

[6] 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, no 4, pp. 1543-1573, 2003.

[7] A. S. Karunajeewa, U. R. Abeyratne, C Hukins, "Pitch/ airway-response and cepstral analysis of snore sounds for non-contact screening of sleep-apnea", *11th Int. Congress of IUPESM,* Munich, Germany, 7-12 September 2009.

[8] P. G. Norton and E. V. Dunn, "Snoring as a risk factor for disease – an epidemiological survey", *Br. Med. J.,* vol 291, pp. 630-632, 1985.

[9] A. G. Bassiri and C. Guilleminault, "Clinical features and evaluations of obstructive sleep apnoea", *The Principles and Practices of Sleep Medicine* ed T. Kryger and W. D. Roth (Philadelphia, PA: Saunders), 2000.

[10] J. A. Fiz, J. Abad, R. Jane, M. Riera, M. A. Mannanas, P. Caminal, D. Rodenstein, and J. Morera, "Acoustic analysis of snoring in patients with simple snoring and obstructive sleep apnea", *Eur. Respir. J.,* vol 9, pp. 2365-2370, 1996.

[11] J. S. Soler, R. Jane, J. A. Fiz and J. Morera, "Automatic classification of subjects with and without sleep apnea through snore analysis", *Proc. 29th IEEE Conf. on EMBS (Lyon, France 23-26 August),* pp. 6093-6096, 2007.

[12] M. Cavusoglu, T. Ciloglu, Y. Serinagaoglu, M. Kamasak, O. Erogul, and T Akcam, "Investigation of sequential properties of snoring episodes for obstructive sleep apnea identification", *Physiol. Meas.*, vol 29, pp. 879-898, 2008.

[13] W. H. Perkins, and R. D. Kent, "Textbook of Functional Anatomy of Speech, Language, and Hearing", Taylor and Francis, London, 1986.

[14] P. Lieberman, "Speech Physiology and Acoustic Phonetics", Collier Macmillan, London, 1977.

[15] W. J. Hardcastle, "Physiology of Speech Production". Academic Press, London, 1976.

[16] U. R. Abeyratne, S. A. Wakwella, and C. Hukins, "Pitch jump probability measures for the analysis of snoring sounds in apnea", *Physiol. Meas,* vol 26, pp. 779-98, 2005.

[17] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Multi-feature snore sound analysis in obstructive sleep apnea and hypopnea syndrome" *Physiol. Meas.,* vol. 32, pp. 83-97, 2011.

[18] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," *IEEE Trans. Acoust. Speech, Signal Processing,* vol. 29, no. 2, pp. 254-272, 1982.

[19] T. Matsui, S. Furui, "Text-independent speaker recognition using vocal tract and pitch information", *Proc. Int. Conf. Spoken Language Processing,* Kobe, vo. 5.3, pp. 137-140, 1990.

[20] R. Rose, R. A. Reynolds, "Text independent speaker identification using automatic acoustic segmentation", *Proc. ICASSP*, pp. 293-296, 1990.

[21] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE,* vol. 77, no. 2, 1989.

[22] X. Huang, A. Acero, H. Hon. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*, Prentice Hall, 2001.

[23] L. E. Baum, J. A. Eagon, "*An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology,*" Bulletin of the American Mathematical Society, vol.73, pp. 360-363 1967.