The automatic recognition and counting of cough

Maria José Pinto Bernal
Department of Biomedical Engineering
Engineering Julio Garavito
Bogotá, Colombia
maria.pinto@mail.escuelaing.edu.co

Diego Fernando Casas Bocanegra
Department of Biomedical Engineering
Engineering Julio Garavito
Bogotá, Colombia
diego.casa-b@mail.escuelaing.edu.co

***Abstract*—Coughing is a common symptom of many respiratory diseases. Many medical publications stress that a system for the automatic, objective, and reliable detection of coughing episodes are essential and very promising for detecting the severity of the pathology, especially in chronic cough disease. Using digital signal processing to calculate characteristic spectral coefficients of sound events, which are then classified into a cough and non-cough events, is very promising for the design, validation, and training of classifiers. To achieve this goal, this paper presents the automatic recognition algorithm that uses an audio signal record. These audio signals were extracted from nine healthy subjects. Feature extraction from all recordings was used as nine characteristics. The performance of an SVM-based algorithm, trained using subject-specific and standardized parameter approaches, was compared during audio recordings in terms of goodness index. The findings indicate that the SVM-based classifier trained using an intra-subject approach showed excellent reliability for the assessment of mean high time, i.e., its intra-subject goodness index (G) was lower than 0.75. In conclusion, the SVM-based method proposed here can be implemented for the recognition and counting of cough.**

***Index Terms*—Cough, no-cough, audio signal, cough features, classifier, support vector machine.**

I. Introduction

Cough is a common but complicated symptom of respiratory diseases. Cough is the most typical symptom for which patients seek medical advice [1]. Population studies reported prevalence of cough to vary between 3% and 40% [2], [3]. Cough is conventionally considered to consist of an initial deep inspiration followed by expiration against a closed glottis that then opens [4], [5]. As a result, a characteristic phonation is formed, which is composed of two distinct components termed first and the second cough sounds [6].

While the recognition of a single cough event is relatively easy, the assessment of cough frequency over a long period remains difficult both for clinical and research purposes. Part of the problem is the paroxysmal nature of cough necessitating recording over a prolonged time in order to generate an accurate estimate of cough frequency [7].

However, the simple recording of cough sound using a microphone and cassette recorder allows for counting of the cough events. However, analysis is very time-consuming even with the application of sound-activated recording or methods for removing silence [8], [9]. Besides, the use of cough recorders that incorporate electromyogram (EMG) [10] or modified Holter monitor [11] require manual reading of the recorded tapes by a trained investigator. Automatic cough recognition from ambulatory multi-channel physiological recordings has been reported [12].

The study presented here features a cough detection system from sound recordings, which reduces the processing time. Hence, we presented the design of a dataset from a phase of characterization of an audio signal presenting cough and cough-free events. The implementation, design, and performance of a machine-learning approach bases on a support vector machine (SVM) are also presented.

II. Materials and Methods

A. Theoretical Approach

One classification strategy has been implemented in the present work for the automatic identification of cough events, drawn from audio signal data coming from recordings. The strategy method may be viewed as a machine-learning algorithm since it requires a training stage and a posterior testing stage. Specifically, the implemented algorithm is based on a support vector machine.

1) Support vector machine classification: The Support Vector Machine (SVM) theory is a new statistical technique and has drawn much attention to this topic in recent years. An SVM is a binary classifier that makes its decisions by constructing an optimal hyperplane that separates the two classes with the largest margin. It is based on the idea of structural risk minimization (SRM) induction principle that aims at minimizing a bound on the generalization error, rather than minimizing the mean square error. For the optimal hyperplane ,$w*x+b=0$, $w \in R^N$ and $b \in R$, the decision function of classifying a unknown point x is defined as:

$$f(x) = sign(wx+b) = sign(\sum_{i=1}^{N_s} \alpha_i m_i x_i * x) \quad (1)$$

where N_s is the support vector number, x_i is the support vector, α_i is the Lagrange multiplier and $m_i \in \{-1,+1\}$ describes which class x belongs to. In most cases, searching suitable hyperplane in input space is too restrictive to be of practical use. The solution to this situation is mapping the input space into a higher dimensional feature space and searching the optimal hyperplane in this feature space. Let $z = \Phi(x)$ denote the corresponding feature space vector with a mapping Φ from R^N to a feature space Z. It is not necessary to know about Φ.

We just provide a function $K(.,.)$ called kernel which uses the points in input space to compute the dot product in feature space Z, that is

$$z_i * z_j = \Phi(x_i) * \Phi(x_j) = K(x_i, x_j) \tag{2}$$

Finally, the decision function becomes

$$f(x) = sign(\sum_{i=1}^{N_s} \alpha_i m_i K(x_i, x) + b) \tag{3}$$

Functions that satisfy Mercer's theorem can be used as kernels. One of the kernel functions is ANOVA Decomposition Kernels (SVAD) [Vapng], which is a way of imposing a structure on multi-dimensional kernels generated as the tensor product of one-dimensional kernels. This gives more accurate control over the capacity of the learning machine (VC-dimension).

B. Experimental Procedure

The experimental procedure was performed to quantify the detection success ratio of the described algorithm within a varied spectrum of cough events. To this end, nine participants were enrolled in this study, forming one group with healthy subjects (H, 4 females, 5 males, 23.22 $\pm$ 1.99 years old, 1.70 $\pm$ 0.046 m, 66.87 $\pm$ 4.29 kg), who had no known asthma, flu or any impairment to record the audio (see Table I for further information);

Table I: Summary of healthy participants (H group). BMI, body mass index.

Subject	Age [Years Old]	BMI [kg/m^2]	Gender
H1	23	22.2	Male
H2	22	20.9	Female
H3	22	22.7	Male
H4	23	24.1	Female
H5	22	23.7	Male
H6	20	26.0	Female
H7	25	20.7	Female
H8	25	23.7	Male
H9	27	23.7	Male

Volunteers were first instructed to perform three times an audio recording at least 60 s with cough events in order to train the SVM-based method. A resting period of at least 2 min was carried out between records.

The training stage of the SVM-based method was conducted through of two different approaches: the well-known intra-subject procedure, hereafter referred to as subject-specific training (SST); and the inter-subject procedure, hereafter addressed as standardized parameters training (SPT). The SST model parameters were trained using a leave-one-out cross-validation applied to the three audio records, whereby two records were used for training and the remaining one for validation [13]. The cross-validation analysis was repeated recursively for each subject so that it was repeated for all trials in turn. Regarding the SPT, a different validation approach was undertaken for each study group during this method, the training dataset for each examined subject was collected from the first two trials of the remaining individuals, while the last trial related to the subject under consideration was used for validation later.

C. Data Processing

Data processing was performed offline using Python software (CWI, Holland), and an Asus VivoBook S15 S510UA (Taipei, China) running Ubuntu 18.04. The audio records obtained from telegram platform were first analyzed in order to identify the sound periods within the recordings. To do this, the signal is analyzed for each frame with a sample frequency of 48 kHz to identify the sound events, extract them, remove those signal fragments with silence and generate a new audio signal (see figure 1)

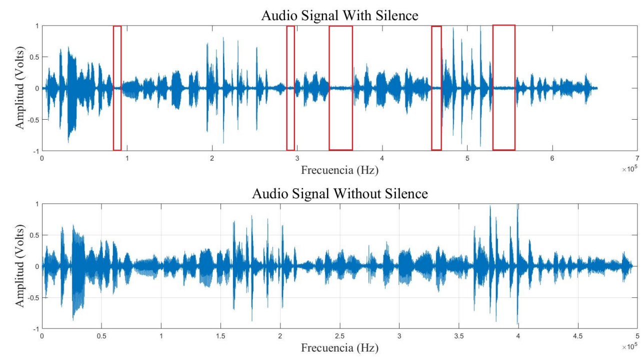

Figure 1: Audio signal with silence and without silence. At the top, the initial signal is presented and there are three wine-red rectangles which refer to signal fragments without sound. In the bottom, the signal is presented with the soundless fragments removed.

The signal is divided into each 1024 sample, and by means of the events found, each sample is classified into two classes: 'cough' or 'without-cough.' It should be noted that the identification of cough start and stop events were performed manually. Then the feature extraction was carried out.

1) Feature extraction: In this study, the Mel-Frequency Cepstral Coefficients was computed from each frame. **Mel-frequency cepstral coefficients** are nonparametric representations of audio signal, which models the human auditory perception system (MFCC). The term "mel" is a unit of measurement of the perceived frequency or pitch of a tone. The mapping between the frequency scale (Hz) and the perceived frequency scale (mels) is approximately linear below 1 kHz and logarithmic at higher frequencies. The suggested formula that approximates this relationship is as follows:

$$F_{mel} = 2595 * log_{10}(1 + \frac{F_{Hz}}{700}) \tag{4}$$

where F_{mel} is the perceived frequency in mels and F_{Hz} is the frequency in Hz.

On the other hand, seven more characteristics related to the each audio signal from each frame were calculated as follows:

1) Room Mean Square (RMS): the signal value (amplitude) is squared, averaged over a period of time, then the square root of the result is calculated.
2) Pitch: estimate fundamental frequency of audio signal.
3) Spectral Centroid: this measure is obtained by evaluating the "center of gravity" using the Fourier transform's frequency and magnitude information, as illustrates the eq.5

$$SpectralCentroid = \frac{\sum_{k=1}^{N} kF[k]}{\sum_{k=1}^{N} F[k]} \tag{5}$$

F[k] is the amplitude corresponding to bin k in DFT spectrum.
4) Entropy: estimated as shows eq.6

$$Entropy = -\sum_{i=1}^{128} P(x_i)\log_2 P(x_i) \tag{6}$$

5) Harmonics: this measure is estimate as multiples of the signal fundamental frequency.
6) Frequency: estimate frequency of audio signal.
7) Width: estimate width of audio signal.

In total, in this study twenty-seven data sets corresponding to the audio signals were generated, i.e., three data sets for each participant.

D. Data Analysis

The performance of the proposed cough detection methods was evaluated through the goodness index (G). G represents a Euclidean distance in the receiver operating characteristic (ROC) space, which poses as a global index of the classifier capability and is based on its sensibility and sensitivity values [14]. The sensitivity, also known as true positive rate (TPR), is computed as follows

$$TPR = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{7}$$

where a true positive is considered if the classifier prediction and the reference value match, in such a way that a cough event was counted as detected even though it was missed by around six samples. Otherwise, such classification is considered a false positive. Likewise, the specificity, also known as true negative rate (TNR), is computed as follows

$$TNR = \frac{True\ Negative}{False\ Positive + True\ Negative} \tag{8}$$

where the non-transitions similarly detected by classifier and reference signal correspond to true negative; otherwise, they have been accounted for by false negatives. The use of TPR and TNR metrics have some implications to bear in mind when comparing performance the classification algorithm. For instance, a cough event detection algorithm that labels all samples of a audio recording with the same cough representation, may score around 25% sensitivity and 75% specificity. This sensitivity score is given because the invariant classifier output eventually matches one cough event, whilst the specificity score comes from the fact that classifying all samples in the same way will produce true negatives for most samples [**?**]. Having these premises in mind, G is expressed as:

$$G = \sqrt{(1-TPR)^2 + (1-TNR)^2} \tag{9}$$

G can assume values between 0 and $\sqrt{2}$, and a classifier can be considered: (i) optimum when $G \leq 0.25$; (ii) good when $0.25 < G < 0.7$; (iii) random if $G = 0.7$; and (iv) bad if $G > 0.70$ [15].

E. Statistical Analysis

The software package SPSS (IBM-SPSS Inc., Armonk, NY, USA) was used for the statistical analysis. The normal distribution of all performance indices was first verified by means of the Shapiro–Wilks test. Since most data did not exhibit a normal distribution, Friedman tests were carried out to find statistically significant differences. Statistical difference was a set at 0.05.

III. Results

Table II illustrates the overall of each characteristic of each participant through its mean and standard deviation according to each class (i.e., no cough and cough).

Table II: Signal audio characteristic (mean ± std) for both class no cough and cough

Signal characteristics	No Cough	Cough
Rms	0.106 ±0.056	0.085 ±0.048
Pitch	130.41 ±0.144	116.51 ±0.056
Spectral Centroid	22.67 ±0.056	31.78 ±0.086
Entropy	0.211 ±0.128	0.198 ±0.125
Harmonics	5.668 ±0.134	6.083 ±0.112
Frequency	-0.001 ±0	-0.0006 ±0
Width	155.55 ±3.45	147.63 ±2.45
First MFCC	5.749 ±0.231	6.099 ±0.318
Fourth MFCC	0.231 ±0.173	0.264 ±1.171

In figure 2 the mean values and standard deviations of the Goodness index (G) for SVM classifier is reported. In addition to that, statistically significant differences among both training classifier are included. For both training methods G values were found to be in the good range ($0.25 \leq G \leq 0.7$), with the best classification performance presented using intra-subject cross-validation (SST). Comparing the SST and SPT procedures, significant differences between the two training approaches were observed. The standard deviations were found to always be less than 0.2, with higher values reached by the SVM-based algorithm trained by means of standardized parameters.

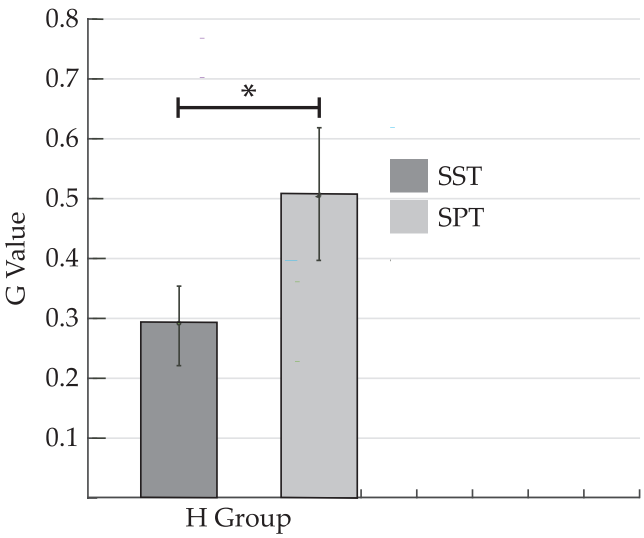

Figure 2: Goodness index (G) for nine subjects volunteers (H) using subject-specific training (SST) and standardized parameter training (SPT). Asterisks indicate statistically significant differences among both methods ($p < 0.05$).

IV. Discussion

In the first place, according to the pre-processing, it was observed how, in the initial audio signal, the fragments without sound were identified, which present a deficient level of amplitude in volts. Therefore, a threshold was established to identify the events with sound and without sound. Similarly, it was observed that those events without sound present a linear behavior. In the pre-processed audio signal, it is observed that these behaviors disappear, the signal is reduced, and finally, the compact signal is obtained with all events that present sound.

According to the feature extraction, it is observed that there is an increase in the average of the data of the following characteristics: RMS, pitch, entropy, frequency, and width, in the events that do not present cough. In contrast, the characteristics: spectral centroid, first and fourth MFCC coefficient, and the harmonics of the signal presented a more excellent value in the mean of the data in the signal events that present a cough.

Finally, regard to the goodness index (G), which involves both specificity and sensibility values in order to precisely evaluate the classifier performance, G values within the good and excellent range were observed for both training methods. Even though a significant difference was found between training techniques, which validates the presumption formulated based on the accuracy values, the inter-subject approach based on standardized data drawn from healthy subjects still has an acceptable segmentation performance (see Figure 2). Apart from that, SPT outperforms SST in terms of training time, since there was a significant difference in the amount of time spent in the training stage between SVM classifier. Therefore, the use of standardized parameters extracted seems to lead to a sufficiently robust trained model that may even match the performance achieved by models constructed with SST without the necessity of continuously carrying out a training stage.

V. Conclusions

This paper features the validation of the automatic recognition algorithm in a system that only involves an audio signal record: an SVM-based algorithm. It is trained through two modalities, namely an intra-subject approach and an inter-subject procedure based on standardized data from the participants. Results in terms of goodness index indicate the best classifier is the SVM-based method trained with intra-subject technique.

For future works, the real-time implementation of the SVM-based detection is expected to be deployed. In doing so, we expect to be able to compare both classifiers (i.e., offline and online classifiers) in terms of latency, i.e., the time needed to online detect each cough phase, since the practical implications of the timing errors in a clinical scenario can be crucial to detecting cough pathologies.

References

[1] S. M. Schappert and C. W. 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*, no. 159, pp. 1–66, feb 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16471269

[2] K. Lai, H. Shen, X. Zhou, Z. Qiu, S. Cai, K. Huang, Q. Wang, C. Wang, J. Lin, C. Hao, L. Kong, S. Zhang, Y. Chen, W. Luo, M. Jiang, J. Xie, and N. Zhong, "Clinical Practice Guidelines for Diagnosis and Management of Cough-Chinese Thoracic Society (CTS) Asthma Consortium." *Journal of thoracic disease*, vol. 10, no. 11, pp. 6314–6351, nov 2018. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/30622806 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6297434

[3] P. Cullinan, "Persistent cough and sputum: prevalence and clinical characteristics in south east England." *Respiratory medicine*, vol. 86, no. 2, pp. 143–9, mar 1992. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1615181

[4] A. H. Morice, G. A. Fontana, M. G. Belvisi, S. S. Birring, K. F. Chung, P. V. Dicpinigaitis, J. A. Kastelik, L. P. McGarvey, J. A. Smith, M. Tatar, and J. Widdicombe, "ERS guidelines on the assessment of cough," *European Respiratory Journal*, vol. 29, no. 6, pp. 1256–1276, mar 2007. [Online]. Available: http://erj.ersjournals.com/cgi/doi/10.1183/09031936.00101006

[5] U. R. Abeyratne, V. Swarnkar, R. Triasih, and A. Setyati, "Cough Sound Analysis - A new tool for diagnosing Pneumonia," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2013, pp. 5216–5219. [Online]. Available: http://ieeexplore.ieee.org/document/6610724/

[6] 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, p. e0162128, sep 2016. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0162128

[7] Y. Shi, H. Liu, Y. Wang, M. Cai, and W. Xu, "Theory and Application of Audio-Based Assessment of Cough," *Journal of Sensors*, vol. 2018, pp. 1–10, 2018. [Online]. Available: https://www.hindawi.com/journals/js/2018/9845321/

[8] S. Subburaj, L. Parvez, and T. Rajagopalan, "Methods of recording and analysing cough sounds," *Pulmonary pharmacology*, vol. 9, no. 5-6, pp. 269–279, 1996.

[9] J. Korpáš, J. Sadloňová, and M. Vrabec, "Analysis of the cough sound: an overview," *Pulmonary pharmacology*, vol. 9, no. 5-6, pp. 261–268, 1996.

[10] J. Hsu, R. Stone, R. Logan-Sinclair, M. Worsdell, C. Busst, and K. Chung, "Coughing frequency in patients with persistent cough: assessment using a 24 hour ambulatory recorder," *European Respiratory Journal*, vol. 7, no. 7, pp. 1246–1253, 1994.

[11] A. Chang, R. Newman, J. Carlin, P. Phelan, and C. Robertson, "Subjective scoring of cough in children: parent-completed vs child-completed diary cards vs an objective method," *European Respiratory Journal*, vol. 11, no. 2, pp. 462–466, 1998.
[12] P. D. Wasserman, *Advanced methods in neural computing*. John Wiley & Sons, Inc., 1993.
[13] J. Taborri, S. Rossi, E. Palermo, F. Patanè, and P. Cappa, "A Novel HMM Distributed Classifier for the Detection of Gait Phases by Means of a Wearable Inertial Sensor Network," *Sensors*, vol. 14, no. 9, pp. 16 212–16 234, sep 2014.
[14] J. Taborri, E. Scalona, S. Rossi, E. Palermo, F. Patane, and P. Cappa, "Real-time gait detection based on Hidden Markov Model: Is it possible to avoid training procedure?" in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*. IEEE, may 2015, pp. 141–145.
[15] J. Taborri, E. Scalona, E. Palermo, S. Rossi, and P. Cappa, "Validation of Inter-Subject Training for Hidden Markov Models Applied to Gait Phase Detection in Children with Cerebral Palsy," *Sensors*, vol. 15, no. 9, pp. 24 514–24 529, sep 2015.