Detection of Explosive Cough Events in Audio Recordings by Internal Sound Analysis

B. M. Rocha, L. Mendes, J. Henriques, P. Carvalho, R. P. Paiva

Abstract— **We present a new method for the discrimination of explosive cough events, which is based on a combination of spectral content descriptors and pitch-related features. After the removal of near-silent segments, a vector of event boundaries is obtained and a proposed set of 9 features is extracted for each event.**

Two data sets, recorded using electronic stethoscopes and comprising a total of 46 healthy subjects and 13 patients, were employed to evaluate the method. The proposed feature set is compared to three other sets of descriptors: a baseline, a combination of both sets, and an automatic selection of the best 10 features from both sets. The combined feature set yields good results on the cross-validated database, attaining a sensitivity of 92.3±2.3% and a specificity of 84.7±3.3%. Besides, this feature set seems to generalize well when it is trained on a small data set of patients, with a variety of respiratory and cardiovascular diseases, and tested on a bigger data set of mostly healthy subjects: a sensitivity of 93.4% and a specificity of 83.4% are achieved in those conditions. These results demonstrate that complementing the proposed feature set with a baseline set is a promising approach.

I. Introduction

Cough is the most common symptom for which patients seek medical advice [1]. It is a natural respiratory defense mechanism to protect the respiratory tract and one of the most common symptoms of pulmonary disease [2]. It can be characterized by an initial contraction of the expiratory muscles against a closed glottis, followed by a violent expiration as the glottis opens suddenly, producing a characteristic sound [3]. The cough sound is usually divided in three phases: an explosive phase, an intermediate period, whose characteristics are similar to a forced expiration, and a voiced phase.

The main goal of this project is to design a method for the automatic recognition and counting of coughs solely from sound recordings, ideally removing the need for trained listeners. Recent technological advances have enabled the development of automated and ambulatory cough monitors, but there are currently neither standardized methods for recording cough nor adequately validated, commercially available, and clinically acceptable cough monitors [4-5]. While some of the most successful approaches [6-7] have used external audio microphones, our proposed method analyzes internal sound captured by stethoscopes placed on the chest wall. The main difficulty with cough detection in audio recordings lies in its efficient discrimination from other audio non-cough events such as speech, laughter, or ambient noise. Following the approach of Drugman et al. [8], we focus on the detection of the explosive phase of cough, as it is characteristic of the beginning of any cough event, while the intermediate phase is very similar to a forced expiration and the voiced phase is not present in all cough events [8]. Cough often occurs as an epoch, where an initial inspiration is followed by a series of glottal closures and expiratory efforts, sometimes with interspersed inspirations [9]. In this paper, we consider each cough event, i.e., each glottal closure and expiratory effort, independently.

In section II we describe the data sets collected for this work and proposed methodology, including the features and classification algorithms used. In section III we present the results and discuss their implications. Finally, conclusions of the work are provided in section IV.

II. Materials and Methods

A. Data Collection

Pulmonary signals were recorded in Naousa, Greece, and Coimbra, Portugal. Both data sets are described in Table I. The recordings were acquired from a total of 46 healthy subjects and 13 patients, all of which had respiratory diseases (e.g., chronic obstructive pulmonary disease (COPD) or asthma) or cardiovascular diseases (e.g., congestive heart failure). Fig. 1 shows the locations on the chest wall that were used for the recordings. The *Naousa* data set contains recordings from all locations, while the *Coimbra* data set only includes recordings from the ones on the posterior chest wall.

TABLE I. Description of Data Sets

Population	Naousa	Coimbra
Stethoscope	3M Littmann 3200	
Sample Rate	4000 Hz	
Bit Rate	16 bits	
# Subjects	9	50
Diagnosis	Patients: 9	Healthy: 46 Patients: 4
Mean signal duration	112 s	60 s
# Explosive Cough events	468	892
# Voiced Cough events	400	207
# Speech events	296	211

Research supported by the EU Project WELCOME (FP7-611223).

B. M. Rocha, L. Mendes, J. Henriques, P. Carvalho, and R. P. Paiva, are with the Center for Informatics and Systems, Department of Informatics Engineering, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal {bmrocha, lgmendes, jh, carvalho, ruipedro}@dei.uc.pt

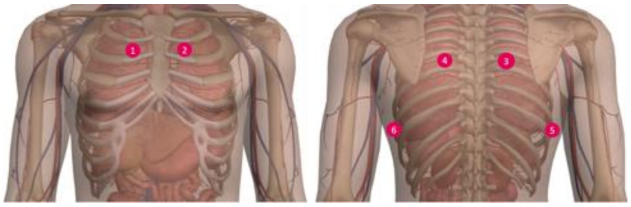

Figure 1. Acquisition locations on the chest wall

During the acquisitions, the subjects were seated and were asked to produce events of cough, speech, laughter, and throat clearing. The physicians who supervised the acquisitions annotated the different events in the timeline and we assigned them to four classes: (1) explosive cough, (2) voiced cough, (3) speech, and (4) other, a class composed of background noises, body rubs, wheezes, crackles, laughter, throat clears, and other artifacts.

B. Pre-Processing

In the pre-processing stage, the audio signal is filtered, using an 8th-order infinite impulse response (IIR) band-pass filter, with 80 Hz (below the lower bound of the typical adult human voice [10]) and 1000 Hz (above the frequency band where most energy of the cough sound is) as the half-power frequencies, and normalized. We then proceed to discard near-silent segments through the following process: given a threshold for length (25 ms) and another for amplitude (5%), segments whose length and amplitude are both below their respective thresholds are classified as near-silent and discarded, i.e., a segment is near-silent if its number of consecutive samples with absolute amplitude below 5% adds up to more than 25 ms. Subsequently, we compute the rms energy in each remaining segment, in 10 ms frames with 80% overlap, to find the onset (threshold: 20%) and ending (threshold: 5%) of each event. These parameters were experimentally obtained and sensitivity analysis proved their robustness. Finally, a vector of event boundaries is fed to the feature extractor. Of the total of 1360 explosive cough events in the data sets, 2.9% (40 events) are lost in this stage, while 0,7% (10 events) of the ones fed to the feature extractor are false positives.

C. Feature Extraction

First, we compute the magnitude spectrum for each event using the Short-Time Fourier Transform (STFT) in frames of 50 ms with 80% overlap. Fig. 2 displays the magnitude spectrum of some of the types of events that occur in these recordings. Then, we extract two types of descriptors: pitch-related features and descriptors of the spectral content. The MIR Toolbox [11] is used to perform fundamental frequency (F_0) estimation and computation of spectral features.

1) Pitch Features

At this stage, the peaks corresponding to the fundamental frequency (F_0) at each frame are estimated. Five pitch features are based on the F_0 vector:

- **Pitch coverage**: ratio of frames where F_0 is detected to total number of frames in the event.
- **Pitch mean, median, standard deviation**: average, median, and standard deviation of F_0, respectively.

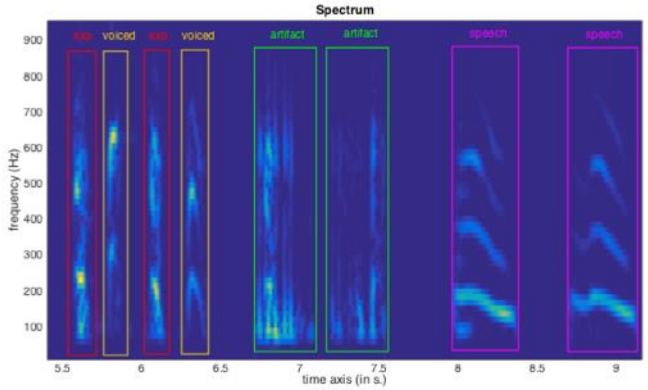

Figure 2. Magnitude spectrum of eight events of classes explosive cough (red), voiced cough (yellow), other (green), and speech (pink)

- **Pitch inharmonicity**: ratio of partials that are not multiple of F_0, taking into account the amount of energy outside the ideal harmonic series [11].

2) Spectral Features

To describe the spectral content over time, we compute the spectral flux as the Euclidian distance between the magnitude spectrum of successive frames.

Three features are extracted from the flux vector:

- **Flux mean, median, and maximum**: average, median, and maximum value of the flux vector, respectively.

Additionally, we compute:

- **Spectral entropy**: Shannon entropy of the magnitude spectrum relative to the length of each event [11].

D. Classification

To classify the events, we use the *Logistic Regression* algorithm from the Weka data mining software [12], transforming the multi-class problem into several 2-class ones with the *1-against-all* method. This algorithm was chosen after validation and comparison with other common machine learning algorithms on a subset of the data.

III. Evaluation

A. Baseline Feature Set

To compare our results, we computed a baseline set of spectral content descriptors and measures of noise, corresponding roughly to the 14 best features proposed by Drugman et al. [6].

1) Spectral Features

To characterize the spectral content of each event, we extract:

- **Mel Frequency Cepstral Coefficients (MFCCs)**: these coefficients offer a description of the spectral shape of the sound [11]; we extract 9 coefficients, ranks 0 to 8.
- **Spectral centroid, spread, skewness, kurtosis**: the 4 standardized moments of the spectral distribution.

2) *Noise Features*

To quantify the level of noise in each event, we compute a number of features that describe it:

- **Spectral flatness**: a measure of the noisiness or sinusoidality of a spectrum [13], computed for the frequency ranges [250-500] and [500-1000] Hz
- **Zero-crossing rate**: number of times the waveform changes sign; the higher it is, the noisier the signal is.
- **Chirp group delay**: phase-based measure proposed in [14] for highlighting turbulences during glottal production.
- **Harmonic to noise ratio (HNR)**: computed for the frequency ranges [0-500] and [0-1500] Hz using the Voice Sauce toolkit [15].

All the features except *Chirp Group Delay* are calculated over the previously extracted magnitude spectrum and F_0 and averaged for each event.

B. *Feature Selection*

The results presented in section III.C are computed using four feature sets:

- **Baseline**: comprises the 19 features described in III.A.
- **Proposed**: consists of the 9 features described in II.C.
- **Combined**: a merger of the 28 features.
- **Filtered**: a selection of 10 features obtained after the following procedure: (1) classification on the training set, (2) removal of misclassified instances, and (3) selection of the best 10 attributes, using the Relief algorithm [16].

C. *Results*

To evaluate the proposed algorithm, we partitioned the data sets in several ways and performed a one sample t-test (available online at [17]) to evaluate the statistical significance of the differences between the feature sets, with *$p<0.01$* (results signaled with * are statistically significantly better than the baseline). First, the combined data sets were split into 3 folds: two of them with 3 *Naousa* subjects and 17 *Coimbra* subjects, and one with 3 *Naousa* subjects and 16 *Coimbra* subjects. This process was repeated 10 times and the results are shown in Table II.

TABLE II. RESULTS ON THE COMBINED DATA SETS

Feature Selection	Sensitivity (SS) %	Specificity (SP) %
Baseline	87.5 (3.7)	76.0 (4.0)
Proposed	91.8 (2.4)*	79.2 (5.2)*
Combined	92.3 (2.3)*	84.7 (3.3)*
Filtered	92.1 (2.6)*	80.8 (4.9)*

Mean (Standard Deviation)

On the combined data sets, the proposed feature set is better than the baseline set for both sensitivity and specificity, while the filtered set, comprised of at least 8 of the 9 proposed features in all folds, is better than the proposed set. The combined set achieves the best results in this experiment. Given the high number of events in the combined data sets, all differences in this experiment are statistically significant. In this case, the contribution of the baseline set is especially important to improve the specificity in the combined set.

Then, to evaluate the robustness of the method specifically for the patients, those 13 subjects were split into 3 folds. This process was repeated 10 times and the results are displayed in Table III.

TABLE III. RESULTS ON PATIENTS

Feature Selection	Sensitivity (SS) %	Specificity (SP) %
Baseline	83.7 (5.8)	78.5 (4.3)
Proposed	88.3 (4.4)*	77.2 (8.0)
Combined	89.1 (3.8)*	84.6 (4.4)*
Filtered	88.0 (3.4)*	76.6 (5.0)

Mean (Standard Deviation)

The performance of all feature sets drops when only patients' cough events are considered. It is interesting to note the significantly lower specificity of the filtered set compared with both the proposed and the baseline sets. We speculate that, in patients with respiratory diseases, the level of noise in the voiced phases varies more than in healthy subjects, making it more difficult to discriminate between explosive and voiced phases of those patients. Therefore, noise features, which are not globally important when considered alone and are not selected by the attribute selection, prove useful for this task.

Next, to evaluate the generalization of the method between data sets, training was performed on the *Naousa* data set and tested on the *Coimbra* data set, and vice-versa. The results of this experiment are presented in Table IV.

TABLE IV. GENERALIZATION ON DIFFERENT DATA SETS

	Test: Coimbra		Test: Naousa	
Feature Selection	SS%	SP%	SS%	SP%
Baseline	37.5	95.3	44.3	72.7
Proposed	70.9*	95.3	89.5*	49.7
Combined	93.4*	83.4	86.0*	71.5
Filtered	93.4*	75.3	83.3*	68.7

These results indicate that the combined set generalizes well when trained on the *Naousa* data set but no feature set is generalizable when trained on the *Coimbra* data set. Furthermore, the filtered set, comprising 7 proposed features plus *MFCCs 4*, *5*, and *8*, yields the same sensitivity as the combined set with about a third of the features. One important feature of the *Naousa* data set is the variability between patients, i.e., although the number of subjects is small, the diversity of diseases affecting those patients might be useful for designing algorithms. Future work should take into account these details.

Finally, to evaluate the generalization of the method between different types of subjects, the algorithm was trained on the healthy subjects and tested on the patients, and vice-versa. Table V shows the results of this experiment.

TABLE V. GENERALIZATION ON DIFFERENT DIAGNOSES

	Test: patients		**Test: healthy**	
Feature Selection	**SS%**	**SP%**	**SS%**	**SP%**
Baseline	47.4	73.2	72.1	68.1
Proposed	89.8*	54.6	83.7*	90.6*
Combined	90.6*	68.6	83.9*	83.8*
Filtered	84.3*	70.9	91.7*	79.3*

This experiment's results seem to confirm that generalizing from patients to healthy subjects might be easier than the opposite. The proposed feature set outperforms the combined set when trained on patients, achieving the same sensitivity and significantly better specificity. Besides, the filtered set, containing 8 proposed features plus *MFCCs 4* and *8*, significantly outperforms the combined set in sensitivity when trained on patients.

Fig. 3 plots two of the most discriminant features in the proposed set: Flux Maximum vs Spectral Entropy, by diagnosis. While the spectral flux is roughly the same for both types of subjects, spectral entropy changes drastically. Theoretically, Shannon entropy is maximal when the input is flat and minimal when there is a single predominant peak [11].

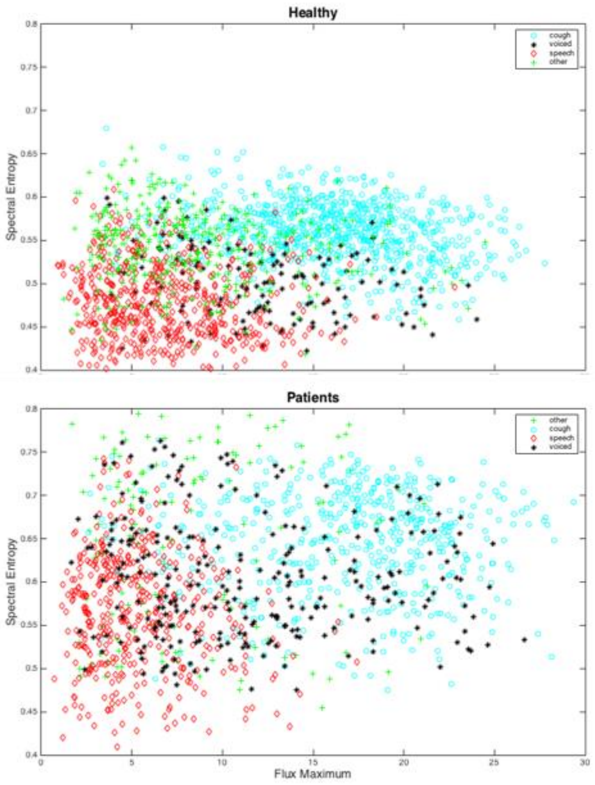

Figure 3. *Flux maximum* vs *Spectral entropy* by diagnosis

The plots in Fig. 3 support the idea that voiced phases in patients are noisier and more difficult to discriminate from explosive phases than in healthy subjects.

IV. CONCLUSION

This paper presents a method for the discrimination of explosive cough events captured with stethoscopes. Two data sets comprising a total of 46 healthy subjects and 13 patients were used and the best results in the analyzed partitions were achieved with a combination of the proposed feature set with a baseline set. Future work should focus on developing robust algorithms that detect cough events in patients with a variety of respiratory and cardiovascular diseases.

REFERENCES

[1] S. M. Schappert and W. C. Burt, "Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 2001-02." *Vital and Health Statistics. Series 13, Data from the National Health Survey*, (159), 1-66, 2006

[2] J. Korpáš and Z. Tomori, "Cough and other respiratory reflexes" 81-104, S. Karger, 1979.

[3] J. N. Evans and M. J. Jaeger, "Mechanical aspects of coughing", *Lung*, 152(4), 253-257, 1975.

[4] J. Korpáš, M. Vrabec, J. Sadlonova, D. Salat, and L. Debreczeni, "Analysis of the cough sound frequency in adults and children with bronchial asthma," *Acta Physiologica Hungarica*, 90(1), 27-34, 2003.

[5] A. Abaza, J. Day, J. Reynolds, A. Mahmoud, W. Goldsmith, W. McKinney, and D. Frazer, "Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function," *Cough*, *5*(1), 1, 2009.

[6] T. Drugman, J. Urbain, and T. Dutoit, "Assessment of audio features for automatic cough detection," In *Signal Processing Conference, 2011 19th European* (pp. 1289-1293). IEEE, 2011.

[7] S. Matos, S. S. Birring, I. D. Pavord, and D. H. Evans, "An automated system for 24-h monitoring of cough frequency: the Leicester cough monitor," *IEEE Transactions on Biomedical Engineering*, *54*(8), 1472-1479, 2007.

[8] T. Drugman, J. Urbain, N. Bauwens, R. Chessini, C. Valderrama, P. Lebecque, and T. Dutoit, "Objective study of sensor relevance for automatic cough detection," *Biomedical and Health Informatics, IEEE Journal of*, *17*(3), 699-707, 2013.

[9] G. A. Fontana and J. Widdicombe, "What is cough and what should be measured?" *Pulmonary pharmacology & therapeutics*, *20*(4), 307-312, 2007.

[10] R. J. Baken and R. F. Orlikoff, "Clinical measurement of speech and voice," *Cengage Learning*, 2000.

[11] O. Lartillot and P. Toiviainen, "A Matlab toolbox for musical feature extraction from audio," *International Conference on Digital Audio Effects* (pp. 237-244), 2007.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, *11*(1), 10-18, 2009.

[13] G. Peeters. A large set of audio features for sound de- scription (similarity and classification) in the cuidado project. 2003.

[14] T. Drugman, T. Dubuisson, and T. Dutoit. Phase-based information for voice pathology detection. In Int. Conf. on Acoustics, Speech and Signal Processing, 2011

[15] Y. L. Shue, P. Keating, C. Vicenik, K. Yu, "VoiceSauce: A program for voice analysis," *Proceedings of the ICPhS XVII*, 1846-1849, 2011.

[16] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," In *Machine Learning: ECML-94* (pp. 171-182). Springer Berlin Heidelberg, 1994.

[17] D. G. Uitenbroek, *SISA Binomial.* Southampton. Retrieved February 17, 2017, from the World Wide Web: http://www.quantitativeskills.com/sisa/distributions/binomial.htm