Detection of Cough and Adventitious Respiratory Sounds in Audio Recordings by Internal Sound Analysis

B.M. Rocha[1], L. Mendes[1], I. Chouvarda[2], P. Carvalho[1], and R.P. Paiva[1]

[1] Centre for Informatics and Systems, Department of Informatics Engineering, Faculty of Sciences and Technology, University of Coimbra, Coimbra, Portugal
[2] Lab of Computing, Medical Informatics and Biomedical Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

Abstract— **We present a multi-feature approach to the detection of cough and adventitious respiratory sounds. After the removal of near-silent segments, a vector of event boundaries is obtained and a proposed set of 126 features is extracted for each event. Evaluation was performed on a data set comprised of internal audio recordings from 18 patients. The best performance (F-measure = 0.69 ± 0.03; specificity = 0.90 ± 0.01) was achieved when merging wheezes and crackles into a single class of adventitious respiratory sounds.**

Keywords— **cough, adventitious respiratory sounds, automatic classification**

I. Introduction

Respiratory diseases cause an immense socio-economic impact and are the third leading cause of death worldwide [1] and a burden to public health systems [2]. Therefore, significant research efforts have been dedicated to improving early diagnosis and routine monitoring of patients with respiratory diseases to allow for timely interventions [3].

A great amount of research has been focused on the auscultation and characteristics of cough and respiratory sounds (RS), as they are valuable indicators of respiratory health and respiratory disorders [4].

Cough is a natural respiratory defense mechanism to protect the respiratory tract and one of the most common symptoms of pulmonary disease [5]. 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 [6]. 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. 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 [7]. In this paper, we consider each cough event, i.e., each glottal closure and expiratory effort, independently.

Respiratory sounds are generally classified as normal or adventitious. Auscultation-based diagnosis and monitoring of respiratory conditions rely heavily on the presence of adventitious sounds and on the altered transmission characteristics of the chest wall. Adventitious sounds are RS superimposed on normal respiratory sounds which can be discontinuous (crackles) or continuous (wheezes). Crackles are discontinuous, explosive, and non-musical adventitious RS that occur frequently in cardiorespiratory diseases [8]. They are usually classified as fine and coarse crackles based on their duration, loudness, pitch, timing in the respiratory cycle, and relationship to coughing and changing body position [9]. Wheezes are musical RS that usually last more than 250ms. They are a common clinical sign in patients with obstructive airway diseases such as asthma and chronic obstructive pulmonary disease (COPD) [10].

The main goal of this work was to design a method for the automatic detection of cough and adventitious RS solely from audio recordings. The automatic detection of cough and adventitious RS has been the subject of many studies in the last decades. Algorithms developed to detect or classify events usually involve two steps; cough and adventitious RS are no exception. The first step is to extract the relevant features that will be used as detection or classification variables. The second step is to use detection or classification techniques on the data, based on the features extracted. The most common features employed in the literature include Mel-frequency cepstral coefficients (MFCCs), spectral features, energy, entropy, and wavelet coefficients. Machine learning algorithms proposed in the literature include empirical rule-based methods, support vector machines (SVMs), artificial neural networks (ANNs), Gaussian mixture models (GMMs), k-nearest neighbors (k-NNs), and logistic regression models [11]. Prior attempts at automated classification of adventitious RS have tried to simplify the problem by focusing on a single type of sound and, to the best of our knowledge, none has tried to classify cough and adventitious RS at the same time.

In section II we describe the data collected for this work and the methodology proposed, 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

Respiratory sounds were acquired at the Papanikolaou General Hospital, Thessaloniki and at the General Hospital of Imathia (Health Unit of Naousa), Greece. Sounds were collected sequentially from six chest locations, as shown in Figure 1. The acquisition of RS was performed on adult and elderly patients. All patients had COPD with comorbidities (e.g. heart failure, diabetes, hypertension). Table 1 provides a description of the data set.

These recordings were acquired as part of the European project WELCOME (Wearable Sensing and Smart Cloud Computing for Integrated Care to COPD Patients with Comorbidities) project and were annotated using Audacity[1] 2.0.6 a free, open source, cross-platform software for recording and editing sounds.

Respiratory sound annotations were performed by three experienced physicians, two specialized pulmonologists and one cardiologist. Annotations discriminated the following sounds: normal (respiratory sound), crackles, wheezes, speech, cough, artifact.

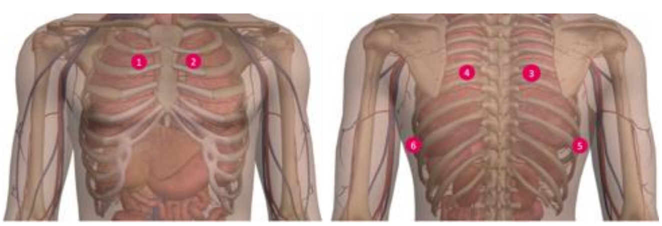

Figure 1. Chest locations for the recording of respiratory sounds.

Table 1. Description of data set

# Patients	18
Average signal duration	106 s
# Cough segments	574
# Crackle segments	184
# Wheeze segments	248
# Speech segments	440
# Other segments	602

[1] http://audacity.sourceforge.net/

B. Pre-processing

In the pre-processing stage, the audio signal is filtered, using an 8th-order infinite impulse response (IIR) high-pass filter at 80 Hz (below the lower bound of the typical adult human voice [12]), and normalized. We then proceed to discard near-silent segments through the following process: given a threshold for length (100 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 100 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.

C. Feature extraction

A total of 42 descriptors were computed in frame windows of 50 ms and 80% overlap. The final number of features computed for each event was 128, corresponding to the median, the maximum, and the standard deviation of each descriptor.

a) Musical features

The MIR Toolbox [13] was used to extract 35 features related to dynamics, timbre, pitch, and harmonic content. Table 2 provides a brief description of the musical features used in this work.

b) Other features

Seven other features were extracted in this work. Chirp group delay is a phase-based measure proposed in [14] for highlighting turbulences during glottal production. Harmonic to noise ratio (HNR) was computed for the frequency ranges [0-500] and [0-1500] Hz using the Voice Sauce toolkit [15]. The information entropy is a measure of the disorder of a system; the maximum of the entropy in each frame was used as a feature. Another computed feature was the maximum of the Teager energy in each frame. The maximum of Katz's fractal dimension of the filter WPST-NST, described in [16], was also calculated. Finally, the wheeze signature in the spectrogram space, thoroughly described in [17], was computed.

D. Classification

Before classifying the events, the data set is partitioned into 10 stratified folds. Then, the training folds are filtered through the following procedure: 1) a class balancer is applied to reweight the instances in the data so that each class has the same total weight; 2) feature selection is performed and each feature is evaluated according to the information gain it provides; 3) each instance of the training set is classified and misclassified instances are removed. Finally, a random forest classifies each event of the test fold. This algorithm was chosen after validation and comparison with other common machine learning algorithms on a subset of the data. This process is repeated 10 times.

Table 2. Description of the musical features

Feature	Description
RMS	Root-mean square energy of the frame
Spectral Centroid	Center of mass of the spectral distribution
Spectral Brightness	Amount of energy above 1500 Hz
Spectral Spread	Variance of the spectral distribution
Spectral Skewness	Skewness of the spectral distribution
Spectral Kurtosis	Excess kurtosis of the spectral distribution
Spectral Rolloff 95	Frequency such that 95% of the total energy is contained below that frequency
Spectral Rolloff 85	Frequency such that 85% of the total energy is contained below that frequency
Spectral Entropy	Complexity of the spectrum
Spectral Flatness	Noisiness of the spectrum
Spectral Roughness	Estimation of the sensory dissonance
Spectral Irregularity	Degree of variation of the successive peaks of the spectrum
MFCC	13 Mel-frequency cepstral coefficients
Zero-crossing Rate	Waveform sign-change rate
Spectral Flux	Distance between the spectrum of successive frames
Chromagram Centroid	Tonal centroid
Chromagram Peak	Peak of the tonal centroid
Key Clarity	Probability of key candidates
Mode	Modality estimation
Harmonic Change Detection Function	Flux of the tonal centroid
Pitch	Pitch estimation
Pitch Inharmonicity	Ratio of partials that are not multiple of the fundamental frequency, taking into account the amount of energy outside the ideal harmonic series
F0	Fundamental frequency estimation

III. Evaluation

Six versions of the data set were used for evaluation: *Complete*, with five classes (cough, wheezes, crackles, speech, other) and no feature selection; *Merged*, with four classes, where wheezes and crackles were merged (cough, adventitious sounds, speech, other) and no feature selection was performed; *Complete 50*, i.e., *Complete* with the best 50 features; *Merged 50*, i.e., *Merged* with the best 50 features; *Complete 20*; *Merged 20*; *Complete 10*; *Merged 10*; *Complete 5*; *Merged 5*. Table 3 shows the sensitivity, specificity, precision, and F-measure for all sets.

Table 3. Results

Data set	Specificity	Sensitivity	Precision	F-measure
Complete	0.90 ± 0.01	0.67 ± 0.03	0.67 ± 0.03	0.67 ± 0.03
Merged	0.90 ± 0.01	0.69 ± 0.03	0.70 ± 0.03	0.69 ± 0.03
Complete 50	0.90 ± 0.01	0.67 ± 0.03	0.67 ± 0.04	0.67 ± 0.03
Merged 50	0.90 ± 0.01	0.69 ± 0.03	0.70 ± 0.03	0.69 ± 0.03
Complete 20	0.90 ± 0.01	0.65 ± 0.04	0.65 ± 0.04	0.65 ± 0.04
Merged 20	0.89 ± 0.01	0.67 ± 0.03	0.68 ± 0.03	0.67 ± 0.03
Complete 10	0.90 ± 0.01	0.62 ± 0.03	0.63 ± 0.03	0.62 ± 0.03
Merged 10	0.88 ± 0.01	0.65 ± 0.03	0.65 ± 0.03	0.65 ± 0.03
Complete 5	0.88 ± 0.01	0.57 ± 0.03	0.58 ± 0.03	0.57 ± 0.03
Merged 5	0.87 ± 0.01	0.61 ± 0.03	0.62 ± 0.03	0.61 ± 0.03

Mean ± Standard Deviation

Regarding specificity, the performance did not change significantly with feature selection or the merging of the adventitious RS classes. Regarding the other metrics, the performance in *Merged* sets is always better than in *Complete* sets and it is especially so in the adventitious RS classes. The removal of features also seems to have less impact in *Merged* sets. Given these results, one can speculate that performing a hierarchical classification where adventitious RS are first merged and then discriminated might improve the performance.

IV. Conclusions

This paper presents a method for the detection of cough and adventitious RS. A data set comprising a total of 18 patients was used to evaluate the performance. The results indicate that future work should employ a hierarchical classifier.

Acknowledgment

The authors would like to thank the health professionals and the patients who have agreed to participate in the data collection process. This work was financially supported by the EU project WELCOME[2] (FP7-ICT-2013-10/611223)

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. World Health Organization. The top 10 causes of death. 2017.
2. Gibson GJ, Loddenkemper R, Lundbäck B, Sibille Y. Respiratory health and disease in Europe: the new European Lung White Book. *Eur Respir J*. 2013;42:559–63.
3. Marques A, Oliveira A, Jácome C. Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: a systematic review. *Respir Care*. 2014 May;59(5):765–76.
4. Earis J, Cheetham B. Current methods used for computerized respiratory sound analysis. *Eur Respir Rev*. 2000; 10(77):586–90.
5. Korpáš, J. and Tomori, Z., Cough and other respiratory reflexes. 81-104. S. Karger, 1979.
6. Evans, J. N., and Jaeger, M. J., Mechanical aspects of coughing. *Lung*, 152(4), 253-257, 1975.
7. Fontana, G. and Widdicombe, J., What is cough and what should be measured? *Pulmonary pharmacology & therapeutics*, *20*(4), 307-312, 2007.
8. Piirila P, Sovijarvi AR. Crackles: recording, analysis and clinical significance. Eur Respir J. Eur Respiratory Soc; 1995;8(12):2139–48.
9. Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. *Ann Thorac Med*. Medknow Publications; 2015;10(3):158.
10. Sovijarvi ARA. Characteristics of breath sounds and adventitious respiratory sounds. *Eur Respir Rev* 2000;10:591-6.
11. Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: A systematic review. *PLoS One*. Public Library of Science; 2017;12(5):e0177926.
12. R. J. Baken and R. F. Orlikoff, Clinical measurement of speech and voice. *Cengage Learning*, 2000.
13. Lartillot, O. and Toiviainen, P. A Matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects* (pp. 237-244), 2007.
14. Drugman, T., Dubuisson, T., and Dutoit, T. Phase-based information for voice pathology detection.” *In Int. Conf. on Acoustics, Speech and Signal Processing*, 2011
15. Shue, Y.L., Keating, P., Vicenik, C., Yu, K., “VoiceSauce: A program for voice analysis.” *Proceedings of the ICPhS XVII*, 1846-1849, 2011.
16. Mendes L, Vogiatzis IM, Perantoni E, Kaimakamis E, Chouvarda I, Maglaveras N, et al. Detection of crackle events using a multi-feature approach. *38th Annu Int Conf IEEE Eng Med Biol Soc. IEEE*; 2016; 3679–83.
17. Mendes L, Vogiatzis IM, Perantoni E, Kaimakamis E, Chouvarda I, Maglaveras N, et al. Detection of wheezes using their signature in the spectrogram space and musical features. *37th Annu Int Conf IEEE Eng Med Biol Soc. IEEE*; 2015.

[2] http://www.welcome-project.eu/