Voluntary Cough Detection by Internal Sound Analysis

Carlos Lúcio*, César Teixeira*, Jorge Henriques*, Paulo de Carvalho* and Rui Pedro Paiva*

*CISUC - Centre for Informatics and Systems of University of Coimbra

Polo II, Pinhal de Marrocos, 3030-290, Coimbra, Portugal

Abstract—**Cough can be defined as a forced expulsive onrush, normally against a closed glottis, producing a characteristic sound. It can be an indicator of many respiratory diseases, and its counting and classification is an important aspect. We propose a method on internal sound signal to automatically identify, count and (partly) qualify cough sounds. Our approach relies on explosive phase detection, because of its acoustic and spectral distinctive characteristics, and its potential for accurate onset detection of cough sounds. The features analyzed, related with tonality, pitch, timbre and frequency, prove to be very relevant in our explosive phase detection approach. Our results show a recall value of 86.6% and a precision value of 84.3%, for a wide testing population with and without respiratory perturbations. The internal sound analysis reveals advantageous in external noise reduction, therefore internal sounds are highlighted and better characterized.**

I. Introduction

Counting and classifying cough sounds automatically for ambulatory monitoring (long-time and continuous) has proven to be an important problem, with several challenges to address. As a common and clinic descriptive symptom of many respiratory diseases[1-4], cough has been vastly explored by the scientific community as a diagnosis marker. In some conditions like pulmonary fibrosis, lung cancer and chronic obstructive pulmonary disease (COPD), the daily life quality depends on the assessment, monitoring and control of chronic cough.

We propose a method for automatic cough detection based on internal sounds analysis, aiming to not only count cough sounds but also to characterize the event by intensity and pattern. For that, we use an explosive phase detection approach, analyzing 50 individuals, which produced 411 voluntary cough epochs (continuous coughing sounds without a 2-s pause), 383.4 s of speech, 24.55 s of laughing and 26 throat clear events, captured by a digital stethoscope.

The paper is organized as follow: The physiology of cough and past attempts to count and classify cough sounds are presented in Section II; Section III describes the experimental setup; the methodology developed is outlined in Section IV, the main results are introduced and discussed in Section V, and the relevant conclusions are presented in Section VI.

II. Physiology and Related Work

The European Respiratory Society (ERS) suggests the definition of cough as a forced expulsive onrush, normally against a closed glottis, producing a characteristic sound, that literature has consistently cloven in three main patterns[5,6]. First, the explosive phase, characterized by an initial burst of sound emerges that in the moment of glottal opening. It provides information about bronchus, inasmuch that the high frequency sound yields in the vibrations produced by the forced air flux in the airway and the bronchial narrowing places. Then, the intermediate phase, steady-state flow with the glottis wide open. It reflects the status of trachea, and the presence of sputum add a characteristic high frequency component to the sound, and is directly related with the duration of this phase. Finally, the voiced phase, where glottis narrows again, with the vocal cords approaching each other. This third phase may not take place, but the occurrence probability in voluntary cough is about 50% higher than in a spontaneous event[7]. Some other definitions add to this 3 main phases the starting inspiration effort phase and the final expiration time next to voiced phase, as glottis recover to the rest diameter[8]. We did not consider these extra phases, since the three main phases are spectrally more relevant, and show regularity in energy distributed in different frequency bands[9].
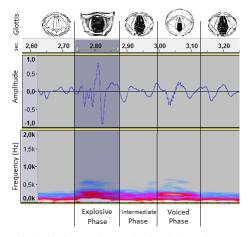

Fig. 1: A typical single cough sound. The darkened area corresponds to explosive phase, followed by the intermediate silence-like phase and the voiced phase, with a slight presentation of the harmonics

Yet, a more careful description of the cough is needed for an accurate classification of each event. We define cough epoch as a time interval with continuous coughing sounds without a 2-s pause. According to the ERS, this definition is valid in clinical terms[10]. For classifying the event, three other characteristics of the cough sound provide important clinical information: the intensity of the event, given by both peak intensity and overall energy released; the pattern of coughing, since either single coughs or fits of coughing serve different

mechanical purposes; and the acoustic properties, which may identify the presence of sputum and wheeze.

Identifying, quantifying and classifying cough has proven to be difficult, not only because of cough variety of phases, patterns and adjustment to pathologies with chronic cough, but also because the aim is to perform ambulatory long term monitoring, which by hand can become a time-consuming and arduous task. There have been attempts to achieve a consensual and reliable system for monitoring cough, with experiments based on both voluntary and pathologic events. Some approaches from the past used audio signals either alone or combined with others[11], [12], but they only enabled to manually spot the cough sounds. Recently, other research works tried to automate the recognition and counting of cough sounds. In [13] and [14], audio signal is coupled with data from different sources, such as: contact microphone, ECG, accelerometers, respiratory inductance plethysmography, among others. However, the acquisition of these signals is difficult to perform in continuous ambulatory settings.

Other attempts mainly use only ambient audio signal. The use of Hidden Markov Models (HMM) to detect cough signals as keyword spotting in continuous audio recordings exhibit promising results[15], although some authors deem that coughs should not be treated as speech, since their acoustic differing characteristics need to be fully considered in the design of algorithms[9]. The best results are also achieved with some manual expert annotation of some cough segments for each person for model tailoring.

Other publications use probabilistic neural networks (PNNs), achieving good results[16]. The key idea was to focus only on the detection of explosive phase of cough. The intermediate phase had proven to be very similar to a forced expiration[17], or in some healthy sputum-free subjects cases to a silence phase, inasmuch the voiced phase may not occur and resembles to a speech sound, while explosive phase is characteristic of the beginning of any cough event and possibly more nonspecific among individuals. Despite this, Drugman et al.[16] did not validate this assumption by manually note the explosive phase, and approached it by the first 60 ms of the total cough event, which is not accurate, since explosive phase length in audio signal could vary from 29.6 ms to 54.6 ms[8]. Additionally, they experiment voluntary cough from only healthy subjects. The Hull Automatic Cough Counter (HACC)[18], uses an event detection logic based on adaptable thresholding. The spectral coefficients are calculated for detected event candidates, which are then classified into cough and non-cough sounds by the use of PNNs. Tests performed only in smoking subjects achieved good results.

The VitaloJAK system uses a contact microphone placed on chest wall, but likewise needs an initial input of voluntary coughs by the subject. In [19], this system was adapted to a physiological approach tailoring. Subjects perform voluntary coughs, which are recorded, from set lung volumes. The same author concludes in [20] that much of the variability in cough sounds within an individual can be explained by the lung volume from which the cough occurs. Acoustic parameters extracted from these voluntary coughs can be then used to interrogate a 24-h sound recording and pick out spontaneous cough sounds. The system obtains a sensitive cough detection in pathologic cough.

Despite those, ERS Committee currently defend that there are at the moment no standardized methods for recording cough nor adequately validated, commercially available, and clinically acceptable cough monitors[10].

III. Experimental Setup

We performed 200 recordings from 36 individuals without any respiratory perturbation (without respiratory perturbations group - WPG) and from 14 individuals with respiratory perturbations (respiratory perturbations group - RPG). For each individual four recordings with 15 s each were saved for future processing. In the group with respiratory perturbations we could find smokers (n=1) and pathologies like asthma (n=3), bronchitis (n=1), rhinitis (n=1) or simple colds (n=8). The total set of individuals obtained a mean of 33.26±14.73 years old, mean body mass index (BMI) of 26.06±4.24 kg/m^2, being 26 males and 24 females. The recording system consisted in a digital stethoscope (3M Littmann, model 3200, St. Paul, Minnesota, USA) with a sampling frequency of 4000 Hz, and the sound signal was acquired on the posterior inferior lobe site of right lung, with individuals seated in a normal hospital room, without any external sound cancellation. In each recording, the person was asked to produce a single cough, a peal of two or more coughs, and around 5 s of one of the following events: speech, laughter and throat clears. The order of these events was randomly and not proportionally varied for each one of the four recordings. The final total audio data consisted of a 50 individual's recordings of 1 minute, containing 411 single cough or peal events, 383.4 s of speech, 24.55 s of laughing and 26 throat clears.

All the audio data were analyzed by an observer, using Audacity v2.0.5 in order to obtain the annotation of explosive phases. The annotations were implemented by listening to the audio to detect the cough sound and then by observing the spectrum signal for onset and offset adjustments.

IV. Methods

We developed an algorithm based on the waveform, spectral and prosody related features. The signal was splitted into 50 ms frames, for which the features were computed using the open-source Matlab toolboxes MIRtoolbox[21] and VOICEBOX[22]. To minimize the inherent errors of parallax on the annotation, it was considered that a frame was part of a given event if more than half of the frame belonged to the noted event.

Our implementation was performed in Matlab. After feature selection, we applied the K-Nearest Neighbors (KNN) pattern recognition algorithm to discriminate between sound frames with and without cough. KNN is a simple and non-parametric approach that is known to be a good choice when data distribution is unknown, or difficult to determine. The algorithm determines the class of a given pattern based on a distance metrics (usually Euclidean), and on the class of surrounding

neighbors patterns. More precisely, the algorithm finds the closest k neighbors by using the chosen distance metrics and the class of a given pattern will be the class of the majority of its neighbors. In this way, the unique training parameters are the distance metrics and the number of neighbors (k). Here, the best k value was selected by considering, iteratively, odd numbers in the range of 3 to 15, for the total set of ranked features. After fixation of the best k, the optimal number of features was determined experimentally by using a simple forward feature selection approach that consists on adding one feature at a time based on their relevance, and stopping when the maximization of the results is obtained. Feature relevance or ranking were performed using WEKA v3.6, and the Relief[23] algorithm. A total of 79 features were analyzed and ranked for training and testing the developed methodology.

Before classification, the set of features is rearranged by Principal Component Analysis method, which uses an orthogonal transformation to convert the set with possibly correlated features into a set of values linearly uncorrelated called principal components, using a 90% value of covariance. The results were then post-processed, merging the detected explosive phase events spaced less than 2 s, an approximation of cough epochs as we defined later. By this approximation, only the last intermediate and voicing phases of last cough sound of each cough epoch will be missed, and so, for the purpose of counting, those will be equivalent to the number of cough epochs. The precision and recall metrics were obtained, frame by frame and by this approximated cough epochs. The mean and standard deviation of onset and offset flaw and the mean peak energy of the signal in the approximated cough epochs were also computed. These procedures were performed for all data, and then repeated for the WPG individuals and for the RPG individuals.

V. RESULTS AND DISCUSSION

We performed tests in the whole dataset (WPG+RPG), and for the WPG and RPG groups alone. As we can see in Table I, for all of the testing groups, the Relief algorithm ranked as the most relevant feature the one obtained with mirkey function of MIRtoolbox, which relates with tonality and gives a broad estimation of tonal center positions and their respective clarity. This feature prove to be very relevant in our explosive phase detection approach. It shows that explosive phases of cough have a tonal center distinct from the tonal center of all other sounds tested, i.e., the tonic elements (which tend to assert their dominance over all others) in explosive phases seem to be distinctive. The ranking for WPG and RPG groups was the same up to the first 11 features. This reveals the relevance of these features for the problem, and also the proximity between the groups. The next features in this groups are by order of importance related with frequency, timbre and spectral analysis. Evo gives an estimate distance between the Fast Fourier Transform of the signal in each successive frames, mfcc4 corresponds to the value of the fourth Mel Frequency Cepstral Coefficient and midi is the mean of the conversion of Fast Fourier Transform of the signal to midi scale of semi-

TABLE I: Feature Ranking for testing groups up to the fouth feature. This features resulted from functions from MIRtoolbox.

	Feature Ranking			
	1st	**2nd**	**3rd**	**4th**
WPG+RPG	key	evo	nofilterbank	midi
WPG	key	evo	mfcc4	midi
RPG	key	evo	mfcc4	midi

tones. For the RPG+WPG group, one new feature appeared. Nofilterbank is related with pitch analysis and calculate the mean of the best pitch of the the discretized note events of the signal data, for no filterbank configuration. Our sound signal reveals flexibility in characterization, with several areas of sound analysis contributing for the identification of the explosive phase.

The results for the testing groups are present in Tables II, III and IV. The achieved measures of onset and offset (mean and STD) reveal some difficulties to pinpoint the initial and final instants of the cough events detected. It can be assumed that the major contribution for this values comes from deviations in detection of the fits of cough approximated events, since those have much longer duration than explosive phases alone. Therefore, in the detection of fits, if one or more explosive phases from the beginning or the ending miss the detection, the flaw is in the order of more than a 3-phase cough sound duration, around 350.7 seconds[8]. The onset vagueness should represent the expected imperfection of the detection. However, the higher values of STD for the onset suggest that the flaw in missed detections of first explosive phase of peal events is balanced with prior detections, which can be justified by detections in the initial forced expiration before the first explosive phase. The higher values in mean offset suggest difficulties in the detection of last explosive phases of peal events, which can be understood by the gradual decrease in the intensity of the signal in explosive phases during the peal event, leading to missed detections in the end of peals. The high value of STD in offset insinuate that this missed detections are also balanced with late detections, which can be caused by detections in intermediate and voicing phases of the last 3-phase cough sound of each peal event. All this values suggest that the windowing of events should be improved.

TABLE II: Results obtained for WPG+RPG analysis, with optimized k=11 and set of 8 ranked features

	Value
Recall frame by frame	42.4%
Precision frame by frame	65.1%
Recall by approx. cough epochs	86.6%
Precision by approx. cough epochs	84.3%
Onset error (mean)	17.3 ms
Onset error (STD)	318.9 ms
Offset error (mean)	83.1 ms
Offset error (STD)	194.3 ms
Recall (fits of cough)	59.6%
Precision (fits of cough)	91.2%
Mean peak energy approx. cough epochs	0.5561 a.u.

TABLE III: Results obtained for WPG analysis, with optimized k=11 and set of 12 ranked features

	Value
Recall frame by frame	42.6%
Precision frame by frame	65.6%
Recall by approx. cough epochs	84.6%
Precision by approx. cough epochs	83.1%
Onset error (mean)	50.2 ms
Onset error (STD)	284.6 ms
Offset error (mean)	232.8 ms
Offset error (STD)	222.8 ms
Recall (fits of cough)	57.7%
Precision (fits of cough)	90.7%
Mean peak energy approx. cough epochs	0.5161 a.u.

TABLE IV: Results obtained for RPG analysis, with optimized k=7 and set of 19 ranked features

	Value
Recall frame by frame	43.1%
Precision frame by frame	60.9%
Recall by approx. cough epochs	89.5%
Precision by approx. cough epochs	81.3%
Onset error (mean)	26.0 ms
Onset error (STD)	164.0 ms
Offset error (mean)	82.8 ms
Offset error (STD)	123.4 ms
Recall (fits of cough)	62.0%
Precision (fits of cough)	86.0%
Mean peak energy approx. cough epochs	0.8174 a.u.

TABLE V: Comparative results of the tests performed

	Recall	Precision
WPG+RPG	86.6%	84.3%
WPG	84.6%	83.1%
RPG	89.5%	81.3%

Recall and precision metrics were also calculated for the number of cough hits detected in fits of cough. This values represent the capability of discriminate how many explosive phases we can detect in a multiple event, providing important information about the pattern of coughing. However, recall values are slightly low.

The energy metrics intended to give a rough estimation of the energy peak intensity of the cough sounds, one of the three characteristics of the cough sound which provide important clinical information. Its relevance is evidenced in the differences of the values for the WPG and RPG groups. The mean peak energy of the signal in the approximated cough epochs detected is substantially higher in RPG group, which is due to a louder cough sound, with high frequencies resulting probably from presence of sputum or wheeze.

The overall results of recall and precision metrics, present in Table V, for the approximated cough epochs detection are similar for the three testing groups, revealing flexibility and applicability of this approach. Notice that achieved results were originated from analyzing concise sets of features, not being necessary large loads of data and computation, something important in continuous monitoring for battery saving. Our results are comparable to mentioned works above, although they were not obtained on the same database. Drugman et al.[16], with the similar explosive phase detection approach, obtained specificity and recall of 88%, for audio signal analysis in voluntary protocols performed by only 20 healthy subjects. They also applied the same method to the signal of a contact microphone over trachea and over thorax, achieving both specificity and recall of 71%. In [13] the signal of a contact microphone placed on the throat is compared to other signal analysis in voluntary protocols for counting cough, performed using 32 healthy subjects, and results also fall short of expectations when compared with audio signal, but they do not use the explosive phase detection approach. To the best of our knowledge, no other works tried the internal sound signal alone for automatic cough counting. However, this signal solves one of the disadvantages of the commonly used audio signal, referred in [13] - the external ambient noise. In internal sound, only the subject's internal acts can confuse the classification. A fully automated cough-counting industrial device, like the VitaloJAK system, seems to support this, since when experimented in 10 subjects with respiratory perturbations achieved a recall of 97.5% and a specificity of 97.7%[20].

Our acquisition protocol was more focused in coughing and speech since it is ethically a requirement to eliminate speech content from acquired audio signals for medical applications.. Furthermore, we advocate that throat clears noisy enough for trigger the classifier are not so bad when counted, as an important respiratory event too.

For classification of the cough sound, the acoustic properties were not analyzed, because the method identified only the first explosive phase, while this characteristics are important to analyze in intermediate and voiced phase, for sputum and wheeze prediction, among other important clinical information. This can be solved with a future improvement in post-processing of identified explosive phases, by methods to identify the characteristic intermediate phase in presence of sputum or finding speech related moments in a neighborhood of the explosive phase. A promising solution could be the classification integrating various events such as speech, throat clears, laughter and sneezes, as well as cough itself, since internal sound seems to characterize those more minutely and with less ambiguity than audio sound. The use of inputs for tailoring parameters by each subject can also improve the efficiency, with the drawback of loss of autonomy. Future work on this method involves trying other classifiers, testing in ambulatory data and the use of new features.

VI. CONCLUSION

Our results refute the idea that internal sound signal analysis provide no significant performances when compared with audio signal analysis, for automatic cough detection. Moreover, internal sound demonstrates advantages in one of principal challenges of audio signal analysis, the external-to-subject noise. As important implication, we have a significant reduction on the captured sound and external speech, relevant in ethic terms. Besides, the subject internal events like cough, speech, laughs, throat clears, forced expirations, etc. seem to be better characterized by this signal. Here, some features related with tonality, pitch, timbre and spectral analysis reveal themselves very descriptive. The internal sound analysis through a contact microphone gathers attention for applications in new generation of smart clothing, seeming to be less inconvenient and less intrusive than a lapel microphone.

The explosive phase detection approach reveals very promising for counting cough, showing flexibility between subjects with different respiratory perturbations and mitigating the inherent difficulty of the variety of patterns in cough. Improvements are needed in post-processing methods to identify intermediate and voiced phase, and thereby obtain the complete cough sound, which can then be analyzed in terms of acoustic properties for better assessing.

ACKNOWLEDGMENT

REFERENCES

[1] K. F. Chung, *Methods of assessing cough and antitussives in man*, Pulmonary Pharmacol., vol. 9, pp. 373–377, 1996.
[2] R. S. Irwin, F. J. Curley and C. L. French, *Chronic cough. The spectrum and frequency of causes, key components of the diagnostic evaluation, and outcome of specific therapy*, A. Rev. Respiratory Disease, vol. 141, no. 3, pp. 640–647, 1990.
[3] R. S. Irwin, L. P. Boulet, M. M. Cloutier, R Fuller, P. M. Gold, V. Hoffstein, A. J. Ing, F. D. Mccool, P. O'Byrne, R. H. Poe, U. B. Prakash, M. R. Pratter and B. K. Rubin, *Managing cough as a defense mechanism and as a symptom. A consensus panel report of the American College of Chest Physicians*, Chest., vol. 114, pp. 133S–181S, 1998.
[4] A. B. Chang, *Causes, assessment and measurement of cough in children*, in Cough: Causes, Mechanisms and Therapy, pp. 57–73, 2003.
[5] J. Korpas, J. Sadlonova, D. Salat and E. Masarova, *The origin of cough sounds*, Bull. Eur. Physiopathol. Respir., vol. 23, pp. 47–50, 1987.
[6] C. W. Thorpe, L. J. Toop and K. P. Dawson, *Towards a quantitative description of asthmatic cough sounds*, Eur. Respir. J., vol. 5, pp. 685–692, 1992.
[7] A. V. Hirtum and D. Berckmans, *Automated recognition of spontaneous versus voluntary cough*, Medical engineering and physics, vol. 24, pp. 541–545, 2002.
[8] P. M. Olia, P. Sestini and M. Vagliasindi, *Acoustic parameters of voluntary cough in healthy non-smoking subjects*, Respirology., vol. 5, pp. 271–275, 2000.
[9] Z. Chunmei, T. Lianfang, L. Xiangyang, M. Hongqiang and Z. Zeguang, *Recognition of cough using features improved by sub-band energy transformation*, BMEI 6th Int. Conf., pp. 251–255, 2013.
[10] A. H. Morice, G. A. Fontana, M. G. Belvisi, S. S. Birring, K. F. Chung, P. V. Dicpinigaitis, J. A. Kastelik, L. P. McGarveye, J. A. Smith, M. Tatar and J. Widdicombe, *ERS guidelines on the assessment of cough*, Eur. Respiratory J., vol. 29, pp. 1256–1276, 2007.
[11] P. Munyard, C. Busst, R. Logan-Sinclair and A. Bush, *A new device for ambulatory cough recording*, Pediatr. Pulmonol., vol. 18, no. 3, pp. 178–186, 1994.
[12] A. B. Chang, R. G. Newman, P. D. Phelan and C. F. Robertson, *A new use for an old holter monitor: An ambulatory cough meter*, Eur. Respiratory J., vol. 10, pp. 1637–1639, 1997.
[13] T. Drugman, J. Urbain, N. Bauwens, R. Chessini, C. Valderrama, P. Lebecque, T. Dutoit, *Objective Study of Sensor Relevance for Automatic Cough Detection*, IEEE J. Biomed. Hea. Inf., vol. 17, no. 3, pp. 1637–1639, 2013.
[14] M. A Coyle, D. B. Keenan, L. S. Henderson, M. L. Watkins, B. K. Haumann, D. W. Mayleben, and M. G. Wilson, *Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease*, Cough Journal, vol. 1, no. 1, 2005.
[15] S. Matos, S. S. Birring, I. D. Pavord, D. H. Evans, *An Automated System for 24-h Monitoring of Cough Frequency: The Leicester Cough Monitor*, Biomed. Eng., IEEE Transactions on, vol. 54, no. 8, pp. 1472–1479, 2007.
[16] T. Drugman, J. Urbain, N. Bauwens, R. Chessini, A. S. Aubriot, P. Lebecque and T. Dutoit, *Audio and Contact Microphones for Cough Detection*, 13th Interspeech An. Conf. Int. Speech Com., 2012.
[17] J. Korpas, J. Sadlonova and M. Vrabec, *Analysis of the cough sound: An overview*, Pulmonary Pharmacol., vol. 9, pp. 261–268, 1996.
[18] S. J. Barry, A. D. Dane, A. H. Morice, and A. D. Walmsley, *The automatic recognition and counting of cough*, Cough, vol. 2, no. 1, pp. 8–17, 2006.
[19] K. McGuinness, J. Morris, A. Kelsall, A. Woodcock, J. A. Smith, *The relationship between cough acoustics and the volume inspired prior to coughing*, Am. J. Resp. Crit. Care Med., vol. 175, pp. A381, 2007.
[20] K. McGuinness, A. Kelsall, J. Lowe, A. Woodcock, J. A. Smith, *Automated cough detection: a novel approach*, Am. J. Resp. Crit. Care Med., vol. 175, pp. A381, 2007.
[21] O. Lartillot, P. Toiviainen and T. Eerola, *MIRtoolbox v1.5*, Finnish C. of Exc. in Interdiscipl. Music Res..
[22] M. Brookes, *Voicebox: Speech Processing Toolbox for Matlab*, Dept. Electrical and Electronic Eng., Imp. C., 2002.
[23] M. Robnik-ikonja, I. Kononenko, *Theoretical and Empirical Analysis of ReliefF and RReliefF*, Machine Learning, vol. 53, no. 1–2, pp. 23–69, 2003.