31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Formant Analysis of Breath and Snore Sounds

Azadeh Yadollahi, *Student Member, IEEE* and Zahra Moussavi, *Senior Member, IEEE*

***Abstract*— Formant frequencies of snore and breath sounds represent resonance in the upper airways; hence, they change with respect to the upper airway anatomy. Therefore, formant frequencies and their variations can be examined to distinguish between snore and breath sounds. In this paper, formant frequencies of snore and breath sounds are investigated and automatically grouped into 7 clusters based on K-Means clustering. First, formants clusters of breath and snore sounds of all subjects were investigated together and their union were calculated as the most probable ranges of the formants. The ranges for the first four formants which span the main frequency components of breath and snore sounds were found to be $[20-400]Hz$, $[270-840]Hz$, $[500-1380]Hz$ and $[910-1920]Hz$. These ranges were then used as priori information to recalculate the formants of snore and breath sounds separately. Statistical *t–test* showed the 1^{st} and 3^{rd} formants to be the most characteristic features in distinguishing the breath and snore sounds from each other.**

I. INTRODUCTION

Snoring is a common symptom and about 50% of the adults suffer from snoring [1, 2]. Snoring degrades the sleep quality of the snorer, the bed partner and other members of the household; hence, leads to somnolence, daytime sleepiness [3], impaired performance at work, higher risk of accidents [4], and diseases such as ischaemic brain infraction, systemic arterial hypertension, coronary artery disease and sleep disturbance [5, 6]. Snore sounds are also found as one of the earliest symptoms of obstructive sleep apnea (OSA).

Snore and breath sounds can be recorded with a microphone which is attached to the patient's neck or forehead or is hung near patient's head. Given the size of data recorded during sleep, the first step in any diagnosis based on acoustical analysis is to classify snore and breath sound automatically. In previous studies, Hidden Markov model (HMM) [7], energy of the recorded signal [8], and combination of zero crossings and signal's energy [9] have been used for classification of breath and snore sounds. The accuracies of these methods have been reported between 82% to 97% [7-9]. However, one should note that in all the above mentioned studies, the microphone was hung in the air above the patient's head, which means the recorded signal mostly included snore and ambient sounds.

This study was supported by TeleCommunication Research Labs (TRLabs), Canada.

A. Yadollahi is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada, R3T 5V6. She is also a research assistant in TRLabs Winnipeg, 100-135 Innovation Drive, Winnipeg, MB, Canada. azadeh@ee.umanitoba.ca

Z. Moussavi is with the Department Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada, R3T 5V6. She is also an adjunct scientist in TRLabs Winnipeg, 100-135 Innovation Drive, Winnipeg, MB, Canada. mousavi@ee.umanitoba.ca

Snoring sounds are generated in the upper airways either directly by the vibration of different structures such as soft palate, epiglottis, or the collapsible walls of the airways or by the turbulence of airflow near the narrowing of the airways [10-13]. Spectrum of snore sound signals has two characteristic components of pitch and formant. Pitch is the fundamental frequency of snore sounds' vibrations while formant frequencies represent resonance frequencies of the airways. Hence, the formant frequencies change with respect to the upper airways anatomy. On the other hand, respiratory breathing sounds convey important information on the pathology and physiology of the airways [14, 15] and their investigation during sleep may reveal useful information about changes in the breathing pattern of the patient. Therefore, in this study we have recorded tracheal respiratory sounds to have both snore and breath sounds.

One of the most frequently used methods for estimation of formant frequencies is based on linear predictive coding (LPC) [16-19]. In this method, the spectrum of the signal is approximated and the peaks represent the formant frequencies. Thus, having some preliminary information about the frequency ranges of each formant is very helpful. This information not only simplifies the search for formant frequencies, but more importantly, for the cases that one of the formants is not detectable, it helps to make sure the missed formant is not be misplaced with higher or lower formants. In [20], the formant frequencies of snore sounds were investigated manually and five frequency ranges of $[0-300]Hz$, $[300-700]Hz$, $[700-1400]Hz$, $[1400-1900]Hz$ and $[1900-2500]Hz$ were selected as the corresponding ranges of the first 5 formants, respectively.

In this study the formant frequencies of breath and snore sounds were investigated in detail and the optimum frequency range of each formant was found automatically using K–means clustering. Then, the formant frequencies of breath and snore sounds were calculated based on the predefined ranges of each formant. These formants were investigated to find the most significant formants for differentiating breath sounds from snore sounds.

II. METHOD

A. Data

Tracheal respiratory and snore sounds were recorded from 15 (3 females) patients at Health Sciences Center Sleep Disorders Clinic (Winnipeg, Canada) simultaneously with their full polysomnography (PSG) study. Subjects were recruited randomly and no limitations in terms of age, gender, BMI or AHI were applied. The sounds were recorded with a Sony (ECM-77B) microphone placed over the neck of the patient

TABLE I

PATIENTS' DEMOGRAPHIC INFORMATION.

Parameter	Age ($\mu \pm \sigma$)	BMI ($\mu \pm \sigma$)	AHI ($\mu \pm \sigma$)
Average	52.3 ± 15.2	35.1 ± 4.6	33.9 ± 42.3
Range	$[25 - 87]$	$[30.1 - 48]$	$[0.9 - 126]$

on the suprasternal notch. Sound signals were amplified and lowpass filtered with the cutoff frequency of 5 kHz using Biopac (DA100C) amplifiers. The amplified signals were digitized by National Instruments data acquisition module (NI9217) with a sampling rate of 10240 Hz. A LabView based software was developed to record and save digitized signals on a laptop machine. To synchronize our recording device with PSG system, the clock of laptop and the PSG were synchronized and the start time on our recording system was automatically saved in a text file; this information was later used to retrieve the exact time of different events and associate them with the PSG based information. The patients demographic detailed information is shown in Table I.

B. Signal Analysis

The recorded sounds were first highpass filtered with a Butterworth filter of order 5 and cutoff frequency of $20Hz$ to remove low frequency noises including motion artifacts. The snore and breath sounds segments were extracted manually by listening to the sounds and investigating them in time–frequency domain. From the segments including snore, the periods for which the snore sounds dominated breath sounds were selected and marked as pure snore segments. For every available sleep position, including left, right, supine (lying, face up) and prone (lying face down), at least 5 min of data was investigated with the same procedure. In total, 1636 snore segments and 3059 breath segments at different sleeping positions were selected from all subjects.

1) Formants range calculation: For every sound segment, linear predictive coding (LPC) was used to find the formant frequencies [19]. Since the duration of the segments can be as long as 2 seconds, the sounds are not stationary in the whole segment. Therefore, in every segment, the sound signals were windowed with a Hamming window of $20ms$ with 50% overlap. In each window, the signal was estimated by an autoregressive (AR) model. In [18], it was shown that the optimum order of the AR model for formant estimation has a strong correlation with the sound sampling rate and for sampling rates of $Fs \in [6 - 18]kHz$, the optimum order would be $M = Fs(kHz) + \gamma$ where $\gamma = 4, 5$. In this study, sound signals were recorded with a sampling rate of $10kHz$. Therefore, an AR model of order 14 was used to estimate the formants frequencies. The roots of the AR model were calculated and angles of the complex roots with positive real values were estimated, which represent the formant frequencies. Therefore, for each window, up to 7 formants in the frequency range of $[20 - 5000]$ Hz were found.

The formants of all subjects were aggregated, and K–means clustering was used to find the partitions, and group the formants frequencies of snore and breath sounds segments. K–means clustering is one of the most commonly used methods for unsupervised partitioning of data into K clusters [21-23]. In this method, K initial partitions are selected, then the partitions boundaries and centroids are updated iteratively:

- Step 1: Finding average of data in each partition and set it as the centroid of the cluster.
- Step 2: Generating the new partitions by assigning each data point to the cluster for which the Eucledian distance to the cluster centroid is minimum.
- Step 3: Finding the error as:

$$Err = \sum_{j=1}^{K} \frac{E\left[x(j) - m_j\right]^2}{E\left[x(j)\right]^2}, \tag{1}$$

  where $E[.]$ is the average function, $x(j)$ represents data points in cluster j and m_j is centroid of cluster j. The mean square error ($E\left[x(j) - m_j\right]^2$) in each cluster is dependent to the energy of the formants in the cluster. Therefore, it is normalized by the energy of the formants in the cluster; hence, the error values in the clusters with larger formant frequencies do not dominate the overall error.
- Step 4: If the difference between error values of current iteration and previous iteration is more than 10^{-5}, the method would be continued from step 1.

Since the maximum number of formant frequencies in each window was 7, the number of clusters in K–means algorithm was set to $K = 7$. On the other hand, K–means clustering is sensitive to the selection of initial partitions and it may end to a local minimum. Therefore, it was repeated with 20 different initial partitions and the partition with the minimum normalized mean square error (Eq. 1) was selected.

After finding the formant frequency ranges of snore and breath sounds segments, the results were used to find the optimum frequency ranges for the sound segments (Rng_{SnBr}). For each formant, the new frequency range was specified as the union of the corresponding formant of snore and breath sounds to cover the formants of both sounds.

2) Formant estimation: To investigate the changes in the formant frequencies of snore and breath sounds, the results of Rng_{SnBr} were used to calculate the formant frequencies again. For every sound segment, the formant frequencies were calculated in windows of $20ms$ (as mentioned in section II-B.1). Then in every window and for each frequency range of Rng_{SnBr}, the formant frequency that lied in the range, was assigned as the corresponding formant frequency. If there were more than one frequency component in the range, the smaller one was determined as the formant frequency. After finding the formants of all windows of the sound segment, for every formant ($F1 - F7$), the median of the formant values in different windows of the sound segment were determined as the formant frequencies of that segment. For every subject, the formant frequencies of the breath and snore segments were averaged in all segments, and the standard *t–test* was performed to find the most significant

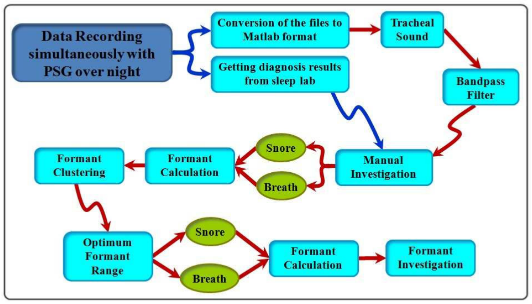

Fig. 1. Schematic of the proposed method.

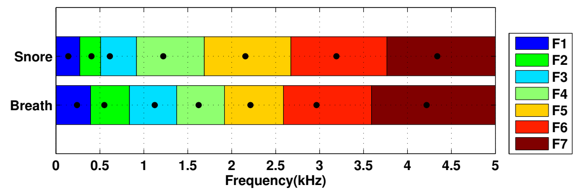

Fig. 2. Results of K–means clustering in finding the formants' frequency ranges of breath and snore sounds. Black dots show the average of the formants in each frequency range.

formants for distinguishing breath and snore sounds from each other. Figure 1 shows the schematic of the proposed method.

III. Results and Discussion

First, snore and breath segments were clustered separately by K–means clustering to investigate the differences between the sounds, and the results are shown in Fig. 2. As can be seen, the right bounds (F_H) of the first four formant clusters of the breath segments are higher than those of the snore segments.

Formant frequencies represent the effects of resonance in generation of sound signals. The results of investigating the acoustical properties of snore sounds have shown that the main peaks in their spectrum are below $1000Hz$ [24, 25, 13, 10]. On the other hand, the peaks in the power spectrum of normal tracheal sounds are found to be in higher frequency ranges of up to $1800Hz$ [14]. These studies justify our findings that formants of breath sounds have higher frequencies than those of snore sounds.

K-means clustering is a fast and automatic method to group data into clusters, and find the partitions of the clusters blind to any priori information. Therefore, in order to validate the results of K–means clustering, the frequency ranges were also estimated manually and independently from the K–Means clustering results. It was found that the ranges of K–Means clustering included the manually detected ranges (they were slightly wider).

The optimum frequency ranges of breath and snore sounds segments were calculated as the union of the ranges which were found for each sound separately (Rng_{SnBr}). Figure 3 and Table II show the results in detail. Using this method, the acquired ranges include the ranges of different formants

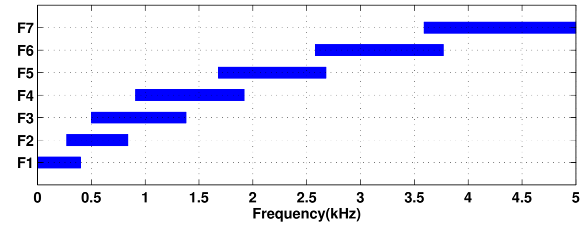

Fig. 3. Results of formants' frequency ranges after combining the information from breath and snore sounds of all subjects.

TABLE II

Results of formants' frequency ranges of snore and breath sounds of all subjects based on K–means clustering after combining the ranges. F_L and F_H are the lower and upper limits of each frequency range, respectively.

Formant	$F1$	$F2$	$F3$	$F4$	$F5$	$F6$	$F7$
$F_L(Hz)$	20	270	500	910	1680	2580	3590
$F_H(Hz)$	400	840	1380	1920	2680	3770	5000

of both sounds and cover the variations due to the differences in generation of snore and breath sounds.

Sola–Soler et. al. [20] calculated the formant frequencies of snore sounds of 16 subjects, and manually inspected the ranges of $[0-300]Hz$, $[300-700]Hz$, $[700-1400]Hz$, $[1400-1900]Hz$ and $[1900-2500]Hz$ as the frequency ranges of the snore formants. Investigating the results shown in Table II, it is evident that the ranges of Rng_{SnBr} include the ranges given in [20] for snore sounds. In addition, the proposed method in this study is fully automatic; thus, it is not biased to the observer's skill and it can be applied to the data of a large group of subjects to have more reliable ranges.

Our goal in this study was to find the optimum frequency ranges of snore and breath sounds for distinguishing the two sounds from each other automatically. Hence, the results of Rng_{SnBr} of the first stage were used to recalculate the formants of snore and breath sounds as mentioned in section II-B.2. Then, the formants were averaged among the subjects. Figure 4 displays the average and standard deviation of the first four formants frequencies of the snore and breath segments. The main frequency components of breath and snore sounds are in the frequency range of below $1500Hz$ [25, 14]. Hence, hereafter only the first 4 formants are investigated which span this frequency range. The results show that $F1$ and $F2$ frequencies of the breath sounds are greater than those of the snore sounds, while for $F3$ the relationship is reversed.

The student *t–test* analysis was performed on the formant frequencies of breath and snore sound segments and the corresponding p–values are shown in Table III. Based on the results, it is evident that $F1$ and $F3$ are significantly different between the breath and snore segments. Therefore, $F1$ and $F3$ formants can be considered as promising features for classifying snore sounds from breath sounds automatically. Furthermore, this method may be investigated further for

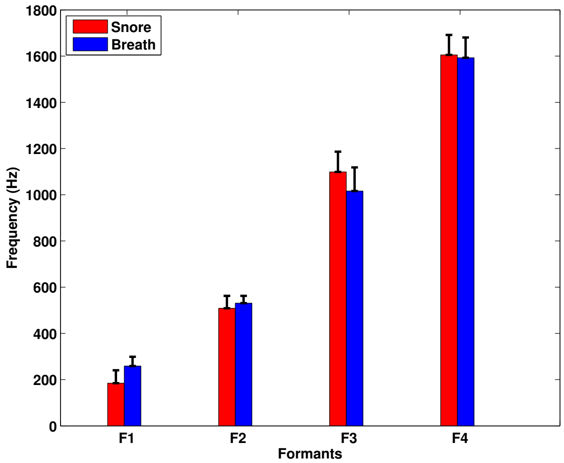

Fig. 4. Average and standard deviation of formant frequencies of snore and breath segments in 3 groups of SS, SAD and ALL.

TABLE III

RESULTS OF P–VALUES OF T–TEST BETWEEN THE FORMANTS OF SNORE AND BREATH SOUNDS IN DIFFERENT GROUPS OF SUBJECT. * REPRESENTS THE SIGNIFICANT VALUES.

Formants	$F1$	$F2$	$F3$	$F4$
p–value	0.0003*	0.1793	0.0244*	0.7009

classifying simples snorers from OSA patients.

ACKNOWLEDGEMENTS

This study was supported by TeleCommunication Research Labs (TRLabs), Canada. The authors would like to acknowledge and thank Dr. Eleni Giannouli, Wayne Thompson and Health Sciences Center Sleep Disorders Clinic (Winnipeg, Canada) staff for the invaluable help and assistance in clinical data recording.

REFERENCES

[1] E. Lugaresi, F. Cirignotta, C. Coccagna, and C. Piana, "Some pidemiological data on snoring and cardiorespiratory disturbances," *Sleep*, vol. 3, pp. 221–224, 1980.

[2] P. Norton and E. Dunn, "Snoring as a risk factor for disease: an epidemiolocial survey," *Br Med J*, vol. 291, pp. 630–632, 1985.

[3] R. Rakel, "Clinical and societal consequences of obstructive sleep apnea and excessive daytime sleepiness," *Postgrad Med.*, vol. 121(1), pp. 86–95, 2009.

[4] F. Pizza, S. Contardi, S. Mondini, L. Trentin, and F. Cirignotta, "Daytime sleepiness and driving performance in patients with obstructive sleep apnea: comparison of the mslt, the mwt, and a simulated driving task," *Sleep*, vol. 32(3), pp. 382–91, 2009.

[5] N. Takama and M. Kurabayashi, "Influence of untreated sleep-disordered breathing on the long-term prognosis of patients with cardiovascular disease," *Am J Cardiol.*, vol. 103(5), pp. 730–4, 2009.

[6] G. Hamilton, I. Meredith, A. Walker, and P. Solin, "Obstructive sleep apnea leads to transient uncoupling of coronary blood flow and myocardial work in humans," *Sleep*, vol. 32(2), pp. 263–70, 2009.

[7] W. Duckitt, S. Tuomi, and T. Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiol. Meas.*, vol. 27, pp. 1047–56, 2006.

[8] 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, pp. 841–53, 2007.

[9] A. Karunajeewa, U. Abeyratne, and C. Hukins, "Silence-breathing-snore classification from snore-related sounds," *Physiol Meas.*, vol. 29(2), pp. 227–43, 2008.

[10] R. Beck, M. Odeh, A. Oliven, and N. Gavriely, "The acoustic properties of snores," *Eur Respir J.*, vol. 8(12), pp. 2120–8, 1995.

[11] J. Grotberg and S. Davis, "Fluid-dynamic flapping of a collapsible channel: sound generation and flow limitation," *J. Biomech.*, vol. 13, pp. 219–30, 1980.

[12] G. Liistro, D. Stanescu, and C. Veriter, "Pattern of simulated snoring is different through mouth and nose," *J. Appl. Phys.*, vol. 70, pp. 2736–41, 1991.

[13] J. Perez-Padilla, E. Slawinski, L. Difrancesco, R. Feige, J. Remmers, and W. Whitelaw, "Characteristics of the snoring noise in patients with and without occlusive sleep apnea," *Am Rev Respir Dis.*, vol. 147(3), pp. 635–44, 1993.

[14] H. Pasterkamp, S. Kraman, and G. Wodicka, "Respiratory sounds, advances beyond the stethoscope," *Am J Respir Crit Care Med*, vol. 156, pp. 974–987, 1997.

[15] A. Sovijarvi, L. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, and J. Earis, "Characteristic of breath sounds and adventitious respiratory sounds," *European Respiratory Review Journal*, pp. 591–596, 2000.

[16] R. W. Schafer and L. R. Rabiner, "System for automatic formant analysis of voiced speech," *The Journal of the Acoustical Society of America*, vol. 47, Issue 2B, pp. 634–648, 1970.

[17] B. Atal and S. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.*, vol. 50(2), pp. 637–655, 1971.

[18] J. Markel, "Digital inverse filtering-a new tool for formant trajectory estimation," *IEEE Transactions on Audio and Electroacoustics*, vol. 20(2), pp. 129–137, 1972.

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

[20] J. Sola-Soler, R. Jane, J. Fiz, and J. Morera, "Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea," in *IEEE-EMBS*, Cancun, Mexico, 2003, pp. 2527–2530.

[21] A. K. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22(1), pp. 4–37, 2000.

[22] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, No. 2, pp. 129 – 137, 1982.

[23] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Springer, 1991.

[24] F. Dalmasso and R. Prota, "Snoring: analysis, measurement, clinical implications and applications," *Eur Respir J*, vol. 9, pp. 146–159, 1996.

[25] J. Fiz, J. Abad, R. Jane, M. Riera, M. Mananas, 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. 2365–70, 1996.