33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis

Brian H. Tracey, Germán Comina, Sandra Larson, Marjory Bravard, José W. López, and Robert H. Gilman

Abstract— **In regions of the world where tuberculosis (TB) poses the greatest disease burden, the lack of access to skilled laboratories is a significant problem. A lab-free method for assessing patient recovery during treatment would be of great benefit, particularly for identifying patients who may have drug-resistant tuberculosis. We hypothesize that cough analysis may provide such a test. In this paper we describe algorithm development in support of a pilot study of TB patient coughing. We describe several approaches to event detection and classification, and show preliminary data which suggest that cough count decreases after the start of treatment in drug-responsive patients. Our eventual goal is development of a low-cost ambulatory cough analysis system that will help identify patients with drug-resistant tuberculosis.**

I. Introduction

Although effective therapies have been available for many decades, tuberculosis (TB) remains among the world's most deadly infectious diseases. The emergence of multi-drug resistant (MDR) and extremely drug-resistant (XDR) TB is a major cause for concern. MDR TB accounts for roughly 3.6% of all TB cases, but accounts for as much as 28% in some regions [1].

In much of the developing world, diagnosis of pulmonary TB is made exclusively by sputum smear due to lack of access to skilled laboratories and culture-based methods. Patients are generally put on first-line therapy, and treatment failure (whether due to the presence of MDR TB, or other reasons) is only discovered 4-6 months later. Patients who fail treatment continue to be infectious, spreading disease to others and running increased risks of morbidity and mortality.

A lab-free method to identify patients who are failing treatment would be of great benefit to clinicians without access to laboratory culture. Previous studies have indicated that cough rates (counts/hour) drop by roughly 50% in the first two weeks of treatment for patients who are responding to treatment [2], thus providing a potential means for detecting treatment failure. The physical infrastructure for an automated cough monitoring system is relatively simple (low-cost recording units and access to computing power, either locally or via telecommunications).

This work builds on recent progress in developing low-cost ambulatory systems that can record for extended time periods [3,4]. We also build on past work in automated cough counting [5-7]. Fully automated analysis, while key for protecting patient privacy, is challenging as patient recordings often include a large amount of environmental noise. This is particularly true in our dataset, where patients are recorded going about their daily activities, and extraneous noise (speech, traffic, bangs, etc.) is common.

In this paper we report on a pilot data collection and development of a cough detection algorithm. For this phase of the project, our goal is to develop an algorithm that is sufficiently accurate to evaluate the clinical utility of TB patient cough analysis. In the long term, we seek to develop a fully automated cough analysis system.

Similar to previous work [5-7], our algorithm consists of event detection followed by cough classification based on time-frequency analysis features commonly used in speech processing. We discuss several event detection approaches, including a novel shape-based detector, and present results from several candidate classifiers. While a full algorithm validation is ongoing, initial results are presented below.

II. Methods

A. Data collection

The data collection was conducted at Hospital Nacional Dos de Mayo in Lima, Peru. This public national tertiary referral hospital also operates as a community hospital for the surrounding inner-city area. This study received IRB approval from Hospital Dos de Mayo, Associacion Benefica Prisma (Lima, Peru), and Johns Hopkins University. Subjects were provided transportation to the hospital and nutritional supplementation throughout the study. The dataset consists of a series of 24-hour acoustic recordings, made using a Marantz PMD 620 handheld recorder and an Audio-Technica AT899 sub-mini microphone attached at the patient's lapel. A 24-hour recording was made before start of treatment (day 0) to establish a cough baseline, and 24-hour recordings were made at each subsequent visit (days 21, 30 and 60 after start). Additionally, extended recordings during the first 14 days are available for many patients.

B. Tracey is with Tufts University, Medford MA 02155 USA (phone: 617-519-4390; fax: (617) 627-3220; e-mail: btracey@eecs.tufts.edu).

G. Comina is with Laboratorio de Ingeniería Física, Facultad de Ciencias, Universidad Nacional de Ingeniería, Rimac, Lima, Perú.

S. Larson is with Michigan State University, College of Osteopathic Medicine, East Lansing, MI USA.

M. Bravard is with the Dept. of Internal Medicine, Massachusetts General Hospital, Boston MA 02114.

J. Lopez are with the Unidad de Epidemiología, Hospital Nacional Dos de Mayo, Lima, Perú.

R. Gilman is with the Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

The study includes drug-susceptible TB patients, MDR TB patients, and HIV/TB patients (HIV is a complicating factor in treatment of TB). 62 patients have completed the study. Of those, 54 patients are HIV- (5 MDR TB, 49 drug-susceptible) and 8 are HIV+ (5 susceptible, 1 negative, 1 pending). Extensive clinical information is linked to this dataset, such as patient weight, temperature, sputum smear microscopy results, MODS culture results, CD4 count (for HIV+ subjects) and a description of symptoms at each visit.

An example of recorded cough, plotted using our analysis software, is shown in Figure 1. As seen in the figure, coughs typically exhibit an initial explosive phase with a very sharp increase in energy during as air is released. A second phase follows in which voicing (vocal cord vibration) may occur.

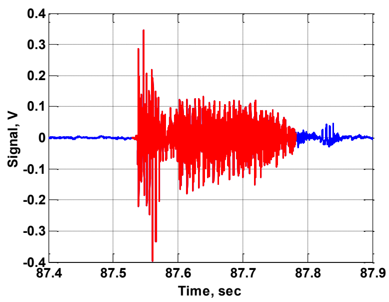

Figure 1. Example of prototypical cough.

B. Algorithm description

Figure 2 shows our overall algorithm flow, which is generally similar to [5-7]. In analyzing a recorded signal, the first step is event detection. We describe two approaches to event detection below. Once an event is detected, it is broken up into 50% overlapped, 32 ms long frames. And acoustic features for each frame are calculated. Similar to [5-7] we calculate the Mel frequency cepstral coefficients (MFCC) as the primary features for acoustic analysis. These are given by:

$$c(n) = \sum_{k=1}^{K} log(S_k) \cos\left(\frac{n\pi}{K}(k-0.5)\right), n = 1,2, \dots N \qquad (1)$$

where N is the number of coefficients to be calculated. The S_k, are the outputs of K filterbanks, found by weighted sums of the short-time Fourier transform magnitude over a set of frequency bands with center frequencies chosen to approximate the human auditory system response [8]. These coefficients and their first and second time derivatives are then used to form the feature vector:

$$\vec{f}^T = [\vec{c} \quad \dot{\vec{c}} \quad \ddot{\vec{c}}] \qquad (2)$$

As will be described below, the features for individual frames are fed into a classifier (trained offline) to determine whether or not the frame has the characteristics of cough. The decisions for individual frames are then combined into an overall decision for the event. The features above were also used by [5-7].

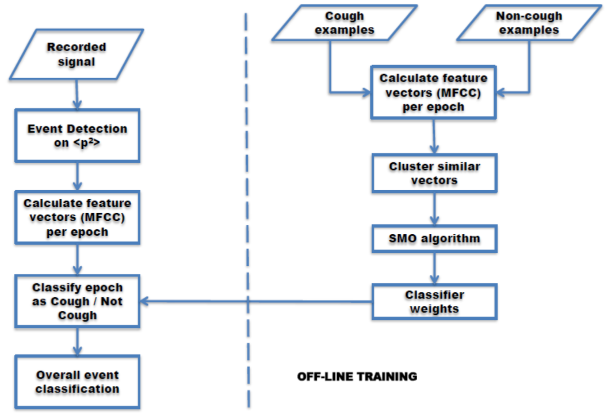

Figure 2. High-level flowchart of cough detection algorithm

Event detection

In our first implementation, similar to [5-7], acoustic events within each file were detected based on signal energy. A signal envelope was estimated by squaring the input signal and smoothing it with a running average filter (a boxcar filter of length 0.05 s). The detector then identified events as regions of time where the signal exceeds a fixed threshold. If two events were separated by a very small gap in time (0.1 s), we merged them together into one event.

While reasonably effective, this approach suffers from several shortcomings, as seen in Figure 3. In some cases (left panel) the energy threshold is not triggered until partway through the event. This is undesirable as the initial portion of the event is important for cough classification. Another common occurrence, seen in the right panel, is that the energy detector triggers on speech signals. These signals can have significant energy, but generally exhibit a much more gradual increase in signal energy than is the case for cough. In the large majority of cases, speech signals are correctly classified as non-cough, but detecting speech events does create a small number of false positives.

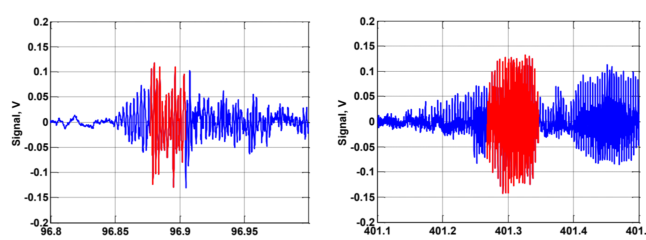

Figure 3. Example issues with simple energy detector. The threshold may miss the start of the acoustic event (left) and is frequently crossed during speech events (right), increasing the chance of misclassification.

We therefore developed an improved detection scheme in which we search for a rapid increase in signal energy above a time-varying estimate of the ambient noise. The time-varying noise estimate is helpful as our dataset was collected in a wide range of acoustic environments; looking for a rapid rise in energy is helpful in screening out speech events.

The smoothed signal energy is first calculated, as above. A background noise estimate is found by calculating the 10th percentile of energy in a sliding 20 sec time window. A fixed multiplier is added to approximate the median noise during quiet regions. To reduce computation, percentiles were calculated for a subset of windows with 75% overlap and then interpolated to each time sample. Parameters for the noise estimation were determined empirically.

An initial set of events is then found by finding all times when the signal energy exceeds the noise floor by a fixed threshold (9 dB). The start and end samples for each event are found, as well as the peak energy in the event. For loud events (peak > 9 dB above noise) this underestimates the event duration. Therefore, the event start is adjusted earlier until the signal energy either stops decreasing or reaches the noise floor. A similar adjustment is made for the endpoint. If two events become overlapped as a result of these adjustments, the events are combined. The onset is next validated by calculating the median energies for 1/8 sec before and after the event onset. If the energy increase after onset is > 6 dB, the detected onset is considered to be valid.

Off-line classifier selection and training

A subset of the recorded data was used as training data. Because of data availability at the time training was done, the training data is mainly taken from two male subjects.

During training, an event detector was used to identify events, which were manually reviewed classified as 'cough', 'not cough', or 'unclear'. To build a wide library of cough and non-cough events, the simple energy threshold detector described above was used, with a low threshold to allow sensitive detection of events. The training data contained 418 cough events, 1980 'not cough' events, and 75 'unclear' events. The 'unclear' events were generally combinations of coughs and other sounds (groans, speech, etc.) such that they were felt to be unsuitable for training purposes.

Once each event was classified, the MFCC features were calculated as described above. The 'cough' or 'noncough' label from the overall event was then applied to all frames extracted from the event. In total, 13,429 cough frames and 43,925 non-cough frames were available for training after this step. As a note, this approach does introduce some possibility for misclassification; for example, an overall event might be classified as 'cough' but an individual frame within that event may contain mainly non-cough sounds. The large number of frames used for training was felt to mitigate the effects of any individual misclassified frames.

The large number of frames in the training set increases the computational load for machine learning algorithms. To address this, we used a previously developed 'divide and conquer' clustering method suitable for large datasets [9]. In this clustering approach a subset of the MFCC vectors is first analyzed. Vectors with high correlation, defined as a Pearson correlation coefficient of 0.95 or better, are clustered together. A second subset of vectors is extracted, and any of the new vectors that have high correlation with the previously identified clusters are assigned to them. Additional clusters are formed using the remaining feature vectors. This process is repeated until the full dataset is clustered. After clustering, a single vector (the cluster center) is used to represent each cluster. For our training data, this yielded 2074 vectors, representing half cough events and half non-cough (because there were more non-cough clusters, small non-cough clusters were discarded).

Classifier training was done using the Weka 3.6 software package [10], which implements different machine learning algorithms. We compared performance of neural networks (multilayer perceptrons, or MLP), support vector machines (SVM), and sequential minimal optimization (SMO), a variant of the SVM approach. Performance results obtained using 10-fold cross-validation is shown in Table 1. While the multilayer perceptron gave slightly better results, the performance differences were not large and the SMO approach was chosen for ease of implementation.

	MLP	SMO	SVM
Correct cough	906	903	949
Correct non-cough	924	888	875
False alarm cough	113	149	162
Missed cough	131	134	88
Overall accuracy	88.2%	86.4%	88%

Table 1. Frame-by-frame classification results for training dataset.

On-line event classification

During on-line processing, detected events are split into frames and acoustic features are calculated as discussed above. These features and the SMO-generated feature weights are used to classify each frame as cough or not-cough. A final processing step is required to combine the frame-by-frame results in order to classify the overall event. A previously used approach [5] is to average all frames within an event; if most of the frames are found to be cough, the overall event is labeled as cough. We found better performance by first identifying the 1/3 of contiguous frames that have the most 'cough-like' scores, then averaging the classifier outputs for those frames. This was found to help in cases where the detected event contains a mix of cough and other vocalizations.

III. Results

An initial review of algorithm performance was made by running the algorithm on a randomly selected set of 30 files from 10 patients, each 30 minutes long. Each event detected by the algorithm was manually reviewed and marked as cough or non-cough. Based on this review, the algorithm sensitivity (percent of events manually identified as coughs, and labeled as such by the algorithm) was 81%. The average false positive rate across files was 3.3 / hour; the median false positive rate was 1 / hour. In all but four cases, the false positive rate was below 4/hour; however, in some especially noisy files the algorithm performed poorly, with a maximum false alarm rate of 27/hour.

This methodology does not capture cases in which coughs were missed by the algorithm's event detection logic. A more detailed validation is therefore in progress. In this review, manual cough assignments were made by two nurses for comparison to the algorithm. Coughs for which both nurses are in agreement were used for validation.

Figure 4 shows cough count vs. time for a set of 28 patients with drug-sensitive TB. The observed decrease in cough count (roughly 50% over the first 2-3 weeks) is consistent with earlier results from Loudon *et al*. However, significant scatter is seen in individual patient records. For example, patient 111's cough decreases steadily over time, but patient 129 shows a decrease followed by a spike at day 30. Without manual review, it is unclear whether this increase is real or is due to false positives.

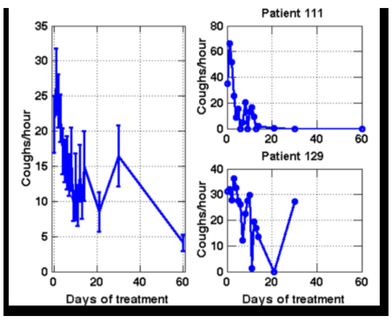

Figure 4. TB cough count over time. Left panel shows mean and standard error calculated over a set of 28 drug-susceptible patients. Right panel shows results for two individual patients, one with unambiguous recovery and another with more scatter in the data.

IV. Discussion

We have developed a cough analysis system that builds on previous approaches [5-7] and have applied it to a cohort of drug-sensitive TB patients. In a preliminary review, the algorithm shows 81% sensitivity and an average of 3.3 false alarms/hours; a more complete validation is ongoing.

Figure 4 shows both the promise of cough-based TB monitoring and the need for further development. A decrease in cough count is clear, but before the algorithm can be clinically applied it is important to improve performance so we are confident that fluctuations in cough count are real. This will require improvements in algorithm performance, though metrics that quantify when a recording is too noisy for reliable analysis would also be valuable.

Several concepts introduced above may prove useful for other cough analysis approaches. Previous methods [5-7] require some degree of manual review. Our event detection approach is useful in the context of manual review, as it reduces the number of speech events that are detected as cough candidates. The 'divide-and-conquer' clustering algorithm described here may prove useful for applications where the classifier is updated as new data become available, as it only requires storing a single representative of each previously identified cluster.

Several possibilities exist for improving algorithm performance. In the work to date, our event detection logic accounts for the expected cough shape (rapid energy increase at onset) but the classifier does not (although some information about neighboring time periods is encoded in the MFCC time derivatives). A natural approach is to investigate use of the Hidden Markov Model (HMM), as in [7-8], which can track the time evolution of the cough sound. The HMM or other approaches that analyze the entire event may have performance advantages, especially for isolated cough events. On the other hand, it is possible that the frame-by-frame analysis used here may have advantages when the detected events contain overlapped coughs and other sounds. Improvements might also be possible by assigning different categories of non-cough noises to different classes, rather than lumping them into a single 'non-cough' class as is done here. Finally, we are investigating new sensors that may be more robust to noise.

V. Conclusions

In this paper we have described data collection and initial algorithm development in support of a cough analysis system for tracking the recovery of pulmonary tuberculosis patients. The final goal is a low-cost monitoring system that could be used in areas where laboratory facilities are not accessible, and could alert physicians to patients who may have drug-resistant tuberculosis. We have developed an initial cough detection algorithm, and have on-going work to fully validate and improve algorithm performance. Preliminary results show a decrease in cough count over time for drug-susceptible patients, suggesting that cough count may provide a useful marker of patient recovery.

Acknowledgment

This work was funded in part by National Institutes of Health award 5D43TW006581 "Infectious Diseases Training Program in Peru". JL and MB were supported by NIH Fogarty Foundation fellowships.

References

[1] World Health Organization (WHO),"Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response," WHO/HTM/TB/2010.3.

[2] R.G. Loudon and S.K., Spohn, "Cough frequency and infectivity in patients with pulmonary tuberculosis," Am Rev Respir Dis.,1969;99:109–11.

[3] J.A. Smith and A. Woodcock, "New Developments in the Objective Assessment of Cough," Lung, 2008; 186(1):S48-S54.

[4] J.A. Smith, "Ambulatory methods for recording cough," Pulmonary Pharmacology & Therapeutics, 2007; 20:313-318.

[5] S.J. Barry, A.D. Dane, A.H. Morice, A.D. Walmsley, "The automatic recognition and counting of cough," Cough, 2006; 2:8.

[6] S. Matos, S.S. Birring, I.D. Pavord, D.H. Evans, "Detection of cough signals in continuous audio recordings using hidden Markov models," IEEE Biomed Eng 2006; 53: 1078–1083.

[7] S. Matos, S.S. Birring, I.D. Pavord , D.H. Evans, "An automated system for 24-hour monitoring of cough frequency: the Leicester cough monitor," IEEE Trans Biomed Eng 2007; 54:1472–1479.

[8] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html, by Michael Brooks, Imperial College, London.

[9] B.H. Tracey, N. Lee, and S. Turaga, "Cluster analysis and robust use of full-field models for sonar beamforming," J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006, pp. 2635-2647.

[10] I.Witten and E. Frank, *Data Mining*, 2nd Edition, 2005, Morgan Kaufmann.