35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Neural network based algorithm for automatic identification of cough sounds

Swarnkar V, Abeyratne U R, *Senior Member IEEE, Yusuf Amrulloh,* Craig Hukins, Rina Triasih, Amalia Setyati

***Abstract*—Cough is the most common symptom of the several respiratory diseases containing diagnostic information. It is the best suitable candidate to develop a simplified screening technique for the management of respiratory diseases in timely manner, both in developing and developed countries, particularly in remote areas where medical facilities are limited. However, major issue hindering the development is the non-availability of reliable technique to automatically identify cough events. Medical practitioners still rely on manual counting, which is laborious and time consuming. In this paper we propose a novel method, based on the neural network to automatically identify cough segments, discarding other sounds such a speech, ambient noise etc. We achieved the accuracy of 98% in classifying 13395 segments into two classes, 'cough' and 'other sounds', with the sensitivity of 93.44% and specificity of 94.52%. Our preliminary results indicate that method can develop into a real-time cough identification technique in continuous cough monitoring systems.**

I. Introduction

Cough is a natural protective mechanism. It helps clearing the secretions from the respiratory tract and prevents entering of noxious particles into the respiratory system. It is generally defined as the sudden expulsion of air accompanied with typical sound[1]. Chronic cough is commonest symptom of many respiratory diseases[2], including pneumonia, asthma, bronchitis etc. The prevalence of cough in Europe and USA varies between 9 – 33%[3]. This number worsens in under-developed and developing countries. In 2004, an estimated 17% of children (age under 5), died of pneumonia alone [4] and 98% of those were from under developed and developing[5] countries.

Cough is often present as earliest symptom in almost all the respiratory diseases. As such it can be a useful tool to develop a new approach for respiratory disease screening. A new technology which is cost effective, easily accessible and can be used for community screening, so that the patients with respiratory diseases can be identified and treated in timely manner.

In the past several researchers [6, 7] have recorded and analyzed cough sounds. However, most common problem hindering the analysis is the unavailability of a reliable method to automatically identify cough sounds from other biological sounds, such as speech, laugh and other background noises. Often analysis relies on age old method of manually listening and scoring cough sounds. This not only brings subjectivity but is also laborious and time consuming job.

In order to overcome the problems of manual scoring several researchers have proposed automatic cough detection system. Based on acoustic features from the cough sounds (such as Linear Predictive Coding (LPC) and Cepstral coefficients) and training probabilistic neural network, a Hull Automatic Cough Counter (HACC) [8] was developed. HACC achieved the sensitivity of 80% and specificity of 96% in cough segmentation. The drawback of this system is that it requires a reference cough and non-cough signals for detection[9]. Similar to this, Matos et al [7] using the Cepstral coefficients and spectral features developed a Hidden Markov Model for cough classification. They achieved a median sensitivity and specificity of 85.7% and 94.7% respectively. The limitation of this system is its performance depends on the energy level of the signal [9].

Another automated cough counter system which claims to achieve the sensitivity and specificity in the range of 98% is VitaloJAK system[9]. However critical drawback of this system is that it uses contact microphones placed on the chest wall for sound recording. In addition to this VitaloJAK algorithm depends on the explosive phase of the cough for its detection. This limits its usefulness only for quantitative analysis.

Despite inherit subjective nature, and tedious job, human scoring of cough is still considered the golden standard. To address this issue, in this paper we propose a new method for automatic recognition of cough events from the continuous sound recordings. The method is based on the computation of characteristic features from sound segments and then training a neural network to classify them into two classes 'Cough (CG)' or 'Other sounds (OS)'. The proposed method is free from sound intensity and essentially depends on the temporal variation of acoustic features within a given time frame. In next section we describe our method in detail.

This work has been funded by the grant from Bill & Mellinda Gates Foundation, awarded to Dr. Abeyratne.

II. METHOD

A. Clinical Data Acquisition and recording system.

The clinical data acquisition environment for this work is Respiratory Medicine Unit of The Prince Alexandra Hospital, Australia and Sardjito Hospital, GadjahMadaUniversity, Indonesia. Our subject population includes individuals with infective lung disease defined as presence of cough, sputum and increased breathlessness. Table 1 lists the inclusion and exclusion criteria. All patients fulfilling the inclusion criteria were approached. An informed consent was made. Patients were recruited within first 12 hours of their admission. After the informed consent continuous sound recordings were made for next 4-6 h.

TABLE 1
INCLUSION AND EXCLUSION CRITERIA USED IN THE STUDY

Inclusion Criteria	**Exclusion Criteria**
- Patients with symptoms of chest infection : At least 2 of - Cough - Sputum - Increased breathlessness - Temperature >37.5° - Consent	- Advanced disease where recovery is not expected eg terminal lung cancer - Droplet precautions - NIV required - No Consent

The recordings were made in the general adult ward of the hospital. Patient shared the room with 5 other patients separated by curtains. The patients were accompanied by their family members. The attending physician regularly visited the patient, however no confidential information related to the patient was recorded. The common noise present in the recordings were from celling fan, foot step, speech, door banging, trolley movement and other ambiguity noises from outside the room.

The cough sounds were acquired with a high fidelity, computerized data acquisition system. A matched pair of low-noise microphones having a hypercardioid beam pattern (Model NT3, RODE, Sydney, Australia) were used to capture the sound signals. The nominal distance from the microphone to the mouth of the patient was 50 cm, but could vary from 40 cm to 70 cm due to patient movements. A professional quality pre-amplifier and A/D converter unit (Model Mobile-Pre USB, M-Audio, California, USA) was used for sound signal acquisition. We used a sampling rate of 44.1 kHz with a 16 bit resolution. The system had a 3 dB bandwidth of 20800 Hz at the sampling rate used.

B. Feature Extraction

[1]. Let $x[k]$ denotes the k^{th} sample of the discrete time sound signal. In this paper we model $x[k]$ as :

$$x[k] = y[k] + b[k] \tag{1}$$

Where y[k] represents the cough sound signal and $b[k]$ is the background noise including, breathing, speech, silence, laughing and other duvets noises. Our target in this paper is to identify $y[k]$ segments of the sounds.

[2]. Segment x[k] into blocks of size t=100 ms. Let $x^i[k]$ represents the i^{th} segment of $x[k]$.

[3]. In the speech processing pre-emphasis filter is commonly used to boost the higher-frequencies and to correct the -6 dB/octave in the spectrum of the sound signal. Hence, this filter given by Eq. (2) was used at the pre-processing stage of the sound data.

$$x^i[k] = x^i[k] - \alpha x^i[k-1] \tag{2}$$

Where

$$\alpha = e^{\left(-2\times3.14\times\frac{F}{Fs}\right)} \tag{3}$$

The cut-off of the pre-emphasis filter can be set by altering F in eq. 3. In speech processing application typically α is set to 0.96, therefore this value was used in this application.

[4]. To compute the features we sub-segmented $x^i[k]$ into 'j' sub-segments each of $t1$ = 10ms. Let $x^{ij}[k]$ represents the j^{th} sub-segment of the i^{th} segment. Size of $t1$=10ms will give us j=10 sub-segments. For each filtered sub-segment $x^{ij}[k]$ we computed following features

(a) Non-Gaussianity Score (NGS) – NGS gives the measure of non-gaussianity of a given segment of data. The normal probability plot can be utilized to obtain a visual measure of the gaussianity of a set of data. The NGS of the data segment $x^{ij}[k]$ can be calculated using eq. 4. Note that in eq. 4, $g(k)$ and y(k) represents the probabilities and $\bar{g}$ and $\bar{y}$ represents the mean of the reference normal data and the analyzed data, respectively. Detail method of computing NGS can be found in [10].

$$NGS = 1 - \left(\frac{\sum_{k=1}^{N} g[k] - \bar{g}}{\sum_{k=1}^{N} y[k] - \bar{y}}\right) \tag{4}$$

(b) Formants frequencies – In speech analysis formants are referred as the resonance of the human vocal tract. They are manifested as the amplitude peak in the LPC spectrum of the acoustic signal. We included the 1st four formant frequencies (F1, F2, F3, F4) in our feature set. Past studies in the speech and acoustic analysis have shown that F1-F4 corresponds to various acoustic features of upper airway. We computed the F1-F4 by peak picking the LPC spectrum. For this work we used 14th order LPC spectrum and its parameters were determined via Yule-Walker autoregressive method along with the Levinson-Durbin recursive procedure[11].

(c) Log Energy(LogE) – The log energy for every sub-segment was computed using eq. 5

$$LogE = 10log_{10}\left(\varepsilon + \frac{1}{N}\sum_{k=1}^{N}(x^{ij}(k)^2)\right) \quad (5)$$

where ε in (%) is and arbitrarily small positive constant added to prevent any inadvertent computation of the logarithm of 0.

(d) Zero crossing (Zcr) – The number of zero crossings were counted for each sub-segment.

(e) Kurtosis (Kurt) – The kurtosis is a measure of the peakedness associated with a probability distribution of sub-segment $x^{ij}[k]$, computed using (6)

$$kurt = \frac{E(x^{ij}[k] - \mu)^4}{\sigma^4} \quad (6)$$

Where μ and σ is the mean and stand deviation of the sub-segment $x^{ij}[k]$ respectively.

(f) Mel-frequency cepstral coefficients (MFCC)- MFCCs are commonly used in the speech analysis systems[12]. They represent the short term power spectrum of an acoustic signal based on a cosine transform of a log power spectrum on a non-linear mel-scale of frequency. We included the 12 MFCC coefficients in our feature set.

(g) Bispectrumscore(Bscore) – The 3rd order spectrum of the signal is known as the bispectrum. The bispectrum of a segment $x^{ij}[k]$ can be estimated using (7)

$$B^{ij}(\omega1, \omega2) = \gamma . H(\omega1) . H(\omega2) . H^*(\omega1 + \omega2) \quad (7)$$

where H(.) is the 1-d Fourier transform and γ is the skewness of the sub-segment $x^{ij}[k]$. From the bispectrum of each sub-segment we computed the 1-d slice (inclined to the w1-axis by 45 degrees and passes through the origin) and arranged them consequently to form a 2-d matrix BSG(size of BSG is 10xF; 10 = number of sub-segments, F = fft length used). From this matrix we computed the B-score using (8)

$$Bscore = \frac{1}{F}\sum_{j=1}^{10}\sum_{f=1}^{F} BSG > \mu \quad (8)$$

where μ in (8) is mean of BSG.

[5]. From each 100 ms segment, a vector containing 201 features (120 from MFCC; 40 – Formant frequency; 10 each from NGS, LogE, Zcr, Kurt; and 1 from Bispectrum) was formed. This feature vector was then supplied to a simple pattern recognition neural network to classify it into either of the two classes, 'Cough (CG)' or 'Other Sound (OS)'.

III. Results

A. *Clinical database*

Sound database consisted of three patients. Table II gives the patients diagnostic details and duration of sound data recorded. For the analysis each audio recordings were listened to identify and individually label start and end of cough events and other sounds. These labels were used both for training and for testing the neural network. The overall data consists of 733 events (13395 100 ms segments); 342 Cough events (2546 segments) and 391 other sound events (10849 segments). From each patient all the cough sound events plus some randomly selected events containing only other sounds (speech, other ambient sounds) were selected.

TABLE II
Clinical Database.
CG - Cough Segments; OS - Other Sounds

Sub	Age	Sex	Diagnosis	Rec. Duration (hr:min)	Segments included in analysis (CG/OS)
1	48	M	Lung Infection	6:38	1309/3270
2	68	M	COPD with secondary pulmonary infection	2:05	854/3279
3	75	F	Pulmonary oedema	4:15	383/4330

B. *Network structure & identification results*

Neural network was consisted of a feed forward network with 1 input layer of 201 neurons, 4 hidden layers each of 50, 20, 10 and 5 neurons and 1 output layer of 1 neuron. Sigmoid transfer function was used in all the layers. Network was trained using Levenberg-Marquardt back-propagation algorithm. We used 60% of segments (8037) for training, 10% (1339) to validate and 30% (4019) to test the network.

Figure 2 shows the results of the neural network based identification of cough events in four different types of continuous test data segments of short duration. According to the fig.2, trained neural network identified all the cough events in the 3 segments (1, 2 & 4) with only two false positives. In segment 3, where only ambient noise was present Network performed with 100% accuracy. Table 3 shows the contingency table, comparing the automatic classification of cough segments using neural network with that of manual segmentation. Overall we achieved an accuracy of 98.24% in classifying 13395, 100 ms segments into 'cough' and 'other sounds' with sensitivity of 93.44% and specificity of 95.52%.

TABLE III
CONTINGENCY TABLE

		Test Method		
		cough	**Other sounds**	
Manual Segmentation	**cough**	2358	188	**2546**
	Other Sounds	48	10801	**10849**
		2406	**10981**	**13395**

IV. CONCLUSION

In this paper we presented a novel neural network based technique to automatically identify the cough segments from the continuous sound recording. We computed different mathematical features, generally used in speech processing to form a large feature vector. We trained a feed-forward neural network to classify this feature vector either into cough or other sound. We achieved an accuracy of 98% in classifying 13395 segments from 3 patients. Proposed method can be easily deployed to automatically detect cough sounds from continuous recording, aiding in the assessment of patients with respiratory infection by giving statistics such as cough frequency and mean duration. However, method has to be tested on the larger and prospective database before it can be used in practical application.

REFERENCES

[1] Korpá, J., J. Sadloová, and M. Vrabec, *Analysis of the cough sound: an overview*. Pulmonary Pharmacology, 1996. **9**(5-6): p. 261-268.

[2] McGarvey, L., *Idiopathic cough, prevalence and underlying mechanisms*. Pulmonary Pharmacology & Therapeutics, 2004. **17**(6): p. 435-439.

[3] Chung, K.F. and I.D. Pavord, *Prevalence, pathogenesis, and causes of chronic cough*. The Lancet, 2008. **371**(9621): p. 1364-1374.

[4] *The Global Burden of Disease 2004 Update*. World Health Organization, 2008.

[5] *UNICEF annual report 2009*. http://www.unicef.org/publications/files/UNICEF_Annual_Report_2009_EN_061510.pdf, 2009.

[6] Ayman, A., et al., *Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function*. Cough, 2009. **5**.

[7] Matos, S., et al., *Detection of cough signals in continuous audio recordings using hidden Markov models*. Biomedical Engineering, IEEE Transactions on, 2006. **53**(6): p. 1078-1083.

[8] Barry, S.J., et al., *The automatic recognition and counting of cough*. Cough, 2006. **2**(1): p. 8.

[9] Smith, J.A., *Cough: Assessment and Equipment*. http://dev.ersnet.org/uploads/Document/bb/WEB_CHEMIN_2559_1194522793.pdf.

[10] Ghaemmaghami, H., U. Abeyratne, and C. Hukins. *Normal probability testing of snore signals for diagnosis of obstructive sleep apnea*. 2009: IEEE.

[11] Oppenheim, A.V., R.W. Schafer, and J.R. Buck, *Discrete-time signal processing*. Vol. 1999. 1989: Prentice hall Englewood Cliffs, NJ:.

[12] Zheng, F., G. Zhang, and Z. Song, *Comparison of different implementations of MFCC*. Journal of Computer Science and Technology, 2001. **16**(6): p. 582-589.

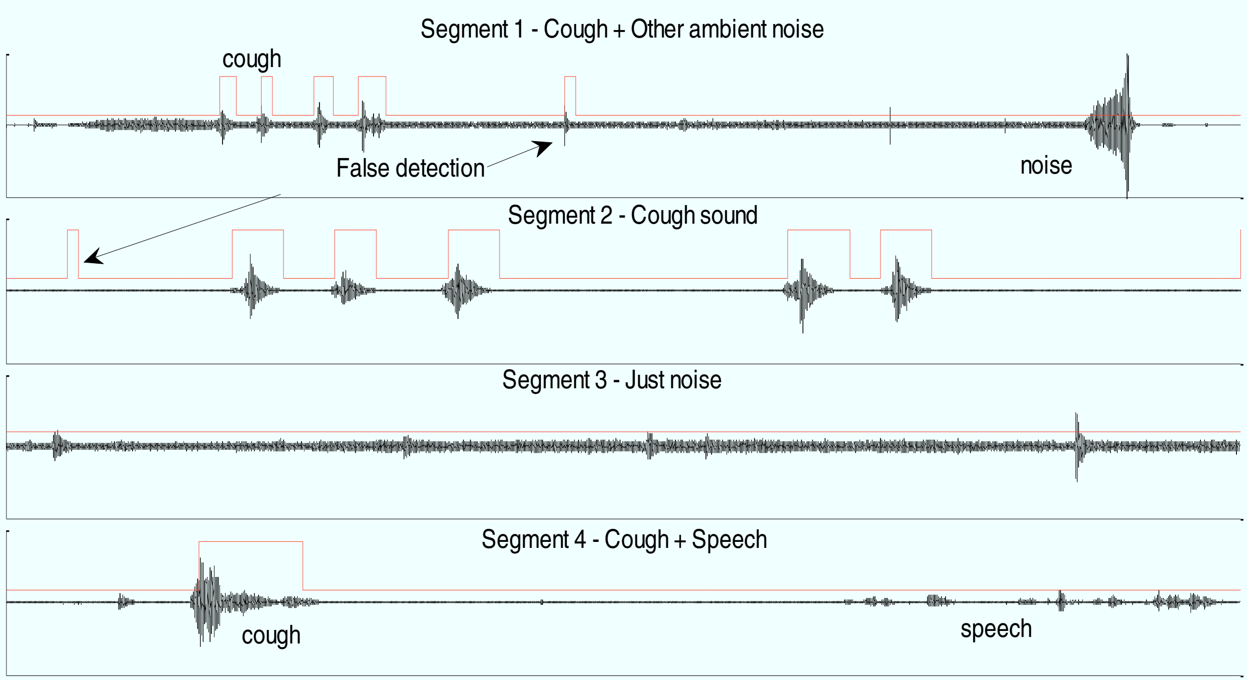

Fig 2. Neural Network based identification of cough events in continuous sound segments of short duration. Trained network identified portion of cough events with very high accuracy while discarding other sounds.