Discriminant Feature Vectors for Characterizing Ailment Cough vs. Simulated Cough

Ravi Shekhar Jha[1], Vishwanath Pratap Singh[2]
Indian Institute of Technology Guwahati
Assam, India
r.jha@iitg.ac.in[1], s.vishwanath@iitg.ac.in[2]

Vinay Kumar Mittal[3]
Indian Institute of Information Technology Chittoor
Sri City, Andhra Pradesh, India
vkmittal@iiits.in[3]

***Abstract*— Cough is the powerful mechanism of human body to clear the central airways. Cough is often triggered by the mucus that drains down the back of the throat. An infection in the lungs or upper airway passages can cause *cough*. This paper describes the characteristics of *ailment* (cold) cough with respect to *simulated* (healthy) cough sound signals. Standard signal processing methods are used to extract the features from the acoustic signal of cough sounds that characterize the cough sounds. Analysis of cough sounds is carried out using the instantaneous fundamental frequency (*F0*) as a source feature, Mel frequency cepstral coefficients (*MFCCs*) as filter features, and signal energy as a combined feature of both the source of excitation and the vocal tract system. First four formants of the human speech production mechanism are also used for characterizing the *ailment* cough vs. *simulated* cough. Further a Support Vector Machine (*SVM*) classifier is used for the automated classification of the ailment cough sound from the simulated cough sound signal. Encouraging results are obtained. The proposed approach can have potential applications in assisted diagnosis of diseases, based upon cough sound signal analysis.**

Keywords—Cough, simulated cough, ailment cough, MFCC, DFT, DCT, instantaneous fundamental frequency, formants

I. INTRODUCTION

Cough is a powerful natural respiratory defence mechanism. It clears the central airways in human breathing system, and is one of the most common symptom of pulmonary diseases. The characteristics of the cough sounds are dependent on the flow and configuration of the tissue elements involved. These characteristics are likely to change during cough, as the physiological configuration changes dynamically. The changes do not occur frequently in healthy speakers. But it is an important symptom in many respiratory diseases. The temporal pattern of changes in the *cough sound* can be analysed in *three phases*: (i) starting with an explosive *burst* and opening of glottis, (ii) a period of *noisy sound* and slow decay of the noise (because of reduction in cough flow due to glottal closure), and (iii) *transient* forms of cough [1]. These sounds can provide enough information to a medical doctor, so as to distinguish between wet and dry cough, and ailment cough vs. voluntary cough etc. Hence, the ability to characterize the cough sounds of unhealthy speaker may be helpful in assisted diagnosis of the ailment. Few sporadic studies have been carried out towards understanding the cough sounds. About two decades back, the cough sounds were described for asthma patients [2]. In that study, mean frequency, standard deviation in it and other statistical measures of asthmatic cough sounds were used for the acoustic description of asthmatic cough sounds [2]. Automated detection of cough sounds using Hidden Markov Model was attempted earlier [3]. Acoustic analysis of cough sounds was attempted few years back [4]. In our earlier work, a preliminary analysis of *ailment cough* sounds and *simulated cough* sounds was carried out in [5], using a cough sound database, named IIITS CSSD, especially collected for the study [6]. However, the production characteristics of cough sounds need to be examined in detail, in order to aim for the automated classification of cough sounds signals.

This paper examines the production characteristics of *cold cough* with respect to *simulated cough*. A cough speech sound database IIITS CSSD has been used in this paper [6]. Standard signal processing tools are used to extract the discriminating features that characterize the cough sounds. Feature of excitation source (fundamental frequency i.e., F0), and of human vocal tract system (MFCC), along with a combined feature of both source and feter (i.e., signal energy) are used for characterization and classification of *ailment (cold) cough* sounds and *healthy* (*simulated*) *cough* sounds. The aim of this study is to classify the *ailment cough* sounds and *simulated,* i.e., healthy cough sounds, based on the feature vectors consisting of these 3 features. A machine learning approach is proposed for *real-time classification* of cough sounds. First, a tandem HMM approach combines the sequence modelling capabilities of the HMM with the high-accuracy context-dependent discriminative capabilities of an artificial neural network. The model is trained using the minimum cross entropy criterion. Next an SVM–GMM vector approach uses the noise adaptive kernels for better approximating the KL divergence between distributions of features in different voice segments. First, this study is carried out for *ailment* cough vs *simulated* cough. The encouraging results also highlight the need of collection of sufficient data of cough sounds of various diseases, such as throat and pulmonary diseases, for the extended study in future.

This paper is organized as follows. Section II discusses the details of the cough speech sound database used, i.e., IIITS CSSD. Section III discusses the signal processing methods and features extracted using those methods. Classification techniques and the classifier used are discussed in Section IV. Observations and performance results of the classifier are

978-1-5090-2597-8/16/$31.00 ©2016 IEEE

Table I: *Cough Database Summary*

1	Total# of Speakers	59
2	Total# of *Cough* Sound Samples	275
3(a)	Total# of *Ailment (cold) Cough* Samples	123
3(b)	Total# of *Ailment* (other than cold) Cough sound Samples	35
3(c)	Total# *Simulated* (*Healthy*) *Cough* Samples	117
4	Total# of *Normal Speech* Samples	23

discussed in Section V. Section VI gives the summary and scope of further work on the analysis of the cough sounds.

II. DATABASE USED FOR CLASSIFICATION

In this paper, a cough speech sound database has been used. The IIITS CSSD contains 117 *simulated cough* sound samples, 123 *cold cough* sound samples, and 35 other ailment cough sounds such as *asthma*, *physical disorder* etc., along with 20 *normal* speech samples. Thus IIITS CSSD contains total 295 sounds of cough and normal speech. Data collected includes, a sentence spoken by the person in *normal speech*, along with his *cough sounds*. *Normal speech* was recorded as reference speech of the speaker. Table I gives the details of the cough sounds database. In few cases, the noise or overlapping speech recorded with the cough sound samples, was removed using 'Wave surfer' tool, while preprocessing those signals in the database. Thus our database contains total 240 *ailment* (cold) cough vs. *simulated* cough sounds, including 123 ailment (cold) cough and 117 simulated cough sound. Out of 240, total 120 randomly selected cough sound samples were used for training, and 60 sounds were used for validation. Remaining 60 cough sound samples were used for testing. Hence, all the three sets i.e., training, validation and testing sets are disjoint sets.

The cough data was down-sampled to a sampling frequency of 8 kHz, with 8 bit/sample coding, before use. Then cough sound signals were preprocessed by passing through a first order FIR filter. Output of the FIR filter is given by:

$$x[n] = s[n] - s[n-1] \quad (1)$$

where, *s[n]* is input to the filter and *x[n]* is preprocessed signal. The signal *x[n]* is the *first order difference of s[n]*.

The human speech production system of speech signals are non-stationary for long duration but can be assumed as stationary for short duration (10-30 ms). Hence, the cough sound data is used as frames. Each cough sound file is divided into short time segments i.e., signal frames, which usually overlap with each other. In this analysis, each cough sample file is divided into 10 ms frames, with frame shift of 3 ms. The short-time signal energy ($\Sigma x^2(n)$) is computed for each frame, and energy threshold based unvoiced/silence frame removal is carried out for each cough sound sample. Then feature extraction is carried out on each signal frame using the speech signal processing methods.

III. FEATURES AND SIGNAL PROCESSING METHODS

Features such as instantaneous fundamental frequency (F0), MFCC's, signal energy (E) and first four formants (F1, F2, F3 and F4) are used as feature vectors. The signal preprocessing methods for extracting these features are discussed in this section, along with the brief discussion on the features extracted from the acoustic signals of cough sound and also from normal speech, for comparison purpose.

A. *Instantaneous Fundamental Frequency (F0)*

In this paper, the *instantaneous fundamental frequency* is obtained using three different signal processing methods so as to validate the signal processing methods. First, autocorrelation analysis is performed on short segment of cough sound, and the interval (T0) between two successive peaks is obtained using peak detection algorithm. Second, *Linear Prediction (LP)* analysis is carried out on short segments of cough sounds, to obtain LP residual of each frame, from which the interval T0 (=1/f0) between two peaks is calculated, in order to get fundamental frequency of the cough sounds. Third, *autocorrelation analysis of LP residual is carried out* to obtain the F0 [5]. All these three methods are discussed in details in [5]. Fig 1 and Fig 2 shows an illustration of the *instantaneous fundamental frequency (F0)* of *simulated* cough and *ailment* cough sounds, obtained using all three methods.

A. MFCC

The *Mel frequency cepstral (MFC)* represents the *short term* power spectrum of a signal, on Mel scale of the frequency. Hence, *MFCCs* are the coefficients of *MEL*-frequency spectrum. Process of *MFCC* calculation consists of three stages. In first stage, *hamming* window is applied on a pre-processed signal, obtained in section III. Then DFT of each frame is calculated using:

$$X[k] = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-j\frac{2\pi}{N}nk} \quad (2)$$

where, *x[n]* is the discrete signal in the time-domain, obtained after preprocessing as discussed in section II, and *w[n]* is the applied window (in time domain). Here *m* is *discrete* and *w* is *continuous* variable. *N* represents number of samples in each frame and *m* is frameshift. Since, frequency range in DFT spectrum is wide and voice signal does not follow the linear scale, the Mel filter bank is applied on the DFT of cough sound signal. Using the Triangular filters, with unity magnitude, the weighted sum of the spectral components is computed next. Output of the each such filter is the sum of filtered spectral components. For a given frequency (in Hz), the Mel-frequency is computed using:

$$F(Mel) = 2595\log_{10}[1 + \frac{f}{700}] \quad (3)$$

Finally in stage 3, DCT is applied on the output of the Mel filter bank using eq (2), in order to get the Mel-frequency cepstral coefficient, i.e., MFCC [7].

C. Signal Energy

Signal frame energy is used as a combined feature of the human vocal tract system and the excitation source. Signal energy for a signal frame can be obtained as:

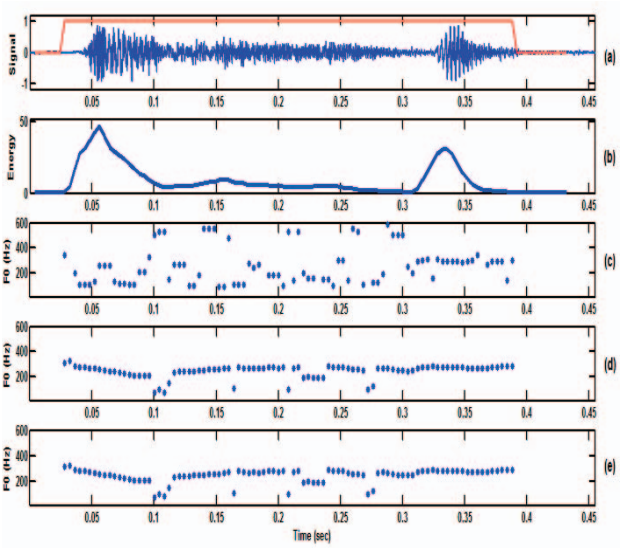

Fig1. Illustration for *simulated cough* sound (a) *signal waveform, (b) energy contour, and F0 contours using (c) LP residual (d) autocorrelation of signal and (e) autocorrelation of LP residual.*

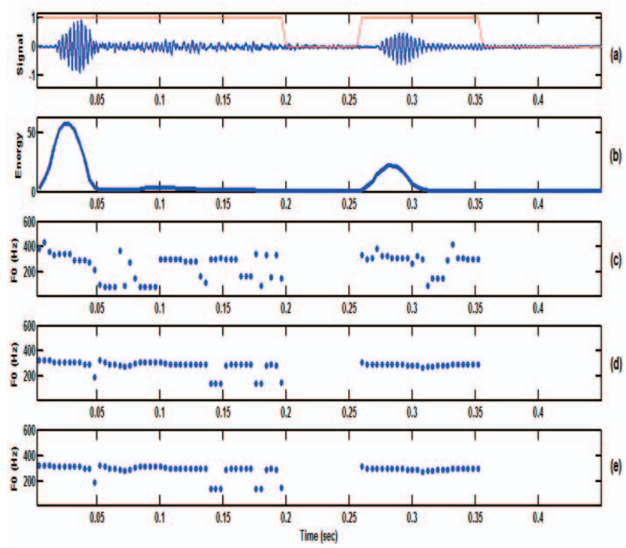

Fig2. Illustration of *ailment cough* sound's (a) *signal waveform, (b) energy contour, and F0 contours using (c) LP residual (d) autocorrelation of signal and (e) autocorrelation of LP residual.*

$$E(i) = \sum_{n=0}^{N-1} x_i^{\,2}[n] \tag{4}$$

where, x_i and *E(i)* represents the signal and signal energy of i^{th} frame, respectively. Energy contour of *simulated* and ailment (cold) cough has been illustrated in Fig 1(b) and 2(b), respectively.

D. Formants

In general, the speech signal can be modeled as a convolution of the excitation sequence and the impulse response of the vocal tract system, that is time-varying. But for the short-time duration of 10-40 ms, it can be considered as stationary system, for carrying out *linear prediction (LP)* analysis. In *z domain*, the vocal tract system can be modelled as *all pole model* [9] it can be expressed as:

$$H(z) = \frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \tag{5}$$

Here, *A(z)* is reciprocal of *H(z)* and is called *Inverse Filter* , that is used for obtaining the LP residual of the speech signal. Formants can be obtained using the LP spectrum of the speech signal. Peaks in the LP spectrum correspond to formants F1, F2, F3 and F4, respectively [9, 10].

IV. PROPOSED METHOD OF CLASSIFICATION

Support vector machine (SVM) is used for classification of *simulated* cough and *ailment* cough sounds. This section discusses about the *SVM* classifier and the parameters used, in detail. *SVM* is an efficient supervised non-parametric machine learning classifier, which gives very good results for classifying the linearly separable data. If the data is non-linear, then data can be projected onto higher dimensions, in order to make that linearly separable. Various kinds of kernels can be used, for projecting the data into higher dimensional space, to make the data linearly separable. Mainly polynomial, exponential, and *Radial basis functions (RBFs)* kernels are used for classifying the non-linear data. In this paper, we have used *Radial Basis Function (RBF) kernel*. RBF kernel nonlinearly maps the samples into a higher dimensional space, when the relation between class labels and attributes is nonlinear [8]. In this paper, *Gaussian radial basis functions* is used, which can be given by the expression:

$$K(x_i, x_j) = e^{-\gamma\left\|x_i - x_j\right\|^2} \tag{6}$$

Separation between the two hyperplanes can be given by:

$$\begin{aligned} W^T X_i + b_0 &= 1 \\ W^T X_i + b_0 &= -1 \end{aligned} \tag{7}$$

where *W* is, *weight vector* and b_0 is a real Constant. Since *training set* contains finite number of samples, the classification equations should satisfy:

$$\begin{aligned} W^T X_i + b_0 &> 1, \quad \forall\, i \ and \ Y_i = +1 \\ W^T X_i + b_0 &< -1, \quad \forall\, i \ and \ Y_i = -1 \end{aligned} \tag{8}$$

Where $Y_i = 1$ indicates that X_i belongs to class I, and $Y_i = -1$ indicates that X_i belongs to class II. Then equation (8) can be converted to a binary linear-class separation problem, given by:

$$Y_i[W^T X_i + b_0] > 1, for\ \forall\, i \tag{9}$$

The *separation* between two hyperplanes is given by:

$$m = \frac{2}{\|W\|} \tag{10}$$

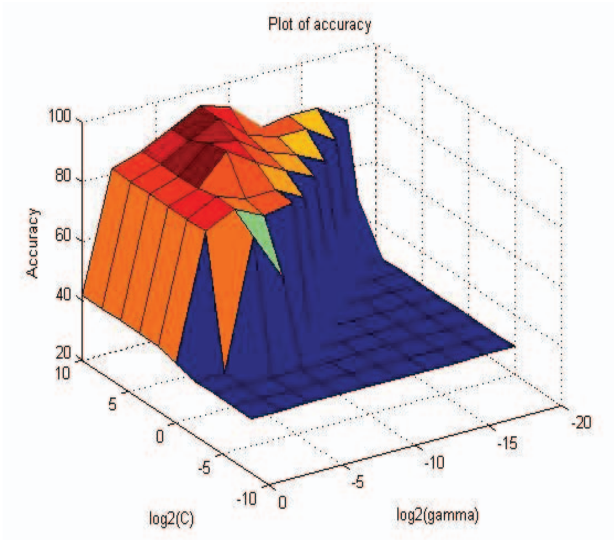

Fig3. Accuracy vs. parameters (log $_2$C and log$_2$ γ) plot for SVM. Here, γ *is a RBF kernel constant and C* indicates the misclassification penalty.

where, *m* is the *margin of* class separation *boundary*. Maximum separation gives *best classification efficiency*. This requirement can be achieved by solving the constraint optimization problem given as:

$$\text{Min} \quad \frac{1}{2} W^T W$$

subject to $Y_i[W^T X_i + b_0] > 1, \ for \ \forall i$ (11)

where W is weight vector and b0 is a real constant. Then Lagrangian technique can be used for solving eq (11) and obtaining the optimal value of W and b0. *Grid search algorithm* is used for finding the optimal value of RBF kernel parameters, namely cost C and γ [8,11]. Here, γ is RBF kernel *constant* and C is *penalty* of misclassification.

V. PERFORMANCE EVALUATION

Effectiveness of the classification (discussed in section IV) using the feature vectors (discussed in section III), is evaluated using two *performance evaluation measures*. The (i) *confusion matrix* and (ii) *efficiency vs parameter* plot are used for assessing the two class classification problem, i.e., classifying the acoustic signals of cough sound due to *ailment (cold)* vs. *simulated cough*. *Confusion matrix* of the classification is given in Table II.

The effectiveness of the binary classification between *aliment* vs. *simulated* cough is also evaluated, using the accuracy plot vs. parameters γ and C, A 3-D plot can be obtained by plotting the *accuracy* plot vs. γ and C. Here γ is the RBF kernel constant and C is the penalty of misclassification. Accuracy vs parameters (γ, C) plot obtained for the SVM classification of simulated vs cold cough sound is shown in Fig. 3. Here, Ref: represents *actual class* of the signal, Pred: represents *predicted class* of signal and *CI*: represents *confidence interval*. Then *efficiency* of classification for SVM classifier can be computed using:

$$Eff. = \frac{Total\,no.Of\ correctly classified samples}{Total\,no.of\ test samples} \tag{12}$$

Table II: *Confusion Matrix of the Classification*

95 % CI: (0.8608, 0.9896)		
	1: *Simulated Cough*	0: *Ailment Cough*
Ref / *Pred*	0	1
0	23	3
1	0	34

Hence, the efficiency of the SVM classifier, used here, is

$$Eff. = \frac{57}{60} \times 100 = 95\% \tag{13}$$

VI. SUMMARY AND CONCLUSIONS

In this paper, three types of features, the instantaneous fundamental frequency F0, signal energy (E), and first four Formants (F1, F2, F3 and F4) are used as constituents of the feature vectors. The changes in these features are observed for *cold cough* with respect to those for *simulated cough*, which is used here as a reference. Different signal processing methods and tools are used for extraction of these features, from the cough sounds signals. Instantaneous fundamental frequency is obtained using three different signal processing methods, for validation of the F0 obtained. Support vector machine (classifier) is used for the classification of the *ailment* vs *simulated* cough. Various methods are used for parameter tuning of the classifier, in order to improve the efficiency of the classification. *Grid search algorithm* is used, in order to obtain the optimal parameters of the SVM classifier.

The main aim of the study is to classify the *ailment* cough vs *healthy* cough, and to compare the changes in the features of the excitation source and the vocal tract system. The overall classification efficiency of the SVM classifier achieved is 95%. Classification efficiency is around 90%, when only the MFCC features are used as a feature vector. This efficiency is improved by 5% when MFCC is concatenated with instantaneous fundamental frequency, signal energy and formants features. It indicates that the classification using machine learning tools give better results, when all the speech production characteristics, i.e., source and system are used for the feature vectors. The study can be helpful for other similar works using SVM classifier, for binary classification of (i.e. linear/nonlinear) separation of the data for diverse future applications.

REFERENCES

[1] William Thorpe, Miranda Kurver, Greg King and Cheryl Salome, "Acoustic analysis of cough," *Seventh Australian and New Zealand Intelligent Information Systems Conference*, pp. 391–394, Nov. 2001.

[2] C.W. Thorpe, L.J. Toop and K.P. Dawson, "Towards a quantitative description of asthmatic cough sounds," *Europian Respiratory Journal*, pp. 685- 692, 1992.

[3] Sergio Matos, Surinder S. Birring, Ian D. Pavord and David H. Evans, "Detection of Cough Signals in Continuous Audio Recordings Using Hidden Markov Models," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1078–1083, June 2006.

[4] Xiaodan Zhuang, Xi Zhou, Mark A. Hasegawa-Johnson and Thomas S. Huang, "Real-world acoustic event detection," *Pattern Recognition Letters* 31 (2010) 1543–1551.

[5] Vishwanath Pratap Singh, J.M.S Rohith, Vinay Kumar Mittal, "Preliminary Analysis of Cough Sounds," *IEEE 12th International Conference (E3-C3) INDICON*, New Delhi, 17-20 Dec 2015.

[6] Vishwanath Pratap Singh, JMS Rohith, Yash Mittal, Vinay Kumar Mittal, "IIITS CSSD: A Cough Speech Sound Database," 22nd National Conference on Communication 2016 (NCC 2016), 4-6 March 2016.

[7] Lawrence Rabiner, Biing H. Juang and B. Yegnanarayana, "*Fundamentals of Speech Recognition*," chapter third, pp. 88–113. Pearson Education *Inc.* , New Delhi, India, Indian subcontinent adaptation, first edition, 2009.

[8] Richard O. Duda, Peter E. Hart and David G. Stork, "*Pattern Recognition*," Chap 3-9, pp. 100-270, 2nd Edition, *Wiley*, October 2000.

[9] John Makhoul, "linear Prediction: A Tutorial Review," *Proceedings of The IEEE*, vol. 63, no. 4, Apr 1975.

[10] B.Yegnanarayana, "Pole-Zero Decomposition of Speech Spectra," *Technical Report*, *CMU-CS-79-101*, 1979.

[11] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, "A Practical Guide to Support Vector Classification," *Technical Report*, Department of Computer Science, National Taiwan University, 2003.

[12] Vinay Kumar Mittal and Bayya Yegnanarayana, "Analysis of production characteristics of laughter", *Computer Speech and Language*. vol. 2015, no. 30, pp. 99-115, Nov 2014.