

Classification of Vibratory Patterns of the Upper Airway During Sleep*

Hisham Alshaer^{1,5}, Frank Rudzicz^{2,3}, Tiago H. Falk⁴, Wen-Hou Tseng¹, and T. Douglas Bradley¹

Abstract—Upper airway (UA) narrowing and collapse during sleep results in obstructive sleep apnea (OSA). We hypothesize that vibratory patterns of snoring can distinguish simple snorers from those with OSA. Samples of breath sounds were collected from 7 snorers without OSA and 5 with OSA. Snoring pitch (F_0) contours were found using the robust algorithm for pitch tracking (RAPT). The OSA snoring contours showed fluctuating patterns as compared to the smoother patterns of simple snorers. This suggests that snoring reveals the underlying instabilities of UA tissue in OSA. Conditional random fields, a statistical sequence classifier, gave 75% accuracy in distinguishing the 2 groups.

I. INTRODUCTION

Obstructive Sleep apnea (OSA) is a common breathing disorder affecting approximately 7% of the adult population [1]. It is characterized by repetitive partial (hypopnea) or complete (apnea) cessation of breathing during sleep for intervals of 10-90 seconds as a consequence of partial or complete collapse of the upper airway (UA), respectively. These events are accompanied by intermittent hypoxia and are terminated by sudden arousals that cause sleep disruption. They alternate with episodes of hyperventilation, during which loud snoring occurs. As a consequence of intermittent hypoxia and sleep disruption, patients suffer from daytime sleepiness and impaired cognitive performance. Because of these effects, patients with OSA are at higher risk for motor vehicle accidents than subjects without OSA; e.g., in the US, OSA-related motor-vehicle collisions are estimated to have caused 1,400 deaths and \$15.9 billion in associated costs in one year [2]. Over time, OSA also increases the risk of developing hypertension heart failure and stroke by 2 to 4 fold compared to subjects without OSA [3], [4]. OSA is therefore a major public health problem whose diagnosis and treatment could have a very substantial beneficial medical and public health impact [5].

A. Pathophysiology of OSA

In patients with OSA, upper airway collapse is caused by the normal attenuation of neural activation of UA dilator muscles at the transition for wakefulness to sleep superimposed upon an anatomically narrowed UA [6]. Conditions

that lead to upper airway narrowing and predisposition to OSA include obesity, tonsillar enlargement, and retrognathia. The predominant site of UA collapse during sleep in patients with OSA resides within the pharynx between the posterior aspect of the hard palate and the glottic inlet. This comprises the naso-, retropalatal-, oro-, and hypo-pharynx [7]. Despite UA obstruction and reduced or absent airflow, the respiratory centre continues to generate respiratory drive.

B. Snoring as an Indicator of Upper Airway Dynamics

When narrowing of the UA takes occurs during sleep, airflow induces vibration of the UA tissues resulting in the snoring sounds [8]. Therefore, vibratory patterns of snoring are expected to be influenced by the biomechanical properties of the UA, such as the tissues laxity and UA narrowing. Snoring that takes place in the absence of UA collapse might thus be distinguishable from snoring that takes place during more profound narrowing as in OSA. The goal of this study, therefore, is to study and classify the vibratory features of snoring from subjects with obstructive events and compare them to snoring from subjects without OSA.

II. METHODS

A. Data Acquisition

We recruited patients who were referred to the Toronto Rehabilitation Institute Sleep Research Laboratory with suspected OSA for overnight polysomnography (PSG). During PSG, thoracoabdominal movements were measured by respiratory inductance plethysmography, airflow by nasal pressure cannulae, and arterial oxyhemoglobin saturation (SaO_2) by oximetry. Apneas were scored as a drop in sum of thoracoabdominal movement from baseline by 90% lasting 10 seconds and hypopneas were defined as a 50% to 90% reduction in thoracoabdominal sum lasting 10 seconds. An OSA disorder was defined as an $AHI \geq 10$. Snorers without OSA were defined as subjects with an $AHI < 10$.

Breath sounds were recorded simultaneously with PSG using an electret microphone embedded in a face frame at a sampling rate of 16 kHz as described in [9]. A group of 5 subjects with OSA (G_{OSA}) and 7 simple snorers (G_{SN}) without OSA were included. Since breathing is absent during apneas, snoring is also absent. Therefore, we examined sound segments with hypopneas rather than apneas in G_{OSA} , where there is some ventilation and therefore where snoring may be present. A research technician used PSG clinical scoring results in order to locate portions of the night where hypopneas took place in the G_{OSA} . In the G_{SN} , the technician identified segments with snoring by listening. Subsequently, 3 to 5 segments from various part of the night were isolated

*This project has been supported by the Ministry of Research and Innovation of Ontario, MaRS Innovation, Ontario Centre of Excellence, and Johnson and Johnson Inc. Toronto Rehabilitation Institute receives funding from the Ontario Ministry of Health and Long-Term Care. Dr Alshaer received the NSERC scholarship.

¹ Sleep Research Laboratory, ² Toronto Rehabilitation Institute, University Health Network, ON, Canada; ³ Department of Computer Science, University of Toronto, ON, Canada; ⁴ University of Quebec, Montreal, QC, Canada; ⁵ Institute of Biomaterials and Biomedical Engineering, University of Toronto, ON, Canada. hisham.alshaer at uhn dot ca

from the breath sounds data of each subject. Each segment (S) was 5-7 minutes long containing either predominantly hypopneas or normal breathing with snoring.

B. Estimation of Snoring Pitch (F_0)

Previously, autocorrelation-based methods have been used to detect snoring based on its semi-periodic nature [10]. In this work we adopt an autocorrelation algorithm for identifying snores given their fundamental frequency (F_0), which is determined by the robust algorithm for pitch tracking (RAPT) [11]. RAPT employs the normalized cross-correlation function (NCCF), which is similar to the autocorrelation function, but has the advantage of eliminating the dependence of the summation boundary on the lag index and normalizes the sum by dividing it by the magnitudes of the multiplied signals [11]. Subsequently, RAPT selects a pitch candidate generated from the NCCF using dynamic programming over all frames within the analysis window. By choosing the minimum-cost path between the peaks of the NCCF, analysis criteria can be defined based on the typical properties of the signal, such as the suppression of large jumps to harmonics of the true F_0 [11].

The RAPT is applied to each segment (S) using the parameters. RAPT is set to detect F_0 between 20-300 Hz of successive 50 ms windows, where each analysis window is classified as being periodic, if contains a valid F_0 , or aperiodic otherwise. Non-snoring breath sounds such as normal inspirations and expiration are aperiodic sounds that don't have an F_0 and thus will not be included in the subsequent analysis. On the other hand, each snoring episode with sound periodicity will yield a pitch contour sequence (P). P sequences are subsequently isolated and fed into the sequence classification algorithms in section II-C.

C. Pattern Classification of Pitch Contour Sequences

In order to test the differentiability of pitch contours between snores associated with obstructive hypopneas and snores not associated with hypopneas, two types of pattern recognition techniques were tested: hidden Markov models (HMMs), conditional random fields (CRFs), and hidden conditional random field (HCRFs). HMMs and HCRFs are sequences classifiers that model long-term changes in an input signal by estimating the dynamics of hidden or unobserved variables that conditions that signal. The advantage of using them is that they model temporal variability without explicitly quantifying the nature of that variability.

1) *Hidden Markov models (HMMs)*: HMM's categorize observable temporal data sequences according to unobserved discrete variables that have an underlying connected-state structure. In this work we use continuous HMMs that observe a univariate continuous space O where the likelihood of a sequence of observations, \mathbf{o} , given a current state $q_i \in Q$ is

$$B_i(o) = \sum_{m=1}^M \omega_{i,m} \frac{1}{(2\pi)^{d/2} |\Sigma_{i,m}|^{1/2}} \exp \left[-\frac{1}{2} (o - \mu_{i,m})^\top \Sigma_{i,m}^{-1} (o - \mu_{i,m}) \right] \quad (1)$$

where $d = 1$ is the number of dimensions in the data, M is the number of component Gaussians in each state, $\omega_{i,m}$ is the weight of the m^{th} Gaussian in state q_i (subject to $\sum_{m=1}^M \omega_{jm} = 1$), $\mu_{i,m}$ is its mean, $\Sigma_{i,m}$ is its covariance, and $|\Sigma|$ is the determinant of Σ .

HMM parameters, Φ , consists of $B_i(o)$, the state transition matrix $A(q_i, q_j)$ describing the *a priori* probability of transitioning from state q_i to q_j and an initial state distribution π_i . Since states are unobserved, their values are inferred during the automatic adjustment of these parameters using the Baum-Welch algorithm. During testing, we do not use the traditional Viterbi algorithm since we are not interested in the state sequence; rather, we train one HMM for each of our two classes (i.e., snoring without OSA and snoring with OSA), and compare the likelihoods of test sequences \mathbf{o} in each of these models separately. These likelihoods are performed with the Forward algorithm given the respective parameterization Φ ,

$$P(\mathbf{o}; \Phi) = \sum_{\forall \mathbf{q}} P(\mathbf{q}; \Phi) P(\mathbf{o} | \mathbf{q}; \Phi), \quad (2)$$

which sums over all possible sequences of hidden states \mathbf{q} .

2) *Conditional random fields*: The conditional random field (CRF) and the hidden conditional random field (HCRF) are also sequence classifiers but differ from the HMM in that its estimation of the distribution of a class label \mathbf{l} (i.e., normal or apneic) given the observation data does not model the observation prior $P(\mathbf{o})$, as shown in Eq. 3. The CRF does not model intrinsic sequential substructures using hidden states as the HCRF does, but both assign labels to the entire sequence [12].

In CRFs, the parameter set θ defines the weights ($\theta_k \in \theta$) applied to *feature functions* f_k of the graphical model, which are analogous logarithms of the conditional state and observation probabilities in HMMs and are initialized randomly [13]. In the HCRF, the likelihood of a particular label \mathbf{l} of an observation sequence \mathbf{o} given some parameterization θ must be computed over all possible sequences of hidden states (where \mathbf{q} is a particular state sequence), where each state q_i comes from the set \mathcal{Q} of states. In other words,

$$P(\mathbf{l} | \mathbf{o}, \theta) = \sum_{\mathbf{q}: q_i \in \mathcal{Q}_i} P(\mathbf{l} | \mathbf{q}, \mathbf{o}, \theta) P(\mathbf{q} | \mathbf{o}, \theta), \quad (3)$$

where $P(\mathbf{q} | \mathbf{o}, \theta)$ is the standard conditional random field formulation that defines state and transition functions [12], [13], namely

$$P(\mathbf{q} | \mathbf{o}, \theta) = \frac{\exp(\sum_k \theta_k F_k(\mathbf{q}, \mathbf{o}))}{\sum_{\mathbf{r}} \exp(\sum_k \theta_k F_k(\mathbf{r}, \mathbf{o}))}, \quad (4)$$

where $F_k(\mathbf{q}, \mathbf{o})$ is the sum over all state transition feature functions applicable to \mathbf{q} and observation feature functions applicable to \mathbf{o} .

Given a training set of labeled sequences (\mathbf{o}_i, \mathbf{l}) where $i = 1..N$, we apply conjugate gradient ascent to find the optimal

parameter values $\theta^* = \arg \max_{\theta} L(\theta)$ given the following objective function:

$$L(\theta) = \sum_{i=1}^N \log P(\mathbf{I} | \mathbf{o}_i; \theta) - \frac{1}{2\sigma^2} \|\theta\|^2, \quad (5)$$

which is the log-likelihood of the parametrization given by the conditional log-likelihood of each training sequence $\log P(\mathbf{I} | \mathbf{o}_i; \theta)$ and the Gaussian prior likelihood of θ with variance σ^2 . If the parameter space θ is uniformly distributed, as we assume here, σ^2 approaches infinity and we discount the second term.

D. Experiments

We perform a 5-fold cross validation of the given data. In each fold, 80% of the P sequences are randomized for training and 20% randomized for validation.

III. RESULTS

We isolated 18 segments with hypopneas and 21 segments with simple snoring.

A. Qualitative Description of Snoring Pitch Contours

Pitch contours originating from snore episodes in the G_{SN} were found to be relatively flat over time with a minor drop towards the end. In other words, each snoring episode had relatively flat F_0 course throughout the episode (bottom panel of Fig. 1). On the other hand, most snoring pitch contours originating from G_{OSA} showed a fluctuating (i.e. ‘wavy’) course over time. That is, F_0 values oscillated over the course of individual snores as illustrated in the examples in Fig. 1.

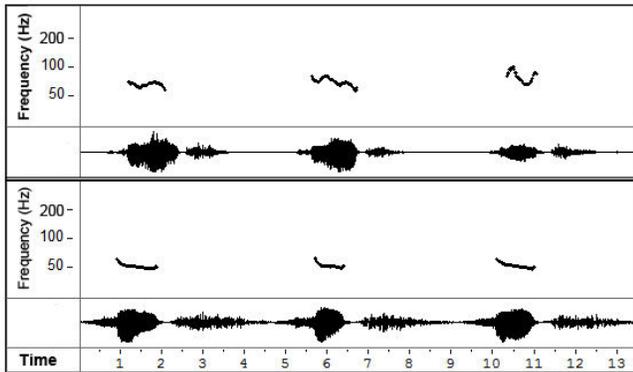


Fig. 1. Exemplary pitch (F_0) contours originating from 3 breathing cycles concatenated from a subject with OSA (above) and from a subject without OSA (below). Individual segments of pitch contour are plotted against frequency (Hz). Each of which originates from a single snore corresponding to the sound waveform underneath. Obstructive snores show a wavy pattern compared to the flatter pattern of simple snores.

B. Performance of Sequence Classifiers

The proportion (between 0 and 1) of successful classifications of previously unseen snores for each of the 5 folds across each of the parameterizations of the HMM, CRF, and HCRF models are shown in Tables I and II. The highest mean accuracy achieved by HMM was 55%. The CRF model, however, had a mean accuracy of 75%. Clearly,

the conditional random field classifier outperforms the hidden Markov model in this scenario, across all parameterizations, with left-tailed heteroscedastic $t(16.5776) = -23.15, p < 0.00001, CI = [-\infty, -0.2247]$. A two-way analysis of variance in the hidden Markov model results shows no significant effect of the number of states ($F(1,3) = 0.89, p = 0.4819$) or of the number of mixtures ($F(1,3) = 1.04, p = 0.4215$). Similarly, a two-way analysis of variance in the hidden CRF results shows no significant effect of the window size ($F(1,1) = 0.08, p = 0.7937$) or of the number of states ($F(1,3) = 0.19, p = 0.898$). That these models do not offer significant variation based on their parameters indicates that the superiority of the (H)CRF-type relative to the HMM is based on the underlying modeling structure. Furthermore, it suggests that more parsimonious parameters might be more useful in practice, as their resulting computations are quicker.

IV. DISCUSSION

In this study, we have shown that intra-snore F_0 values, or pitch contours, follow different time courses in simple snorers and in snorers with OSA and that their patterns could be reliably classified using sequence classification algorithms, with up to 75% accuracy using CRF. F_0 represents the speed at which UA tissue flaps collide resulting in the semi-periodic snoring sounds. In this semi-periodic signal, each cycle is a result of a single tissue collision i.e. 1 closure and opening. Therefore, it stands to reason that the speed of tissue collision will be influenced by the biomechanical properties of the colliding tissue flaps of the UA. These properties include flexibility of the UA tissues, the distance between the vibrating flaps, and the amplitude of the driving force. In this context, the driving force is the inspiratory airflow created by the negative airway pressure during inspiration. During inspiration, in which snoring takes place, the pressure inside the lung drops and UA below atmospheric reaching the minimum in the middle of inspiration (-1 mm Hg) and back to atmospheric pressure at the end of inspiration.

In order to understand UA biomechanics, researchers suggested that the human UA can be modelled as a Starling resistor, which consists of a collapsible segment in the pharynx surrounded by 2 solid segments; the nose from above and glottis from below [14]. The collapsible segment is the part that narrows and gets occluded in patients with OSA due to attenuated neuromuscular tone and the negative inspiratory pressure. In patients with OSA, it has been shown that the UA requires less negative pressure to collapse than in subjects who snore but don’t have OSA [14]. This means that the UA of simple snorers, although partially narrowed, is still more stable during sleep in the face of the inspiratory negative pressure than in OSA. On the other hand, the UA of an OSA patient is less stable and is prone to narrowing and collapse under the same conditions. This could explain the noticeably wavy course of snoring pitch sequences in the G_{OSA} as compared with the more constant pitch time sequence in the G_{SN} . It can be postulated that with the ensuing inspiration, the progressively negative airway

TABLE I

RESULTS OF THE 5-FOLD CROSS VALIDATION OF HMM WITH MEAN AND VARIANCE ACROSS THE 5 FOLDS

	Q=1				Q=2				Q=3				Q=4			
	M=1	M=2	M=3	M=4												
Fold 1	0.567	0.453	0.579	0.476	0.499	0.312	0.375	0.499	0.481	0.642	0.484	0.539	0.444	0.527	0.559	0.650
Fold 2	0.352	0.387	0.424	0.433	0.553	0.375	0.542	0.693	0.413	0.450	0.372	0.496	0.636	0.470	0.453	0.441
Fold 3	0.453	0.579	0.648	0.582	0.527	0.479	0.536	0.550	0.410	0.605	0.487	0.395	0.539	0.527	0.467	0.610
Fold 4	0.321	0.352	0.453	0.513	0.456	0.544	0.639	0.593	0.662	0.519	0.570	0.567	0.556	0.539	0.507	0.447
Fold 5	0.355	0.513	0.453	0.427	0.490	0.458	0.470	0.441	0.427	0.473	0.424	0.499	0.513	0.424	0.395	0.610
mean	0.410	0.457	0.511	0.486	0.505	0.434	0.512	0.555	0.479	0.538	0.468	0.499	0.538	0.497	0.476	0.552
var	0.010	0.008	0.009	0.004	0.001	0.008	0.010	0.009	0.011	0.007	0.006	0.004	0.005	0.002	0.004	0.010

Q: the number of hidden states used; M: the number of Gaussians that are used to model the observation probabilities at each state

TABLE II

RESULTS OF THE 5-FOLD CROSS VALIDATION OF CRF AND HCRF WITH MEAN AND VARIANCE OF THE 5 FOLDS

	CRF		HCRF							
	W=3	W=5	W=3				W=5			
			Q=2	Q=3	Q=4	Q=5	Q=2	Q=3	Q=4	Q=5
Fold 1	0.75072	0.74785	0.74212	0.73639	0.73926	0.75072	0.74499	0.73926	0.73639	0.72493
Fold 2	0.72493	0.74785	0.75358	0.73926	0.73639	0.72206	0.73352	0.70774	0.7192	0.75645
Fold 3	0.75358	0.75645	0.77937	0.75931	0.75645	0.76791	0.75931	0.75358	0.76504	0.73066
Fold 4	0.80229	0.73066	0.73066	0.74499	0.70487	0.76218	0.73926	0.72493	0.73926	0.75645
Fold 5	0.72493	0.72779	0.70487	0.69628	0.67908	0.70487	0.71633	0.73066	0.71347	0.7192
mean	0.75129	0.74212	0.74212	0.735246	0.72321	0.741548	0.738682	0.731234	0.734672	0.737538
var	0.000999	0.000152	0.000759	0.000552	0.000955	0.000732	0.000248	0.000289	0.000409	0.000314

W = the window size, which consists of the number of adjacent samples used as an input vector to models e.g. W=5 means a subsequence of 5 measurements are concatenated together as an input.

pressure acts on the pliable UA tissues of the OSA patients causing narrowing of the UA and bringing the vibrating flaps closer together. In turn, this modulates the speed of tissue collision and eventually give rise to the wavy pitch contours observed in OSA.

To our knowledge, this is the first study to implement RAPT for snoring analysis. RAPT offers the ability to track snoring signals in short windows, which provided excellent time resolution of the pitch contours. Snoring pitch has been examined earlier by Abeyratne et al [10] who found that snores in OSA patients are characterized by intra-snore pitch discontinuities and thus have higher pitch jump probability as compared to simple snores [10]. Our results are in line with that work in term of the instability of snoring pitch in patients with OSA. However, our work extends that of Abeyratne et al. [10] by deploying the CRF and HCRF sequence classification algorithms that characterize the relationship among all pitch values in the entirety of each sequence, without the need to explicitly identify the discriminatory feature(s). In this study, the direction and depth of F_0 fluctuations were not determined, yet CRFs yielded 75% accuracy.

Future work will incorporate larger sample size, independent measures for UA biomechanics, and alternative classification schemes.

REFERENCES

- [1] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of obstructive sleep apnea: a population health perspective," *Am. J. Respir. Crit. Care Med.*, vol. 165, pp. 1217–1239, May 2002.
- [2] A. Sassani, L. J. Findley, M. Kryger, E. Goldlust, C. George, and T. M. Davidson, "Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome," *Sleep*, vol. 27, no. 3, pp. 453–8, 2004.
- [3] T. D. Bradley and J. S. Floras, "Sleep apnea and heart failure: Part I: obstructive sleep apnea," *Circulation*, vol. 107, no. 12, pp. 1671–8, 2003.
- [4] M. Arzt, T. Young, L. Finn, J. B. Skatrud, and T. D. Bradley, "Association of sleep-disordered breathing and the occurrence of stroke," *Am J Respir Crit Care Med*, vol. 172, no. 11, pp. 1447–51, 2005.
- [5] R. S. Leung and T. D. Bradley, "Sleep apnea and cardiovascular disease," *Am J Respir Crit Care Med*, vol. 164, no. 12, pp. 2147–65, 2001.
- [6] F. Lopez-Jimenez, F. H. Sert Kuniyoshi, A. Gami, and V. K. Somers, "Obstructive sleep apnea: implications for cardiac and vascular disease," *Chest*, vol. 133, pp. 793–804, Mar 2008.
- [7] J. E. Remmers, W. J. deGroot, E. K. Sauerland, and A. M. Anch, "Pathogenesis of upper airway occlusion during sleep," *J Appl Physiol*, vol. 44, no. 6, pp. 931–938, Jun 1978.
- [8] A. K. Ng and T. S. Koh, "Analysis and modeling of snore source flow with its preliminary application to synthetic snore generation," *IEEE Trans Biomed Eng*, vol. 57, no. 3, pp. 552–560, Mar 2010.
- [9] H. Alshaer, A. Levchenko, T. D. Bradley, S. Pong, W. Tseng, and G. F. Fernie, "A System for Portable Sleep Apnea Diagnosis Using an Embedded Data Capturing Module," *J Clin Monitor Comp*, in press.
- [10] U. R. Abeyratne, A. S. Wakwella, and C. Hukins, "Pitch jump probability measures for the analysis of snoring sounds in apnea," *Physiological Measurement*, vol. 26, no. 5, pp. 779–98, 2005.
- [11] D. Talkin, *A Robust Algorithm for Pitch Tracking, a chapter in Speech Coding and Synthesis*. Elsevier Sciences, 1995, ch. 14, pp. 495–518.
- [12] L.-P. Morency, A. Quattoni, and T. Darrell, "Latent-dynamic discriminative models for continuous gesture recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.
- [13] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML-2001)*. Morgan Kaufmann, San Francisco, CA, 2001, pp. 282–289.
- [14] I. C. Gleadhill, A. R. Schwartz, N. Schubert, R. A. Wise, S. Permutt, and P. L. Smith, "Upper airway collapsibility in snorers and in patients with obstructive hypopnea and apnea," *Am. Rev. Respir. Dis.*, vol. 143, no. 6, pp. 1300–1303, Jun 1991.