2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN) | 978-1-6654-0362-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/BSN51625.2021.9507017

CoughBuddy: Multi-Modal Cough Event Detection Using Earbuds Platform

Ebrahim Nemati[1], Shibo Zhang[2], Tousif Ahmed, Md. Mahbubur Rahman[1], Jilong Kuang[1], Alex Gao[1]

***Abstract*—There has been an extensive amount of study on cough detection using acoustic features captured from smartphones and smartwatches in the past decade. However, the specificity of the algorithms has always been a concern when exposed to the unseen field data containing cough-like sounds. In this paper, we propose a novel sensor fusion algorithm that employs a hybrid of classification and template matching algorithms to tackle the problem of unseen classes. The algorithm utilizes in-ear audio signal as well as head motion captured by the inertial measurement unit (IMU). A clinical study including 45 subjects from healthy and chronic cough cohorts was conducted that contained various tasks including cough and cough-like body sounds in various conditions such as quiet/noisy and stationary/non-stationary. Our hybrid model was evaluated for sensitivity and specificity in these conditions using leave one-subject out validation (LOSOV) and achieved an average sensitivity of 83% for stationary tasks and an specificity of 91.7% for cough-like sounds reducing the false positive rate by 55%. These results indicate the feasibility and superiority of fusion in earbuds platforms for detection of cough events.**

Keywords—Sensor fusion, template matching, DTW.

I. Introduction

Ubiquitous computing is entering every part of our life. Mobile phones and wearable devices are carried by users almost anywhere at any time, enabling passive monitoring of various health conditions. According to a recent report from lung health institute, pulmonary disease is one of the major leading causes of morbidity and mortality globally [1]. Coughing is a ubiquitous symptom of pulmonary disease, especially for patients with COPD and asthma. Cough frequency and its characteristics are often used to monitor disease activity. There has been a tremendous amount of research work in detecting cough events from recorded audio [2, 3]. Researchers have also tried to characterize these coughs and enable disease detection/prediction [4]. The need for a robust mobile cough counter that could passively monitor patient's lung health seems more serious than ever with the current COVID-19 pandemic. The key reason for low adoption of these algorithms is the poor precision of detection when the algorithms are exposed to in-field data containing unseen cough-like sounds. Various methods to tackle this issue were utilized such as inclusion of labor-demanding labeled field data, utilizing data augmentation techniques and utilizing generative and similarity-based algorithms [5, 6]. However, none of these were able to provide robust detection of cough events while keeping the specificity of the algorithm high in the presence of cough-like sounds. The reason for this limitation is that most of these existing works are unimodal focusing only on audio signals which leads to high false positive rate, e.g., bystanders' cough or a dog barking can be detected as the user's cough. In this work we are relying on the fusion of the audio and IMU signals captured on a newly developed earbud platform. A cough is only detected if cough audio signature is accompanied by the head motion signature captured by the IMU sensors. Our key contributions are:

- To the best of our knowledge, this is the first cough detection algorithm designed for earbuds platform utilizing both audio and IMU.
- A novel sensor fusion algorithm is proposed that combines the acoustic-based classification and the DTW-based cough IMU signature identification to best address the specificity issue.
- The evaluation of the model was done on a dataset that was collected in various conditions of quiet vs. noisy and stationary vs. non-stationary containing cough-like sounds not seen in the training.

In the next section, we will provide a summary of previous related works in the domain. Section III describes the novel hybrid algorithm in details and section IV explains the dataset and the training process to generate the models. Section V presents the results of the evaluation and eventually, a conclusion and discussion of the future work will be provided.

II. Related Works

Research domain for cough event detection from the audio signal from mobile devices has been extensively explored. Most of the previous works rely on feature engineering where specific acoustic features need to be extracted from the raw audio signal. For example, Matos et al. [7] extracted MFCC features to train HMM models. Larson et al. [8] build random forest model using spectrogram-based features and Amoh [9] ran deep learning models on spectrogram images. Most of these works used datasets that included cough and a few other body sounds such as speech and breathing and tried to classify between them. Many of them did not explore the viability of the model when applied to real field data. Xu et al. explored augmentation method to simulate the noisy environment in the field and evaluate the model but did not provide any results on

* Research supported by Samsung Research America (SRA) with partnership of a clinical research organocation for subject recruitment.
E. Nemati is the corresponding author (e.nemati@samsung.com).

1. Research Engineer, Digital Health Lab (DHL), Samsung Research America (SRA), mountain view, CA, 94043
2. PhD Candidate, Northwestern Uni., 633 Clark St, Evanston, IL 60208

the precision of the model when exposed to cough-like sounds such as throat clearing and bystander cough [6]. Alvarez et al. utilized the Hu moments while capturing noisy field data to evaluate robustness of the algorithm [5]. They reported a very high specificity, but a big portion of their noise data was already seen in the training set.

To tackle the problem of high false positive rate (FPR), unlike previous methods in the literature, we propose a hybrid structure of both classification and matching algorithm. Matching algorithms do not need including of all the different sources of cough-like sound in the training. However, they are poor when dealing with rich data such as audio and video signals due to the information loss. But they have been widely used for IMU signals for activity recognition. Use of IMU signal for cough requires placement of IMU somewhere close to the center of mass to capture the cough motion. Earbud platform seems to be the perfect platform for this purpose. Fig. 1 depicts the raw audio signal and its spectrum as well as the accelerometer signal for cough and cough-like sounds such as speech and throat clearing. As can be seen from Fig. 1, while the audio signature of cough could be similar to a cough-like sound, their motion signatures are different.

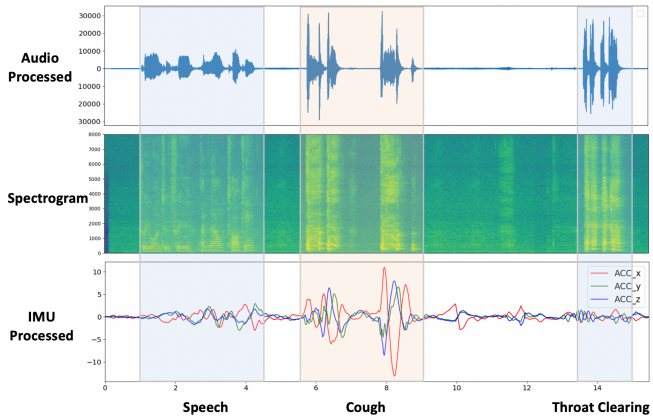

Fig. 1. Acoustic and motion signatures for cough and cough-like events

III. Fusion Architecture for Cough Detection

Feature-based classification algorithms require collection of enough data from different labels of sounds to robustly detect coughs with low false positive rate. Matching algorithms on the other hand rely on the similarity of the input and a template. Therefore, distance threshold could be picked relying only on a limited set of negative and positive samples. Summarizing a video or audio signal to a template, mandates extreme reduction of dimensionality. However, for motion signals, the required information could often be summarized in a narrowband-filtered signal around the frequency of the motion of interest without much loss of information. While matching algorithms mitigate the need for collecting large number of negative samples from a diverse set of classes, they are essentially slow algorithms. Therefore, in our algorithm we exercise matching only when a cough is detected from the audio model. Fig. 2 presents the proposed fusion algorithm architecture. In the first layer, the audio signal is processed where silence episodes are filtered out and a classifier differentiates cough from speech and noise. When a cough is detected, DTW is activated to verify or reject the cough based on its similarity to template's IMU motion signature.

A. Audio-based Cough Classification

The audio-based classifier in the fusion algorithm utilizes a similar structure to the one proposed in Nemati et al [3]. The audio is first preprocessed using a low-pass filter with a corner frequency of 20 KHz (defined by the frequency content of cough signal). Then the signal is normalized using min/max values of the sensor. Preprocessed signal is then passed through a sliding window with 0.4 second window size (matching to average duration of a single cough) and 0.1 second jump size. Each audio chunk is then transformed into the feature domain including temporal and spectral features. Among the features, "SPlevel" feature which represents the area under the curve of the signal amplitude is employed to filter out the silence (and low volume) portion of the signal. The rest is fed to a classifier which differentiates "cough" from "speech" and "noise". Postprocessing is then done to smooth the labels and identify the start and end index of each cough episode. The data used to train the classifier is limited to speech, cough and background noise. Therefore, the audio model is not able to differentiate between cough and cough-like sounds such as bystander cough. Therefore, the specificity of the audio portion of the fusion algorithm should theoretically be low when introduced to the field data. However, we mitigated this issue by using the IMU-based template-matching module described next.

B. IMU-based DTW Template Matching

Elastic distance measurement (esp. DTW) is shown to achieve unbeatable performance for template matching problems [10]. While cough head motion seems to have a unique signature compared to many other activities (Fig. 1), one can imagine the possibility of similar head motion occurring in the normal daily life. Therefore, the key in using DTW for cough detection is to only use it when there is a level of certainty that a cough has happened. That is provided by the audio model in our algorithm. To preprocess the IMU signal, a Butterworth high-pass filter with a corner frequency of 0.3 is used to remove the DC baseline and motion noise. Then a rolling average smoothing filter with the width of 10 samples is employed to remove the high-frequency glitches. The IMU

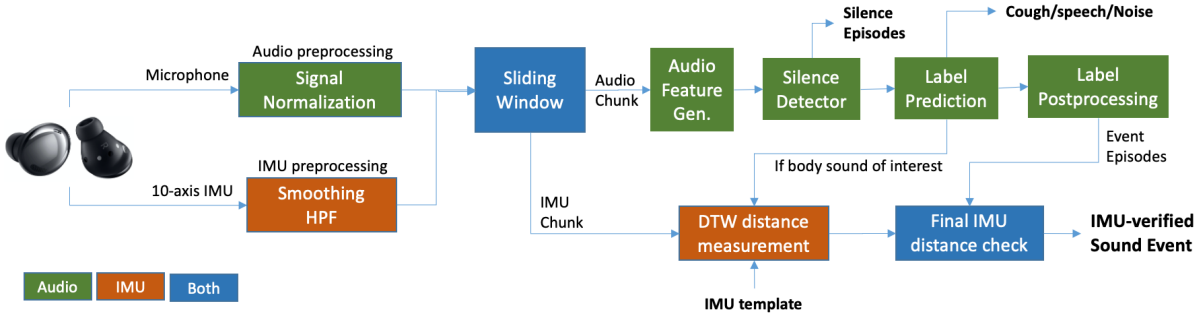

Fig. 2. Fusion cough detection algorithm architecture

chunk selected by the audio channel is fed to a sliding window with 0.1s jump size and 0.2s window size. The DTW distance is then measured for each window and the minimum is used to make the decision whether a cough exists inside the chunk.

IV. Training

To train and evaluate our model, we conducted a study to collect cough and breathing data from a large number of subjects performing coughing and other tasks in various stationary and non-stationary conditions within different background noise profiles while wearing earbuds. The focus of this work is only on the cough and not the breathing.

A. Study Design

Through a collaboration between Samsung Research America (SRA) and a clinical research organization (CRO), 45 participants were recruited under ethic committee approval, including 30 healthy and 15 chronic cough subjects. The study was approved by the Institutional Review Board (IRB) with protocol number IAA-2112. The cough portion of the study included two phases: Cough onboarding and in-home data collection. The onboarding was done virtually due to COVID-19 infection concerns. The data was collected using Samsung Earbuds pro paired with a Samsung S20 phone (as gateway). The phone received 16 KHz audio and 50 Hz IMU signals and sent them to the server. Buds' firmware was designed to send 1-axis audio and 10-axis IMU (accelerometer, gyroscope and quaternion). Table 1 lists the onboarding tasks and their duration. Coughs were collected in different situations for a thorough evaluation of the model. Non-cough tasks were specifically chosen as scenarios where the model would be most vulnerable to false positives. For the background noise we asked the user to play a YouTube links of a fan (as white noise) or, TV and crowded area (as colored noise).

TABLE I. Cough Onboarding Tasks

Cough Tasks	Non-Cough Tasks
Stationary quiet: Sitting, Lying down while coughing occasionally (each 30 sec)	Scripted speech (1 min)
Stationary with background noise: White, Colored (each 30 sec), coughing occasionally	Cough-like sounds: Bystander coughing, Throat clearing, Laughing, Eating, Drinking (each 30 sec)
Stationary with music played in buds (30 sec) while coughing occasionally	
Yoga with and without background noise (each 45 sec) while coughing occasionally	
Walking while talking (1 min) while coughing occasionally	Free head motion while talking (30 sec)

B. Annotation Platform

To generate the ground truth, the cough task from the subjects were manually listened to and segmented by trained annotators using "audacity" toolbox. Similarly, annotators picked the timestamp of the IMU peaks associated with each cough by visualizing both audio and IMU signals. Out of 45 subjects, 9 subjects were excluded as the cough peaks were not visually observable as some subjects did not sit still while coughing (against what they were asked to do) or provided coughs that didn't sound natural leading to unnatural head motion when coughing.

C. Audio Model Training

After annotation, cough audio is filtered and normalized. After preprocessing, 47 features were generated including well-established temporal (mean, median, std, skewness, kurtosis, zcr, SPlevel and quartile range) and spectral (MFCC, chroma, centroid, spread, rolloff, flatness, etc.) by sliding through cough segments. The cough features are coming from "sitting" and "lying down" tasks while speech features come from sliding through speech task. Non-cough parts of the stationary tasks in Table 1 were used to create the "others" class. Fig. 3 provides an overview of the audio pipeline for training the model. Ten-fold cross validation was used to find the optimum number of features through a stepwise feature inclusion process and model selection was done through LOSOV using the selected features.

D. DTW Template Training

To train the IMU cough templates, annotated IMU segments from the "sitting" task were fed to the pipeline provided in Fig. 3 by picking a window around the IMU peaks associated with each single cough after applying a LPF with 0.3 Hz corner frequency and a 10-sample moving average smoothing filter. Cough templates for each subject are then aligned and the personalized template was generated by accumulating and averaging them (Fig. 4). The distance threshold was picked as the average of distance values between cough tasks and cough-like tasks followed by a grid search around the value.

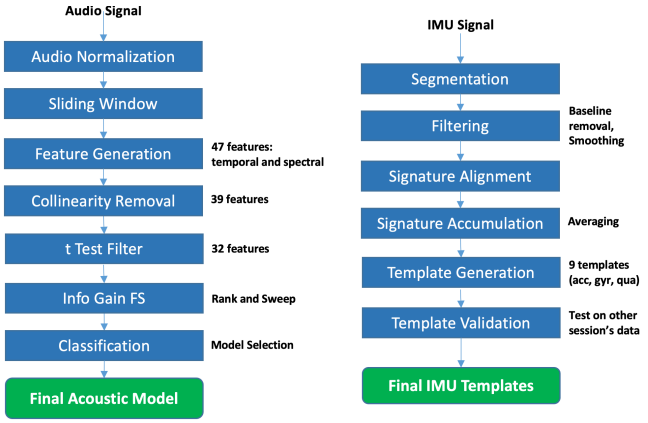

Fig. 3. Training pipeline for audio and IMU modules

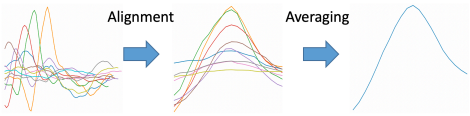

Fig. 4. Accelerometer signal through template generation process

V. Results

A total number of 5296 "cough" feature rows were generated from 1073 coughs. A total of 6750 "speech" and 3815 "others" feature rows were also generated. A total of 32 features were selected using collinearity and t-test feature selection methods and then ranked based on their info gain values. A total of 14 features were ultimately picked through a stepwise feature inclusion process as more features didn't lead to better F1 score. In the next step, for each subject, we evaluated the sensitivity of the model using all the cough tasks (Table 1 column 1) by excluding the subject from training set. We also evaluated model's specificity using cough-like tasks (Table 1 column 2) with the addition of yoga and walking session where a high FPR is probable due to body motion.

FPR is defined as the total number of instances detected as cough in non-cough tasks, divided by the total number of windows in non-cough tasks. Specificity is defined as the number of true negative instances (TN) divided by (TN+FP). As can be seen from Fig. 5, the audio-based model is very sensitive (90+% recall values). However, false positive rate is more than 10% for many of the cough-like sounds and up to 44.8% for the "eating" task. This clearly shows the specificity limitation for the audio-based model.

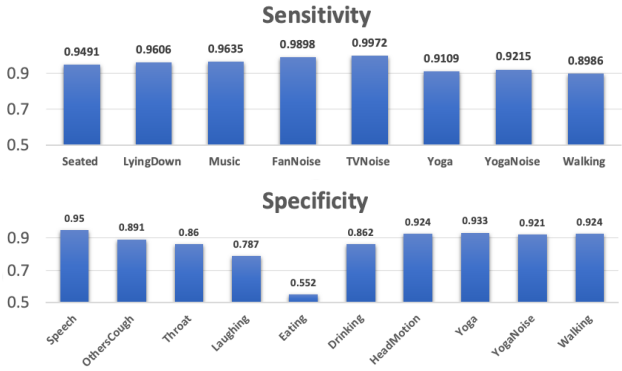

Fig. 5. Audio-based model Sensitivity (top) and Specificity (bottom)

In the next step IMU is fused to reduce the FPR. DTW is only enabled when at least two consecutive coughs are detected from the audio model. A grid search for the window size of the template was done with values of 0.1, 0.2, 0.3 and 0.4 and different combination of sensor modalities. Gyroscope signal and template window size of 0.2 gave the best results. The distance from all 3 axes were combined a. The threshold value was learned to be 350 for gyroscope (GYR) signal and grid search with 4 points around this value with step sizes of 25 was done to find the optimum threshold. Fig. 6 shows the sensitivity and specificity results for these threshold values. Sensitivity is provided only for tasks that contained cough while specificity is provided for only tasks with high chance of false positive. As can be seen, sensitivity goes up with threshold while specificity goes down. The optimum point was found where product of the two values maximizes (325) where the sensitivity is 83% for stationary tasks (a drop of 14.2% compared to baseline audio model). However, specificity improves considerably (more than 90% for most of cough-like tasks). For a cough counter in the field, having low FPR is more important than a high recall. Table 2 provides a comparison between the fusion model and the baseline audio model showing 55% FPR reduction compared to the baseline.

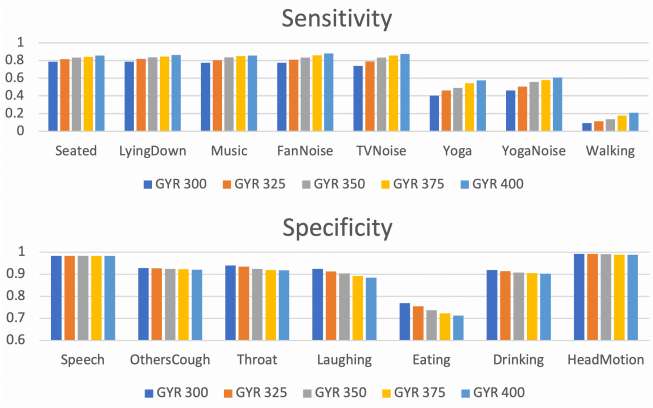

Fig. 6. Fusion model Sensitivity (top) and Specificity (bottom)

TABLE II. FPR VALUES FOR AUDIO-BASED AND FUSION MODELS

Tasks	Audio FPR (%)	Fusion-based FPR (%)	FPR Reduction (%)
Speech	5.0	1.7	66.0
Bystander cough	10.9	7.3	33.0
Throat Clearing	14.0	6.5	53.6
Laughing	21.3	8.7	59.2
Eating	44.8	24.5	45.3
Drinking	13.8	8.5	38.4
Head motion	7.6	0.8	89.5

VI. CONCLUSION AND FUTURE WORKS

Utilizing a hybrid fusion algorithm comprising of an audio-based classification model and a DTW-based templating matching model, we implemented the first multi-modal cough detection algorithm for earbuds platform. The fusion model on the other hand achieved an average sensitivity of 83% (a reduction of 14.2% over the baseline) for stationary tasks while leading to an average reduction of 55% for FPR (an average specificity of 91.7%) for cough-like tasks. Gyroscope proved to be most promising for the matching module. The model, however, is still not sensitive enough for non-stationary tasks. Body motion is polluting the IMU signal and disinfecting from it could be a potential future work. Another future work could be to propose an automatic method for annotation of the IMU peaks from the labels of the cough within the audio. Finally, testing the proposed algorithm on actual field data seem necessary and perhaps vital for further evaluation of the algorithm.

REFERENCES

[1] Lung Institute, "The Cost of Lung Disease" Available: https:-//lunginstitute.com/blog/the-cost-of-lung-disease/, Accessed May 2019.

[2] S. Matos, S. S. Birring, I. D. Pavord, H. Evans. "Detection of cough signals in continuous audio recordings using hidden Markov models," IEEE Transactions on Biomedical Engineering, 53, 1078-1083, 2006.

[3] E. Nemati, M. M. Rahman, V. Nathan, J. Kuang. "Private audio-based cough sensing for in-home pulmonary assessment using mobile devices." In Proceedings of the 13th Body Area Networks. 2018.

[4] E. Nemati, et al. "Estimation of the Lung Function Using Acoustic Features of the Voluntary Cough." In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 4491-4497. IEEE, 2020.

[5] J. Monge-Álvarez, et al. "Robust detection of audio-cough events using local hu moments." IEEE journal of biomedical and health informatics 23, no. 1 (2018): 184-196.

[6] X. Xu, et al. "Listen2Cough: Leveraging End-to-End Deep Learning Cough Detection Model to Enhance Lung Health Assessment Using Passively Sensed Audio." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5, no. 1 (2021): 1-22.

[7] S. Matos, S. S. Birring, I. D. Pavord, H. Evans. "Detection of cough signals in continuous audio recordings using hidden Markov models," IEEE Transactions on Biomedical Engineering, 53, 1078-1083, 2006.

[8] E. C. Larson, et al. "Accurate and privacy preserving cough sensing using a low-cost microphone." In Proceedings of the 13th international conference on Ubiquitous computing, pp. 375-384. 2011.

[9] J. Amoh, and K. Odame. "Deep neural networks for identifying cough sounds." IEEE transactions on biomedical circuits and systems, 10, no. 5 (2016): 1003-1011.

[10] M. Abdullah, and E. Keogh. "Extracting optimal performance from dynamic time warping." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2129-2130. 2016.