Classification of Cough Events

By Uyen Diep (Sally), ECE '17

In parts of the world with limited access to laboratory resources, a reliable lab-free tool to track recovery of patients with respiratory diseases, particularly tuberculosis (TB), is very much needed. Cough analysis as part of an automated cough monitoring system for patients has been developed to achieve this goal, and is usually divided into two main stages, event detection and event classification. This overview discusses the widely used method for classification of cough event from past studies and one potentially new method, as well as analyzes their advantages and limitations against each other.

Background

Despite being a treatable and curable disease, TB is one of the top 10 death causes worldwide with millions of new cases [1] every year. The emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) has increased the difficulty of treating TB as more limited resources are required. In developing countries with lack of access to skilled laboratories, having a lab-free tool to monitor patients cough for early detection of treatment success or failure will be of great benefit.

The assessment of cough frequency and intensity from continuous ambulatory recordings has remained difficult for both clinical and research purposes as cough varies substantially from person to person, making the manual examination of recordings a very slow and tedious job. Several research has then been done to develop an automated and reliable cough detection algorithm to detect cough sounds. The goal is to build a system with high sensitivity to reduce the processing time and eliminate the need for trained observers to manually analyze recording data to some extent. The metrics for quantifying cough is usually coughs/hour (cough count), cough episodes (cough events closely spaces in time) or cough epoch (cough events separated by less than 2 seconds) [2].

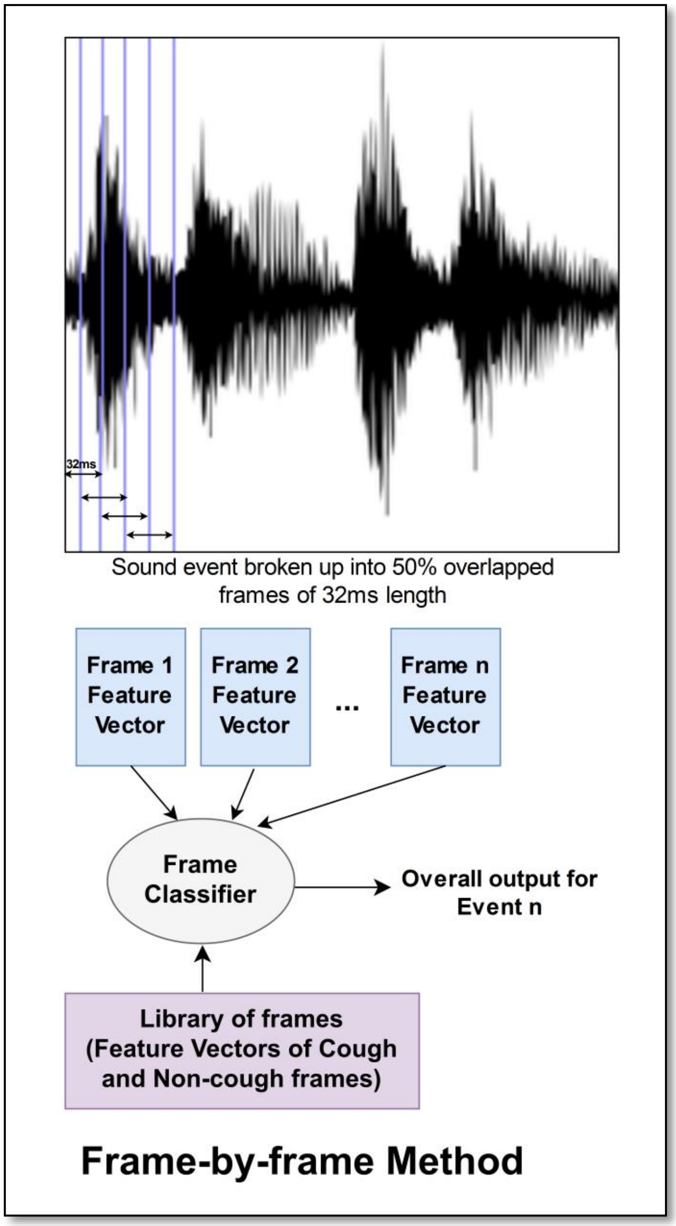
Cough analysis using digital signal processing presented in previous approaches [2-4] consists of two main stages, event detection and event classification. The signal energy is first calculated and compared to a threshold to detect sound events (high energy) and discard unsound events (low energy), then these sound events are classified into cough and non-cough events (such as speech or other noises). This paper will focus on methods of classification of cough events in the second stage, given a detected non-silent input signal.

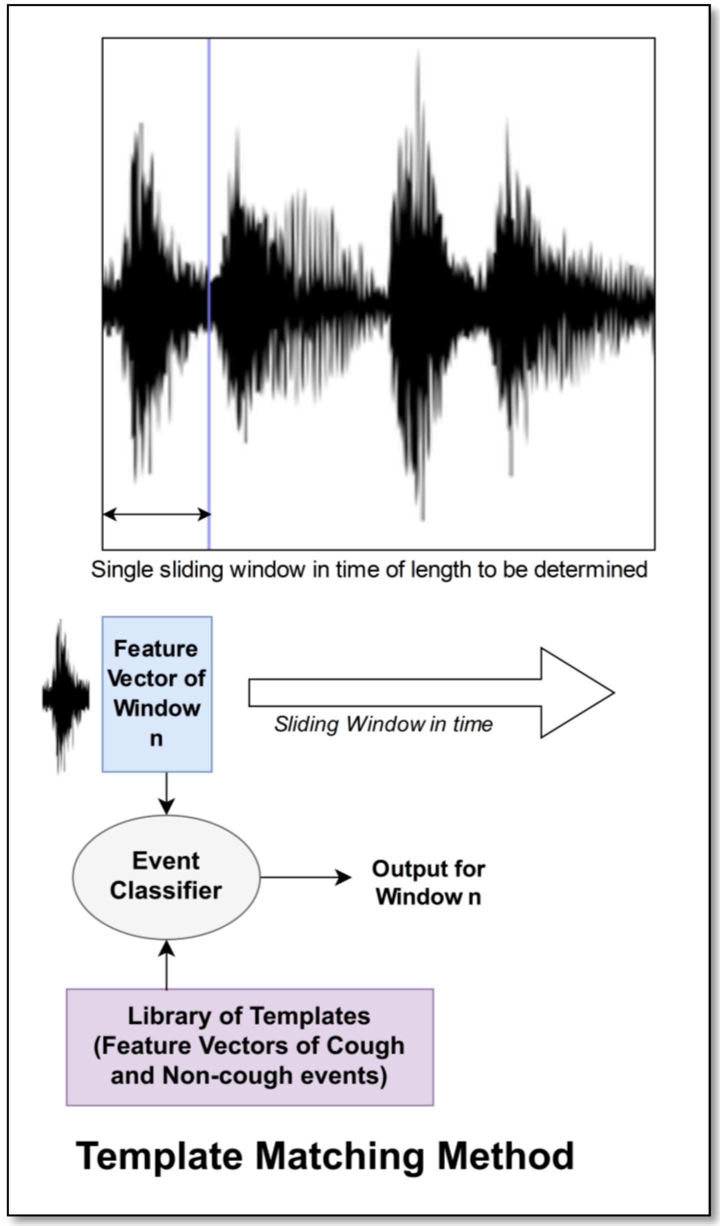
The Frame-by-frame Method

In this method, sound events are broken up into overlapped time frames of 32ms length [2] or 16ms length [4]. The features of each frame are calculated and fed into a classifier to determine if the frame has characteristic of a cough. The results are then combined to make an overall decision for the event.

Feature Extraction

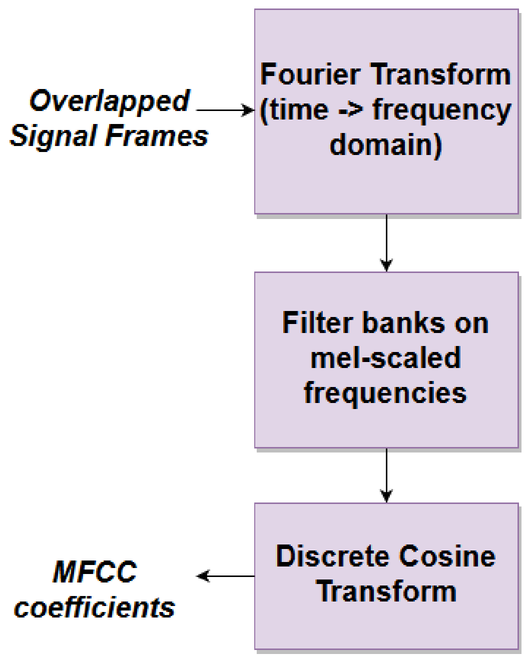
The source-filter model [5] is an important model for speech recognition, and can be applied to cough analysis. The model describes the speech production





as a generation of a sound source (vocal cords) mainly responsible for the pitch of a sound, that is then shaped by a filter (vocal tract) to actually produce different sounds. These two processes are independent of each other, i.e. the filter for the same word in a high pitch female voice and a low pitch male voice should be the same. Since speech analysis often cares about the different sounds or words generated rather than the pitch, the first step is to separate these two processes and characterize the filter, or feature extraction.

In other words, the feature of a sound in an individual frame can be characterized by using a parametric representation of the filter shape within that frame. Mel frequency cepstral coefficients (MFCC) is a widely used parameterization method in speech recognition applications to perform this task. MFCC are coefficients that make up the power spectrum of a sound on a mel scale frequency, which is a scale of pitches that reflects how people hear different tones. The calculation steps are implemented to transform the features into K coefficients (usually K = 13) to represent all the speech samples within that frame. To improve the recognition performance, the first and second time derivatives of the MFCCs are also derived and combined with the zeroth-order coefficients into a single feature vector of K = 39 parameters represented each time frame [2,4].



Mel-frequency cepstrum calculation steps

Event-classification with off-line training

A library of cough features for individual frames is first created. Previously recorded data used as training data is manually reviewed and labeled as coughs or non-coughs. The feature vectors of the frames are then calculated to comprise a large library of cough feature vectors stored offline. To reduce the size of the dataset and reduce the number of computations, similar data (high correlation) is clustered into one group represented by a single feature vector.

Each input frame fed into this classifier will be compared to the feature vectors in the library to determine whether the frame has characteristic of a cough, represented by a probability value. In order to perform fast pattern recognition and decision-making, different machine learning algorithms (such as probabilistic neural network [3]) can be used to store feature vectors and trained to effectively classify cough events. The results for all individual time frames within a sound event are then combined to get the overall decision for that event. To avoid the misclassification when the detected event contains a mix of cough and other sounds, the overall decision is made by only looking at 1/3 of contiguous frames with the highest probabilities [2].

Although this method provides a robust classification tool independent of amplitude (loudness), its performance can be further improved by taking into account the rapid increase in signal energy at the onset of a cough. Therefore an approach that can consider the changes between frames or make decision based on an entire event may have performance advantages.

The Template Matching Method

In this method, a complete sound event is analyzed by template matching. This approach has been used for automated detection of rat ultrasonic vocalizations at 50 kHz and is capable of extracting >90% of the recorded signals [6].

Instead of breaking signals into distinct frames, the input signal is scanned and compared to a template over time. Since cough and speech signals are broad band (distributed over a variety of frequencies), the method performed in [6] where templates are only within a specific range of frequencies for ultrasound is not applicable. One possible approach is to create a template with extracted cough feature vectors for a whole cough event and compare to those calculated from the input events at each time lag. This allows the detector to quantify the acoustic similarity between sound event and the template, and a cough is detected if the similarity (correlation between two vectors) is above a fixed threshold.

The performance can be increased by having a set of templates stored in a library automatically updated when new cough data is found. It is also noted that template matching fails with low signal-to-noise ratio because the presence of noise can distort the signal significantly, so the human scoring still plays an important role in this case.

The limitations of this method are the inflexibility of threshold and difficulty in creating cough templates given the immense variability in cough sounds. The accuracy of the detector also decreases significantly with the presence of ambient noise [6] so input signals need to be cleaned up before going into the detector. It is possible that this problem can be fixed by using different recording hardware and setup that is less susceptible to ambulatory noises.

Discussion

Given a continuous recording with lots of overlaps of cough sounds and impulse noises, the frame-by-frame approach yields better result as each frame is analyzed separately; however the relationship between each frame is not quite utilized. On the other hand, template matching approach may give better performance using fewer computations for isolated cough events, however the templates need to capture the variability of cough sounds.

One approach to combine the advantages of both methods is the use of Hidden Markov Model (HMM) as a classifier. MFCCs are still calculated for each signal frame, but the Markov model can be trained to take into account the time-varying characteristics of cough events to detect the occurrence of cough sounds [4]. Other improvements include the use of more robust microphones to raise the signal-to-noise ratio, or more complex event classification methods to identify and reject similar cough sounds such as throat clearing or sneezing.

Lastly, one major change in hardware implemented by the Periwinkle team this year is the integration of multiple sensors in the system, including 1 microphone and 1-4 accelerometers. Hence a new cough detection algorithm that accounts for multiple inputs needs to be developed. This addition of sensors will be advantageous for both event detection and classification, as a cough event is now better defined by multiple different sets of features.

References

1. World Health Organization. (2016). *Global tuberculosis report 2016*. World Health Organization.

2. Larson, S., Comina, G., Gilman, R. H., Tracey, B. H., Bravard, M., & López, J. W. (2012). Validation of an automated cough detection algorithm for tracking recovery of pulmonary tuberculosis patients. *PloS one*, *7*(10), e46229.

3. Barry, S. J., Dane, A. D., Morice, A. H., & Walmsley, A. D. (2006). The automatic recognition and counting of cough. *Cough*, *2*(1), 1.

4. Matos, S., Birring, S. S., Pavord, I. D., & Evans, H. (2006). Detection of cough signals in continuous audio recordings using hidden Markov models. *IEEE Transactions on Biomedical Engineering*, *53*(6), 1078-1083.

5. Klautau, A. (2005). The mfcc. *doi: http://www. cic. unb. br/~ lamar/te073/Aulas/mfcc. pdf*.

6. Barker, D. J., Herrera, C., & West, M. O. (2014). Automated detection of 50-kHz ultrasonic vocalizations using template matching in XBAT. *Journal of neuroscience methods*, *236*, 68-