SUBJECT INDEPENDENT IDENTIFICATION OF BREATH SOUNDS COMPONENTS USING MULTIPLE CLASSIFIERS*

Hisham Alshaer[1], *Aditya Pandya*[2], *T. Douglas Bradley*[1,3], *and Frank Rudzicz*[1,4]

[1] University Health Network, Toronto Rehabilitation Institute, ON, Canada;
[2] Department of Physics, Ryerson University, Toronto, ON, Canada;
[3] Centre for Sleep Medicine and Circadian Biology,
[4] Department of Computer Science, University of Toronto, ON, Canada

ABSTRACT

Breath sounds have been shown very valuable for diagnosis of obstructive sleep apnea. In this study, we present a subject independent method for automatic classification of breath and related sounds during sleep. An experienced operator manually labelled segments of breath sounds from 11 sleeping subjects as: inspiration, expiration, inspiratory snoring, expiratory snoring, wheezing, other noise, and non-audible. Ten features were extracted and fed into 3 different classifiers: naïve Bayes, Support Vector Machine, and Random Forest. Leave-one-out method was used in which data from each subject, in turn, is evaluated using models trained with all other subject. Mean accuracy for concurrent classification of all 7 classes reached 85.4%. Mean accuracy for separating data into 2 classes, snoring and non-snoring, reached 97.8%. To our knowledge, these are the highest accuracies achieved in automatic classification of all breath sounds components concurrently and for snoring, in a subject independent model.

***Index Terms*—** Breath Sounds, Inspiration, Expiration, Snoring, Obstructive Sleep Apnea, Pattern Classification

1. INTRODUCTION

Breath sound analysis, especially of snoring, has been used to diagnose sleep-related respiratory disorders such as obstructive sleep apnea (OSA) [1, 2, 3]. OSA is a common respiratory condition, affecting approximately 7% of the adult population [4]. OSA is characterized by repetitive interruptions of breathing during sleep, each lasting for 10–90 seconds. Obstructive hypopneas and apneas results from partial or complete collapse of the upper airway, respectively. OSA results in intermittent drops in blood-oxygen saturation and sudden arousals from sleep that cause sleep fragmentation and poor sleep quality. Patient with OSA, therefore, suffer from excessive daytime sleepiness and impaired cognitive performance and thus are at higher risk for motor vehicle accidents which results in thousands of fatalities every year [5]. OSA increases the risk of developing hypertension, heart failure, and stroke by 2 to 4 fold compared to individuals without OSA [6, 7]. OSA is therefore a major public health problem whose diagnosis and treatment could have a substantial impact on public health and health care cost [8].

OSA influences breath sounds due to its impact on the upper airway.The upper airway narrowing and collapse in OSA causes the airflow to induce vibration of the upper airway tissues whose common audible manifestation is known as snoring [9]. It has been found that snoring that takes place in simple snorers has different harmonic and frequency distribution [10] and different temporal regularity [1] in OSA than non-apneic snorers. Even in the absence of snoring, the acoustic signatures of inspiratory and expiratory sounds are also influenced by the degree of upper airway narrowing [11, 12].

1.1. Relation to Prior Work

Different techniques for identifying respiratory conditions have used different components of breath sounds depending on the specific goal of the analysis. Many studies focused on snoring alone to detect OSA [10, 1]. We have previously examined non-snoring inspiratory sounds for the effects of upper airway narrowing similar to that in OSA [11]. The isolation of inspiratory sounds in that study was performed manually. Manual annotation, however, is a labor-intensive task and non-practical in real life applications, especially when data is recorded overnight for 6 to 8 hours long per patient. Therefore, an automatic and accurate system is required to classify reliably breathing sounds in new patients. There has been several attempts towards achieving this goal. In our previous work, we have accurately identified breath-

*This project has been supported by the Ministry of Research and Innovation of Ontario, MaRS Innovation, Ontario Centre of Excellence, FedDev via the Ontario Brain Institute, 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 and the Ontario Brain Institute Entrepreneurial Fellowship. The authors acknowledge the help of Wen-Hou Tseng and Richard Hummel in data processing and analysis.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

ing phases (inspiration and expiration) [13, 14], but did not take into account the abnormal sounds such as snoring and other incidental sounds from the surrounding environment. Studies by other groups have shown accuracies in identification of snoring and related sounds, such as breathing and silence, ranging between 82% to 93.2% from ambient microphones [15, 16, 17, 18]. In those studies, all non-snoring breath sounds were treated as as one class, without separating inspiration from expiration.

In this study, we extend the results achieved by us and others to encompass all possible breathing sounds including individual breathing phases. We hypothesize that discriminant classifiers can accurately identify relevant breath sounds in the data of new subjects given appropriate temporal and frequency acoustic features. The goal of this work was to develop a subject-independent system for identifying all breath sounds that could be deployed for practical usage.

2. METHODS

2.1. Data Acquisition and Labelling

Breath sounds were recorded from 11 subjects during overnight sleep using an electret microphone in front of the face embedded in a small open mask at a sampling rate of 16 kHz as described in [19]. Five-minute segments (L) were extracted from the first, middle, and last third of the overnight recording of each subject. A total of 33 segments, yielding 165 minutes of data, were extracted. An experienced annotator listened to each segment and manually identified each sound unit as one of: inspiration, expiration, snoring, wheezing, not-audible, and other-noise. Wheezing is a high pitch musical sound. Although it can rarely be detected at the upper airway level, it was given a separate class for completeness. Snoring is typically an inspiratory phenomenon, yet expiratory snoring can also takes place in rare cases, in which case it should have distinct acoustic characteristics. Therefore, a separate class for expiratory snoring was created to yield a total of 7 classes. For simplicity, the mere term 'snoring' herein will refer to inspiratory snoring alone since it represents the majority of snoring episodes.

2.2. Feature Extraction

Each L was segmented using a moving window of 64 ms with 50% overlap, herein referred to as (W). From each W, the following 10 features were extracted:

Periodicity: Autocorrelation-based methods have previously been used to detect snoring based on its semi-periodic nature [3]. In this work, we adopt an autocorrelation-based algorithm known as the *robust algorithm for pitch tracking* (RAPT) [20]. RAPT calculates the periodicity of W as a value between 0 and 1, denoting total randomness and complete periodicity respectively.

Frequency bands ratio: We have shown that expiration has most of its energy concentrated in the frequency band below 400Hz and vice versa for inspiration [14]. In this work, this feature is calculated as the ratio of frequency bin magnitudes below 400 Hz to those above 400 Hz.

Spectral centroid: This indicates the 'center of spectral mass', which is perceptually related to the 'brightness' of a sound and is given by:

$$\frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{1}$$

where $x(n)$ represents the magnitude of bin number n, and $f(n)$ represents the center frequency of that bin.

Flatness: This indicates whether the spectral distribution is smooth or spiky, and results from the ratio between its geometric and arithmetic means, given by:

$$Flatness = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} \tag{2}$$

where $x(n)$ represents the magnitude of bin number n.

Shannon entropy is a measure of uncertainty of a random variable X given by:

$$H(x) := -\frac{\sum_{i=1}^{n} p(x_i) log_b p(x_i)}{log(length(p))} \tag{3}$$

where, $p(x_i)$ the probability mass function of X.

Zero crossing rate is the number of zero crossings (both positive and negative) present in the segment normalized by the length of the segment.

Uniformity: This measures the uniformity of the negative peak amplitudes of a windowed segment. The peaks are obtained by estimating the local maximas and minimas. The uniformity value is then defined by:

$$U = \frac{std(peak_{neg})}{mean(peak_{neg})} \tag{4}$$

where $peak_{neg}$ are peaks following negative zero crossings.

Shimmer: Shimmer is standard time domain feature used in speech processing that is quantified as:

$$Shimmer = \frac{\sum_{i=1}^{N-1} |X(n) - X(n+1)|}{N-1} \tag{5}$$

where X is obtained by a 3^{rd}-order median filtering of W.

Click factor: Clicks are defined as the sharp loud peaks resulting from tissue collision as happens with snoring, wether periodic or non-periodic. Clicks manifest as transient wide frequency bands in the spectrogram, analogous to a step function. To identify this pattern, a pre-emphasis filter is applied to spectra of W and short-time spectrograms is obtained (window size=256 points [16ms] with 75% overlap). The frequency bins are then summed, which converts the spectrogram from a 2 dimensional to a 1 dimensional waveform. The latter is then de-trended to remove local offsets and the resulting waveform is herein defined as K. The roughness of K reflects the occurrence of sharp transients in the time domain (clicks). The click factor is quantified as: $C = mean((10 \times K)^2)$.

Relative energy: The ratio of the root mean square of W to the positive maximum signed 16 bit integer level was defined as the 'relative energy' level for a given segment. This is an expression of a signal's energy as a proportion of its allowed maximum amplitude.

2.3. Pattern Classification

Three classification methods are compared in this work: naïve Bayes (NB), support vector machine with sequential minimal optimization (SVM), and random forest (RF). These three methods differ greatly in the optimization of the parameters, with both the SVM and RF models optimizing (separate) discriminative criteria. The naïve Bayes classifier assumes conditional independence between its features, the SVM is a parametric classifier that provides highly non-linear decision boundaries given particular kernels. A polynomial kernel of degree 2 is used herein. RF is an ensemble classifier that returns the mode of the class predictions across several decision trees. In this work, the RF uses the standard Breiman algorithm with 5 trees. Parameters for the SVM kernel and RF are set empirically.

2.4. Experiments

In order to ensure generality of the trained models, we implemented the leave-one-subject-out cross validation method (LOSOCV). Here all subjects data sets except one are used for training, which is used for validation to obtain the accuracy of identifying sound classes in that 1 subject. The process is then repeated in such a way that each subject is used for validation exactly once. The total accuracy is averaged over individual scores. This approach tends to preclude effects of over-fitting to the training data. This was done twice – once for the 7-way classification of the classes described in section 2.1 and once for the binary classification of snoring versus all other sounds. Snoring was selected from among the other classes as an exemplary class, since it has been shown as clinically relevant signal of interest in many previous studies. Yet, any one of the other classes, such as inspiration or expiration etc., can be equally chosen for this problem.

3. RESULTS

The mean values for the 10 acoustic features and their distributions for 3 of the 7 classes are displayed in Figure 1. Many of these features were in the expected ranges for the relevant classes, as discussed latter. Accuracies across classifiers and participants are shown in Table 1. The mean accuracy for distinguishing the 7 classes concurrently ranged between 79.7% with NB and 85.40% with SVM. On the other hand, when the problem was reduced to a 2-class problem (snoring vs not-snoring), performance improved remarkably to between 94.9% with NB and 97.8% with RF. Expectedly, the discriminative classifiers outperform NB, but the unprecedented high accuracy of the latter may indicate that the selected features are indeed highly informative to the identification of snores.

4. DISCUSSION

In this work, we have shown that individual components of breath sounds can be identified with a high degree of accuracy using a totally subject-independent scheme. Accuracy for identifying all breath and related sounds reached 85.4% in the 7-class problem and 97.8% in the binary identification of snores. Validation was performed using a LOSOCV scheme, which shows that these results can be duplicated in practical situations in which a trained system can be used to classify breath sounds in new subjects. To our knowledge, this is the first work that reports classification of all breath sounds including snoring (inspiratory and expiratory), inspiration, expiration, wheezing, in addition to non-respiratory noises. As for snoring alone, this study achieved the highest classification accuracy to date, according to our knowledge.

The current work expands the approaches of previous studies that deployed an ambient microphone by identifying all possible breath sounds components rather than snoring alone. Although snoring has been the focus of most studies that aimed at detecting OSA, other non-snoring breath sounds still carry important information about the upper airway dynamics. For example, the upper airway is prone to narrowing and collapse during inspiration more than expiration, due to the negative lung pressure [11]. Both expiratory and inspiratory phases can be used for accurate tracking of breathing rate and activities [13]. For all these reason, detection of breathing components is needed for research and clinical purposes.

We have selected a group of features that characterize the physical nature of breath sounds. Figure 1 displays the distribution of the 10 features across the most common breath sounds. Periodicity for example was highest in snoring. Periodicity arises with snoring due to collision of tissue pliable flaps of the upper airway, in contrast to the other sounds that

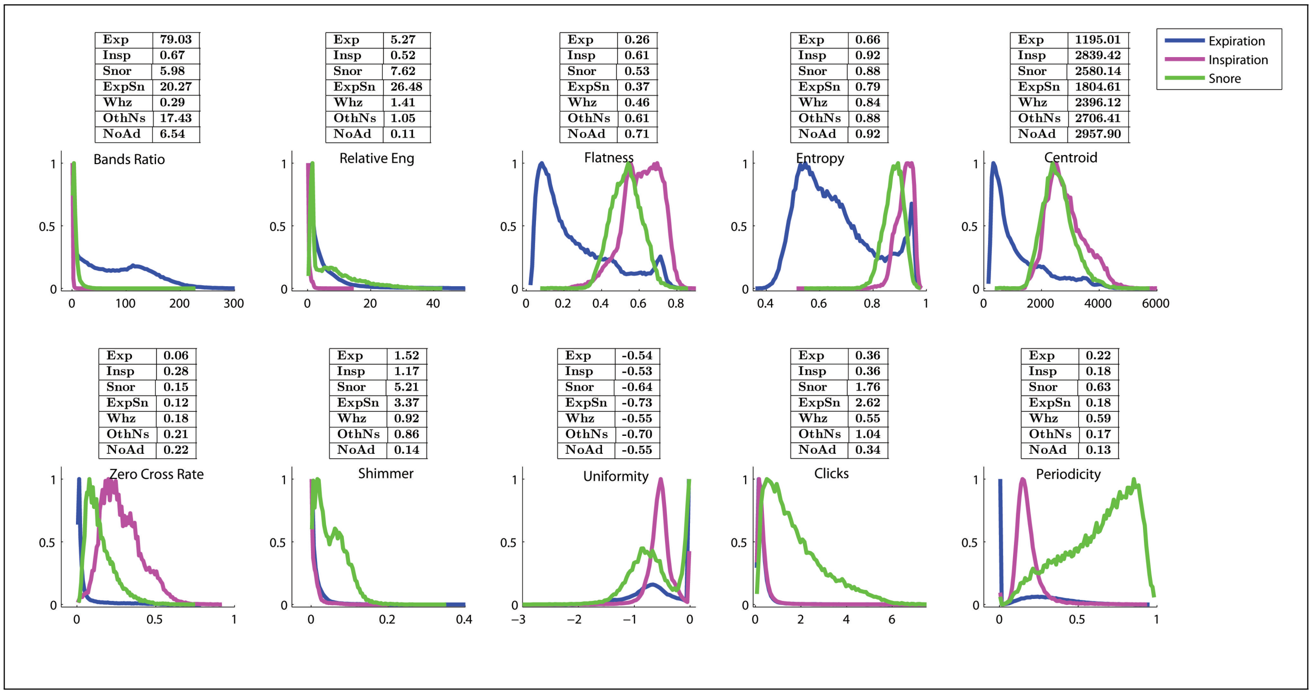

Fig. 1. Display of the 10 features showing graphical distributions (histograms) of the 10 features across the 3 main sound classes: expiration, inspiration, and snoring. Each figure is accompanied by its feature mean values in the 7 sound classes. Each distribution curve was normalized to unity to facilitate visualization.
Exp: Expiration; Insp: Inspiration; Snor: Snoring; ExpSn: Expiratory Snoring; Whz: wheezing; OthNs: Other Noise; NoAd: Not Audible

Table 1. Average and Median Accuracy achieved by LOSOCV for the 11 subjects

7-Class Classification Problem, Snoring vs Non-Snoring														
Classifier	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Avg	Med	Std
NB	52.2	76.2	90.3	65.2	88.9	80.3	84.1	87.9	89.3	75.6	86.6	**79.7**	**84.1**	11.9
SVM	67.7	86.5	91.1	70.7	93.1	86.5	89.5	91.3	86.1	84.8	91.5	**85.4**	**86.5**	8.44
RF	56.4	87.1	84.3	74.6	91.4	87.8	88.7	92.1	86.6	85.6	89.6	**84.0**	**87.1**	10.3
2-Class Classification Problem														
Classifier	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Avg	Med	Std
NB	88.0	94.7	98.0	94.0	99.1	97.5	92.4	99.5	95.5	88.2	97.0	**94.9**	**95.5**	4.0
SVM	92.1	97.2	97.7	93.8	99.8	97.9	97.9	99.8	91.4	94.5	99.6	**96.5**	**97.7**	3.0
RF	85.0	98.3	94.2	97.3	99.9	98.5	97.8	100.0	94.7	97.1	99.8	**97.8**	**98.0**	2.0

P1...11: Patients 1 to 11; **Avg**: Average; **Med**: Median; **Std**: Standard Deviation; **NB**: Naive Bayes; **SVM**: Support Vector Machines; **RF**: Random Forests

are turbulent in nature. The click factor was also highest in snoring since it captures sharp tissue collisions regardless of periodicity. On the other hand, expiration is characterized by concentration of spectral energy in the lower bands, which resulted in remarkably lower spectral centroid value than other classes. Naturally, relative energy was lowest in the 'not audible' class.

Table 1 shows noticeable differences in performance across subjects. Subject P1, for example, had an especially low accuracy across all classifiers. The data of this subject contained atypically numerous instances of the class 'expiratory snoring', which is not a commonly occurring sound. This suggests that larger sets of representative training data would be useful in improving the robustness of this method; a limitation that should be addressed in future works.

Conclusion: This study is the first to present a comprehensive yet practical and pragmatic classification models that takes in consideration all breath sounds and have proven to achieve high performance in new unseen subject data. Future studies will expand the work to include more representative subjects and examine the binary classification for other classes besides snoring.

5. REFERENCES

[1] M. Cavusoglu, T. Ciloglu, Y. Serinagaoglu, M. Kamasak, O. Erogul, and T. Akcam, "Investigation of sequential properties of snoring episodes for obstructive sleep apnoea identification," *Physiol Meas*, vol. 29, no. 8, pp. 879–98, 2008.

[2] A. K. Ng, T. S. Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran, "Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?," *Sleep Med*, vol. 9, no. 8, pp. 894–898, 2008.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] H. Michael, S. Andreas, B. Thomas, H. Beatrice, H. Werner, and K. Holger, "Analysed snoring sounds correlate to obstructive sleep disordered breathing," *European Archives of Oto-Rhino-Laryngology*, vol. 265, no. 1, pp. 105–113, 2008.

[11] H. Alshaer, M. Garcia, M. H. Radfar, G. R. Fernie, and T. D. Bradley, "Detection of upper airway narrowing via classification of LPC coefficients: Implications for obstructive sleep apnea diagnosis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 681–684.

[12] A. Yadollahi, H. Alshaer, M. H. Radfar, and T. D. Bradley, "Relationship of respiratory sounds to alterations in the upper airway resistance," in *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, San Diego, USA, Sep 2012, pp. 3648–51.

[13] H. Alshaer, G. R. Fernie, and T. D. Bradley, "Monitoring of breathing phases using a bioacoustic method in healthy awake subjects," *J Clin Monit Comput*, vol. 25, pp. 285–294, Oct 2011.

[14] H. Alshaer, G. R. Fernie, and T. D. Bradley, "Phase tracking of the breathing cycle in sleeping subjects by frequency analysis of acoustic data," *International Journal of Healthcare Technology and Management*, vol. 11, no. 3, pp. 163–175, 2010.

[15] W. D. Duckitt, S. K. Tuomi, and T. R. Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiological Measurement*, vol. 27, no. 10, pp. 1047–56, 2006.

[16] A. Yadollahi and Z. Moussavi, "Automatic breath and snore sounds classification from tracheal and ambient sounds recordings," *Med Eng Phys*, vol. 32, no. 9, pp. 985–90, 2010.

[17] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Silence-breathing-snore classification from snore-related sounds," *Physiological Measurement*, vol. 29, no. 2, pp. 227–243, 2008.

[18] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, and T. Akcam, "An efficient method for snore/nonsnore classification of sleep sounds," *Physiol Meas*, vol. 28, no. 8, pp. 841–53, 2007.

[19] 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*, vol. 27(3), pp. 303–11, Jun 2013.

[20] D. Talkin, *A Robust Algorithm for Pitch Tracking, a chapter in Speech Coding and Synthesis*, chapter 14, pp. 495–518, Elsevier Sciences, 1995.