Accurate and Privacy Preserving Cough Sensing using a Low-Cost Microphone

Eric C. Larson[1], TienJui Lee[1], Sean Liu[2], Margaret Rosenfeld[3], Shwetak N. Patel[1,2]

[1]Electrical Engineering,
[2]Computer Science & Engineering,
DUB Institute, University of Washington
Seattle, WA 98195
{eclarson, tienlee, sysliu, shwetak }@uw.edu

[3]Seattle Children's Hospital
Center for Clinical and Translational Research
4800 Sand Point Way NE
Seattle, WA 98105
margaret.rosenfeld@seattlechildrens.org

ABSTRACT

Audio-based cough detection has become more pervasive in recent years because of its utility in evaluating treatments and the potential to impact the quality of life for individuals with chronic cough. We critically examine the current state of the art in cough detection, concluding that existing approaches expose private audio recordings of users and bystanders. We present a novel algorithm for detecting coughs from the audio stream of a mobile phone. Our system allows cough sounds to be reconstructed from the feature set, but prevents speech from being reconstructed intelligibly. We evaluate our algorithm on data collected in the wild and report an average true positive rate of 92% and false positive rate of 0.5%. We also present the results of two psychoacoustic experiments which characterize the tradeoff between the fidelity of reconstructed cough sounds and the intelligibility of reconstructed speech.

Author Keywords

Cough detection, health, mobile phones, signal processing, sensing, privacy.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Algorithms, Experimentation, Performance, Security.

INTRODUCTION

Coughing is a common, distressing symptom that results in significant health care costs, medical consultations, and medication use [15]. According to the U.S. Department of Health and Human Services, coughing is the most frequent symptom mentioned by people when seeking medical advice [7, 19, 38]. Although the importance of diagnosing and managing cough is well recognized [15, 20], evaluation is limited by the lack of objective measures of ambulatory cough frequency and severity [40]. Efforts to develop objective cough monitoring systems date to the 1950s [3], but have intensified in recent years due to technological advances [40]. As such, the medical community has setup guidelines for objective cough monitoring systems [14, 20, 30]. They recommend using systems that have automated cough recognition, can distinguish cough from other sounds, permit 24-hour recording, and provide digital processing of cough recordings; other critical and desirable features include privacy preservation, mobility, compactness, and unobtrusiveness. Although recent advances in audio cough monitoring have begun to make the process more automated [2, 4 26], current systems do not meet all of these requirements. Most worryingly, current audio based methods expose private information about users' speech.

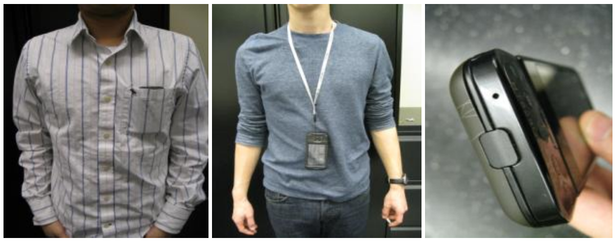

Figure 1. (Left) Participants carrying the mobile phone in their shirt pocket or using a neck strap. (Right) The built-in microphone facing up in the direction of the mouth.

In this paper, we describe our mobile phone-based solution for detecting and counting personal coughs. Our system attempts to meet all of the requirements for objective cough monitoring as outlined by the medical community. Our approach reliably detects cough sounds while preserving the audio privacy of the patient and bystanders. We show our method to have three unique contributions to the research community: (1) it advances the current state of the art in audio cough classification (with average true positive rate of 92% and false positive rate of 0.5%) and allows for classification of individual cough sounds within a coughing episode, (2) it also allows for cough sounds to be reconstructed with good fidelity so that they can be assessed by a physician or to further remove false positives, and (3) it simultaneously prevents other sounds from being reconstructed with useful fidelity (i.e., speech sounds are unintelligible). Our methodology can be used as a general frame-

UbiComp' 11, Sep 17 – Sep 21, 2011, Beijing, China.
Copyright 2011 ACM 978-1-60558-431-7/09/09...$10.00.

work for other audio-based UbiComp sensing applications that need to reconstruct sounds of interest while simultaneously suppressing private audio, such as speech.

Our approach uses principal components analysis (PCA) of the audio spectrogram for classification and for the prevention of speech reconstruction. We found that using 10 components was sufficient for classification. For reconstructing cough sounds, we found that 25 components produces a "good" fidelity cough sound while simultaneously disguising 84% of spoken words.

The remainder of the paper is organized as follows: We first motivate our work, highlighting the privacy vulnerabilities of existing audio-based health sensing classification methods. Second, we review and summarize related bodies of work. We then introduce our methodology using eigenvector projections of the audio spectrogram. Finally, we present the results of two experiments: the first evaluates the accuracy of our classifier on *real world* audio data collected from mobile phones for 17 subjects experiencing cough episodes. The second uses subjective ratings and speech dictations to characterize the tradeoff between clarity of reconstructed coughs and the intelligibility of speech.

MOTIVATION

Ambulatory cough sensing combined with the capabilities of modern smart phones is a particularly interesting UbiComp application. The integration can allow individuals to keep a baseline cough frequency on their digital health record and help increase compliance and effectiveness of treatment regimens. Moreover, using mobile phones as sensing platforms could enable large-scale epidemiological studies, such as using cough frequency across cities to identify and track influenza outbreaks. Finally, because of the relatively low cost of commodity phones, this approach may be of particular interest in developing countries where physicians have limited resources to track coughing and lung function (e.g., preventing suffocation fatalities from complications of pneumonia [45]).

Value of Ambulatory and Objective Cough Sensing

Self-report of cough frequency and severity is notoriously unreliable, particularly in patients with chronic respiratory conditions [14, 22, 31]. Thus, objective ambulatory monitoring of cough frequency has the potential for substantial clinical benefits. First, it would allow early detection of respiratory exacerbations in patients with chronic respiratory diseases such as asthma, cystic fibrosis, and chronic obstructive pulmonary disease (COPD)—allowing earlier and therefore more effective treatment. Early intervention in exacerbations of these conditions has been shown to decrease hospitalization rates and improve long-term outcomes, including survival [29, 37, 39, 41]. For example, a potential correlation may exist between cough frequency throughout a daily period and the severity of asthma in patients, even before other warning signs are present [8]. Secondly, objective cough monitoring would allow improved evaluation of treatment efficacy in many diseases, including tuberculosis, pneumonia, bronchiolitis, asthma, cystic fibrosis and COPD, with the associated ability to change treatment algorithms as needed. Finally, there is great interest in developing cough monitors as sensitive endpoints for clinical trials of therapies in patients with many of these diseases [21]. Cough patterns have been used as a metric to evaluate various treatments regimens, in conditions such as gastro esophageal reflux and COPD [17, 40].

The potential utility of ambulatory cough quantification, however, has been limited to date by the cumbersome and expensive monitoring systems available. Formative research on the benefits of cough detection is in its infancy—better tools for automatic cough detection are sorely needed in order to establish larger, more conclusive studies. Current studies, for example, can only recruit 15-20 participants because of the overhead in extracting cough events reliably [27, 40].

Challenges with Existing Cough Sensing Approaches

Researchers have recently created systems that use machine learning to automatically detect coughs from recorded audio streams [2, 4, 13, 26, 27]. The most limiting criterion of audio approaches is robustness to outside noise sources. Because continuous audio streams are high data rate, the specificity of the systems must be extremely high. Otherwise the number of false positives may exceed the actual cough rate of the individual. To gain such high specificity, these systems do not use *one* model for cough classification—in the case of one algorithm, they need *hundreds* of models [4, 26]—or require models to be adjusted for each user. This is a common limitation of audio based classification methods: while the feature extraction may be able to run on a mobile device, the classification may be better suited to reside on a server somewhere in the cloud or database, where these algorithms and models can be continually updated. Moreover, some of the existing algorithms opt to be semi-automated to reduce the false positive rate. A trained annotator cleans the dataset by listening to segments of audio where the algorithm thinks a cough occurred.

In this way, data transfer, cloud computing, and semi-automation can all compromise the privacy of recorded audio. Audio based cough detection systems, however, have historically not focused on maintaining the privacy of an individual's recordings. As such, little work has been done on optimizing the tradeoffs between privacy vulnerabilities and algorithmic methods. Table 1 outlines several classification architectures and the tradeoffs for each. Architectures which use non-invertible transforms, for example, do not provide audio for the physician to evaluate, nor do they provide a mechanism for semi-automation to reduce false positives. Architectures which use ubiquitous speech features, such as mel-frequency Cepstral coefficients (MFCCs), potentially expose a wide range of private speech information including content, identity, and prosody [46].

The advantages of our approach are four-fold:

(1) The features that are transmitted correspond to the weights of components, and the actual components

Architecture	Privacy Tradeoffs	Clinical Tradeoffs
Send raw audio to classification server	(-)Health professionals have access to private audio	(+)Physician can listen to cough sounds (+)Reviewer can remove false positives
Send non-invertible features (like MFCC's)	(-)MFCCs are ubiquitous to speech recognition and carry information about speech content	(-)Physician cannot listen to cough sounds (-)False positives cannot be removed
Send unique, non-invertible features	(+)Private audio is protected	(-)Physician cannot listen to cough sounds (-)False positives cannot be removed
Our algorithm	(+)Speech is unintelligible	(+)Physician can listen to cough sounds (+)Reviewer can remove false positives

Table 1. Tradeoffs between different audio cough sensing system architectures.

needed to reconstruct the audio only exist on the phone and the server.

(2) These features generalize across subjects so that no initial calibration is needed.

(3) Once the weights arrive at the server, they can be used to reconstruct the cough event with good quality, allowing physicians to listen to and diagnose the cough sounds, in addition to allowing the system to be semi-automated to further reduce false positives, if needed.

(4) The patient need not worry about health professionals inadvertently hearing private conversations since speech is wholly disguised.

RELATED WORK

Our related work falls into five categories: (1) mobile phone-based health applications, (2) general cough detection, (3) audio-based cough detection, (4) audio privacy, and (5) eigenvector feature selection in machine learning.

Mobile Phone-based Health Applications

Many ubicomp researchers have investigated the possible use of mobile phones as a sensing and/or feedback platform to enhance people's health conditions. Chiu *et al.* created a mobile phone-based system called the Playful bottle [10]. They used the phone's built-in camera and accelerometer to detect how much water people consume during a day and persuade them to drink healthy quantities of water using hydration games. Consolvo *et al.* proposed a UbiFit garden system that employs a pager-size sensor called MSP (Mobile Sensing Platform) to recognize people's activity levels and provide feedback using a garden display on their mobile phone [12]. AsthmaMD is an iPhone application that allows people to quickly and easily log their asthma activity, medication, and causes of an asthma attack in the form of a diary [1]. This aggregated, anonymous data is sent to the cloud, providing researchers with information regarding the causes and external variables contributing to asthma and other illnesses. We share similar goals with these research projects and applications–using a commodity mobile phone as the sensing platform for collecting and distributing data to health professionals and, in doing so, increasing people's awareness of their own health conditions.

General Cough Detection

The most common technique for estimating cough frequency is to have patients self report using numeric scoring (0-5) or Visual Analog Scoring (VAS) [36]. However, numerous studies have shown self-report to be highly inaccurate, including the study presented in this paper (see [40] for a summary). The number of coughs a patient self-reports, for instance, has been shown to be significantly influenced by the placebo effect and the patient's perception of their cough severity [40]. Moreover, patients cannot accurately track trends in their cough frequency from hour to hour, or when they are asleep. For these reasons, more objective methods have been developed for counting coughs. Many require expensive, cumbersome equipment (e.g., the chest wall device in [25]) or require paid annotators to listen to recordings and manually annotate cough sounds [40].

There are a number of sensing systems that automatically assess cough frequency. Dating back to the 1950s, researchers developed methods of measuring thoracic pressure changes (airflow from the mouth) in order to obtain unbiased measures of cough frequency [3]. Semi-autonomous methods exist that require individuals to parse through a pre-segmented list of possible cough sounds. The "Overnight Cough Monitoring System" is such a system, which attached an air-coupled microphone to the chest wall or over the trachea when the participant is sleeping [16]. Kraman *et al.* created another accelerometer-based system that placed an accelerometer at the participant's chest wall [25], but required researchers to manually count coughs based on the visualization of the accelerometer data.

The VivoMetrics Lifeshirt [13] is a commercial product that incorporates various physiological sensors to monitor breathing rate, heart rate, activity, posture, and skin temperature. It can be used to detect coughs with an extra throat microphone and the existing sensor array. The reported true positive rate is 78.2% and the false positive rate is 0.4%. VitaloJAK [27] is another commercial product that uses a piezoelectric sensor attached to the chest wall to detect coughs. The reported true positive rate was 91.3-99.5% after an initial calibration. Each method suffers from the same paradigm: users must wear specialized sensors on the chest wall or around their body, which adds expense, is cumbersome, and limits the system in ambulatory settings.

Audio Based Cough Detection

Audio based systems, on the other hand, are easy to deploy in an ambulatory setting and can be made extremely low cost. These systems have become more accurate and lower cost over recent years. Barry *et al.* created a system called the Hull Automated Cough Counter (HACC) [2], using a lapel microphone and wearable recording device. The feature set used was motivated by speech recognition; namely, mel-frequency cepstral coefficients (MFCCs) and linear

Algorithm (Author)	Sensing	Subjects	Recording Environment	Automation	Initial Calibration?	Mean True Positive Rate	Mean False Positive Rate	Mean False Alarms / Hr
LifeShirt	Throat Mic. +sensor array	N=8	Lab, 24 hours	Automatic	Yes	78%	0.4%	Not reported
VitaloJak	Piezo Sensor	N=10	Lab, 24 hours	Automatic	Yes	97.5%	2.3%	Not reported
HACC	Lapel Mic.	N=15	Clinic, 1 hour	Semi	Yes	80%	4%	Not reported
LCM (Matos)	Lapel Mic.	N=19	In Wild, 6 hours	Semi	Yes	71-82%	Not reported	13
LCM (Birring)	Lapel Mic.	N=19	In Wild, 2-6 hours	Semi	Yes	91% *	<1%	2.5ǂ
Our algorithm	Phone Mic. on necklace	N=17	In Wild, 2-6 hours	Automatic	No	92%	0.5%	17

Table 2. Summary of related work in audio based cough detection. *It is not clear if these rates are reported with or without a 95% energy threshold. ǂThese rates are reported after review by an annotator.

predictive cepstral coefficients (LPCC). They applied a Neural Network classifier and achieved an 80% (55-100%) true positive rate and 4% (2-8%) false positive rate. However, they recorded audio signals in an outpatient clinic for only one hour per person, which is a relatively controlled and noise-reduced environment.

Similarly, Matos *et al.* created a system called the Leicester Cough Monitor (LCM) [26], which uses a lapel microphone with a portable audio recorder. They used MFCCs (with derivatives) as features to a Hidden-Markov Model (HMM). Their average true positive rate was 71% (50% - 99%) and a false alarm rate of 13 cough events per hour (false positive rate not reported). After applying an energy threshold to discard low intensity coughs, the average true positive rate for LCM could be boosted to 82% and false alarms reduced to 2.5 events per hour. However, the tradeoff was to discard on average 29% (6-72%) of the cough events for each subject, and the energy thresholds were required to be computed per individual. Recently, LCM has reported a true positive rate of 91% and false positive rate ~1% [4]. However, this has received unfavorable criticism by the medical community [28], who point out that their most recent publications are not forthcoming about whether the true positives are reported with or without the energy threshold and the system is only evaluated on a small subset of their audio data. They also point out that to get such a low false positive rate, the system requires hired annotators to listen to the low confidence coughs and the annotators must provide a portion of hand segmented cough examples in order to prime the algorithm. As such, the system should actually be coined as semi-automated.

Our system uses principal components analysis (PCA) and a random forest classifier. It has comparable accuracies to existing detection algorithms (92% mean true positive rate), but does not require any automation in order to prime or retrain the models. We note that a direct comparison between our approach and HACC or LCM is impossible. Many of these systems consider their algorithms as propriety, so there is limited information on many of the actual details. Instead, we must opt to compare algorithms on the published accuracies, albeit different datasets. Table 2 summarizes and compares the classification rate of ambulatory cough detection algorithms.

Audio Privacy

Prior work in audio privacy has largely dealt with hiding certain cues about the speakers and conversations around them so that a machine learning algorithm cannot reconstruct valuable information from the feature sets. It is generally accepted that MFCCs are poor features for maintaining privacy, as they reveal not only speech, but also inflection, and prosody [46]. As such Wyatt *et al*. have devised audio features that can successfully hide speech intelligibility, while simultaneously providing cues for prosody and recognition of conversations [46]. Most of the related audio privacy work attempts to preserve certain quantities while providing poor features for modern speech recognizers [32]. Chen *et al*., on the other hand, use linear prediction to replace vowels in speech, while keeping environmental noises such as cars and running water intelligible to subjects [9]. Our work in this paper, similar to [9], attempts to make the speech unintelligible, but also make it possible to reconstruct cough sounds. Our methodologies however, are quite different.

Eigenvector Feature Selection

The most common application of eigenvectors in machine learning is called principal components analysis (PCA). PCA uses orthogonal components (i.e., eigenvectors) of a particular feature space to reduce dimensionality. Components can be sorted in terms of their corresponding Eigenvalue, which ranks the components by how much variation they can explain in the data. Traditional PCA is limited by the assumptions that the optimal transformation of the feature space is linear and orthogonal, which is not true in general. Even so, PCA has been successfully applied in many domains, the best known of which is face recognition (i.e., Eigenfaces [42]) and gene mapping dimensionality reduction [23]. The use of PCA on audio spectrograms is not new. Pinkowski successfully used PCA to develop a model of the spectrogram for different English vowels sounds [35]. Our work also uses PCA on spectrograms, except our model is made for coughing sounds.

PHYSIOLOGY OF COUGHING

This section provides a background on the physiology behind the cough reflex and the generation of cough sounds. We also discuss how coughs manifest in an audio stream using spectrograms, motivating the design of our model.

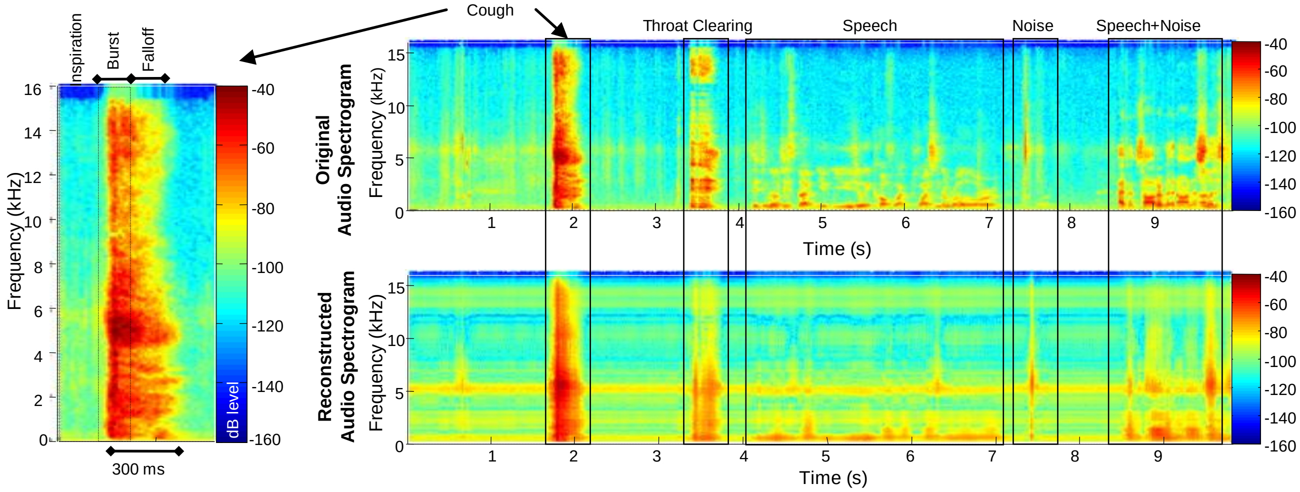

Figure 2. (left) An example cough spectrogram. (top right) An example spectrogram of cough and non-cough audio sounds. (bottom left) An example of the reconstructed spectrogram using principal components analysis.

Cough Reflex

The irritation of afferent cough receptors in the airways triggers the cough reflex. Once triggered, the cough reflex consists of four phases: (1) an initial deep inspiration and glottal closure, (2) contraction of the expiratory muscles against the closed glottis, (3) a sudden glottis opening with an explosive expiration, (4) a wheeze or "voiced" sound [24, 40]. During the initial inspiration, the glottis temporarily narrows and closes. Previous studies show that the glottis closure and expulsive phase can be repeated several times without any inspiration [34].

Audio Analysis

The third and fourth phases of the cough reflex manifest audibly as a cough sound. All cough sounds share common attributes: relatively loud intensity, a quick burst of sound, and a predictable duration and falloff. These are illustrated in Figure 2, where it is easy to identify the cough sound from the audio spectrogram. Notice the overall energy is very strong relative to the surrounding environment and that the initial burst of air causes significant energy well into the 15 kHz range. Despite these common aspects, previous studies [40] show that the pathological processes in the lungs can determine the characteristics of coughing sounds, depending on how lung tissue and vocal resonances are affected. This is why people with different pathological conditions can have different cough sounds.

Figure 2 also shows a close-up of the cough spectrogram. Notice the substantial amount of high frequency content after the burst of sound. This is due to the fourth stage of the cough reflex in which the remaining air from the initial impulse is pushed out of the vocal tract [24, 40]. The differences in cough sounds between different people are largely isolated to the segment of time during this fourth stage. Lastly, notice the cough duration is approximately 300 ms. Empirical measures show coughs range in duration from 300-500 ms [24]. In our approach we leverage the fact that the first 150 ms of a cough sound corresponds only to the explosive phase of the cough reflex and is generally consistent across observers. We only model this explosive stage of the cough reflex so that our model can generalize across observers.

Quantifying Coughs

According to the European Respiratory Society [30], coughing can be quantified in four different ways: (1) *Explosive cough sounds*: the number of characteristic explosive cough impulses. (2) *Cough seconds*: the number of seconds per hour containing at least one explosive cough sound. (3) *Cough breaths*: the number of breaths containing at least one cough. (4) *Cough epochs*: the number continuous cough sounds where each cough is separated by no more than two seconds. The effectiveness of any one of these metrics over the other is still ongoing research. We focus on *explosive cough sounds* because the other three measures can be inferred from this quantity.

DATA COLLECTION AND ANNOTATION

We now turn our attention to creating a labeled audio corpus for evaluating our approach. Participants experiencing coughing episodes were recruited and told to wear a mobile phone (either in shirt pocket or around the neck) that continually recorded their surrounding audio throughout their normal day. These audio streams were manually annotated, labeling all events as one of nine categories, described below in detail.

Data Collection

Seventeen participants, symptomatic of cough before enrollment, were recruited from a local health center. Participants ranged in age from 18 to 60 years old, and 7 of them were females. Their causes of cough include common cold (n=8), asthma (n=3), allergies (n=1) and chronic cough (n=5) due to various reasons including smoking. Table 3 shows these demographic details.

The participants first came to the lab and we explained the recording process. We provided each with an Android G1 mobile phone and asked them to carry the mobile phone around their neck or in their shirt pocket and continue their daily routines. The audio recording software was turned on to record all of the sounds around the participants at 32 kHz using 16 bits. Figure 1 shows our experimental set-up. In all, 72 hours of audio were recorded, resulting in 2,558 cough sounds which occur inside 1,016 coughing episodes. This cough dataset is comparable in size to the datasets used in prior work, however our dataset incorporates a more diverse set of cough events [2, 13, 27]. See Table 2 for a summary of other dataset used in prior work.

Annotation

In order to obtain accurate annotations of the 72 hours of audio recordings, we recruited 6 linguistic students to manually annotate the audio recordings using Praat [5], a scientific software tool for analyzing phonetic sounds and annotate them. The linguists were asked to annotate all sounds with one of the following nine labels: cough (n=2558, 12.2 min), speech (n=5404, 15.8 hr), laughter (n=819, 14 min), breathing (n=522, 11.2 min), throat-clearing (n=1210, 10.23 min), sneezing (n=53, 35 sec), sniffing (n=1289, 9.2 min), other people's cough (n=901, 5.66 min), and environmental noise (n=7296, 28.5 hr). A total of 278 hours were spent completing the annotation of the 72 hours of audio recordings. It took almost 3.9 hours to annotate each hour of audio recording because 62% of the recordings contained ambiguous audio activity which often required the annotators to listen to the clip multiple times. In addition, some ambiguous sounds required discussion among the annotators, researchers, and experts before it was classified into one of the nine categories. If a conclusion could not be made, the event was labeled as unknown and excluded from further analysis. After a sound was labeled by an annotator, it was then verified by a researcher for accuracy. This careful attention was taken to ensure that the annotations were as accurate as possible.

Self Report

We also explicitly asked subjects to keep track of how many times they coughed during the recording process and report this to us after they returned. No correlation existed between self report and the annotated number of coughs. As shown in Table 3, when comparing self report with the annotations, the average difference was 22.8 cough sounds per hour with a standard deviation of 33 cough sounds per hour (i.e., most subjects severely underreported their cough occurrences). The minimum difference was 6 cough sounds per hour while the maximum difference was 149 cough sounds per hour. This highlights the inherent inaccuracy of self report even when patients are primed to monitor their cough for part of the day.

Subject Demographics and Dataset		
# Subjects	17	7 Female,10 male
Age Range	18 – 60,	μ.=27, *mode*=25
Diagnosis	3 Asthma, 5 Chronic, 8 Cold, 1 Allergy	
Audio Recorded per Subject	3 – 6.5 hrs	μ =4.2, *mode*=3
Coughs per Subject	33 – 894,	μ =150, *mode*=79
Coughs/Hour	10 – 178,	μ =33, *mode*=15
Difference from Self Report	6 – 139 cough/hr	μ =22.8, *mode*=20
Total Coughs	2558 coughs	1016 epochs

Table 3. Demographic information and number of coughs collected of all the participants.

ALGORITHM METHODOLOGY

We use our cough event corpus to inform the design of our cough detection algorithm. Recall that our design objectives were to create an accurate classifier that used features which could also be used to reconstruct the cough sounds, but could not be used to reconstruct other sounds, such as speech. A key element of the design is to retain the fidelity of coughs so that a physician can asses them for diagnosis, and so that false positives can be removed audibly, if needed. During evaluation, we divide the corpus into five folds across the participants. Each fold contains all of the audio data for three or four participants.

The algorithm can be divided into four parts: (1) cough model generation, (2) event extraction, (3) cough classification, and (4) cough reconstruction. Before any processing we remove all audio data from the corpus for participants in a given fold. In subsequent discussion we refer to the removed data as the test fold and the remaining data as the training fold. In this way, the cough detection for a single participant's coughs was not trained upon using any audio data from that participant.

Cough Model Generation

During this step of the algorithm we create a cough model for the training fold using PCA on the audio spectrogram [18]. Recall that our audio is recorded at a 32 kHz sampling rate. We first take the magnitude spectrogram of the entire audio sequence (using a hamming window size of 16 ms, 50% overlap, and 512 point FFT). We then select, at random, 40 annotated coughs from each participant in the training fold. For each cough we place the first 150 ms of cough into a single column vector and normalize the vector. We then concatenate each column vector to create a matrix of cough spectrograms, $\boldsymbol{X}$. PCA is run on $\boldsymbol{X}$, yielding a matrix of eigenvectors (components). We save the N components with the largest eigenvalues. This becomes our cough model, $\widehat{\boldsymbol{X}}_N$, where the subscript, N, denotes the number of components in the model. This is similar to the approach used in the face recognition algorithm called "Eigenfaces" [42]; the components in our cough model are analogous to "Eigenface" components.

We then reconstruct the training data spectrograms from $\widehat{\boldsymbol{X}}_{10}$ saving the projection weights used for reconstruction, along with the residual error of the reconstructed spectrogram (11 features). In addition, we also calculate three energy measures of the spectrogram. Namely, the mean decibel energy of the entire FFT, the mean decibel energy of the FFT coefficients above 16 kHz, and below 16 kHz. The energy values, component weights, and residual error are hereafter referred to as our feature set (14 features). For classification of the cough sounds, we found that our model only needed 10 components for reliable classification.

However, the fidelity of the reconstructed coughs can be increased by using a larger number of components, at the expense of making speech more intelligible. We quantify this tradeoff later.

Event Extraction

We use the feature set of our training fold to create a simple extraction algorithm based on thresholds. The algorithm prunes the audio stream, searching for potential events. For each feature, we calculate the threshold at which 98% of the coughs in the training fold are retained regardless of the false positives. We then select five features (out of all 14) with the lowest false positive rates on the training fold. The thresholds of these five features are saved so that they can be used in pruning the audio of the test fold during evaluation. On average, the event extraction retained 96% of the coughs in the training fold, while letting between 5% and 16% of other audio through.

Cough Classification

After event extraction, we train a random forest (RF) classifier [6] on the feature set of extracted events in the training fold. All features are used to train the classifier. The RF classifier is set to weight cough errors more during the building of the forest, and the majority voting threshold for the cough class is set three times lower than non-cough sounds. The max number of trees was set to 500 (as over fitting is rarely a problem in RF classifiers). These parameters, however, were not investigated extensively because of the prohibitive time involved in training and evaluating the models. We ran a few variations on a small subset of the data (about two hours of audio from two participants) and found these parameters to work well. Many different classifiers could be used—we use a RF classifier because it has empirically been shown to work as well as support vector machines and neural networks, but is less sensitive to parameter variation [6].

Cough Sound Reconstruction

For audio reconstruction (if the actual cough sound needs to be replayed), we use the optimal PCA reconstruction method, designed to minimize mean-squared error between the reconstruction and vector of interest [33]. That is, for a given spectrogram, we stack 150 ms of adjacent columns into a normalized column vector, $\boldsymbol{a}$. The reconstruction is

$$\hat{\boldsymbol{a}} = \hat{\boldsymbol{X}}_{\mathrm{N}}\hat{\boldsymbol{X}}_{N}^{\boldsymbol{T}}(\boldsymbol{a}-\bar{a})+\bar{a}$$

where $\bar{a}$ is the mean of $\boldsymbol{a}$ and $\hat{\boldsymbol{X}}_{N}^{\boldsymbol{T}}(\boldsymbol{a}-\bar{a})$ are the projection weights (an N element row vector). This provides us with an estimate of the overall spectrogram magnitude for a 150 ms segment of audio. We then remove the stacking and normalization, and reapply the phase of the original spectrogram. We then convert the spectrogram back to the time domain signal by: (1) performing the inverse short time Fourier transform (ISTFT) and (2) applying an inverse hamming window. Figure 2 shows an example reconstruction. We note that to perform reconstruction a mobile platform must send the mean, $\bar{a}$, normalization constant, and phase of the spectrogram in addition to the projection weights. These quantities do not increase the privacy vulnerabilities of the system because phase and mean have historically been found to have little use in speech or speaker recognition. In fact, most recognition algorithms ignore phase or only use it in conjunction with the magnitude [e.g., 32, 45].

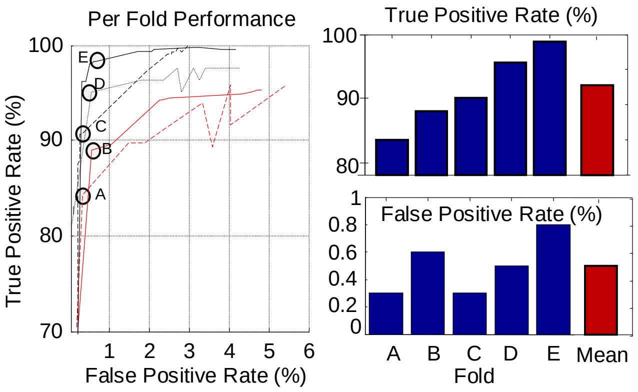

Figure 3. Receiver operating characteristic (ROC) for models trained on five folds of our dataset. "Good" tradeoff points in the ROC are also graphed to the right in column form.

CLASSIFICATION PERFORMANCE

We evaluate our random forest classification method using five-fold cross-validation, across subjects, as discussed previously. We obtain receiver operating characteristic (ROC) curves for each test fold by varying parameters of the RF models. This results in changes to the false positive rate and true positive rate that can be plotted to show the tradeoffs in adjusting each RF model. Figure 3 shows the tradeoff in performance for each fold. We define a true positive to be any explosive cough sound that contains at least two consecutive classifier identifications in the same 300 ms window (this is the finest resolution we can achieve as dictated by our PCA and spectrogram implementations). We define a false positive to be any consecutive classifier identifications that do not occur within 10 seconds of an actual explosive cough sound. This is divided by the total number of overlapping 10 second windows in our test folds. This is a standard measure of false positives used in cough detection [4, 26]. Notice that all but one fold rapidly approaches a 90% true positive rate while false positives are below 1%. Also note that we report true and false positive rates using all audio from the database, *not* just events that are pruned from the "event extraction" phase.

Detection Rate and False Alarms

If we take the "best" performance of each classifier to be the highest true positive rate achieved while simultaneously having a false positive rate less than 1% we achieve sensitivities of 85%, 89%, 91%, 96%, and 99% for each fold with a mean of 92%. The respective false positive rates are 0.3%, 0.6%, 0.3%, 0.5%, and 0.8% with a mean of 0.5%. These points are marked in Figure 3 with a circle and graphed across each fold, A-E. When comparing to other cough detection algorithms, however, the false positive rates cannot be directly compared. This is because each algorithm uses a slightly varied sampling window of the audio data. Instead, the false alarms per hour are a better indication of the number of falsely classified sounds one can expect. The respective number of false alarms per hour for each fold is 7, 20, 17, 29, and 12 with a mean of 17. We also note that one can reduce this value by trading off model parameters. For instance, the mean false alarms per hour can be reduced to 10.7 by trading off a mean true positive rate of 82%, which is comparable to the state of the art in automated cough detection [4, 26, 27]. Alternatively, a trained annotator can review the collected events and discard false positive sounds, as is done with some of the existing cough detection systems. For our dataset this would require reviewing, on average, 1.4 minutes of audio per hour of recording. This time could be further reduced by having the annotator only review identifications that had low confidences from the classifier.

Confusions

An analysis of the false positives in the dataset reveals that coughs are most often confused for “noise” labels (56% of the actual false positive) and for “speech” labels (43% of the false positives). The remaining 1% of false positives are distributed among “breathing” and “laughter” events. This further stresses the importance of disguising or suppressing the audio for false positives. Roughly half of the false positives to be reviewed would be speech that the wearer never intended to be heard, even by a medical professional.

PRIVACY AND FIDELITY CHARACTERIZATION

We now turn our discussion to assessing how well the components actually disguise speech, while also preserving the fidelity of received coughs. We designed two psychoacoustic experiments aimed at quantifying each design goal. Although our classifier only required 10 bases for high precision, we used 5, 10, 15, 25, and 50 bases in the reconstruction to fully characterize the privacy/fidelity tradeoff.

Experiment 1: Speech Intelligibility

During the first experiment, listeners were asked to dictate the words in 8 segments of speech from our audio corpus. Each segment was 5 seconds long and contained a different speaker (4 male, 4 female). The complication of the speech audio ranged from 4 words up to 21, in various environmental settings from study group conversations, people walking and talking, to quiet settings with almost no environmental noises. Listeners performed the test using a custom interface. They were allowed to listen to the speech segment as many times as they wanted and review and change their dictations at any time. Listeners were also encouraged to give their best guess at the speech, even if they felt it was not wholly intelligible. For instance, if they could only spot a few keywords, they were asked to write those dictations in and any words that they could possibly discern from the context of the keyword.

Each listener heard either a clean version of the speech segment (baseline) or one of five reconstructed versions using 5, 10, 15, 25 or 50 components. 24 observers dictated the speech with, on average, each degraded recording being listened to by 4 listeners. Each listener reported that they were English speakers (n=20) or fluent in the English language (n=4). After completion, two reviewers (one author, one third-party) analyzed all the dictations assessing the number of words that listener’s dictated incorrectly. There was good agreement between reviewers; out of the 1,932 words looked at by each reviewer there were only 32 discrepancies.

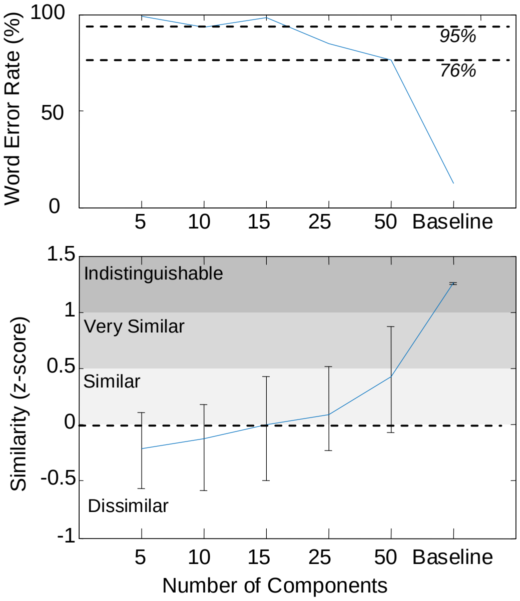

Figure 4. (Top) Word error rate plotted versus the number of model components. (Bottom) Similarity scores (z-scores) of reconstructed cough sounds; error bars are the interquartile range.

Intelligibility Results

Figure 4 shows the results of the word error rate as a function of the number of eigenvectors used in reconstruction. The privacy of speech is well protected when using 15 or less components (i.e., >95% of the words are unintelligible). At 25 components, the word error rate begins to drop. However, even at 50 components, the listener’s are still only able to spot keywords in the phrases for a total word error rate of 76%. We found on average, that longer speech segments were more difficult to reconstruct. Even at 50 components, if the number of words in the phrase was greater than 15, one can expect a word error rate of about 92%. The shorter phrases (less than 7 words) can sometimes be guessed by spotting keywords and inferring surrounding words, with error rates of about 65% using 50 components.

Experiment 2: Assessing Cough Fidelity

In the second experiment users performed a single stimulus two-alternate forced choice (2-AFC) task in which they were presented with two cough sounds. One was a clean version of the original cough sound, and the other was a reconstructed version of the same cough sound. The order of the cough sounds was randomized. Users where given two tasks: (1) they were asked to judge which cough sounded more natural, and (2) they were asked to rate the similarity of the coughs in terms of fidelity on a five point scale which ranged from “very different,” “somewhat different,” “somewhat similar,” “very similar,” to “same cough sound.” If they decided the coughs were the same, their judgment of which cough sounded more natural was discarded. Preferences were entered using radio buttons that defaulted to a position of “not set” for each 2-AFC trial. The listeners could go back and forth between tests (the interface remembers and displays their answers as they go backwards and forwards through each forced choice) and they could listen to any trial as many times as they wished even as they reviewed their responses. The listeners were informed that the cough sounds were of the same recording, but with varying degrees of distortion. For some trials the listener was presented with two reference cough sounds.

A total of 12 cough recordings were chosen for the experiment (6 male, 6 female) which occurred in a variety of environments from conference rooms and atriums to outdoor settings. The length of the audio segments ranged from 1-2 seconds. Half of the sounds were chosen at random from the corpus and half were selected by hand to be sure a wide range of cough types were chosen. Each cough sound was reconstructed using 5, 10, 15, 25, and 50 components, resulting in 60 different cough recordings in addition to the 12 original. One reason we chose a small, but diverse number of cough sounds was so we could investigate whether any particular cough is always reconstructed with high or poor fidelity using a relatively small number of listeners. Each listener reviewed 30 cough sounds per session. After the experiment, we converted each observer’s similarity selections into z-scores. We removed the first five responses from each listener to account for any learning curve the listener may have had. In all, 27 listeners participated for a total of 810 subjective ratings. On average, each degraded cough sound was rated by 13 listeners.

Fidelity Results

Figure 4 shows the results of the second experiment. The mean similarity (z-score) is plotted versus the number of components used in the reconstruction. Error bars are shown around the 25th and 75th percentiles of the z-scores. In all cases the participants could distinguish which cough sounds had been reconstructed and which were natural (i.e., no cough sound was rated as indistinguishable). This result is not wholly surprising. Manipulating the spectrogram often results in audible artifacts that are easy for the listener to cue in upon [32, 46]. The degree of similarity, then, becomes the quantity of interest.

The threshold of similarity was found to be about 15 components. Coughs reconstructed with less than 15 components were found to almost always be rated as dissimilar. Coughs with 25 to 50 components were found to almost always be rated similar or very similar. A Wilcoxon rank test [43] can be used to validate when there is a significant jump in the median similarity score (at 95% confidence). The test reveals that there is a significant difference in the median scores for using 5-10 components versus 25-50. There is also a significant difference for cough sounds with 5-25 versus 50 components. All other comparisons reveal the null hypothesis cannot be rejected with high confidence.

These findings reveal, at 15 components, one can expect about half the coughs to have “good” fidelity while simultaneously making almost all speech unintelligible. Higher similarities can be achieved, however, if one is willing to sacrifice the privacy of some keywords in their conversations. At 50 components, for example, the resulting cough will almost always be “similar” or “very similar,” while making only certain words intelligible. It is likely that the context of the conversation will remain hidden to a listener. This is in contrast to the cough event detection, which only required 10 components.

DISCUSSION AND LIMITATIONS

It is of note that the classifier never confused a bystander’s cough for a cough of the subject. This is likely because of the mobile phone placement. One of the reasons we chose to deploy the mobile phone around our participants’ neck or in the shirt pocket is because these positions provide a short distance between the microphone and subject’s mouth and should receive the best audio quality. However, they might not be the best positions in terms of comfort. According to our participants, the best positions in terms of comfort would be the pants pocket or in bags/purses. Audio signals recorded at these positions will likely require the use of a lapel microphone or Bluetooth headset.

One limitation of this study is that we only look at the details of a cough detection algorithm, without a complete system. We are currently porting our feature extraction algorithm over to a T-Mobile G1 mobile phone platform (rather than an offline calculation). Our initial cough extraction is threshold based and should easily run on a mobile device. This simple extraction algorithm prevents us from sending frivolous features to our server for classification. However, this does not account for *battery life* of the phone, only the real time aspects. We are currently looking at hardware accelerations and algorithmic approximations that can extend the battery life of a system running our algorithm. Additionally, we are exploring using our algorithm on phones as they charge overnight to assess the cough frequency patterns of subjects while they sleep. Traditionally, listening devices in the bedroom or at home have been considered invasive, but our algorithm, because it preserves speech privacy, may mitigate these concerns.

This last point brings up an interesting opportunity for future work. Our algorithm has been shown to preserve speech privacy, but there are other forms of private audio information that may still be vulnerable. For instance, environmental noises may still be intelligible, exposing location or activities. Also, it is unclear whether our algorithm can hide *who* is speaking in an audio stream and the gender of speakers. These are important avenues for future research.

Although we were able to show that subjective listening tests reveal increasing sound similarity, our work is largely informed by working with pulmonologist. More work is still needed to find the effective tradeoff for cough quality from the medical community. It is important to quantify the number of components that can capture key features of the cough that physicians consider important for diagnosis.

CONCLUSION

In this paper, we present a mobile phone-based sensing and inference technique for detecting and counting individual coughs using a commodity mobile phone's built-in microphone in a real world environment. We implement PCA on audio spectrograms for classification and for retaining cough fidelity, while simultaneously disguising and suppressing speech sounds. Our system attempts to address all of the design goals for an objective cough monitoring system as outlined by the medical community: accuracy, low false positives, privacy preservation, mobility, compactness, and unobtrusiveness [14, 20, 30].

REFERENCES

1. AsthmaMD, http://www.asthmamd.org
2. Barry SJ, Dane AD, Morice AH, and Walmsley AD (2006). The automatic recognition and counting of cough. *Cough*.
3. Bickerman HA and Itkin SE (1958). The effect of a new bronchodilator aerosol on the airflow dynamics of the maximal voluntary cough of patients with bronchial asthma and pulmonary emphysema. *J Chron Dis*.
4. Birring SS, Fleming T, Matos S, Raj AA, EvansDH, and Pavord ID (2008). The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough. *Eur Respir J*.
5. Boersma P and Weenink D. Praat: doing phonetic by computer, http://www.praat.org
6. Breiman, L (2001). Random Forests. *Machine Learning*, 5–32.
7. Cerry DK, Hing E, Woodwell D, and Rechtsteiner E (2006). *National Ambulatory Medical Care Survey: Summary*.
8. Chang AB and Gibson PG (2002). Relationship between cough, cough receptor sensitivity and asthma in children. *Pulm. Pharmacol. Ther.*
9. Chen F, Adcock J, and Krishnagiri J (2008). Audio Privacy: Reducing Speech Intelligibility while Preserving Environmental Sounds. *MM'08*.
10. Chiu MC, Chang SP, Chang YC, Chu HH, Chen CC, Hsiao FH, and Ko JC (2009). Playful Bottle: a Mobile Social Persuasion System to Motivate Healthy Water Intake, *Proceedings of UbiComp*.
11. Chung K and Pavord I (2008). Prevalence, pathogenesis, and causes of chronic cough. *The Lancet*, 1364-1374.
12. Consolvo S, Mcdonald D, Toscos T, Chen M, Froehlich J, Beverly H, Klasnja P, LaMarca A, Legrand L, Libby R, Smith I, and Landay J (2008). Activity Sensing in the Wild: A Field Trial of UbiFit Garden, *Proceedings of CHI*.
13. Coyle MA, Keenan DB, Henderson LS, *et al.* (2005). Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease. *Cough*.
14. Decalmer SC, Webster D, Kelsall AA, McGuinness K, Woodcock AA, and Smith JA (2007). Chronic cough: how do cough reflex sensitivity and subjective assessments correlate with objective cough counts during ambulatory monitoring? *Thorax*, 329-34.
15. Gibson PG, Chang AB, Glasgow NJ, Holmes PW, Katelaris P, Kemp AS, Landau LI, Mazzone S, Newcombe P, Van Asperen P, and Vertigan AE (2010). CICADA: Cough in Children and Adults: Diagnosis and Assessment. *Australian cough guidelines summary statement*. *Med J Aust*, 265-71.
16. Gross V, Reinke C, Dette F, and Koch R (2007). Mobile nocturnal long-term monitoring of wheezing and cough. *Biomed Tech*,73–76.
17. Hoglung NJ and Michaelson M (1905). A method for determining the cough threshold with, some preliminary experiments on the effect of codeine. *Acta Physical Scand*, 168-173.
18. Hotelling H (1933). Analysis of a Complex of Statistical Variables Into Principal Components, *Journal of Educational Psychology*, 417-520.
19. Information Resources Inc. (2008) *OTC Market Size Statistics*.
20. Irwin RS, Bolser DC, Braman SS, *et al.* (2006). Diagnosis and Management of Cough Executive Summary: *ACCP Evidence-Based Clinical Practice Guidelines*. *Chest*. Suppl. 1S–23S.
21. Kerem E, Hirawat S, Harmoni S, Yaakov Y, Shoseyov D, Cohen, *et al*. (2008). Effectiveness of PTC124 treatment of cystic fibrosis caused by nonsense mutations: a prospective phase II trial. *Lancet*, 719-27.
22. Kerem E, Wilschanski M, Miller NL, Pugatsch T, Cohen T, Blau H, Rivlin J, Shoseyov D, Reha A, Constantine S, Ajayi T, Hirawat S, Elfring GL, Peltz SW, and Miller LL (2011). Ambulatory quantitative waking and sleeping cough assessment in patients with cystic fibrosis. *J Cyst Fibros*.
23. Khan J, Wei1 J, Ringnér M, Saal L, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu C, Peterson C, and Meltzer P (2001). Classification and diagnostic prediction of cancers using gene expr-ession profiling and artificial neural networks. *Nature Medicine*.
24. Korpas J, Sadlonova J, and Vrabec M (1996). Analysis of the cough sound: an overview. *Pulm Pharmacol*, 261–268.
25. Kraman SS, Wodicka GR, Pressler GA, Pasterkamp H (2006). Comparison of lung sound transducers using a bioacoustic transducer testing system. *J Appl Physiol*, 469–476.
26. Matos S, Birring SS, Pavord ID, and Evans DH (2006). Detection of cough signals in continuous audio recordings using hidden Markov models. *IEEE Trans Biomed Eng*, 1078–1083.
27. McGuinness K, Kelsall A, Lowe J, Woodcock A, and Smith JA (2007). Automated cough detection: a novel approach. *Am J Resp Crit Care Med*
28. McGuinness K, Morice A, Woodcock A and Smith J (2010). The Leicester Cough Monitor: a semi-automated, semi-validated cough detection system? *European Respiratory Journal*, 529-530.
29. Miravitlles M, Murio C, Guerrero T, and Gisbert R (2002). Pharmacoeconomic evaluation of acute exacerbations of chronic bronchitis and COPD. *Chest*, 1449-55.
30. Morice AH, Fontana GA, Belvisi MG, Birring SS, Chung, KF, Dicpinigatis PV, Kastelik JA, Smith JA, Tatar M, and Widdicombe J (2007). *ERS guidelines on the assessment of cough*, *Eur Respir J*.
31. Newcombe PA, Sheffield JK, Juniper EF, Petsky HL, Willis C, and Chang AB (2010). Validation of a parent-proxy quality of life questionnaire for pediatric chronic cough (PC-QOL). *Thorax*. 819-23.
32. Pathak M (2010). *Privacy Preserving Techniques for Speech Processing*. PhD Thesis. Carnegie Melon.
33. Pearson K (1901). On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, 559-572.
34. Piirila P and Sovijarvi ARA (1995). Objective assessment of cough, *Eur Respir J*, 1949-1956.
35. Pinkowski B (1997). Principal Components Analysis of Spectrogram Images. *Pattern Recognition*, 777-787.
36. Raj A and Birring S (2007). Clinical assessment of chronic cough severity, *Pulmonary Pharmacology and Therapeutics*, Vol. 20, Issue 4.
37. Sanders DB, Hoffman LR, Emerson J, Gibson RL, Rosenfeld M, and Redding GJ (2010). Return of FEV1 after pulmonary exacerbation in children with cystic fibrosis. *Pediatr Pulmonol*., 127-34.
38. Schappert S, Burt C, and Ed E (2006). Ambulatory Care Visits to Physician Offices, Hospital Outpatient Departments, and Emergency Departments: United States 2001-02, *Vital Health Stat 13*, 1-66.
39. Seemungal TA, Donaldson GC, Bhowmik A, Jeffries DJ, and Wedzicha JA (2000). Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 1608-13.
40. Smith JA and Woodcock A (2008). New development in the objective assessment of cough, *Lung*, 186.
41. Spencer S and Jones PW (2003). Time course of recovery of health status following an infective exacerbation of chronic bronchitis. *Thorax*, 589-93.
42. Turk M and Pentland A (1991). Face recognition using eigenfaces. *Proc. IEEE Conf on Comp. Vision and Patt. Rec.* 586–591.
43. Wilcoxon F (Dec 1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 80–83.
44. Witten I. and Frank E (2005). Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco.
45. World Health Organization (2001). *Cough and cold remedies for the treatment of acute respiratory infections in young children*. Geneva, Switzerland: World Health Organization.
46. Wyatt D, Choudhury T, and Blimes J (Aug 2007). Conversation Detection and Speaker Segmentation in Privacy Sensitive Situated Speech Data. *Proceedings of Interspeech*.