INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

A dual source-filter model of snore audio for snorer group classification

Achuth Rao M V[1], *Shivani Yadav*[2], *Prasanta Kumar Ghosh*[1]

[1]Electrical Engineering, Indian Institute of Science, Bangalore 560012, India
[2] BioSystems Science and Engineering, Indian Institute of Science, Bangalore 560012, India

achuthraomv@ee.iisc.ernet.in, shivanir@be.iisc.ernet.in, prasantg@ee.iisc.ernet.in

Abstract

Snoring is a common symptom of serious chronic disease known as obstructive sleep apnea (OSA). Knowledge about the location of obstruction site (V- Velum, O- Oropharyngeal lateral walls, T-Tongue, E-Epiglottis) in the upper airways is necessary for proper surgical treatment. In this paper we propose a dual source-filter model similar to the source-filter model of speech to approximate the generation process of snore audio. The first filter models the vocal tract from lungs to the point of obstruction with white noise excitation from the lungs. The second filter models the vocal tract from the obstruction point to the lips/nose with impulse train excitation which represents vibrations at the point of obstruction. The filter coefficients are estimated using the closed and open phases of the snore beat cycle. VOTE classification is done by using SVM classifier and filter coefficients as features. The classification experiments are performed on the development set (283 snore audios) of the MUNICH-PASSAU SNORE SOUND CORPUS (MPSSC). We obtain an unweighted average recall (UAR) of 49.58%, which is higher than the INTERSPEECH-2017 snoring sub-challenge baseline technique by $\sim$3% (absolute).

Index Terms: snore classification, source-filter model, sparsity

1. Introduction

Snoring is a prevalent disorder affecting 20-40% of the general population [1]. It is a marker of obstructive sleep apnea (OSA) which lead to hypertension [2], myocardial infarction [3] and stroke [4]. Sleep quality of bed partners affected by the simple snoring leads to social nuisance even though it is harmless for snorer.

Snoring is caused by the airway narrowing or obstruction during breathing [5][6]. Based on the location of airway narrowing or obstruction along the vocal tract, the snorers are broadly divided into four groups [5], namely V- Velum (palate), including soft palate, uvula, lateral velopharyngeal walls; O- Oropharyngeal lateral walls, including palatine tonsils; T- Tongue, including tongue base and airway posterior to the tongue base; E- Epiglottis. The classification of these four groups often referred as VOTE classification. The corresponding locations are shown in Figure. 2 in the work by Qian et. al [7].

Key to better clinical results is a treatment exactly targeting the area in the upper airways where the snoring sound is generated in a patient. Only moderate success has been shown by various conservative and surgical methods attempting to improve or cure snoring. Now a days, drug induced sleep endoscopy (DISE) is widely used to identify location of obstructions and form of vibrations [5] in order to evaluate the obstruction level and pattern. But DISE is expensive, time consuming and strenuous for patients. Therefore, treatment using acoustic based analysis and classification of snore sounds becomes useful [1].

There are several works in the literature that try to classify the four snorer groups (or their subsets). Pevernagie et. al [1] pioneer in using snore audio to classify the snorer group. They demonstrated the similarity and dissimilarity between snoring and speech. Miyazaki et al. [8] observed that the fundamental frequency of snoring can be used to distinguish between palatal/non-palatal snoring. Hill et al. [9] found that the crest factor (the ratio of peak to root mean square value of a time-varying signal) was significantly higher for palatal snorers. Agrawal et al. [10] showed that snoring sound generated by the palate and tongue are characterized by low and high peak frequency respectively. Qian et. al [7] did VOTE classification by using multiple features such as spectral frequency, sub-band energy ratio, mel-scale frequency cepstral coefficients (MFCC), empirical mode decomposition based features and wavelet features (WF). They found that MFCC and WF features are best suited for snorer group classification.

The aforementioned literature focuses on evaluating certain well-known acoustic features for snorer group classification. None of them tried to model the generative process of the snore directly. In this paper, we propose a model similar to the source-filter model of speech [11], which approximates the generative process of the snore audio. Based on the model, we propose a three step method for snorer group classification. The first step is to estimate the proposed model parameters from the snore audio. The second step is to extract the features from estimated model parameter. Finally, we use SVM based classifier to classify the extracted features into four-snorer groups (VOTE). The classification experiments are performed on the development set (283 snore audios) of the MPSSC. We obtain an unweighted average recall (UAR) of 49.58%, which is higher than the INTERSPEECH-2017 snoring sub-challenge baseline technique by $\sim$3% (absolute).

2. Proposed snorer group classification

Figure 1 shows the block diagram, summarizing the steps involved in the proposed snorer group classification task, which consists of four main blocks. Given a snore audio $s[n]$, and snore generative process represented by block $\mathcal{M}$, model parameters (θ) are estimated by a parameter estimator. The feature extractor computes features (f) from the model parameters (θ). In the final step, we classify the snorer group (VOTE) based on the computed features (f).

The source-filter (SF) model of speech assumes that the speech is generated by exciting a filter by an impulse train or a white Gaussian noise [11]. The filter represents the vocal tract and excitation represents the airflow through the glottis. The impulse train excitation is for the voiced sounds, which approximates the quasi periodic oscillations of the glottal flow. The white Gaussian noise excitation is for the unvoiced sound, which approximates the glottal flow when the vocal folds are open. Several modifications to the SF model have been proposed in the past, such as branched tube model for the nasal sounds [12, 13] and transmission model for the respiratory tract [14, 15]. We hypothesize that a similar model comprising fil-

http://dx.doi.org/10.21437/Interspeech.2017-1211

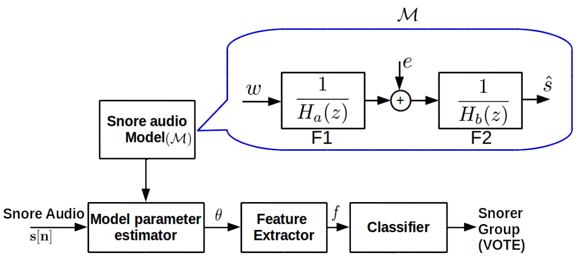

Figure 1: *Steps in the proposed snorer group classification.*

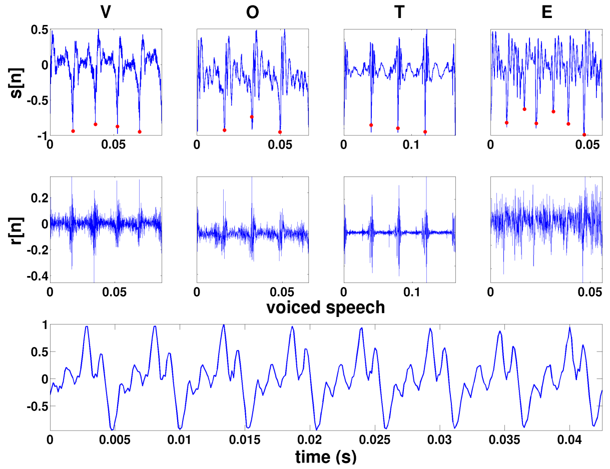

Figure 2: *Waveform level comparison of different snorer audio (VOTE) with the voiced speech. First row shows the snore audio ($s[n]$) from four snorer groups, the red dots indicate the detected SBCIs and second row shows the corresponding $r[n]$.*

ter and excitation holds good for snore audio. The comparison among snore audio for different snorer groups and voiced speech is shown in the first and third row of Figure 2. It is evident from the figure that : 1) The waveform in all four groups are quasi periodic [16][17] and impulsive in nature, similar to the voiced speech 2) $s[n]$ noise level increases from group V to E , while the voiced speech has the least noise level 3) time period of the voiced speech is lower than that of snore audio. We call one period of snore audio as snore beat cycle and the location of impulse is referred to as snore beat closing instant (SBCI). This term is similar to the glottal closure instant (GCI) used in speech [18].

Our hypothesis for generative process of snore audio involves two main components: 1) Two acoustic tubes: one tube from lungs to the obstruction location and second is from obstruction location to lips/nose. 2) Two excitation sources: the first excitation at lungs is white noise similar to the unvoiced speech sounds, which excites the first acoustic tube. Second excitation at the location of obstruction is modelled by sum of airflow from first tube and impulse excitations. Impulse excitation represents the quasi-periodic obstruction which is similar to voiced speech. Based on the above hypothesis, we propose a dual source-filter model which models the snore audio using two acoustic tubes and two excitation sources.

2.1. Proposed dual source-filter model for snore audio

The first acoustic tube from lungs to obstruction location is modelled as one all-pole filter (F1), which is excited by wide sense stationary white Gaussian noise ($w[n]$). Second acoustic tube from airway obstruction to lips/nose is modelled as another all-pole filter (F2), excited by sum of impulse train ($e[n]$) and the first filter output. Output of the second filter $\hat{s[n]}$ approximates the observed snore audio $s[n]$. As snore sound could come through both lips and nose, a pole-zero filter would be a good choice for F2, although, in this work, we approximate it by an all-pole filter. The block diagram of the snore audio model ($\mathcal{M}$) is shown in the blue box in Figure. 1. Thus, $s[n]$ can be written as:

$$s[n] \approx w[n] * h_c[n] + e[n] * h_b[n], \; 0 \leq n \leq N-1 \qquad (1)$$

where $h_c[n] = h_a[n] * h_b[n]$, h_a and h_b are the impulse responses of F1 and F2 respectively, $\mathbf{w} = [w[0], w[1], w[2], \ldots, w[N-1]]^T$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, I)$, $*$ indicates the convolution operator and $[.]^T$ indicates the transpose operator. In the Z-domain the eq. 1 can be written as:

$$S(z) = \frac{W(z)}{H_a(z)H_b(z)} + \frac{E(z)}{H_b(z)} \;, \qquad (2)$$

where $W(z)$, $E(z)$, $\frac{1}{H_a(z)}$ and $\frac{1}{H_b(z)}$ are the Z-transform of $w[n]$, $e[n]$, $h_a[n]$ and $h_b[n]$ respectively. Let $H_a(z) = 1 + \sum_{k=1}^{L-1} a_k z^{-k}$ and $H_b(z) = 1 + \sum_{k=1}^{M-1} b_k z^{-k}$, where a_k and b_k are coefficients of all-pole filters F1 and F2 respectively. The Z-transform of the $h_c[n]$ is given by

$$H_c(z) = \frac{1}{H_a(z)H_b(z)} = \frac{1}{1 + \sum_{k=1}^{M+L-1} c_k z^{-k}} \;,$$

Given the snore audio $s[n], 0 \leq n \leq N-1$, the parameters of the model, $\theta = \{a_k, c_k\}$ are estimated.

2.2. Proposed model parameter estimation method

The direct estimation of model parameters (θ) from the snore audio is, in general, computationally intractable. Hence, to make it computationally tractable we assume impulse response of F2 (h_b) decays faster as compared to the period of e. The above assumption makes the effect of h_b dominant around SBCIs and less dominant in the middle of the snore beat cycle. Therefore, we weigh different parts of the snore beat cycle based on the SBCIs and estimate the filter coefficients. This process is similar to the method proposed by Airaksinen et. al [18] for glottal inverse filtering of speech signal to remove the effect of glottal source in estimating the impulse response of the vocal tract.

Model parameter estimation process is summarized in Figure 3. The main steps involve detecting the SBCI locations, constructing the two windows to attenuate the effect of source/filter and estimating the filter coefficients from the windowed signal. These steps are described in detail below.

1. Spectral tilt estimation and removal: In this step, we estimate the spectral tilt by using a single order linear prediction model on the given snore audio. The estimated coefficient is denoted by t. The snore audio is inverse filtered to get the preprocessed snore audio denoted by $s[n]$.

2. SBCI estimation: The SBCIs are similar to GCIs of the speech waveform. Hence we use GCI estimation method proposed by Prathosh et. al [19] on the snore audio. The m-th SBCI location is denoted by $ci[m]$, $1 \leq m \leq N_{ci}$, where N_{ci} denotes the number of SBCIs in the snore audio $s[n]$.

3. Windows Construction: We construct two windows $w_1[n]$ and $w_2[n]$ such that the first window attenuates the samples in the middle of the snore beat cycle and second window

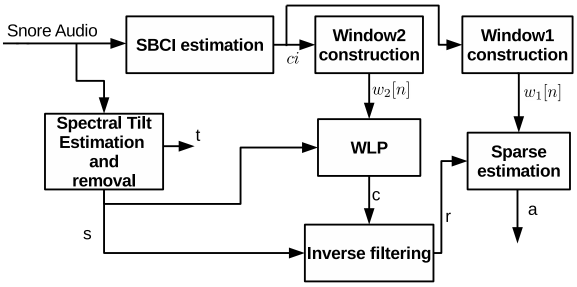

Figure 3: *Steps in the proposed model parameter estimation.*

attenuates the samples around the SBCIs. The proposed windows are defined as:

$$w_1[n] = \begin{cases} 1 - d & ci[m] - L[m] \leq n \leq ci[m] + R[m], \\ d & \text{otherwise} \end{cases}$$

$w_2[n] = 1 - w_1[n]$ and $2 \leq m \leq N_{ci} - 1$, $R[m] = \eta \times (ci[m+1] - ci[m])$, $L[m] = \eta \times (ci[m] - ci[m-1])$ where η is the factor controlling the percentage of snore beat cycle selected by $w_1[n]$ and d is the attenuation factor for windows.

4. Weighted Linear Prediction (WLP): The preprocessed snore audio $s[n]$ from eq. 1 is multiplied with the window $w_2[n]$ to get $s1[n]$ as shown below:

$$\begin{aligned} s1[n] &= s[n] \times w_2[n] \\ &= (w[n] * h_c[n]) \times w_2[n] + (e[n] * h_b[n]) \times w_2[n] \end{aligned} \tag{3}$$

As the effect of $h_b[n]$ is assumed to be not dominant in the $w_2[n]$, the second term is approximately zero. Thus, eq. 3 can be written as:

$$s1[n] = s[n] \times w_2[n] \approx (w[n] * h_c[n]) \times w_2[n] \tag{4}$$

Z-transform of $h_c[n]$ is $H_c(z) = \frac{1}{1 + \sum_{k=1}^{M+L-1} c_k z^{-k}}$. The error function to estimate $\mathbf{c}$ is given by

$$E(\hat{\mathbf{c}}) = \sum_{n=M+L}^{N-1} \left(s[n] - \sum_{i=1}^{M+L-1} s[n-i]\hat{c}_i \right)^2 w_2[n] \tag{5}$$

The minimum mean squared estimate of $\hat{\mathbf{c}}$ is same as the WLP [20] given by

$$\left(\sum_{n=M+L}^{N-1} w_2[n] \mathbf{s_n} \mathbf{s_n}^T \right) \mathbf{c} = \sum_{n=M+L}^{N-1} w_2[n] s[n] \mathbf{s_n}$$

where $\mathbf{c} = [c_1, c_2, \ldots c_{M+L-1}]$, $\mathbf{s_n} = [s[n-1], s[n-2], \ldots s[n-M-L]]$.

5. Inverse filtering: Given the estimate of $\mathbf{c}$ or $H_c(z)$, we remove the effect of $H_c(z)$ from the snore audio using the inverse filter $\frac{1}{H_c(z)}$ and get $r[n]$. From eq. 2 the inverse filtered output can be written as:

$$R(z) = \frac{S(z)}{H_c(z)} = S(z)H_a(z)H_b(z) = W(z) + E(z)H_a(z) \; , \tag{6}$$

Note that the $H_a(z)$ is an FIR filter with impulse response $\mathbf{a} = [1, a_1, a_2, \ldots, a_{L-1}]$, n-th element of $\mathbf{a}$ is denoted by $a[n]$. In the time domain, eq. 6 can be written as:

$$r[n] = w[n] + e[n] * a[n] \; , \tag{7}$$

6. Sparse estimation: Given the inverse filtered signal $r[n]$, the filter coefficient vector $\mathbf{a} = [1, a_1, a_2, \ldots a_{L-1}]$ needs to be estimated. Eq. 7 can be compactly written as:

$$\mathbf{r} = A\mathbf{e} + \mathbf{w} \; , \tag{8}$$

where $\mathbf{r} = [r[0], r[1], \ldots, r[N-1]]^T$, A is a convolution matrix with $\mathbf{a}$ as the first row, $\mathbf{w}$ is the white Gaussian noise vector. $\mathbf{e} = [e[0], e[1], \ldots, e[N-1]]^T$, is an impulse train and hence, it is sparse. $\mathbf{a}$ contains the filter coefficients corresponding to the filter from lungs to obstruction point (F1). In the snorer group classification problem, the obstruction point is unknown, that makes the order of first filter to be unknown. Therefore, we impose a sparsity constraint on $\mathbf{a}$ to automatically infer the order of the filter. The maximum likelihood estimate of $\mathbf{a}$ and $\mathbf{e}$ from eq. 8 with sparse (l_1-norm) constraint on $\mathbf{e}$ and $\mathbf{a}$ is given by ,

$$\mathbf{a}^*, \mathbf{e}^* = \arg\min_{\mathbf{e},\mathbf{a}} ||\mathbf{r} - A\mathbf{e}||_2^2 + \lambda||\mathbf{e}||_1 + \beta||\mathbf{a}||_1 \; , \tag{9}$$

where, β and λ control the sparsity of the $\mathbf{a}$ and $\mathbf{e}$ respectively. From eq. 8, we see that the effect of $\mathbf{a}$ is dominant around the impulse location of $e[n]$. It is also evident in $r[n]$, from the second row of the Figure 2. It is clear that impulse response of $H_b(z)$ decays fast and is primarily centered around a SBCI. We use temporal weighting $w_1[n]$ to only capture the samples around SBCI. Hence, eq. 9 can be written as:

$$\mathbf{a}^*, \mathbf{e}^* = \arg\min_{\mathbf{e},\mathbf{a}} ||W_1(\mathbf{r} - A\mathbf{e})||_2^2 + \lambda||\mathbf{e}||_1 + \beta||\mathbf{a}||_1, \tag{10}$$

where, W_1 is a diagonal matrix with entries $\sqrt{w_1[n]}$. Since both $\mathbf{e}$ and $\mathbf{a}$ are unknown, the direct optimization of eq. 10 is difficult. Therefore, we use the alternate minimization method [21] to optimize one variable given the other. The initial value of $\mathbf{a}$ is found by moving average estimation of the signal $r[n]$ using the Durbin method [22]. Thus, given a snore signal $s[n]$, the model parameters $\theta = \{a_k, c_k\}$ are estimated.

2.3. Feature extraction and classification

Given a snore recording, we extract model parameters in two different ways: 1) We extract model parameters using middle ζ percentage of the entire snore audio to get $\theta^{overall} = \{\mathbf{a}^{\mathbf{overall}}, \mathbf{c}^{\mathbf{overall}}\}$. 2) We divide the recording into frames of length N_f, shift N_s to estimate the model parameters and tilt (t) in each frame (θ^{frame}) and compute the functionals such as mean,variance and median of the parameters across all frames to get $\theta^{mean} = \{\mathbf{a}^{\mathbf{mean}}, \mathbf{c}^{\mathbf{mean}}\}$, $\theta^{var} = \{\mathbf{a}^{\mathbf{var}}, \mathbf{c}^{\mathbf{var}}\}$, $\theta^{median} = \{\mathbf{a}^{\mathbf{median}}, \mathbf{c}^{\mathbf{median}}\}$, t^{mean}, t^{var} and t^{median} respectively.

We use SVM classifier to identify the four snorer groups by using extracted features. In general, SVM[1] classifiers are binary and discriminative. However, a four-class SVM classifier is built using pair-wise coupling strategy [23].

3. Experiments and results

We use the MPSSC database to evaluate the proposed method. The details of the dataset and the train/development/test split details can be found in [24]. The number of events per class in the database is strongly unbalanced, with 85% of samples from the classes V and O, 11% E-events and 5% T-snores. This is in line with the probability of these categories in normal sleep.

[1] We experimented with the Deep Neural Networks classifier, but it performed poorly on the development set.

3.1. Feature extraction and Hyper parameter optimization

The window attenuation factor (d) is set to $1e-5$ [18] and the fraction of snore beat cycle (η) for $w_1[n]$ is set to 0.3. The order of filter for both optimization in eq. 5 and 10 is set to 25. We use ζ=50 to estimate the $\theta^{overall}$. We use N_f=900 and N_s=64 to estimate the θ^{frame}. The sparsity parameters $\lambda = 5e-3$ and $\beta = 5e-4$ are found experimentally using 3-fold cross validation within the training set. The optimization problem in eq. 10 is solved using the interior point method [25].

We implement SVM classifiers using the LIBSVM toolkit [26]. The values of complexity (C) and variance γ of the radial basis function (RBF) are found using 3-fold cross validation on the train set and the grid search. To handle the class imbalance, we use the class weights of 1, 2, 21, 6 for V, O, T, E groups respectively, which are same as the up-sampling factors used by the baseline [24]. To compare the proposed model with the single source filter model, we extracted LPC (linear predictive coefficients) of order 50 using entire snore audio denoted by $AR^{overall}$. We also extract frames wise LPC coefficients using N_f=900 and N_s=64 and compute the same set of functionals defined in the section 2.3 and denoted by AR^{mean}, AR^{var} and AR^{median}.

3.2. Results and discussion

The classifier is trained using the train set and evaluated on the development set. Table 1 shows the UAR for different combinations of features and the classifier settings. Out of all combinations of features and classifier settings, the best eleven combinations are only reported in the table. From the table it can be seen that the filter coefficients $\mathbf{c}^{\mathbf{overall}}$ with the RBF kernel performs 9.18% and 3% (absolute) better than the COMPARE functional (CF) and COMPARE BoAW (CBoAW) methods respectively [24]. This indicates that the proposed dual source-filter based approach is able to classify the snorer groups using only 25-dimensional feature vector unlike a large number of features used in CF and CBoAW [24]. The filter coefficients $\mathbf{a}^{\mathbf{overall}}$ performs worse compared to the baseline. Hence, no results are reported with the $\mathbf{a}^{\mathbf{overall}}$ feature. The combination of $\mathbf{c}^{\mathbf{var}}$ and $\mathbf{t}^{\mathbf{var}}$ achieves better UAR than the baseline with linear and RBF kernel. The LPC coefficients and its functional with RBF kernel result in a lower UAR than the CF indicating that the single filter may not be sufficient to model the snore audio.

Table 2 shows the confusion matrix and the classification accuracy (%) for each snore group obtained by using CF, CBoAW and features #3, #7 and #9 from Table 1– the three best performing feature-classifier combinations. It can be observed from the table that classification accuracies for T and E classes with CF and CBoAW are low because of the highly imbalanced dataset. On the other hand our proposed feature #9 shows an improvement of $\sim$ 41% and $\sim$27% in classifying the T group as compared to CF and CBoAW respectively. An improvement of $\sim$58% and $\sim$34% from CF and CBoAW, respectively, is displayed by using feature #7 in classifying E-group, even though the average UAR decreased after adding t-feature. The misclassification for E to V class is reduced in all proposed feature sets. This could be because the obstruction location for E and V groups are far apart which results in significantly different filter coefficient estimates. The classification accuracies of V and O groups is decreases in all the proposed features because the obstruction point for V and O are closer, which results in the similar filter coefficient estimates.

We further train the classifier using train and development sets and test on the test set. The proposed feature#3 results in an UAR of 52.75%, which is 5.75% lower than that of CF and 2.85% higher than CBoAW. Classification accuracies for V, O, T, E classes are 54%, 41%, 37% and 78%, respectively.

The $\mathbf{c}^{\mathbf{overall}}$ feature shows the highest UAR on the development set. A possible explanation is, $\mathbf{c}^{\mathbf{overall}}$ represents the transfer function from lungs to the lips/nose, which represents the overall shape of the vocal tract that also includes the obstruction location. We observe that the $\mathbf{a}^{\mathbf{overall}}$ feature performs poorly because of the slow convergence of alternative minimization algorithm used for optimizing eq. 10, which fails to select the right model order for the first filter. We also notice that the proposed model is very sensitive to SBCI detection accuracy. The UAR can be further improved by modifying the existing GCI detection method for the snore audio that has more noisy structure and a lower fundamental frequency.

Table 1: *Unweighted Average Recall (UAR) for different sets of features and classifiers on the development set. The baseline numbers are shown in gray and the proposed methods that performed better than the best baseline (CBoAW) are shown in different color. '+' indicates the feature fusion. For all proposed methods the C parameter lies between 1e+4 to 1e+6 and γ parameter between 1e-4 to 1e-1.*

#	feature	SVM kernel	UAR(%)
1	$\mathbf{a}^{\mathbf{overall}}$	Linear	29.80
2	$\mathbf{c}^{\mathbf{overall}}$	Linear	44.80
3	$\mathbf{c}^{\mathbf{overall}}$	RBF	**49.58**
4	$\mathbf{a}^{\mathbf{overall}} + \mathbf{c}^{\mathbf{overall}}$	RBF	40.65
5	$\mathbf{c}^{\mathbf{var}}$	Linear	44.32
6	$\mathbf{c}^{\mathbf{median}}$	Linear	41.37
7	$\mathbf{c}^{\mathbf{var}} + \mathbf{t}^{\mathbf{var}}$	Linear	**47.22**
8	$\mathbf{c}^{\mathbf{mean}} + \mathbf{t}^{\mathbf{mean}}$	Linear	44.79
9	$\mathbf{c}^{\mathbf{var}} + \mathbf{t}^{\mathbf{var}}$	RBF	**47.08**
10	$\mathbf{c}^{\mathbf{mean}} + \mathbf{t}^{\mathbf{mean}}$	RBF	41.81
11	$\mathbf{c}^{\mathbf{median}} + \mathbf{t}^{\mathbf{median}}$	RBF	44.33
12	COMPARE BoAW	Linear	46.6
13	COMPARE functional	Linear	40.6
14	$AR^{overall}$	RBF	35.89
15	AR^{mean}	RBF	32.31
16	AR^{median}	RBF	30.2
17	AR^{var}	RBF	27.35

Table 2: *Confusion matrix for the baseline and the proposed feature set. % indicates the class wise classification accuracy.*

	CF					BoAW					#3					#7					#9				
	V	O	T	E	%	V	O	T	E	%	V	O	T	E	%	V	O	T	E	%	V	O	T	E	%
V	123	32	4	2	76	128	30	1	2	79	105	44	7	5	65.2	107	38	6	8	67.3	104	43	4	10	64.6
O	24	50	0	1	67	38	37	0	0	49	43	30	0	2	40	48	22	0	5	29.3	41	25	0	9	33
T	0	12	1	2	6.6	3	9	3	0	20	6	3	6	0	40	4	7	3	1	20	4	3	7	1	47
E	22	6	0	4	13	13	6	1	12	37	5	10	0	17	53	8	1	0	23	71.8	5	10	0	17	53

4. Conclusion

In this paper, we propose a dual source-filter model for the generative process of the proposed snore audio and present the corresponding parameter estimation method. We compute features from model parameters to classify the snorer group. The proposed method performs better than the CF and CBoAW by 9.18% and 3% respectively on the development set. We also find that the proposed method is able to discriminate the E/T-classes well from the V/O class. Two major limitations of the model are the independence among the two filters and the restriction of the second filter to be all-pole. Future work includes modifying the model to include the interaction between the filters and improve the estimation method with a pole-zero filter model.

5. References

[1] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep medicine reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[2] P. E. Peppard, T. Young, M. Palta, and J. Skatrud, "Prospective study of the association between sleep-disordered breathing and hypertension," *New England Journal of Medicine*, vol. 342, no. 19, pp. 1378–1384, 2000.

[3] J. M. Marin, S. J. Carrizo, E. Vicente, and A. G. Agusti, "Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study," *The Lancet*, vol. 365, no. 9464, pp. 1046–1053, 2005.

[4] H. K. Yaggi, J. Concato, W. N. Kernan, J. H. Lichtman, L. M. Brass, and V. Mohsenin, "Obstructive sleep apnea as a risk factor for stroke and death," *New England Journal of Medicine*, vol. 353, no. 19, pp. 2034–2041, 2005.

[5] N. Charakorn and E. J. Kezirian, "Drug-induced sleep endoscopy," *Otolaryngologic Clinics of North America*, vol. 49, no. 6, pp. 1359–1372, 2016.

[6] V. JamesAbdullah, I. Koutsourelakis, M. Ravesloot, D. L. Y. Lee, S. C. N. Ha, C. A. van Hasselt, and N. de Vries, "Drug-induced sleep endoscopy," *Advanced Surgical Techniques in Snoring and Obstructive Sleep Apnea*, p. 43, 2013.

[7] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Classification of the excitation location of snore sounds in the upper airway by acoustic multi-feature analysis," *IEEE Transactions on Biomedical Engineering*, vol. PP, no. 99, pp. 1–1, 2016.

[8] K. I. K. T. S. Miyazaki, Yoshiaki Itasaka, "Acoustic analysis of snoring and the site of airway obstruction in sleep related respiratory disorders," *Acta Oto-Laryngologica*, vol. 118, no. 537, pp. 47–51, 1998.

[9] P. Hill, B. Lee, J. Osborne, and E. Osman, "Palatal snoring identified by acoustic crest factor analysis," *Physiological measurement*, vol. 20, no. 2, p. 167, 1999.

[10] S. Agrawal, P. Stone, K. McGuinness, J. Morris, and A. Camilleri, "Sound frequency analysis and the site of snoring in natural and induced sleep," *Clinical Otolaryngology*, vol. 27, no. 3, pp. 162–166, 2002.

[11] G. Fant, "Acoustic theory of speech production: with calculations based on x-ray studies of russian articulations," vol. 2, 1971.

[12] K. Schnell and A. Lacroix, "Analysis of nasals and nasalized vowels based on branched tube models," *11th European Signal Processing Conference*, pp. 1–4, 2002.

[13] K. A Schnell and Lacroix, "Parameter estimation of branched tube models by iterative inverse filtering," *14th International Conference on Digital Signal Processing*, vol. 1, pp. 333–336, 2002.

[14] C. M. Ionescu, I. Muntean, J. Tenreiro-Machado, R. De Keyser, and M. Abrudean, "A theoretical study on modeling the respiratory tract with ladder networks by means of intrinsic fractal geometry," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 2, pp. 246–253, 2010.

[15] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, "An acoustic model of the respiratory tract," *IEEE transactions on biomedical engineering*, vol. 48, no. 5, pp. 543–550, 2001.

[16] L. Huang, "Mechanical modeling of palatal snoring," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3642–3648, 1995.

[17] L. Huang, S. J. Quinn, P. D. Ellis, and J. E. F. Williams, "Biomechanics of snoring," *Endeavour*, vol. 19, no. 3, pp. 96–100, 1995.

[18] M. Airaksinen, T. Raitio, B. Story, and P. Alku, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 3, pp. 596–607, 2014.

[19] A. Prathosh, T. Ananthapadmanabha, and A. Ramakrishnan, "Epoch extraction based on integrated linear prediction residual using plosion index," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2471–2480, 2013.

[20] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilised weighted linear prediction," *Speech Communication*, vol. 51, no. 5, pp. 401–411, 2009.

[21] T. F. Chan and C.-K. Wong, "Convergence of the alternating minimization algorithm for blind deconvolution," *Linear Algebra and its Applications*, vol. 316, no. 1-3, pp. 259–285, 2000.

[22] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, vol. 46, no. 3/4, pp. 306–316, 1959.

[23] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[24] B. Schuller, S. Steidl, A. Batliner, E. Bergelson, J. Krajewski, C. Janott, A. Amatuni, M. Casillas, A. Seidl, M. Soderstrom *et al.*, "The interspeech 2017 computational paralinguistics challenge: Addressee, cold & snoring." *Proceedings INTERSPEECH 2017, ISCA, Stockholm, Sweden, 2017.*

[25] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale 1-regularized least squares," *IEEE journal of selected topics in signal processing*, vol. 1, no. 4, pp. 606–617, 2007.

[26] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.