Nocturnal Cough and Snore Detection in Noisy Environments Using Smartphone-Microphones

Sudip Vhaduri[1], Theodore Van Kessel[2], Bongjun Ko[2], David Wood[2],
Shiqiang Wang[2], and Thomas Brunschwiler[3]
[1] University of Notre Dame, Notre Dame, IN 46556, USA
[2] IBM Watson Research, Yorktown Heights, NY 10598, USA
[3] IBM Research - Zurich, 8803 Rueschlikon, Switzerland
[1]{svhaduri}@nd.edu,[2]{tvk,bongjun_ko,dawood,wangshiq}@us.ibm.com,
[3]{tbr}@zurich.ibm.com

Abstract—**The reporting on nocturnal sounds like cough and snore is not only relevant to follow the progress of respiratory diseases of patients but also to assess the quality of sleep of subjects. In this study, we discuss an audio analysis approach to count individual cough events and the duration of snore sounds in presence of air-conditioner noise through recordings of a smartphone and computationally efficient classifiers. A new audio data set of cough and snore sounds was acquired from 26 subjects. Energy threshold-based segmentation was applied to identify cough or snore events in the original low noise dataset. A k-nearest neighbor classifier was trained to merge cough phases belonging to the same cough event, to derive the proper ground-truth labeling. The original audio signal was augmented by the superposition of air-conditioner noise, with a signal-to-noise ratio of -40dB to 40dB, to enrich the training set of the binary classifier. Nine out of 40 *mel-frequency cepstral coefficients* in combination with the logarithm of energy from an entire cough or snore event were computed. Various classifiers, such as k-nearest neighbor (k-NN), rule-based classifier, decision tree, random forest, naive Bayes, and support vector machine were benchmarked against each other. The k-NN classifier with $k = 1$ resulted in the highest F_1 scores of .85 and .88 in the binary classification task using generalized and personalized models, respectively, considering noise augmented samples. These results underline the potential of smartphones to objectively report on patient symptoms through audio recordings at night.**

Index Terms—**audio analytics, cough, snore, noise, MFCC features, binary-classification, smartphone**

I. Introduction

Cough is a common symptom of many respiratory diseases. It is a three-phase expulsive motor act, characterized by the inspiratory, followed by a forced expiratory phase against the closed glottis, with a sudden opening of the glottis and thus, rapid expiratory airflow phase, which can end with a further partial glottis closure phase. As a result, up to three distinct acoustic phases can be observed in a cough event: Φ-1) explosive phase, Φ-2) intermediate phase and Φ-3) voiced phase [1]–[4]. The entirety of consecutive cough events, with a time spacing of less than 2 s is defined as a cough episode (Figure 1) [5].

Various methods of quantifying coughing were reported: a) cough events; b) cough seconds; c) cough breaths; d) cough episodes and e) cough intensity. Most studies consider cough frequency, defined as cough events by time interval, as the objective metric to report on cough symptoms. However, cough intensity seems to be a better predictor of patients' quality-of-life [6].

In current medical practice, cough symptoms are reported by patients themselves through questionnaires, such as the Leicester Cough Questionnaire (LCQ), Cough-Specific Quality-of-Life Questionnaire (CQLQ) [6] or as part of disease progress questionnaires (e.g. Chronic Obstructive Pulmonary Disease (COPD) Assessment Test (CAT)). However, the subjective patient reports do not correlate well with objective cough recordings, particularly nocturnal coughs [7]. Thus, the development of objective outpatient monitors are important for cough symptom reporting in order to monitor progress of diseases such as COPD [8] or asthma [5] in patients.

Various sensor modalities, such as contact and audio-microphones, thermistors, accelerometers, electrocardiograms (ECGs), piezoelectric belts were explored to detect cough. Drugman et al. demonstrated best performance with the audio-microphone compared to other modalities [9]. Audio-microphones are preferred from a usability perspective as well, allowing cough monitoring in a outpatient setting through available smartphones.

First, audio-based cough recorders were implemented in dedicated devices including free-field microphones [5], [10]. The Hull Automatic Cough Counter (Castlefield Hospital, Hull, UK) is one of them and is based on an artificial neural network classifier and achieves a binary-classification sensitivity and specificity of 80% and 96%, respectively. A comprehensive overview about such devices is provided by Shi et al. [6].

Next generation cough detectors take advantage of advances and the wide availability of smartphones. However, they need to consider the compromised computational performance and microphone quality for real-time edge classification. Hao et al. explored the classification accuracy of sounds like cough and snore considering the microphone characteristics of various smartphones [11]. Sound feature computation is essential to result in a performant, but efficient algorithm. Monge-Alvarez

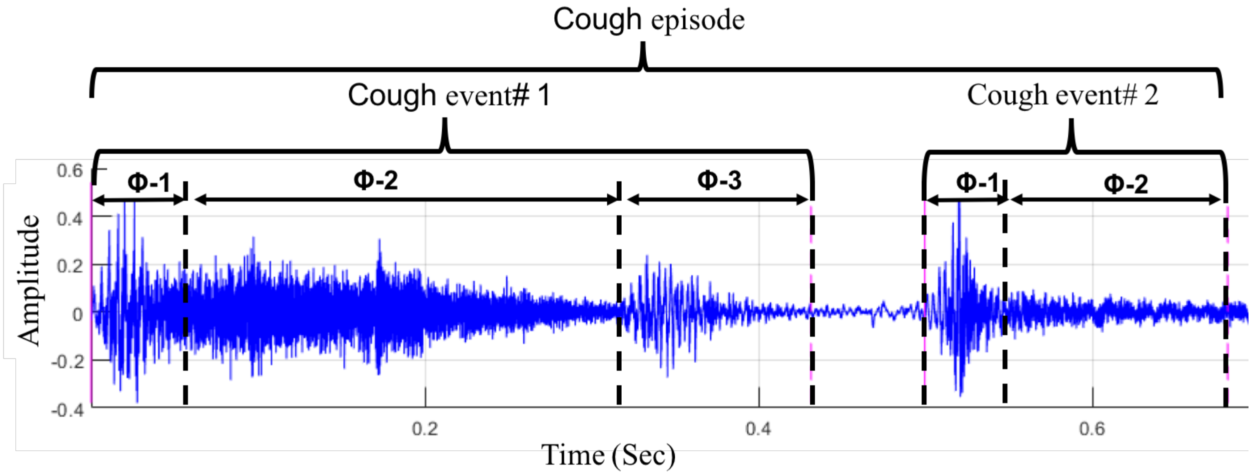


Fig. 1: Three-phase (left) versus two-phase (right) cough events. (Φ-1: explosive phase, Φ-2: intermediate phase, and Φ-3: voiced phase). These two cough events represent a cough episode.

et al. [12] reported highest accuracies using Hu Moments as audio features. However, the computational cost is more than an order-of-magnitude higher than for *mel-frequency cepstral coefficients* (MFCC). Shallow classifiers are prefered in real-time smartphone classification, compared to deep-learning models, due to their low computational cost. Support-Vector-Machines (SVM) and k-Nearest Neighbours (k-NN) are the best candidates and can be further optimized, as described in [13].

In this paper, we report on the exploration of nocturnal cough and snore classifiers to report on disease progress and subject's quality-of-sleep [14] based on smartphone microphones. Background noise (i.e. air-conditioner sound) is considered at various *signal-to-noise ratios* (SNR). Automated ground-truth labeling is performed by energy-based thresholding, followed by cough phase classification to identify individual cough events. Finally, efficient feature extraction (MFCC) and classifiers (k-nearest neighbor, rule-based classifier, decision tree, random forest, naive bayes, and support vector machine) were benchmarked, with respect to their performance on the augmented noisy data set.

II. Data Collection and Ground-Truth Labeling

A. Audio Dataset with Superposition of Noise

A data collection campaign was performed to acquire cough and snore sounds from 26 healthy individuals (3 female and 2 smokers) with average age 42.1 (± 10.6) years, average weight 77.2 (± 11.8) kilograms, and average height 1.78 (± 0.07) meters. The recording was performed with RecForge II (Dje073) through the front microphone of a Samsung Galaxy A3 (2016) at a distance of $80\,\mathrm{cm}$ from the subjects, with a sampling rate of $44.1\,\mathrm{kHz}$, a precision of 32bit and manual gain control. Subjects were ask to perform three forced cough episodes, the first containing one, the second containing two and the last containing three cough events and to perform forced snore sounds in a periodic fashion. Manual ground-truth labeling was then performed in Audacity (The Audacity Team) to partition the 284 cough and 191 snore events in time.

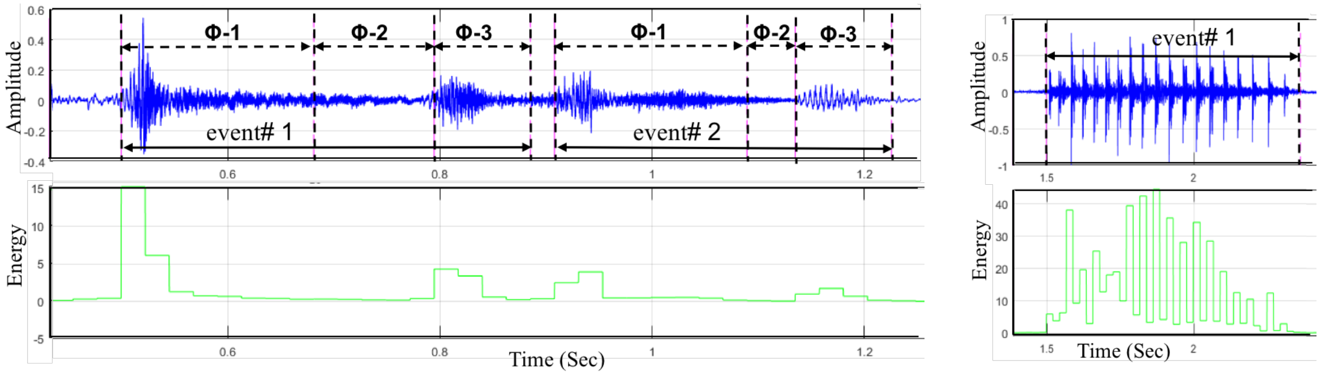
A typical background noise at night is the air-conditioning system. Thus, we considered the superposition of air-conditioning noise to the original cough and snore signal. The power spectrum of the three sounds was computed. It was found that coughs and snores have similar power spectra with an average peak frequencies of 360 ± 45 Hz and 180 ± 25 Hz, respectively, but differ substantially from the air-conditioner noise. Therefore, we decided to construct a binary classifier to differentiate cough and snore events with the superposition of air-conditioner noise at various signal-to-noise ratios.

Signal-to-noise ratio (SNR) is defined by the following equation:

$$
\begin{aligned}
SNR &= \frac{P_{signal}}{P_{noise}} \\
\Rightarrow SNR &= \left(\frac{A_{signal}}{A_{noise}}\right)^2
\end{aligned}
\tag{1}
$$

Where P_{signal} and P_{noise} are the power of the signal (cough or snore) and noise, respectively. A is the amplitude, defined as the root mean square of the power. Power P is computed as $P = \frac{\sum_{n=0}^{L-1} |s[n]|^2}{L}$, where $s[n]$ is the value of a time domain signal at time n and L is the number of samples in a window/frame. SNR can also be expressed in the logarithmic scale defined as $SNR_{dB} = 10log_{10}(SNR)$. For the model training and testing, we vary the SNR of the cough and snore signal to the air-conditioner noise in logarithmic scale, i.e. $SNR_{dB}^{target} \in \{-40, -30, -20, -10, 10, 20, 30, 40\}$. The two audio files are superimposed according to Equation 2 to result in the test and train signal.

$$
A_{signal} = A_{signal} + \left(\frac{SNR^{act}}{SNR^{target}}\right)^{1/2} \times A_{noise} \tag{2}
$$



(a)

(b)

Fig. 2: Energy threshold-based ground-truth collection for (a) cough and (b) snore events.

B. Ground-Truth Collection by Cough Event Segmentation

To find cough events from our collected audio recordings, we first segment an entire clip into sliding windows of 227 ms length. Each window consists of 10 frames (i.e. each 22.7 ms long), with an offset of 22.7 ms between windows. For each frame we compute the energy (Equation 10) of the waveform. Next, we use a pair of threshold values on frame energy to mark the start and end of a cough event [15]. We tuned the parameters with respect to the hand labels and have found 0.5 Joule and 0.3 Joule as optimal values for the pair of energy thresholds.

Comparing the hand labels with the automatic labeling, we observe that most of the events are correctly segmented by the threshold pair. However, there are cases where Φ-1 and Φ-3 of three phase cough events are segmented separately (Figure 2a), resulting in two instead of one cough event. Thus, an more sophisticated approach including classification besides the segmentation is required to identify the cough phases and to merge them into single cough events. This will be discussed in next paragraph. While applying the threshold pair for snore audio clips, we obtain snore events which consist of multiple periodic patterns as shown in Figure 2b. In the energy plot (Figure 2) a staircase function in time is observed since we compute the energy for every frame, as a representation for all values in that frame.

We build a k-Nearest Neighbours (k-NN) classification model to automatically detect Φ-1 and Φ-3 of a cough signal. We compute 40 *mel-frequency cepstral coefficients* (MFCC) features (details in Section III-A) and consider Φ-1 to be the positive class) and Φ-3 to be the negative class of 73 three-phase cough events. In our experiment, we vary the neighbor counts, $k \in \{1, 3, ..., 73\}$ and the Minkowski distance measure (Δ) [16] with order, $p \in \{0.1, 0.2, \ldots, 0.9, 1.0, 2.0, \ldots, 5.0\}$ defined as:

$$\Delta = \left(\sum_{i=1}^{40} |x_i - y_i|^p \right)^{\frac{1}{p}} \tag{3}$$

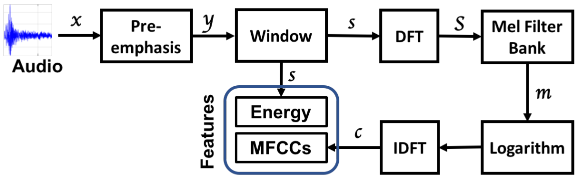


Fig. 3: Sequence of operations to result in the MFCC from an audio signal [17].

Where x_i and y_i are the i^{th} mel-frequency cepstral coefficients from test sample X and training sample Y. We achieve an average accuracy of 95.21% while using leave-one-pair-out validation for $k = 1$ and $p = 0.7$. Once we find two consecutive Φ-1 and Φ-3 parts, we merge them to form three phase cough events, which concludes the ground-truth collection.

III. Cough and Snore Classifier Engineering

A. Feature Computation and Selection

The most common method to extract spectral features in speech recognition are *mel-frequency cepstral coefficients* (MFCC) [18]–[22], with frequency bands adapted to the human perception. Figure 3 depicts the sequence of operations to perform to compute MFCCs from an audio signal.

In the pre-emphasis step a first-order high-pass filter is applied. For a time domain input signal x, signal value at time n, i.e. $x[n]$ and $0.9 \le \alpha \le 1$, the filter is defined as:

$$y[n] = x[n] - \alpha x[n-1] \tag{4}$$

Next, the entire signal y is divided into smaller windows and signal value at time n, i.e. $y[n]$ is extracted by multiplying the value of the Hamming window at time n, i.e. $w[n]$ using the following equation:

$$s[n] = w[n]y[n]. \tag{5}$$

The *Hamming* window is defined as below:

$$w[n] = \begin{cases} 0.54 - 0.46 \times cos(\frac{2n\pi}{L}) & 0 \leq n \leq L-1 \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where L is the number of samples in the window.

Next, from the discrete-time windowed signal we extract the spectral information for discrete frequency bands using the *discrete Fourier transform* (DFT). For a windowed signal $s[n] \dots s[m]$ the output for each of N discrete frequency bands is a complex number $S[k]$, representing the magnitude and phase of that frequency component in the original signal:

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-j \frac{2\pi}{N} kn} \tag{7}$$

The *Fast Fourier transform* (FFT) is commonly used to compute the *Discrete Fourier transform* (DFT).

Next, we warp the frequencies output by the DFT onto the mel scale, which mimics the physiology of human hearing. The mapping between the Hz to the mel scale is linear below 1000 Hz and logarithmic above 1000 Hz. For actual frequency f in Hz, the perceptual frequency m in mel scale is determined by the following formula:

$$m = 2595 \times log_{10}(1 + \frac{f}{700}) \tag{8}$$

We implement the transformation by creating a bank of filters that represent the energy from each frequency band, 10 filters spaced linearly below 1000 Hz and the rest of the filters are spaced logarithmically above 1000 Hz. After this mapping, we take logarithm of each mel spectrum values.

Next, we compute the cepstrum from the mel spectrum to extract 40 cepstral coefficients (MFCC). The cepstrum is the *inverse discrete Fourier transform* (IDFT) of the *log* magnitude of the DFT of a signal defined as:

$$\begin{aligned} c[n] &= \sum_{n=0}^{N-1} log\left(\Big|\sum_{n=0}^{N-1} s[n] e^{-j \frac{2\pi}{N} kn}\Big|\right) e^{j \frac{2\pi}{N} kn} \\ \Rightarrow c[n] &= \sum_{n=0}^{N-1} log(S[k]) e^{j \frac{2\pi}{N} kn} \end{aligned} \tag{9}$$

We also compute the logarithm of energy, i.e. $log_{10}(E)$, where (E) is the energy in the time-domain using the following equation, which is the same as the energy computed in frequency-domain according to Parseval's theorem [23].

$$E = \sum_{n=0}^{L-1} |s[n]|^2 \tag{10}$$

For every cough or snore event, we consider the entire event as one window and compute our candidate set of features consists of 41 features, i.e. $log_{10}(E)$ and 40 MFCCs.

Finally, we perform feature selection using the wrapper with rule-based classifier JRip and the best first search approach [24]. We find $log_{10}(E)$ and MFCC# 1, 3, 5, 6, 7, 8, 13, 15, and 27 as our most significant 10 features, which we will use for future modeling. This means, that eight out of nine MFCC features correspond to frequencies of less than 1.3 kHz.

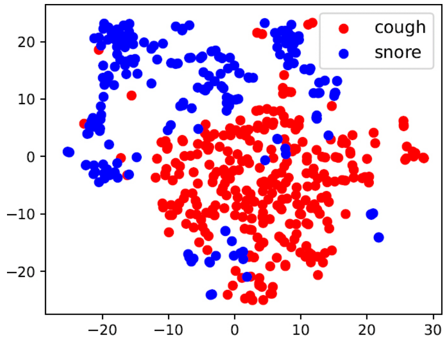


Fig. 4: Projection of the cough and snore samples without noise into the two dimensional plane by t-SNE.

The clustering of the 284 cough and 191 snore samples without noise is depicted in Figure 4. The samples are projected from the 10 dimensional feature space onto the two dimensions by t-distributed Stochastic Neighbor Embedding (t-SNE). The majority of the two classes are clearly separated, but some samples are overlapping with the cluster of the other class.

B. Performance Measures

To evaluate the performance of different cough detection models we consider the following measures:

Accuracy (ACC), which is the fraction of predictions that are correct:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \tag{11}$$

True positive rate (TPR), which is the fraction of coughs that are correctly detected by a model:

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

True negative rate (TNR), which is the fraction of snores that are correctly detected by a model:

$$TNR = \frac{TN}{FP + TN} \tag{13}$$

True positive rate (TPR) and true negative rate (TNR) are also called ***sensitivity*** (SEN) and ***specificity*** (SPC), respectively.

False positive rate (FPR), which is the fraction of coughs that are detected as snores by a model:

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \tag{14}$$

F_1 ***score*** is the harmonic average of the precision and recall, where an F_1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. It is calculated as below:

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{15}$$

TABLE I: Summary of representative classifiers from different classification families

Classifier (parameters)	ACC	Kappa	RMSE	TPR	FPR	F_1 score	AUC-ROC
DT (J48)	92.25	0.84	0.27	0.92	0.08	0.92	0.94
RF	96.13	0.92	0.19	0.96	0.04	0.96	0.99
JRip	90.32	0.81	0.29	0.90	0.09	0.90	0.93
k-NN ($k = 1$, Euclidean)	97.36	0.94	0.16	0.97	0.03	0.97	0.97
NB	90.84	0.82	0.26	0.91	0.09	0.91	0.96
SVM (poly. kernel, $d = 2$, $C = 1$)	94.37	0.89	0.24	0.94	0.05	0.94	0.94
SVM (rbf kernel, $\gamma = 1.5$, $C = 1$)	95.78	0.92	0.21	0.96	0.04	0.96	0.96

Root-mean-square error (RMSE) is a measure of the differences between values predicted by a model or an estimator and the values observed. It is calculated as:

$$RMSE = \left(\frac{\sum_{i=i}^{N} (\hat{y}_i - y_i)^2}{N} \right)^{\frac{1}{2}} \tag{16}$$

Where y_i is the observed value for the i^{th} observation and $\hat{y}_i$ is the predicted value.

Cohen's kappa (κ) measures the agreement between two raters, each classifying N items into C mutually exclusive categories. It is defined as:

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \tag{17}$$

where p_o is the relative observed agreement among raters (identical to accuracy), and p_e is the hypothetical probability of chance agreement. κ can be 1 or 0 if there is a complete agreement or no agreement among raters, respectively.

Area under the ROC curve (AUC-ROC) is equal to the probability that a classification model will rank a randomly chosen positive sample higher than a randomly chosen negative one, where receiver operating characteristic (ROC) curve is the plot of the *true positive rate* (TPR) or *sensitivity* against the *false positive rate* (FPR) or (1 - *specificity*) at various threshold settings. AUC-ROC values close to 1 are better as 1 represents a perfect performance.

C. Classifier Selection

Variouss classifiers including k-nearest neighbor (k-NN), rule-based classifier (JRip), decision tree (DT), random forest (RF), naive bayes (NB), and support vector machine (SVM) with different parameter settings were benchmarked against each other. For the SVM models, we use (1) "Polynomial kernel" function and (2) "Gaussian or Radial Basis Function" (RBF) [25]–[27]. These kernel functions are defined as:

$$\begin{aligned} &Poly.Kernel, K(x_i, x_j) = (1 + \gamma x_i^T x_j)^d \\ &RBFKernel, K(x_i, x_j) = e^{-\gamma x_i^T x_j} \end{aligned} \tag{18}$$

where γ is the "scale parameter", d is the "degree", and x_i and x_j are two feature vectors/windows. Also, we consider the misclassification penalty/cost, $C = 1$. Table I shows the best classifiers from different families along with their best parameter configurations.

Classification results are obtained from 10-fold cross validations performed on a balanced dataset that has 284 cough and 284 snore events, without the superposition of noise. Sampling from the 191 original snore events was performed to balance the dataset. We have observed that the k-NN with $k = 1$ and "Euclidean distance", i.e., $p = 2$ in Equation 3 performs better than other classifiers. Therefore, in the detailed model evaluation, we will only consider the k-NN classifier.

D. Evaluation with Noisy Data

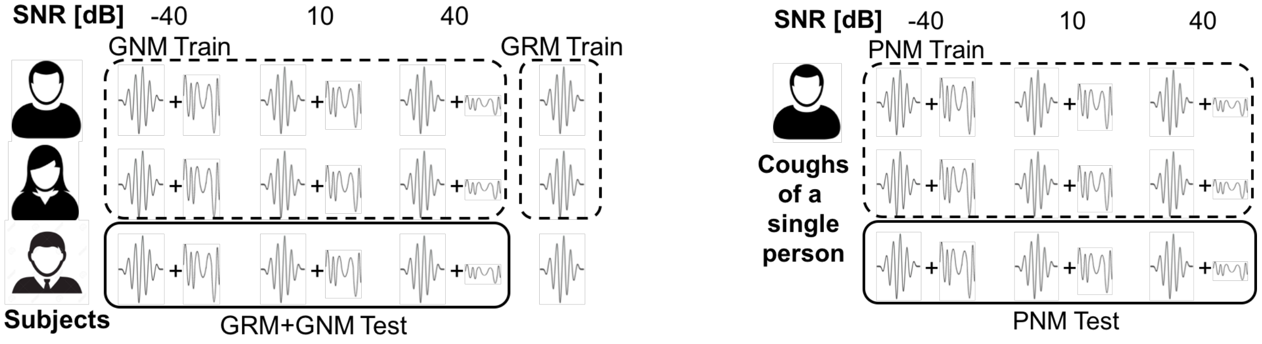
Next, generalized and personalized models are built for a detailed assessment of the binary k-NN classifier on cough and snore data with superimposed noise. The noise superposition was performed according to Equation 2 for the specified SNR levels of -40dB to 40dB on the 191 unique cough and snore events. Our generalized and personalized models are discussed below:

Generalized Raw Model (GRM) is a modeling approach, where training is performed on raw data, but testing is performed on noisy data (Figure 5a). Models are built using features computed from raw events of $N - 1$ (out of N) subjects and one subject is left for testing. For the testing set, features are computed from noise superimposed events. In this approach there are N separate training and testing sets.

Generalized Noise Model (GNM) is a modeling approach, where both training and testing are performed on noisy data (Figure 5a). Models are built using features computed from noise superimposed events of $N - 1$ (out of N) subjects and one subject is left for testing. Equal to the training set, the testing set features are also computed from noise superimposed events. Similar to the previous *GRM* modeling approach, there will also be N separate training and testing sets.

Personalized Noise Model (PNM) is a modeling approach, where both training and testing are performed on the same subject's noisy data (Figure 5b). A pair of cough and snore events with noise superimposed are kept for testing and models are built using the rest of the noise superimposed cough and snore events of a subject. In this way, for each subject with M pairs of cough and snore events, M separate models are built and tested. In our case, we picked five subjects with at least 10 pairs of cough and snore events, and performed this modeling approach.

Table II presents the findings from the three aforementioned modeling approaches. In the table, $\alpha = \left(\frac{SNR^{act}}{SNR^{target}} \right)^{1/2}$. Since each of the four performance measure, i.e. SEN, SPC, ACC, and F_1 score are computed N times; therefore, an weighted average is computed for each performance measure using N values and weights are determined by the number of events that each of the N subjects has. Since these are class



(a) Generalized model depicting train and test

(b) Personalized model depicting train and test

Fig. 5: Visualization of the three modeling approaches.

TABLE II: Summary of k-NN Performance with Noisy Data Testing

	GRM	GNM	PNM
Training	A_{signal}	A_{signal} + ($\alpha \times A_{noise}$)	A_{signal} + ($\alpha \times A_{noise}$)
Testing	A_{signal} + ($\alpha \times A_{noise}$)	A_{signal} + ($\alpha \times A_{noise}$)	A_{signal} + ($\alpha \times A_{noise}$)
SEN (%)	67	90	86
SPC (%)	74	79	90
ACC (%)	67	84	88
F_1 score	0.68	0.85	0.88

imbalance modeling, i.e. each subject has different number of events, SEN, SPC, and F_1 score will be better performance metrics than ACC while comparing different models.

In Table II, we observe that the GNM approach achieves a better performance than the GRM approach, where the train dataset does not include noise, but the test dataset witnesses noise. Therefore, we recommend to include real life noise during training to improve classifier performance.

Similarly, while comparing performance of the PNM approach with the GNM approach, we observe that the PNM modeling approach achieves better performance than the GNM modeling approach in terms of SPC, ACC, and F_1 score. However, there is a slight drop in SEN, which could happen because of low counts for cough events compared to snore events in person-level modeling using unbalanced dataset. Overall, the PNM modeling approach performs better than the GNM modeling approach. Therefore, a new user may start with the GNM model (Figure 5a) and as time passes and the user's smartphone gets more events, the PNM model (Figure 5b) could be trained and used to achieve a better performance.

IV. Conclusion and Future Work

In this study, we reported on the performance of cough event and snore duration counting, based on smartphone recording, considering air-conditioner background noise. A first challenge is, to automatically generate the ground-truth labeling of cough and snore events of the low noise signal. An energy threshold-based approach was chosen to segment the signal into cough and snore events. However, some cough events where split further into cough phases. Thus, a k-NN classifier was needed to properly identify individual cough phases and to merge them into a cough event, irrespective of a two or three-phased cough. Nine low frequency MFCCs and energy features turned out to be most relevant for the classification task. Eight of those MFCC features represent frequecies of less than $1.3\,\text{kHz}$. Thus sub-sampling of the audio data might be feasible to minimize the computational cost and should be studied in the future. Furthermore, we could demonstrate a high performance (.88 F_1 score) in the binary classification of cough and snore, at SNR levels of -40dB to 40dB, considering the training data augmented with air-conditioner noise.

As a next step, we will record nocturnal coughs and snores from COPD patients in a field trial together with our hospital partner to be able to train and test the classifier with real sounds. Further, we will explore the correlation of the objective cough recording with self-reports from patients and will try to identify the clinical relevance of the recordings in relation to patients disease progression. Finally, we will also test various smart-phone models with their specific microphone characteristics considering various noises beyond the air-conditioner to validate the generalization of the model.

Acknowledgment

We would like to express our gratitude to the 26 subjects who volunteered to the cough and snore recording. This study wouldn't have been possible without them. Furthermore, we would like to thank the IBM management to provide funding for the John Kelly internship program, which enabled the collaboration between the University of Notre Dame and IBM Research on the topic.

References

[1] A. Morice, G. Fontana, M. Belvisi, S. Birring, K. Chung, P. Dicpinigaitis, J. Kastelik, L. McGarvey, J. Smith, M. Tatar *et al.*, "Ers guidelines

on the assessment of cough," *European respiratory journal*, vol. 29, no. 6, pp. 1256–1276, 2007.

[2] C. Thorpe, L. Toop, and K. Dawson, "Towards a quantitative description of asthmatic cough sounds," *European Respiratory Journal*, vol. 5, no. 6, pp. 685–692, 1992.

[3] N. S. M. Zawawi, M. P. Robb, and G. A. O'Beirne, "Measuring voluntary cough and its relationship to the perception of voice," *Asia Pacific Journal of Speech, Language and Hearing*, vol. 15, no. 2, pp. 93–109, 2012.

[4] C. Mills, R. Jones, and M.-L. Huckabee, "Measuring voluntary and reflexive cough strength in healthy individuals," *Respiratory medicine*, vol. 132, pp. 95–101, 2017.

[5] A. Proaño, M. A. Bravard, J. W. López, G. O. Lee, D. Bui, S. Datta, G. Comina, M. Zimic, J. Coronel, L. Caviedes *et al.*, "Dynamics of cough frequency in adults undergoing treatment for pulmonary tuberculosis," *Clinical infectious diseases*, vol. 64, no. 9, pp. 1174–1181, 2017.

[6] 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, 2018.

[7] 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.

[8] M. G. Crooks, A. Den Brinker, Y. Hayman, J. D. Williamson, A. Innes, C. E. Wright, P. Hill, and A. H. Morice, "Continuous cough monitoring using ambient sound recording during convalescence from a copd exacerbation," *Lung*, vol. 195, no. 3, pp. 289–294, 2017.

[9] T. Drugman, J. Urbain, N. Bauwens, R. Chessini, C. Valderrama, P. Lebecque, and T. Dutoit, "Objective study of sensor relevance for automatic cough detection," *IEEE journal of biomedical and health informatics*, vol. 17, no. 3, pp. 699–707, 2013.

[10] P. Klco, M. Kollarik, and M. Tatar, "Novel computer algorithm for cough monitoring based on octonions," *Respiratory physiology & neurobiology*, 2018.

[11] T. Hao, G. Xing, and G. Zhou, "isleep: unobtrusive sleep quality monitoring using smartphones," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013, p. 4.

[12] J. Monge-Álvarez, C. Hoyos-Barceló, P. Lesso, and P. Casaseca-dela Higuera, "Robust detection of audio-cough events using local hu moments," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 184–196, 2019.

[13] C. Hoyos-Barceló, J. Monge-Álvarez, M. Z. Shakir, J.-M. Alcaraz-Calero, and P. Casaseca-de La-Higuera, "Efficientk-nn implementation for real-time detection of cough events in smartphones," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1662–1671, 2018.

[14] S. Vhaduri and C. Poellabauer, "Impact of different pre-sleep phone use patterns on sleep quality," in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2018, pp. 94–97.

[15] M. Kaur and A. Kaur, "A review: Different methods of segmenting a continuous speech signal into basic units," *International Journal Of Engineering And Computer Science*, vol. 2, no. 11, 2013.

[16] "Minkowski distance," Available: https://en.wikipedia.org/wiki/Minkowski_distance, Accessed: February 2019, [Online].

[17] D. Jurafsky and J. H. Martin, *Speech and language processing*. Pearson London, 2014, vol. 3.

[18] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE transactions on acoustics, speech, and signal processing*, vol. 28, no. 4, pp. 357–366, 1980.

[19] S. Imai, "Cepstral analysis synthesis on the mel frequency scale," in *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8. IEEE, 1983, pp. 93–96.

[20] S. G. Koolagudi, D. Rastogi, and K. S. Rao, "Identification of language using mel-frequency cepstral coefficients (mfcc)," *Procedia Engineering*, vol. 38, pp. 3391–3398, 2012.

[21] N. Dave, "Feature extraction methods lpc, plp and mfcc in speech recognition," *International journal for advance research in engineering and technology*, vol. 1, no. 6, pp. 1–4, 2013.

[22] B. Logan *et al.*, "Mel frequency cepstral coefficients for music modeling." in *ISMIR*, vol. 270, 2000, pp. 1–11.

[23] "Parseval's Theorem," Available: https://stackoverflow.com/questions/34133680/calculate-energy-of-time-domain-data, Accessed: February 2019, [Online].

[24] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997, special issue on relevance.

[25] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[26] "Kernel function K," Available: https://en.wikipedia.org/wiki/Least_squares_support_vector_machine, Accessed: May 2018, [Online].

[27] "Hyperparameters of the Support Vector Machine," Available: http://philipppro.github.io/Hyperparameters_svm_/, Accessed: February 2019, [Online].