Cough Sound Analysis for Diagnosing Croup in Pediatric Patients Using Biologically Inspired Features*

Roneel V. Sharan, *Member, IEEE*, Udantha R. Abeyratne, *Senior Member, IEEE*, Vinayak R. Swarnkar, and Paul Porter

***Abstract*— This paper aims to diagnose croup in children using cough sound signal classification. It proposes the use of a time-frequency image-based feature, referred as the cochleagram image feature (CIF). Unlike the conventional spectrogram image, the cochleagram utilizes a gammatone filter which models the frequency selectivity property of the human cochlea. This helps reveal more spectral information in the time-frequency image making it more useful for feature extraction. The cochleagram image is then divided into blocks and central moments are extracted as features. Classification is performed using logistic regression model (LRM) and support vector machine (SVM) on a comprehensive real-world cough sound signal database containing 364 patients with various clinically diagnosed respiratory tract infections divided into croup and non-croup. The best results, sensitivity of 88.37% and specificity of 91.59%, are achieved using SVM classification on a combined feature set of CIF and the conventional mel-frequency cepstral coefficients (MFCCs).**

I. Introduction

Croup is a viral infection of the respiratory tract and is common in children. A two-year Australian study in children aged 0–14 years suggests that croup is managed in the general practice about 154,000 times per year and is most prevalent in children aged 1–4 years [1]. Similarly, as summarized in [2], croup affects more than 80,000 children in Canada each year, the second most common cause of respiratory distress in the first 10 years of life.

The infection caused by croup results in an inflammation of the upper airway restricting normal breathing. This results in a "barking" or "croupy" cough usually accompanied by stridor, hoarse voice, and respiratory distress due to airway obstruction [3]. Croup is typically diagnosed in clinical practice relying on this distinctive cough as the primary clinical feature. Physicians listen to cough and make a subjective judgment on the 'croupiness' or 'barkingness' of events. In this paper we propose the use of automated cough sound analysis to diagnose croup.

Our work is inspired from automatic speech recognition (ASR) technology; in particular, we explore the performance of mathematical features inspired by human auditory system. Speech and cough share some similarities in the generation process and physiological wetware used. We propose the use of a feature extraction method that utilizes the frequency selectivity of the human cochlea, applied to time-frequency image of cough sound signals.

One very prominent feature in ASR is mel-frequency cepstral coefficients (MFCC) [4]. Introduced more than three decades ago, the success of MFCC could be attributed to its ability to adequately capture the perceptually relevant aspects of the short-term power spectrum of a speech signal. In our earlier works [5, 6] we illustrated the usefulness of MFCCs (in combination with other features) in diagnosing diseases such as pneumonia.

Most of the dominant frequency components of the cough sound signal lie in the low frequency range. The conventional spectrogram image has evenly spaced frequency components with constant bandwidth which results in suppression of spectral information in the lower frequency range. In this paper, we propose a bio-inspired gammatone filter which offers more frequency components in the lower frequency range with narrow bandwidth and less frequency components in the higher frequency range with wide bandwidth, revealing more spectral information in the time-frequency image as a result [7]. The resulting time-frequency image is referred as a cochleagram and the resulting feature, which captures the statistical distribution, as the cochleagram image feature (CIF) [8].

In this work, we propose the use of CIF of cough sounds for croup diagnosis. Further, we compare the outcomes with those from MFCC analysis, and a combined MFCC-CIF analysis. One of the main targets is to explore how well the human cochlea inspired CIF feature can perform.

We also investigate two different classifier models: the logistic regression model (LRM) and a Support Vector Machine (SVM). The LRM, a well-known linear classifier, has been recently used in cough analysis [5, 6, 9]. We compare the results of LRM against those from support vector machines (SVMs), a nonlinear classifier widely used in speech recognition and sound event detection technology.

II. Feature Extraction

This section describes the feature extraction methods for MFCC and CIF with reference to Fig. 1.

A. MFCC

Firstly, the cough signal is divided into frames and DFT is applied to the windowed frames as

$$X(k) = \sum_{n=0}^{N-1} x(n) w(n) e^{\frac{-2\pi i k n}{N}}, \quad k = 0,1,...,N-1 \quad (1)$$

*Research supported by ResApp Health Limited and The University of Queensland.

Roneel V. Sharan, Udantha R. Abeyratne, and Vinayak R. Swarnkar are with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia. (e-mail: r.sharan@uq.edu.au, udantha@itee.uq.edu.au, vinayak@itee.uq.edu.au).

Paul Porter is a consultant pediatrician with the Princess Margaret Hospital and Joondalup Health Campus in Perth, Australia.

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

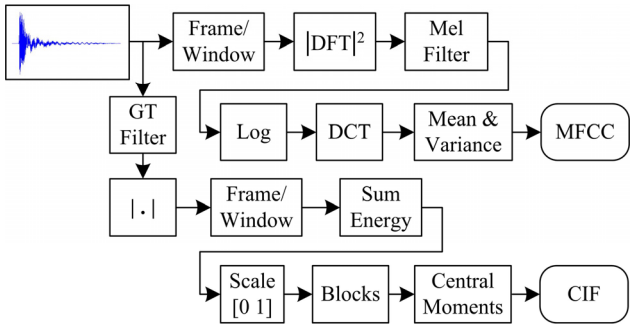

Figure 1. An overview of computing MFCC and CIF.

where N is the length of the window, $x(n)$ is the time domain signal, $w(n)$ is the window function, and $X(k)$ is the k^{th} harmonic corresponding to the frequency $f(k) = kF_s/N$, F_s is the sampling frequency.

MFCCs utilize mel-filter banks, or triangular bandpass filters, which are equally spaced on the mel-scale [10]. The adjacent filters overlap such that the lower and upper ends of the m^{th} filter are located at the center frequency of the $m-1$ and $m+1$ filter, respectively.

The MFCCs are obtained as the discrete cosine transform (DCT) of the log compressed filter bank energies given as

$$c(i) = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} \log\left(E(m)\right) \cos\left(\frac{\pi i}{M}(m-0.5)\right) \qquad (2)$$

where $i = 1, 2, \ldots, l$, l is the order of the cepstrum, $E(m)$ is the filter bank energies of the m^{th} filter, and M is the total number of mel-filters.

B. CIF

In this time-frequency representation, the signal is broken into different frequencies which are naturally selected by the cochlea and hair cells. This frequency selectivity is modeled by the gammatone filter which is a series of bandpass filters with impulse response [11]

$$h(t) = At^{j-1} e^{-2\pi Bt} \cos\left(2\pi f_c t + \phi\right) \qquad (3)$$

where A is the amplitude, j is the order of the filter, B is the bandwidth of the filter, f_c is the center frequency of the filter, ϕ is the phase, and t is the time.

The equivalent rectangular bandwidth (ERB) is used to describe the bandwidth of each cochlea filter in [11] given as

$$f_{c,ERB} = \left[\left(\frac{f_{c,Hz}}{Q_{ear}} \right)^p + \left(B_{min}\right)^p \right]^{1/p} \qquad (4)$$

where Q_{ear} is the asymptotic filter quality at high frequencies and B_{min} is the minimum bandwidth for low frequency channels. The bandwidth of a filter can then be approximated as $B = 1.019 \times f_{c,ERB}$. For this work, we only consider Greenwood's ERB model [12] which was shown to give the best classification performance in [8].

The human cochlea has thousands of hair cells which resonate at their characteristic frequency and at a certain

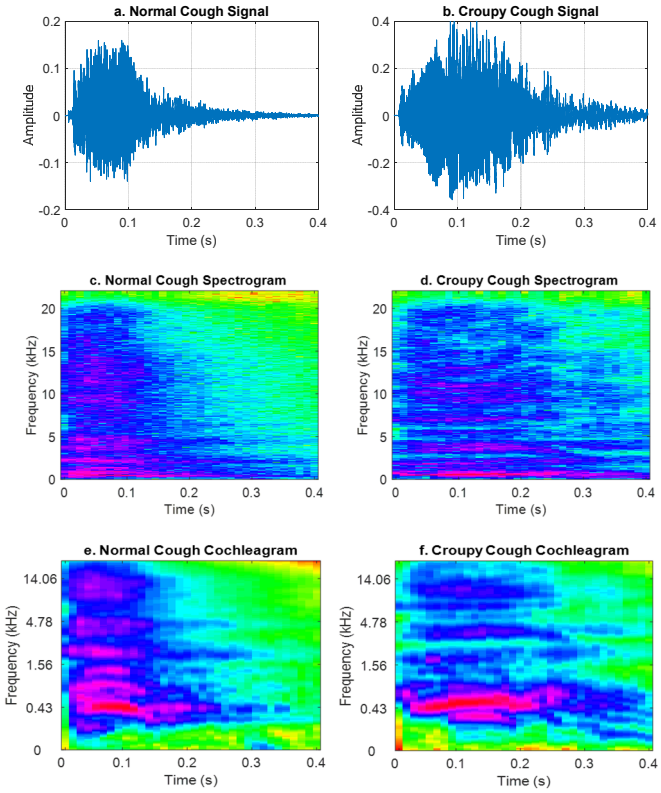

Figure 2. (a), (b): Time domain waveforms of a normal and a croupy cough (c), (d): their spectrograms and (e), (f): corresponding cochleagram.

bandwidth. The mapping between filter index and center frequency is determined as [13]

$$f_{cg} = -Q_{ear}B_{min} + \left(f_h + Q_{ear}B_{min}\right) e^{-gs/Q_{ear}} \qquad (5)$$

where $g = 1, 2, \ldots, G$, G is the number of gammatone filters, f_h is the maximum frequency in the filter bank, and s is the step factor given as

$$s = \frac{Q_{ear}}{G} \log\left(\frac{f_h + Q_{ear}B_{\min}}{f_l + Q_{ear}B_{\min}} \right) \qquad (6)$$

where f_l is the minimum frequency in the filter bank.

Similar to [8] we use a 4th order gammatone filter with four filter stages and each stage a 2nd order digital filter as given in [13]. The gammatone filter was implemented using the Auditory Toolbox for Matlab [14].

A representation similar to the conventional spectrogram image is obtained by smoothing the time series associated with each frequency channel in the gammatone filter and then adding the energy in the windowed signal for each frequency component. The log of the intensity values are then scaled in the range [0 1] for feature extraction. The time domain signal, spectrogram, and cochleagram of a normal and croup cough sound signal are given in Fig. 2. The dominant frequency component, centered around 400 Hz, is suppressed in the spectrogram image but revealed more succinctly in the cochleagram courtesy of more frequency components in the lower frequency range with narrower

bandwidth. The frequency range in both the representations is 0 to 22,050 Hz, which is the Nyquist frequency.

III. Classification

A. LRM

LRM is a regression model where the dependent variable is categorical, the probability of which is estimated using one or more independent variables or features. The dependent variable in this work are croup ($Y = 1$) and non-croup ($Y = 0$). For a given feature vector $\mathbf{F} = [f_1\ f_2\ ...\ f_f]$, the probability that the output is croup ($Y = 1$) can be estimated using the logistic function given as

$$P(Y = 1|_{\mathbf{F}}) = \frac{e^v}{e^v + 1} \tag{7}$$

where

$$v = \beta_0 + \beta_1 f_1 + \ldots \beta_f f_f \tag{8}$$

and $\beta_0, \beta_1, \ldots, \beta_f$ are the regression coefficients. A cough was determined to be croup if its probability was ≥ 0.5.

B. SVM

SVM determines the optimal hyperplane to maximize the distance between any two given classes. Consider a set of l training samples belonging to two classes as $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x}_i \in R^d$ is a d-dimensional feature vector representing the i^{th} training sample, and $y_i \in \{-1, +1\}$ is the class label of $\mathbf{x}_i$. The optimal hyperplane can be determined by minimizing ½$\|\mathbf{w}\|^2$ subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$, where $\mathbf{w} \in R^d$ is a normal vector to the hyperplane and b is a constant. The optimization is solved under the given constraints by the saddle point of the Lagrange functional.

For linearly nonseparable problems, the optimization can be generalized by introducing the concept of *soft margin* [15]. Nonlinear SVM is used in this work which maps the input vector $\mathbf{x}$ to a higher dimensional space $\mathbf{z}$ through some nonlinear mapping $\phi(\mathbf{x})$ chosen *a priori* to construct an optimal hyperplane. The *kernel trick* [16] is applied to create the nonlinear classifier where the dot product is replaced by a nonlinear kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ which computes the inner product of the vectors $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$. A commonly used kernel function is Gaussian radial basis function (RBF), $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\right)$, where $\sigma > 0$ is the width of the Gaussian function.

The classifier for a given kernel function with the optimal separating hyperplane is then given as

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{9}$$

with α_i being the Lagrange multipliers.

C. Patient Classification

After completing cough classification for each patient, the a patient is diagnosed as having croup if one or more coughs are classified as croup, that is,

$$P_D = \begin{cases} 1, & Q \geq 1 \\ 0, & Q = 0 \end{cases} \tag{10}$$

where Q is the total number of coughs classified as croup for a patient.

IV. Experimental Evaluation

A description of the cough sound database used in this work is given first followed by an overview of the experimental setup. Finally, the results using MFCC, CIF, and feature vector combination are presented with LRM and SVM classification methods.

A. Cough Sound Database

The cough sounds were recorded using a smartphone (iPhone) at the Princess Margaret Hospital (PMH) and Joondalup Health Campus (JHC) in Western Australia. The recordings were made in realistic environments. This means the recordings have normal background noise such as people talking, children noise/crying, medical instrument sounds, footsteps, door opening/closing, etc. All signals in the database have a sampling frequency of 44,100 Hz.

The cough sound database has a total of 364 pediatric patients belonging to two classes: croup (43 patients) and non-croup (321 patients). Multiple cough sound signals were manually segmented for each patient, up to 10 coughs per patient. The class croup has patients with croup only and croup plus upper respiratory tract infection (URTI). The non-croup class includes URTI, wheeze (asthma, bronchiolitis, and viral induced), and pneumonia (atypical, bacterial, and viral). All the respiratory tract infections have been clinically diagnosed by clinicians at PMH and JHC using Australian clinical guidelines.

B. Experimental Setup

For all experiments, signal processing is carried out using a Hamming window of 1024 points (23.22 ms) with 50% overlap between frames. The classification performance is measured using Sensitivity, Specificity, Accuracy, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Cohen's Kappa (κ). Except κ, all these values are reported in percentage (%).

All results are reported using leave-one-out-test. That is, all cough sound signals from a single patient are used for testing and all cough sound signals from all other patients are used for training the classifier, making the trained model independent of the test patient. This process is repeated for all patients resulting in the number of trained models equal to the number patients.

For MFCCs, experimentation was performed with different number of mel-filters in the range of 10-50. The best results were obtained at $M = 18$. As such, the feature vector for each frame is 54 dimensional which includes 18 cepstral coefficients plus the first and second derivatives [17]. The final feature vector is a concatenation of the mean and standard deviation values from each dimension resulting in a 108 dimensional final feature vector.

For the cochleagram image, to get the same image resolution along the frequency axis as the spectrogram image,

TABLE I. PATIENT CLASSIFICATION RESULTS USING LRM

	Sens	Spec	Acc	PPV	NPV	κ
MFCC	83.72	82.55	82.69	39.13	97.43	0.44
CIF	93.02	84.11	85.16	43.96	98.90	0.52
MFCC + CIF	88.37	75.70	77.20	32.76	97.98	0.37

TABLE II. PATIENT CLASSIFICATION RESULTS USING SVM

	Sens	Spec	Acc	PPV	NPV	κ
MFCC	81.40	91.59	90.38	56.45	97.35	0.61
CIF	86.05	91.28	90.66	56.92	97.99	0.63
MFCC + CIF	88.37	91.59	91.21	58.46	98.33	0.65

that is $N/2$, the number of gammatone filters, G, is set to 512. The cochleagram image is divided into blocks and second and third central moments are extracted as features in each block. These values are concatenated to form the final feature vector. Various block sizes were experimented with and the best results obtained at a block size of 8×4, along the vertical and horizontal, respectively. This results in a 64 dimensional final feature vector. For all features, each dimension is scaled in the range [0 1] for classification using LRM and SVM.

B. Experimental Results

The results for MFCC, CIF, and a combination of the two features using LRM and SVM classification methods are given in Table I and Table II, respectively.

For LRM classification, with MFCC, a sensitivity and specificity of 83.72% and 82.55% are achieved, respectively. This increases to 93.02% and 84.11% with CIF. While only a marginal improvement is observed in the specificity value from MFCC to CIF, the sensitivity value is almost 10% more. However, no improvement was observed when the two features are combined. This suggests that raw feature combination using LRM is not suitable, at least for the features considered here.

For SVM classification, with MFCC, the sensitivity decreases by 2.32% to 81.40% while the specificity increases by 9.04% to 91.59% when compared to LRM. Similarly, for CIF, the sensitivity decreases by 6.97% and specificity increases by 7.17% and for the combined feature vector, the sensitivity remains same at 88.37% but the specificity increases by 15.89% to 91.59%, which gives the best overall performance.

It should be noted here that a grid search was used in tuning the SVM RBF kernel parameters after which the best combination of sensitivity and specificity values were chosen. This means that while a greater sensitivity might have been possible, the specificity value would have been much lower. In any case, on average, the classification performance of the SVM classifier is seen to be better than the LRM classifier, particularly with raw feature vector combination.

V. CONCLUSION AND FUTURE WORK

A method for diagnosing croup in children using automatic cough sound recognition has been presented in this paper. The proposed feature for use in this work, the CIF, utilizes a gammatone filter which models the frequency selectivity property of the human cochlea. The combination of CIF with MFCC was shown to give the best overall performance using SVM classification.

It should be noted that augmenting the MFCC features with other features and clinical symptoms we could increase the diagnostic accuracy [18]. In future, we will be attempting the same approach on the features and the classifier proposed in this paper.

REFERENCES

[1] J. Charles, H. Britt, and S. Fahridin, "Croup," *Australian Family Physician,* vol. 39, no. 5, pp. 269-269, 2010.

[2] C. L. Bjornson and D. W. Johnson, "Croup in children," *CMAJ: Canadian Medical Association Journal,* vol. 185, no. 15, pp. 1317-1323, 2013.

[3] C. L. Bjornson and D. W. Johnson, "Croup," *The Lancet,* vol. 371, no. 9609, pp. 329-339, 2008.

[4] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 28, no. 4, pp. 357-366, 1980.

[5] U. R. Abeyratne, V. Swarnkar, A. Setyati, and R. Triasih, "Cough sound analysis can rapidly diagnose childhood pneumonia," *Annals of Biomedical Engineering,* vol. 41, no. 11, pp. 2448-2462, Nov 2013.

[6] K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, "Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis," *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 4, pp. 1185-1194, 2015.

[7] R. V. Sharan and T. J. Moir, "Subband time-frequency image texture features for robust audio surveillance," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 12, pp. 2605-2615, 2015.

[8] R. V. Sharan and T. J. Moir, "Cochleagram image feature for improved robustness in sound recognition," in *Proceedings of the IEEE International Conference on Digital Signal Processing (DSP)*, Singapore, 2015, pp. 441-444.

[9] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "A cough-based algorithm for automatic diagnosis of pertussis," *PLOS ONE,* vol. 11, no. 9, pp. 1-20, 2016.

[10] D. O'Shaughnessy, *Speech communication: human and machine.* Addison-Wesley Pub. Co., 1987.

[11] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, "Complex sounds and auditory images," in *Auditory physiology and perception*. vol. 83, Y. Cazals, L. Demany, and K. Horner, Eds. Pergamon, Oxford, 1992, pp. 429-446.

[12] D. D. Greenwood, "A cochlear frequency-position function for several species - 29 years later," *Journal of the Acoustical Society of America* vol. 87, no. 6, pp. 2592-2605, Jun 1990.

[13] M. Slaney, "An efficient implementation of the Patterson-Holdsworth auditory filter bank," Apple Computer, Technical Report 35, 1993.

[14] M. Slaney, "Auditory Toolbox for Matlab," Interval Research Corproation, Technical Report 1998-010, 1998.

[15] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning,* vol. 20, no. 3, pp. 273-297, Sep. 1995.

[16] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, Pittsburgh, Pennsylvania, USA, 1992, pp. 144-152.

[17] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, *et al.*, *The HTK book (for HTK version 3.4).* Cambridge University: Engineering Department, 2009.

[18] ResApp Health Limited (2016). *ResApp Provides Updated Paediatric Clinical Study Results*. [Accessed: Feb, 2017] Available: http://www.resapphealth.com.au