BreathLive: Liveness Detection for Heart Sound Authentication with Deep Breathing

CHENYU HUANG, Hong Kong University of Science and Technology
HUANGXUN CHEN, Hong Kong University of Science and Technology
LIN YANG, Hong Kong University of Science and Technology and Noah's Ark Lab, Huawei Technologies, China
QIAN ZHANG*, Hong Kong University of Science and Technology

Nowadays, considerable number of devices have been proposed to monitor cardiovascular health. To protect medical data on these devices from unauthorized access, researchers have proposed ECG-based and heart sound-based authentication methods. However, their vulnerabilities to replay attacks have recently been revealed. In this paper, we leverage liveness detection to enhance heart sound-based authentication against replay attacks. We utilize the inherent correlation between sounds and chest motion caused by deep breathing to realize a reliable liveness detection system, BreathLive. To be specific, BreathLive captures breathing sounds and chest motion simultaneously, and then eliminates signal delay caused by any imperfections of device components. Next, it extracts a set of features to characterize the correlation between sounds and motion signals, and uses them to train the classifier. We implement and evaluated BreathLive under different attacking scenarios and contexts. The results show that BreathLive achieves an equal error rate of 4.0%, 6.4% and 8.3% for random impersonation attacks, advanced impersonation attacks and advanced replay attacks respectively, which indicates its effectiveness in defending against different attacks. Also the extensive experiments prove the system can be robust to different contexts with a small training set.

CCS Concepts: • **Security and privacy → Security services**; **Mobile and wireless security**; • **Human-centered computing → Ubiquitous and mobile computing**;

Additional Key Words and Phrases: Liveness Detection, Gyroscope, Microphone, Wearable Computing

ACM Reference Format:
Chenyu Huang, Huangxun Chen, Lin Yang, and Qian Zhang. 2018. BreathLive: Liveness Detection for Heart Sound Authentication with Deep Breathing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 12 (March 2018), 25 pages. https://doi.org/10.1145/3191744

*This is the corresponding author

Authors' addresses: Chenyu Huang, chuangak@connect.ust.hk, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; Huangxun Chen, hchenay@connect.ust.hk, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; Lin Yang, lyangab@connect.ust.hk, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, Noah's Ark Lab, Huawei Technologies, Shen Zhen, China; Qian Zhang, qianzh@cse.ust.hk, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

2474-9567/2018/3-ART12 $15.00
https://doi.org/10.1145/3191744

1 INTRODUCTION

Nowadays, considerable devices have been proposed to monitor cardiovascular health by both industry and academia, such as ECG-based heart rate band/strap[8] and PPG-based smartwatch[2] for heart rhythm monitoring, heart sound-based electronic stethoscope[4, 10] and the chest-worn cardiac telemetry system[3] for heart disease diagnosis. Increasing sensitive health and medical data are recorded and stored by these monitoring devices, which increases the potential risk of privacy leakage. It was reported that over 112 million patient digital records missing, stolen or improperly disclosed from healthcare devices and facilities in 2015[7] .

Numerous authentication methods have been proposed to protect these devices from unauthorized access. Widely-adopted PIN(Personal Identification Number), face[23], and iris[9] authentication are unsuitable for most cardiac activity monitoring devices, due to the hardware restrictions of chest-worn devices, i.e., such devices do not have touchscreens or cameras. In recent years, heart sound-based [15] and ECG-based authentication[1] are regarded as the two most appropriate authentication methods, because they leverage existing monitoring signals(EEG, heart sound), thus have no need to upgrade their hardware.

However, these authentication methods are vulnerable to replay attacks in principle. The foundations of the above authentication systems lie in some unique and stable biometrics, including fingerprint, facial features, heart sound and ECG features. Due to the uniqueness and stability of these features, once one copy of features is unfortunately leaked or stolen by attackers, the authentication methods no longer protect the system from unauthorized access. According to literature, ECG-based authentication has been broken in[24]. The attackers used pre-recorded ECG signals stolen from a medical electronic database, and achieved a 62% successful attack rate. For heart sound-based authentication, once heart sound data is leaked through a medical database or unsafe household devices, attackers can spoof the microphone with pre-recorded sounds from the device owner[22] to access the system.

In this paper, we leverage liveness detection to enhance heart sound-based authentication against replay attacks. Liveness detection aims to distinguish live users and pretenders with pre-recorded data, that is, to defend replay attacks. However, the existing solutions are not applicable for heart sound-based authentication. VoiceLive[57] localized different articulation locations of phonemes in human language to achieve liveness detection for voice authentication. Compared with the complicated human vocal system which includes vocal cords, tongue, mouth and nasal cavity, heart sounds are just generated by four heart valves located close to each other. Thus, heart-sound localization could not provide enough entropy to defend against brute-force search attacking. Chen et al.[19] observed that loudspeakers can generate a magnetic field while a real human cannot, and leveraged it for liveness detection. However, it is not convenient for chest-worn devices since for every access, the user needs to take it off and move the device in the air to check the magnetic field variance. Environment noises are also utilized for liveness detection[48, 49, 52], however, these systems are sensitive to dynamic environments, thus have relatively high false positive rates.

Therefore, our system aims to fill the gap of liveness detection for heart-sound based authentication. We utilize the inherent correlation between sounds and chest motion caused by deep breathing to realize a reliable liveness detection system, BreathLive. During one deep breathe, voluntary contraction/expansion of the chest change the air pressure in the lung, and force the air to flow through the trachea into or out from the lung. Different air flow rates lead to breathing sounds of different amplitudes. Gyroscope and microphone on the chest-worn device can measure deep breathing from different aspects. The former measures breathing(chest contraction/expansion) motion, and the latter records incidental breathing sound. We leverage the inherently high-correlation of the above two measures for liveness detection. Beside existing microphone in heart sound authentication system, the widely-used and low-cost gyroscope can be easily be integrated into the system with limited overheads.

Nevertheless, there are still two significant challenges to realize such a liveness detection system. First, according to the state-of-the-art literature, none of the previous works leverage the high-correlation between microphone

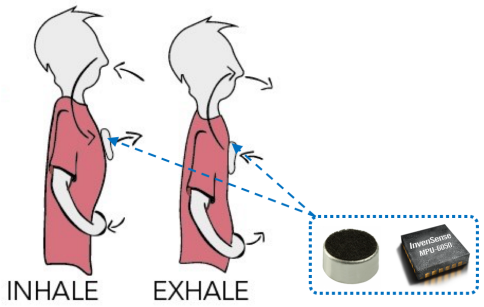

Fig. 1. Brief Illustration of Use Case.

and gyroscope measures of deep breathing for liveness detection. Thus, it is an unknown problem deserving serious research. Second, selecting precise features to quantify the above correlation is a non-trivial work. Feature selection is critical to the accuracy and robustness of a liveness detection system.

To solve the above challenges, we first analyze the deep breathing mechanism and find out two properties beneficial to liveness detection: 1) **Black-box property**. The exact relationship between breathing motion and breathing sound depends on the complicated structure of an individual respiratory system. Thus, it is difficult to infer the exact breathing motion from breathing sounds and vice versa, which significantly reduces the possibility of an attack. 2) **Randomness property**. Individual breathing not only varies across people but also from time to time. It depends on several factors: muscle status, emotion, and external environment[50], which makes precise deep breathing imitation even harder for attackers.

Given these properties, data leakage from the medical database and other unsafe devices makes much less of a threat. First, if only partial data are leaked, either deep breathing sounds or chest movement, based on black-box property, it is hard for attackers to infer the rest of the data so that they fail to break into the system. Second, if both data are leaked, based on randomness property, attackers can access the system only if they find two synchronous signals from the same deep breath(deep breathing sound and corresponding chest movement), which is very unlikely in real life. To sum up, the key insight to defend against replay attack is two signals with an inherent time-dependent (randomness property) and complicated(black box property) correlation — what "liveness" is about, which significantly reduces threats from data leakage of other devices and increases the difficulty of an attack.

Based on the analysis of the deep breathing mechanism, we propose BreathLive, a liveness detection system to enhance heart sound-based authentication against replay attacks. A typical user case is shown in Fig. 1. The user conducts deep breathing while chest-worn devices record related signals. Liveness detection and heart sound authentication will operate simultaneously, which the user is allowed to access only when s/he passes both liveness detection and authentication. In the liveness detection phase, our system extracts features, including correlation coefficient, amplitude ratio and duration ratio to evaluate the correlation between the sound(microphone) and motion(gyroscope) data of breathing, and use a pre-trained classifier to determine whether it is a live user(accept) or a pretender with pre-recorded data(reject).

BreathLive will enhance a heart-sound authentication system to defend against five levels of replay attacks with increasing capability: simple replay attack, gyroscope injection attack, random impersonation attack, advanced impersonation attack and advanced replay attack. The detailed threat models are elaborated in Section 3.2.

This work has three main contributions as follows:

(1) Based on the literature, this work is the first to show the potential of deep breathing for liveness detection, distinguishing between live persons and pretenders with pre-recorded data.
(2) We propose BreathLive, a liveness detection system to enhance heart sound authentication against replay attacks. A set of features are extracted by the designed algorithms to achieve reliable liveness detection.
(3) We implement BreathLive and conduct extensive evaluations under different attacker scenarios and contexts. Evaluation results show BreathLive can defend against five types of replay attacks effectively under various situations. Extension experiments also indicate our system is robust under different contexts with only a small training set.

Paper Roadmap Section 2 introduces the preliminaries of deep breathing. Section 3 presents a feasibility study of our system and illustrates the threat model. Section 4 elaborates the detailed design of BreathLive. Section 5 shows the implementation of the BreathLive system. Section 6 presents the evaluation and result analysis. Section 7 discusses some limitations and future work. Section 9 concludes the paper.

2 PRELIMINARY

In this section, we briefly introduce the operating mechanism of human breath and intrinsic reasons why deep breathing can be leveraged for liveness detection to enhance the security level of heart sound authentication.

2.1 Breathing Mechanism

Breathing is the process of gas exchange between the human body and the environment, which brings in oxygen and flushes out carbon dioxide. There are two phases of normal breathing: *inhalation*(or *inspiration*) and *exhalation*(or *expiration*). Many organs are involved in breathing, including nose, trachea, lungs, rib cage, muscles and diaphragm. During inhalation, the diaphragm(a large sheet of muscle under the lungs) and the intercostal muscles will contract to pull the rib cage upwards and bulge the abdomen out. Such a process extends the volume of the chest cavity, and lowers the pressure in the lung. At the same time, fresh air would pass through the nasal or mouth cavity and trachea, and eventually arrives at the lung. Exhalation is the reverse process of inhalation: the diaphragm and intercostal muscles relax so that the rib cage and abdomen return to their origin position, when the waste air will be flushed out of the body.[5].

Deep breathing, as a special form of breathing, is distinct from normal breathing in the following aspects. First, normal breathing is involuntary and imperceptible by people in most cases, while deep breathing generally occurs due to intended actions by people. Second, in deep breathing, abdomen muscles involve much more actively during inspiration so that the diaphragm is the main engine to drive deep breathing [6]. Third, the movement of the diaphragm and rib cage will be to larger degree in deep breathing than that in normal breathing, thus a larger volume of air will be exchanged in deep breathing.

2.2 Breathing Sound and Chest Movement

Medical researchers have investigated breathing sounds and chest movement for decades. Chest movement caused by muscle and rib cage movement can be measured by inertial sensors[45] on the chest to monitor the respiration cycle[14] and detect sleep apnea[42]. Inertial readings of the chest were revealed to be highly correlated with airflow waveform generated in breathing[14], because the intensity of chest movement has a direct effect on the volume and rate of airflow. In deep breathing, when the diaphragm, abdomen muscles and rib cage jointly change the pressure of the chest cavity and lung to a larger degree, more and faster airflow goes through the trachea and lungs. The larger airflow rate then produces a louder breathing sound. Many theories(linear[34], cube[55]) have been proposed to model and explain the high correlation between airflow rate and breathing sound[25]. However, the exact relationship between breathing sound and airflow in the respiratory system have not been firmly established.

Breathing characteristics vary across different people due to age[26] and gender[26]. Besides, for the same person, breathing characteristics vary across different times during a short term[50] due to different muscle status, emotions[16] and environment[5]. Thus, deep breathing has a randomness property beneficial to liveness detection.

The above analysis of deep breathing provides the following insights for our system:

(1) The degree of chest movement highly correlates to the intensity of breathing sound. In our target scenarios, people wear chest straps/bands to monitor heart rate/heart sound for fitness. Thus, we leverage the chest-worn setting, and utilize a chest-mounted MEMS gyroscope and microphone on the traditional auscultation site to measure the movement of the chest(diaphragm, rib cage) and record breathing sounds respectively.
(2) The exact relationship between breathing motion and breathing sound depends on the complicated structure of the individual respiratory system. Thus, it is difficult to infer the exact breathing motion from breathing sounds and vice versa, which significantly reduces the possibility of an attack.
(3) Breathing characteristics not only vary across people but also different breaths from the same person, which makes precise deep breathing imitation even harder for attackers.

3 FEASIBILITY STUDY & THREAT MODEL

In this section, we first present the feasibility study and then clarify the threat models to our system.

3.1 Feasibility Study

To validate our observations, we collect deep breathing sounds captured by the 3M Littmann Electronic Stethoscope Model 3200[4], and chest movement data from gyroscope on MPU6050 connected with Arduino Uno on two volunteers(A, B). The gyroscope and stethoscope are located together on the traditional auscultation site of chest. Then we smooth the data by calculating the short term energy of both signals(a detailed illustration can be seen in Sec. 4).

Fig. 2 and Fig. 3 show the acoustic signals and gyroscope readings from the deep breathing of two volunteers. The Pearson correlation coefficient between two signals is 0.9386 and 0.9414 for volunteer A and B respectively, while it is only 0.3197 between A's acoustic signal and B's gyroscope reading, and 0.3738 between B's acoustic signal and A's gyroscope readings. The above results validate the high-correlation between the acoustic signal and the gyroscope reading of deep breathing from the same person. More importantly, the correlation coefficients between volunteer A and B are quite low so that our system has the potential of a high security level.

It is also noticed that in Fig. 2 and Fig. 3, the detailed waveforms of deep breathing vary across different times. To verify the randomness property, we separate three deep breaths from the same person, and calculate the pearson correlation coefficient between acoustic signals and gyroscope readings from different deep breaths, e.g., acoustic signal from A's first breath and gyroscope readings from A's second breath, and etc. The average correlation coefficient is 0.6489 for A and 0.5203 for B. Thus, there exists randomness across deep breaths from the same person.

3.2 Threat Model

In the following, we define four attack scenarios for our system. In our scenario, the chest-worn device (e.g. heart rate band or chest straps) would be equipped with our system. Attackers fully understand how our system works and try to access victim's device. Heart and even deep breathing sounds are likely recorded and stored in other devices or at a hospital, thus it is possible that such information has been leaked to attackers. Once attackers get the a victim's device(e.g., the victim lost the device by accident or the attacker stole the victim's device), they will try to spoof the system using leaked information and their impersonation. Four types of attack scenarios are as follows:

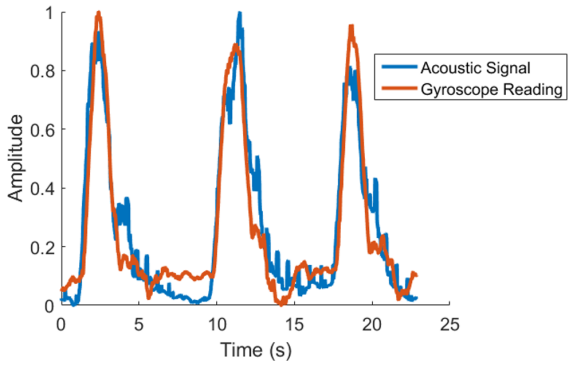

Fig. 2. The sound data and gyroscope data of deep breathing of volunteer A.

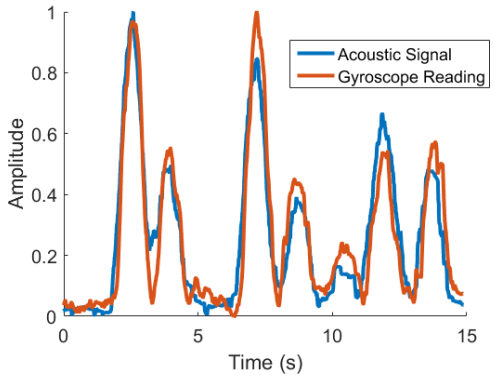

Fig. 3. The sound data and gyroscope data of deep breathing of volunteer B.

Simple Repaly Attack: The attacker simply injects the leaked heart/breathing sound into the device. Such an attack can bypass a heart sound or a breath sound authentication system [24]. However, BreathLive requires two inputs: acoustic signals and gyroscope readings. Even replaying heart/breathing sound with a loud speaker, it is less possible to generate the corresponding gyroscope readings.

Gyroscope Injection Attack: The attacker tries to play the breathing sound that has a direct effect on the gyroscope[39] which is quiet similar to a simple replay attack. The key idea is to vibrate the gyroscope readings with high energy sound.

Random Impersonation Attack: The attacker first wears the device on his own chest, then feeds leaked deep breathing sound of the victim to fool the microphone and conducts deep breathing in person to fool the gyroscope simultaneously.

Advanced Impersonation Attack: The attacker first listens to leaked deep breathing sounds and views the waveform of the victim at the same time repeatedly. The attacker tries to derive useful information from the chest movement pattern including the exact time of inhalation and exhalation. Then s/he mimics the deep breathing behavior(chest movement) based on the derived information to fool the gyroscope and feeds the leaked deep breathing sound of victim to fool the microphone simultaneously.

Advanced Replay Attack: The attacker somehow obtains breathing sounds and chest motion readings from different deep breaths of the victim, e.g. two signals come from different leaked databases or unsafe devices respectively. (In our system, signals will be only used one time for liveness detection and discarded immediately to prevent data leakage.) In this case, the attacker feeds a leaked deep breathing sound and chest motion reading from deep breaths to fool both the microphone and gyroscope simultaneously.

4 SYSTEM DESIGN

BreathLive can be integrated into chest-worn wearables, such as heart rate monitoring band, heart sound monitoring stripe or other household medical devices. It aims to enhance the security level of heart sound authentication by liveness detection with only a microphone and inertial sensor. In this section, we first give an overview of our system and then elaborate on hits detailed design.

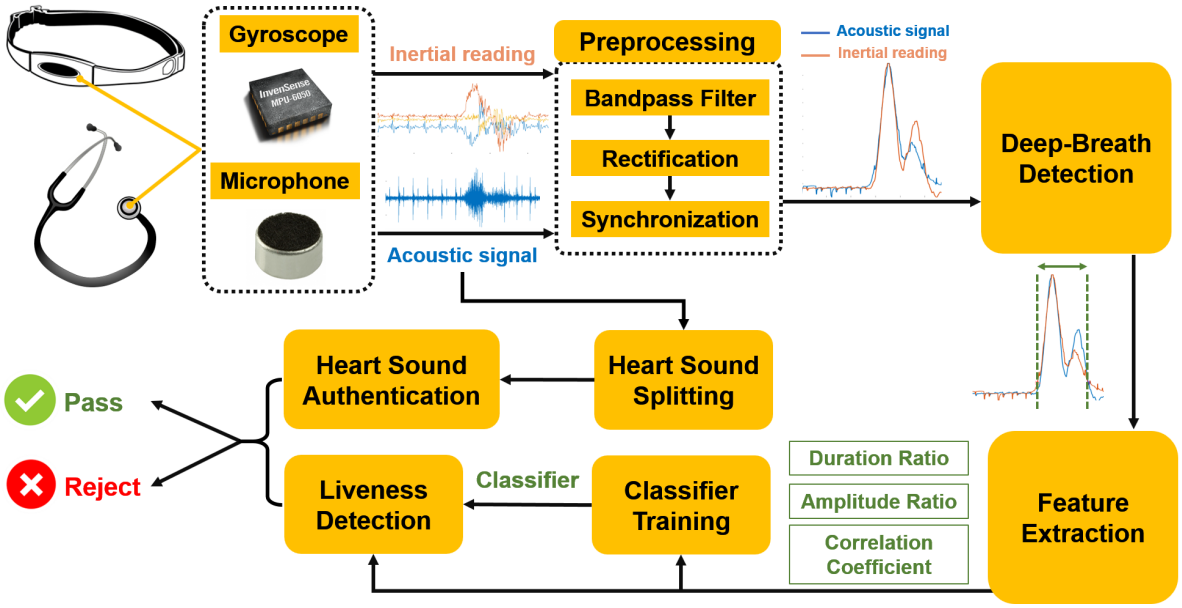

Fig. 4. BreathLive System Overview

4.1 Overview

Fig. 4 shows the overview of our system - BreathLive. The user should conduct deep breathing to access the device. During this process, the microphone captures both heart and deep breathing sounds. Heart sound will be extracted and sent to the heart sound authentication module, which is not the target of our system. Our system focuses on liveness detection, which utilizes the deep breathing sound and corresponding chest movement readings to determine whether it is a live person or a pretender with pre-recorded data.

As shown in Fig. 4, raw sound signals and gyroscope readings firstly pass through a preprocessing module including bandpass filter, short-term energy calculation and synchronization. Preprocessing for both signals are almost the same except some parameter settings. Then we segment the two signals and extract a set of features including duration ratio, amplitude ratio and correlation. In the classifier training phase, the system will train the classifier with features from the owner and other anonymous users, which can be stored in the device before using. In the liveness detection phase, the system will make a decision based on the measured features. Finally, the system would grant user's access only when s/he passes both authentication and liveness detection.

4.2 Heart Sound Splitting

Our liveness detection system is combined with a heart-sound authentication system. If authentication and liveness detection are decoupled, the attacker may pass the liveness detection by himself as a live person and then feed leaked heart-sound signals to fool the authentication — the divide and conquer strategy. To avoid this situation, the system will extract heart-sounds and breathing sounds from the same recordings, and feed them into the authentication and liveness detection respectively.

Our system borrows the method from paper[40] to extract heart sound signals. For each sound recordings, the system first uses 150 as the cutoff threshold and applies Hilbert transformation to get the envelope. Then the peaks of each heart sound will be captured after smoothing. We simply extract the signals which are one second

before and after the peaks as the heart sounds, since there is no need to distinguish different components of heart sound like S1, S2, etc.

We test the algorithm on 480 sound recordings from the deep breathing of 16 volunteers. We define the detection rate as the percentage of recordings where at least one heart sound is successfully extracted. The algorithm achieves a 99.2% detection rate. There are only four failed cases due to environment noise.

4.3 Preprocessing

BreathLive asks the user to conduct deep breathing, then the microphone and gyroscope will capture the deep breathing sounds $S(t)$, and corresponding chest movement readings $G(t)$ respectively. $S(t)$ and $G(t)$ will pass three steps of preprocessing module: bandpass filter, short-term energy calculation and synchronization.

Firstly, we adopt the bandpass butter-worth filter, a widely-used linear filter in bio-signal processing [44] to eliminate noise in the breathing sound and chest movement readings. We set the frequency range of the breathing sound filter from 100Hz to 1200Hz[26]. Filter for chest movement readings is 0.1Hz - 40Hz for two reasons: (1)Most common human body movement fall in the range of 0 - 20Hz[56]. (2) Human's respiration rates are larger than 0.1Hz.

Second, we calculate the short-term energy to tackle following two issues. (1) Two signals have different sampling rates (7800Hz for sound and 350Hz for gyroscope), (2) Breathing sounds contain inherent noise — heartbeat sounds since people can not breath deeply without the heart beating. The frequency of the heartbeat sound(50 - 200Hz) overlaps with that of the breathing sound(100Hz-1200Hz).

The short-term energy of sound signals($S(t)$) is defined as follows:

$$S_{energy}(t) = \frac{1}{T}\sum_{t}^{t+T}|S(t)|$$

For gyroscope data $G(t)$ on three axises, we first calculate the instant energy $\overline{G(t)}_e$as following:

$$\overline{G(t)}_e = \sqrt{G_x(t)^2 + G_y(t)^2 + G_z(t)^2}$$

Then the short-term energy of gyroscope data($G(t)$) is defined as follows:

$$G_{energy}(t) = \frac{1}{T}\sum_{t}^{t+T}|\overline{G(t)}_e|$$

Here, T denotes the windows size which controls the trade-off between smooth envelope and transient variations of the signal. We search the parameter space by cross-validation and find the system achieves the best results when T is set as 0.8 seconds. $S_{energy}(t)$ and ($G_{energy}(t)$) are the inputs to the following steps of our systems.

Although a synchronization line on the Arduino(illustrated in Section 5) has been used to reduce the delay between microphone and motion sensor during data collection, the delay cannot be totally removed due to following three reasons: (1)the imperfection manufacture of the crystal oscillators on two Arduino boards, (2)the lack of timestamps for overhead reduction in order to support the kilohertz sampling on the microphone, (3)the transmission delay between PC and Arduinos which may even be several seconds.

Thus, cross-correlation, the widely-used signal displacement measuring technology, has been adopted to totally synchronize two signals. We first perform dot production of two signal with a sliding window c, the delay r_{delay} can be determined by finding the maximum of results, this process can be expressed as follows:

$$r_{delay} = \arg\max_{t}(S_{energy} * G_{energy})(\tau) = \arg\max_{t}(\sum_{-\infty}^{\infty} S_{energy}(t)^* G_{energy}(t+\tau))$$

Given r_{delay}, two signals can be synchronized and aligned well.

4.4 Segmentation

The goal of the segmentation module is to detect the breathing section, i.e., extract $< begin, end >$ duration of the breathing section on both breathing sound and chest movement readings.

Before segmentation, we conduct normalization to eliminate the effect of different signal amplitudes of different people. Then, the duration search algorithm is applied to preciously determine the $< begin, end >$ duration. First, the two highest peaks in one breathing section represent inhalation and exhalation of deep breathing. They can be found based on several thresholds: mean/standard deviation of energy within a window size and the time between two peaks. We empirically set mean to be 0.4, standard deviation to be 0.1 in 0.1s window size, and time between two peak to be at least 0.2s. After finding these peaks, the algorithm firstly searches backward from the inhalation peak until the energy is lower than a threshold(set to be 0.1 in BreathLive). This timing would be regarded as *begin*. Then it searches forward from the exhalation peak to find *end* using a similar approach.

4.5 Feature Extraction

In this module, we extract the feature vectors $\vec{V}$ from the two signals for classifier training and testing. The key idea of feature selection is to select features that represent the similarity of the two signals — breathing sounds(S) and chest movement readings(G). The feature vectors extracted by BreathLive is $\vec{V}(S(t), G(t)) = (corr(S_{energy}(t), G_{energy}(t)), xcorr(S_{energy}(t), G_{energy}(t)), Mean_{R(t)}, Std_{R(t)}, Max_{R(t)}, Min_{R(t)})$. The extraction details are as follows, and relevant variables are denoted in Fig.5:

(1) *Correlation Coefficients,* $corr(S_{energy}(t), G_{energy}(t))$: This is the most common feature used to measure the relationship between samples. Given two synchronized signals, the pearson correlation is calculated as follows:

$$corr(S_{energy}(t), G_{energy}(t)) = \frac{cov(S_{energy}(t), G_{energy}(t))}{\sigma S_{energy}(t) \sigma G_{energy}(t)}$$

Here, cov denotes the covariance, $\sigma S_{energy}(t)$ denotes the standard deviation of $S_{energy}(t)$ and $\sigma G_{energy}(t)$ denotes the standard deviation of $G_{energy}(t)$. It is worth mentioning that we stretch the duration of sound and gyroscope data to the same length to facilitate the calculation of the Pearson correlation.

(2) *Maximum Cross Correlation,* $xcorr(S_{energy}(t), G_{energy}(t))$*:* Cross correlation measures the similarity between signals and the maximum cross correlation indicates how much the signal are similar to each other.

(3) *Mean, standard deviation, min and max of amplitude ratio:* If the two signal are from the same deep breathing, the statics of the amplitude ratio will remain stable. Although there may be some individual differences in this feature, we find the differences are quite small and can be fully covered with enough training set(the impact of training set size is shown is Sec. 6.6). Given two signals $S_{energy}(t)$ and $G_{energy}(t)$, the amplitude ratio $R(t) = \frac{S_{energy}(t)}{G_{energy}(t)}$ would be calculated. If we adopt amplitude ratio of every sample as features, the feature space is too large due to hundred Hz level signal sampling rate(Gyroscope). Thus, we adopt some statistics data of amplitude ratio — mean, stand deviation, min and max of $R(t)$.

(4) *Duration Ratio, D:* This feature measures the similarity from the time aspect. The duration ratio will be around one if the two signals are from the same deep breath. Given duration of two signal $< begin_S, end_S >$ and $< begin_G, end_G >$, duration ratio $D = \frac{D_S}{D_G}$ would be calculated, where $D_G = end_G - begin_G, D_S = end_S - begin_S$.

Finally, we choose correlation coefficients, mean of amplitude ratio, standard deviation of amplitude ratio, min of amplitude ratio and duration ratio as our features used in the evaluation part. The evaluation of the features can be seen in Sec. 6.5.

In the end, we standardize all the features to have zero-mean which has been widely used in many machine learning algorithms (e.g., support vector machines, logistic regression, and neural networks)[27]. This operation

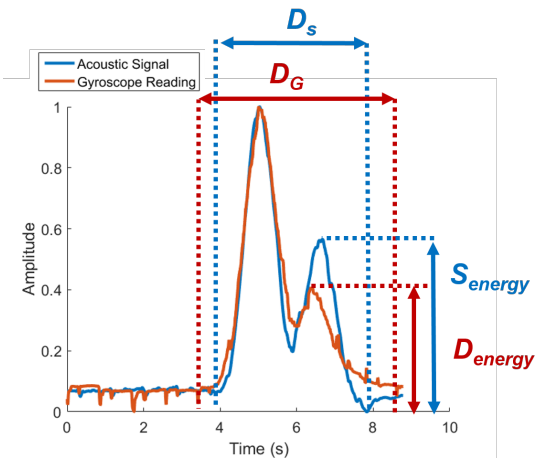

Fig. 5. Some variables including duration D_S, D_G, amplitude S_{energy}, G_{energy} used in the Section 4.5

can make the gradient descent procedure in classifiers quicker and more stable. It can be done by the following equation.

$$x' = \frac{x - \bar{x}}{\sigma}$$

Where x is the original feature vector, $\bar{x}$ is the mean of that feature vector, and σ is its standard deviation.

4.6 Classifier Training

In the enrollment phase, BreathLive trains the classifier using feature vectors from the owner and other people.

The system will label the feature vectors of the same deep breathing from the same person as 1(means accept) and the feature vectors from different deep breathings including the same person but at different times and different person at different times as 0(means reject). Then both of them are used to train the classifier. In our data, the negative labels(label 0) are in a larger quantity than the positive labels, so we oversample the positive data and undersample the negative data with random selection.

We evaluate three light weight classifiers: Logistic Regression(LR), Support Vector Machine(SVM) and Multilayer Perceptron(MLP). The reason for choosing the light weight classifier is the small computation cost and small delay.

Logistic Regression(LR) measures the relationship between categorical dependent variables and one or more independent variables by estimating probabilities using a logistic function. It has been widely used in binary classification tasks[32]. In our system, the max iteration times of the Logistic Regression is set to be 500.

Support Vector Machine(SVM) is a supervised learning model with an associated learning kernel that classifies the data. Different from Logistic Regression, SVM can distinguish data that are not linearly separable with a non-linear kernel function. In our system, the widely used non-linear Gaussian radial basis function (RBF) kernel[33] is adopted, which is defined as follows:

$$k(\vec{V}, \vec{V};) = \exp -\frac{-||\vec{V} - \vec{V}'||^2}{2s^2}$$

Here $\vec{V}$ and $\vec{V}$ are two samples and s defines how far the influence of a single training example reaches. To solve the imbalance problem between positive and negative training data(positive larger than negative in our system),

we set a cost function to weight the data. Soft margin parameter C represents the cost of the penalty factor for misclassification. We choose all the parameters using 5-fold cross validation on the training set (collection process in Sec. 6). Finally, we set s to be 10 and C to be 0.5.

Multilayer Perceptron(MLP) is a feed-forward artificial neural network with the backpropagation learning algorithm. It has three layers: input layer, hidden layer and output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. In our system, we use the common nonlinear function - sigmoid version logistic function. The loss function is the standard cross entropy. We tune the parameter of learning rate $eta1$ of the first layer weight, learning rate $eta2$ of the second layer and the number of hidden nodes h by 5-fold cross validation. Based on the validation performance, we set $eta1$, $eta2$ and h to be 0.5, 0.001 and 7 respectively.

4.7 Liveness Detection

Given a new pair of the deep breathing sound $S(t)$ and chest movement data $G(t)$, BreathLive first extracts feature vector $\vec{V}$ and inputs it into a pre-trained classifier. The classifier determines whether it will accept or reject the current access. As mentioned before, liveness detection only detects whether it is a live user or a replay attacker. The system would grant user's access only when s/he successfully passes both authentication and liveness detection.

5 IMPLEMENTATION

In this section, we illustrate how we build BreathLive prototype.

The system contains the following parts: arduino boards, amplifier, microphone, motion sensor and stethoscope diaphragm. Fig. 6 shows the hardware composition of our system. The stethoscope diaphragm is a plastic disc mounted on a metal head, which can be vibrated by body sounds if placed on the surface of the skin. A motion sensor MPU6050 Six-Axis(Gyrosocpe + Accelerometer) and an Electret microphone are used to capture signals. The former is attached on the surface of the stethoscope to capture chest movement, and the latter is placed inside the stethoscope to capture the sound. Since body sounds are rather low compared to other ambient sounds, we use Max9812, a low-cost 20dB fixed-gain microphone amplifier. To capture both signals with enough sampling rates, two Arduino Uno boards collect data from two sensors respectively. Arduino Uno is a microcontroller board based on the 16MHz ATmega328 with multiple digital pins and analog inputs. The amplified sound signal will be captured by the analog pin of Arduino 1 with the sampling rate of 7800Hz, which is high enough to capture the heart sound and lung sound. MPU6050 communicates with Arduino 2 through I2C communication protocol at the sampling rate of 350Hz. To minimize the delay between the two signals, a synchronization line would be set to trigger Arduino 2, when Arduino 1 begins data sampling. However, due to the imperfection manufactures of crystal oscillators on two Arduinos, the delay can only be reduced but not totally removed. That is the reason why synchronization is needed in our algorithm. A Windows 10 laptop will communicate with two Arduinos simultaneously and analyze the data off-line.

Fig. 7 shows the BreathLive prototype, and Fig. 8 shows the real usage scenario of chest-worn monitor wearables. The prototype can be fixed on the body using a strap. In the enrollment phase, the user needs to train the model by doing a few deep breaths. The system trains the classifier based on not only the data from this user but also from pre-restored data from other users. Then the system can function well to detect whether the current access is a live person or a pretender with pre-recorded sound.

6 EVALUATION

In this section, we first introduce the experiment setup for BreathLive and then test our system to answer the following questions:

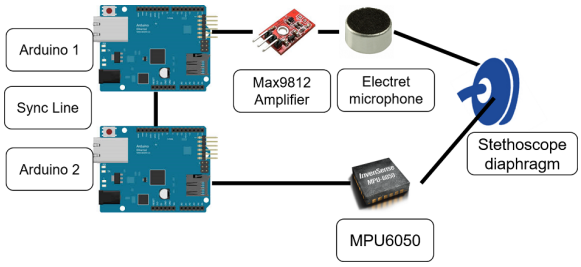

Fig. 6. The hardware diagram of BreathLive.

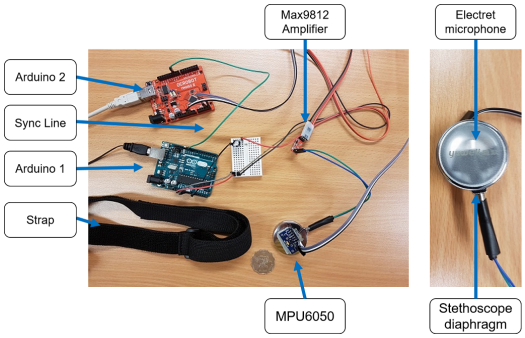

Fig. 7. The real prototype of our system. It is exactly the same structure as the hardware diagram shows.

Fig. 8. The typical user scenario of our system. The left-hand figure shows the device worn on the chest. The right-hand figure shows the approximate position on the rib cage and near the diaphragm. In practical usage, the device should be worn inside the clothes to capture a high quality signal. It is reasonable because most of the chest worn wearables on the market need to contact the skin.

(1) Can the user access the system with high accuracy?
(2) Is our system secure enough to prevent the device from replay attacks?
(3) What are the best features for our system?
(4) How much training data is required by our system?
(5) Is the system robust under different contexts?

6.1 Evaluation Setup

We recruit 16 volunteers(4 females, 12 males) to conduct a comprehensive evaluation. Before doing the experiment, we introduce the usage of our device to all the volunteers. Since BreathLive is a liveness detection system to enhance heart-sound authentication on the chest-worn wearables, the device should be placed on the chest where it can capture heartbeat signals. As shown in Fig. 8, the left side of the chest above the rib cage is an ideal position to capture both heart sound and breathing sound. It is worth mentioning that we do not instruct volunteers specific deep breathing methods, rather ask them to breath deeply as s/he used to do in daily life .

Our evaluation includes three phases: in phase 1, all volunteers act as “good” people to simulate normal device usage in daily life; in phase 2, the data of attackers under different attack scenarios are collected to conduct a

security analysis; in phase 3, the data of normal user under different contexts are collected to check the robustness of our system. We also evaluate the impact of features and training set size on the performance of our system. The evaluation details are illustrated in the following subsections.

6.2 Performance Metrics

In this paper, we adopt six commonly-used metrics to evaluate our liveness detection system.

True Accept Rate(TAR): the probability of identifying a live user as a live user, a higher true accept rate indicates the system is more likely to accept legitimate users.

True Reject Rate(TRR): the probability of successfully identifying an attacker as an attacker, a higher true reject rate indicates the system defends against replay attackers more effectively.

False Accept Rate(FAR): the probability of incorrectly treating an attacker as a live user, a lower false accept rate indicates the system is more unlikely to accept attackers.

False Reject Rate(FRR): the probability of incorrectly treating a liveness user as an attacker, a lower false reject rate indicates the system is more unlikely to refuse legitimate users.

Receiver Operating Characteristic(ROC) curve: ROC curve measures the performance of the binary classifier system under various discrimination thresholds. It can be obtained by plotting the true accept rate with respect to the false accept rate.

Equal Error Rate(EER): The rate where false accept rate equals to false reject rate.

The Area Under the Curve(AUC): The auc is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

6.3 Phase 1: Daily Usage

In the daily usage phase, all volunteers act as normal users to conduct deep breaths in three sessions, each of which contains ten deep breathing. The intervals between each session are at least half an hour. We conduct the experiment under 48dB ambient noise, and volunteers is stand during the deep breathing. In total, there are $10 \times 3 \times 16 = 480$ deep breathing recordings. To avoid intra-session information leakage, the first two breathing sessions per user would be used for training, while the last breathing session would be used for testing. We enumerate all the pairs of gyroscope readings and microphone sounds in the training set. Those pairs of signals from the same deep breathing of the same volunteer are labeled as 1(Accept), otherwise they are labeled as 0(Reject). In the daily usage phase, we only focus on the true accept rate since attackers are not yet involved.

The true accept rates of SVM, LR and MLP are 99.3%, 96.3% and 99.3%. The mis-classification cases in both SVM and MLP come from volunteer 8. The reason is that volunteer 8 has smaller chest movement than the others, which makes the segmentation of gyroscope data imprecise. The mis-classification cases in LR not only come from volunteer 8 but also volunteer 5. The reason is that volunteer 5 produces a small amplitude of the breathing sound due to her small lung capacity. It is obvious that SVM and MLP have better performance than LR, because SVM with RBF kernel and MLP are nonlinear classifiers that fit for separating non-linear features.

6.4 Phase 2: Security Analysis

In the security analysis phase, we first illustrate the experimental details and then show the results of five different attacks: simple replay attack, gyroscope injection attack, random impersonation attack, advanced impersonation attack and advanced replay attack. In these attack scenarios, we are more concerned about whether the attacker can successfully fool the system. Thus, we adopt true reject rate, equal error rate and ROC curve as our evaluation metrics.

Simple Replay Attack: In a simple replay attack, the device is placed on a desk or worn on the attacker's chest, then we feed the system a pre-recorded heart sound signal of the user to check whether it can fool the system.

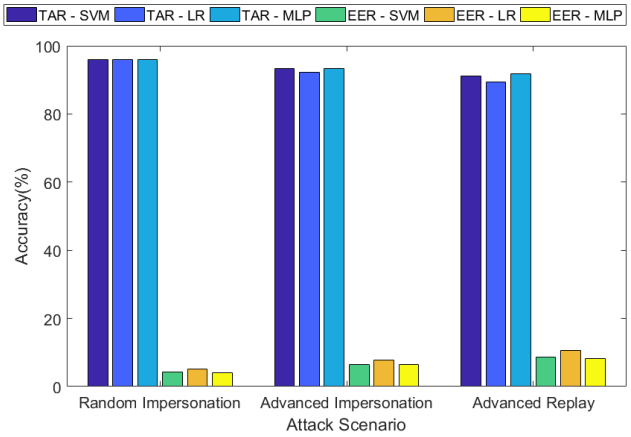

Fig. 9. The true reject rate(TRR) and equal error rate(EER) of SVM and LR under three attack scenarios

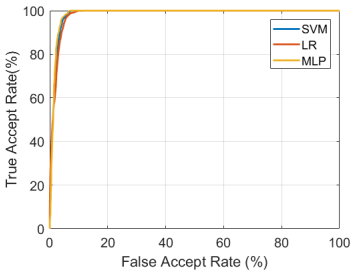

Fig. 10. The ROC curve of SVM and LR in the random impersonation attack. The AUC are 0.985, 0.983 and 0.986 for SVM, LR and MLP respectively.

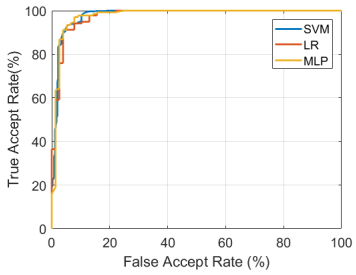

Fig. 11. The ROC curve of SVM and LR in the advanced impersonation attack. The AUC are 0.979, 0.976 and 0.979 for SVM, LR and MLP respectively.

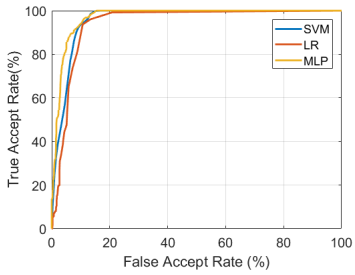

Fig. 12. The ROC curve of SVM and LR in the advanced replay attack. The AUC are 0.959, 0.942 and 0.971 for SVM, LR and MLP respectively.

However, such an attack is highly unlikely to succeed, because the short-term energy of the microphone signal is too low, which makes our algorithm hard to segment $< begin, end >$ duration. Without valid segmentation, our system could not run the classifier, and naturally defeat simple replay attacks. Thus, we only evaluate BreathLive on the rest three attack scenarios in the following.

Gyroscope Injection Attack: Gyroscope injection attack tries to leverage the sound vibration to simulate a gyroscope reading. The attacker first puts the stethoscope diaphragm directly on the loudspeaker, i.e., the distance between the gyroscope and loudspeaker is 0. Then the attacker feeds pre-recorded breathing sounds of the owner to the loudspeaker at different volumes to vibrate the gyroscope and spoof the microphone at the same time. To simulate such a scenario, we adopt the same classification models and data as that in Sec. 6.3, the volume levels of loudspeaker are 65dB, 70dB, 75dB, 80dB, 85dB and 90dB respectively.

Result of Gyroscope Injection Attack: The results show that no cases of gyroscope injection attack succeed. Based on result analysis, we find that the sound lower than 70dB barely has any effect on the gyroscope, i.e., the gyroscope readings remain at the same level as under white noise. The effect become visible when the sound

volume at the range between 70dB and 75dB. However, the short-term energy calculation of the data will totally smooth such effect, which makes our system hard to segment a $< begin, end >$ duration. When the sound volume falls to 80dB to 85dB, our system can extract valid $< begin, end >$ duration of the deep breathing from both sound and gyroscope readings. However, the amplitude of gyroscope readings is so small that all related features are too far away from the normal cases, thus the attacker fails to bypass the system. When the sound volume reaches 90dB or even larger, it exceeds the maximum input limit of our microphone almost all the time within one deep breathing cycle. This leads to low correlation between the short-term energy of two signals. To sum up, gyroscope injection attack fails to bypass our system at various volume levels of loudspeaker.

Random Impersonation Attack: The attacker feeds the victim's deep breathing sound to fool the microphone and performs the deep breathing behavior simultaneously to fool the gyroscope. To simulate such scenario, we adopt the same classification models and data as that in Sec. 6.3. To avoid the intra-session information leakage, first two breathing sessions per user would be used for training, while the last breathing session would be used for testing. All volunteers act as victim and attacker in turn. In one round, one of them acts as the victim and another one acts as the attacker, thus there are $16 * 15 = 240$ victim and attacker pairs in total. The training set includes all volunteers except the attacker(avoiding information leakage from the attacker), those pairs of gyroscope readings and sound signals from the same deep breathing are labeled as 1(denotes accept), otherwise they are labeled as 0(denotes reject). The testing set includes the rest data of victim and the pair of the sound signals from victim and the gyroscope readings from the attacker. The above procedure will run 50 times.

Result of Random Impersonation Attack: Fig. 9 shows the true reject rate, and the equal error rate of the system under three attack scenarios. Fig. 10 shows the ROC curve of the random impersonation attack. The true reject rate is 96.0% for SVM, 95.9% for LR and 96.0% for MLP. The equal error rate is 4.2% for SVM, 5.1% for LR and 4.0% for MLP. The AUCs are 0.985, 0.983 and 0.986 for SVM, LR and MLP respectively.The results show three classifiers achieve almost the same performance. Besides, random impersonation attack has rather low success rate, since the attacker does not know the exact deep breathing behavior of the victim.

Advanced Impersonation Attack: Three of the original 16 volunteers participated in the advanced impersonation attack experiment. Each of them randomly chose six victims out of the other 15 volunteers to mimic their deep breathing based on pre-recorded deep breathing sounds. The attacker would practice the deep breathing behavior of a victim as many times as s/he wants by listening to pre-recorded deep breathing audio. We also told them the exact timing of the inhalations and exhalations with the help of real time waveform display software during playing the audio. When the attacker was ready, s/he would mimic victim's deep breathing behavior 5 times. The pre-recorded deep breathing sound of the victim and the gyroscope readings of the attacker was fed into to the system simultaneously. In this scenario, there are $3 * 6 = 18$ pairs of victims and attackers and $18 * 5 = 90$ test cases in total. The training procedure is the same as that in the random impersonation attack.

Result of Advanced Impersonation Attack: Fig. 9 shows the true reject rate is 93.3% for SVM, 92.3% for LR and 93.4% for MLP, the equal error rate is 6.4% for SVM, 7.7% for LR and 6.4% for MLP. Fig. 11 shows the ROC curve. The AUCs are 0.979, 0.976 and 0.979 for SVM, LR and MLP respectively. The result shows SVM and MLP achieve the best performance. Compared with the random impersonation attackers, the advanced impersonation attackers have much more knowledge of the victim's deep breathing, thus more chance of fooling the system. However, the equal error rate is still small. The reasons are as follows: first, there exists an unknown time difference between the exhalation/inhalation and the increase/decrease of breathing sound waveform. Thus, attackers can not precisely mimic the deep breathing behavior of a victim. Second, given the pre-recorded sounds, attackers may mimic victim's breathing duration, but it is still hard to match the exact amplitude. Third, according to volunteers' feedback, they found hard to produce matched deep breathing behavior based on pre-recorded sounds even after practicing it many times. As mentioned in the introduction section, the black-box property of deep breathing behavior can prevent attackers from exactly mimicing the deep breathing behavior from the pre-recorded breathing sounds.

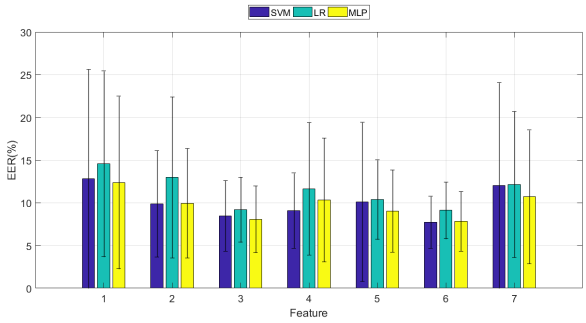

Fig. 13. The EER of different features. The value of Y-axis represents the average of the EER among all the subsets containing the certain feature. In X-axis, 1 to 7 represents different features.

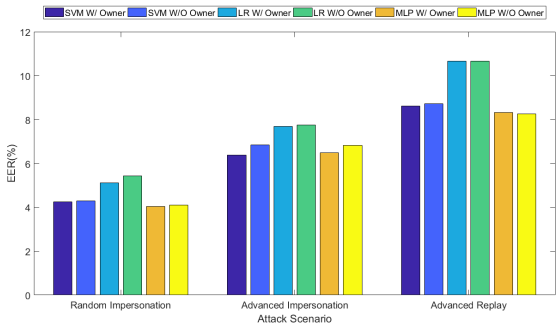

Fig. 14. EER of different classifiers train with and without owner's data under three attack scnearios

Table 1. Top 5 combinations of features with the lowest EER

Classifier	Feature Combination	EER	Classifier	Feature Combination	EER	Classifer	Feature Combination	EER
SVM	[2,3,4,5,6]	4.5%	LR	[2,3,4,5,6]	5.7%	MLP	[1,2,3,4,6,7]	4.3%
	[2,3,4,5,6,7]	4.5%		[2,3,4,6,7]	5.8%		[2,3,4,5,6]	4.4%
	[2,3,5,6]	4.6%		[2,3,4,5,6,7]	5.8%		[1,2,3,5,6,7]	4.5%
	[2,3,5,6,7]	4.7%		[1,2,3,4,6,7]	5.9%		[2,3,4,6]	4.5%
	[1,2,3,4,5,6,7]	4.9%		[1,2,3,4,5,6]	6.0%		[1,2,3,4,5,6,7]	5.6%

Advanced Replay Attack: The attacker feeds the system with gyroscope readings and breathing sounds of the victim collected at different times. We adopt the same classification models as that in Sec. 6.3, and conduct testing with the gyroscope data and sound data from different breathing samples of the same victim.

Result of Advanced Replay Attack: The true reject rate is 91.2% for SVM, 89.4% for LR and 91.7% for MLP, and the equal error rate is 8.6% for SVM, 10.6% for LR and 8.3% for MLP. The AUCs are 0.959, 0.942 and 0.971 for SVM, LR and MLP respectively.The results show MLP and SVM achieve a better performance than LR, because the MLP and RBF kernel of SVM are more suitable for nonlinear features. As mentioned in introduction section, the randomness property in deep breathing behavior can prevent advanced replay attacks. Such slight differences can be learned better by nonlinear classifiers. The reason for successful attacking cases is that there exists some similarity in the gyroscope readings and sound recordings from the different breathing samples of the same user, which is the basis of the advanced replay attack. However, such similarities are still smaller than that of gyroscope readings and sound recordings from the same breathing of the same user. Our classifier can distinguish them by learning the data from the different breathing data of other people. This is not the reason that our system learns something from the specific owner. It is explained in Sec. 6.6 about the impact of the owner's data, where it shows that given enough training dataset, the security level can be guaranteed without the owner's data. For example, MLP can still achieve a high performance with only four other people in the training set.

6.5 Impact of the Features

In this section, we test the various combinations of features and evaluate the importance of these features.

In Sec. 4.5, we define the feature vector $\overline{V}(S(t), G(t))$. However, not all the features play the same important role for liveness detection. We test all the subsets of these features to select the important ones due to a relatively small feature space. The total number of feature subsets is $2^7 - 1 = 127$. All the subsets are evaluated under all three attack methods

Fig. 13 shows the EER of seven features. The value of Y-axis denotes the average EER of each feature among all the combinations under three attack scenarios, and the error bar represents the standard deviation of value. The X-axis values, 1 to 7 represent the seven features, including maximum cross correlation, duration ratio, correlation coefficients, mean of amplitude ratio, standard deviation of amplitude ratio, minimum of amplitude ratio and maximum of amplitude ratio. The results indicate the EER of SVM is 12.5%, 9.4% 8.1%, 8.7%, 9.9%, 7.5%, 11.7% and the standard deviation is 13.6%, 6.0%, 4.0%, 4.4%, 9.7%, 3.1%, 15.2%; the EER of LR is 14.3%, 12.6%. 8.9%, 11.4%, 10.1%, 8.9%, 11.9% and the standard deviation is 10.9%, 9.2%, 3.6%, 7.8%, 4.6%, 3.3%, 8.5%; the EER of MLP is 12.4%, 10.0%. 8.1%, 10.3%, 9.0%, 7.8%, 10.7% and the standard deviation is 10.1%, 6.4%, 4.0%, 7.2%, 4.8%, 3.5%, 7.8%. Table. 1 shows the top 5 combinations of features with the lowest EER for SVM, LR and MLP.

The results show that the 'correlation' and 'minimum of amplitude ratio' are the most important features in all the classifiers. These two features have the smallest EER and standard deviation, and appear in all the top 5 feature combinations. Since our system tries to measure the similarity of two signals, it is reasonable that 'correlation' is an important factor. Moreover, we find that if two signals come from the same deep breathing, the 'minimum of amplitude ratio' is around 1. However, if two signals come from different deep breathing, it is highly probable that two signals have a large gap between their amplitude at a certain time, which results in a small 'minimum of amplitude ratio'. Another interesting point is that 'maximum cross correlation' is not very important as it only appears once in top 5 feature combinations per classifier. This is because 'maximum cross correlation' is mainly affected by the peak values of the two signals, which is not robust enough. 'Maximum of amplitude ratio' is less important than 'minimum of amplitude ratio', because it is hard to totally filter the heartbeat effect on gyroscope readings, which makes the signals fluctuate periodically. Such fluctuations cause a high amplitude ratio value at the beginning and ending of one deep breathing cycle. A more precise segmentation algorithm may make this feature more useful. Based on the results in Table. 1, we finally adopt [2,3,4,5,6] as features in all evaluations since they rank top 2 under all the classifiers.

6.6 Impact of the Training Set

In this section, we test the impact of the training set on our system. To be specific, we want to answer the following two questions: (1) Will the system performance change much with/without training data from the device owner? (2) How many training data/people are required by our system to achieve a satisfying performance?

To answer the first question, we leave the owner's data out of training set and test the system under three attack scenarios. The other evaluation setup is the same as that in Sec. 6.4. Fig. 14 shows that there is only a small improvement by including the owner's data in the training set. It indicates our system can extract the common features among people and has no need to learn individual features from the device owners.

To answer the second question, we change the number of volunteers involved in the training set from 1 to 16, and test the system under three attack scenarios. We also evaluate the system when the training set includes/excludes the owner's data. Other evaluation setups are the same as that in Sec. 6.4. Fig. 15 shows the EER of different training set sizes with the owner's data in the training set. Fig. 16 shows the EER of different training set sizes without the owner's data in the training set. The results show that performance will increase with more volunteers in the training set. However, the EER keeps stable when the size of the training set is larger than a threshold, which depends on the classifier and whether the training set includes or excludes the owner's data. For example, with the owner's data in the training set, the threshold is 4 for LR but 2 for MLP and SVM, while the threshold is around 4 for MLP without the owner's data in the training set. The reason is that with the

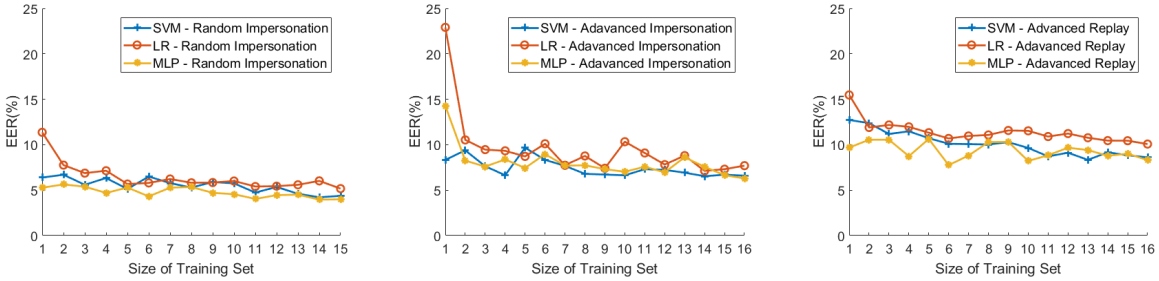

Fig. 15. EER of different training set size when training with owner's data under different attacks. The threshold for EER getting stable are 4, 2 and 2 for LR, SVM and MLP respectively.

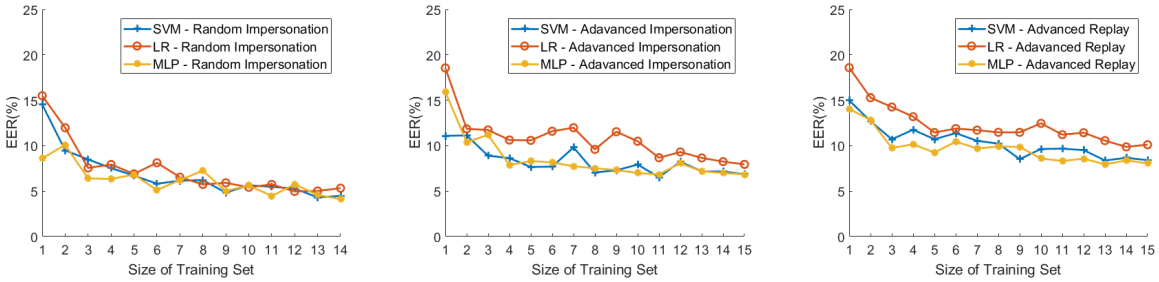

Fig. 16. EER of different training set size when training without owner's data under different attacks. The threshold for EER getting stable are 5, 5, 4 for LR, SVM and MLP respectively.

help of the owner's data, the classifier can more quickly find the appropriate boundary between the negative and positive data. Besides, the evaluation indicates that the system can still perform well with a small training set even without the owner. To sum up, MLP and SVM perform a little bit better at most times, and reach the stable status more quickly than LR. Moreover, the system can reach a stable status quicker by including the owner's data in the training set.

6.7 Phase 3: Context Experiment

In this section we investigate the impact of different physical activities and ambient sound on the performance of our system, i.e., the robustness of our system. Three of the original 16 volunteers participated in these experiments.

To test the impact of different physical activities, we collect 10 deep breathing when the volunteers are sitting, standing, after walking for 3 minutes and after running for 3 minutes respectively in the office. We also collect 10 deep breathing when the volunteers are reclining on beds, sitting, standing, after walking for 5 minutes and after running for 5 minutes respectively at home.

For walking and running in a home scenario, the volunteers ran and walked outside and then conducted breathing tests at home. To evaluate the system's robustness, the training set only includes data of these three volunteers from Sec. 6.3, and the testing set includes the data we collected under the different contexts described above.

Fig. 17 shows the true accept rate of our system. For an office setting, the true accept rate is 96.7%, 100%, 100%, 95.8% for SVM, 95.3%, 100%, 100%, 93.3% for LR and 93.3%, 100%, 96.7% for MLP when sitting, standing, after

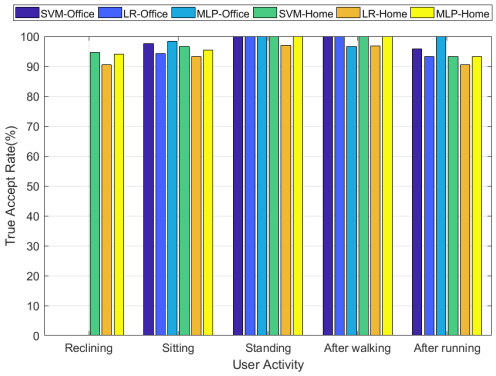

Fig. 17. The true accept rate of three volunteers under different physical activity scenarios.

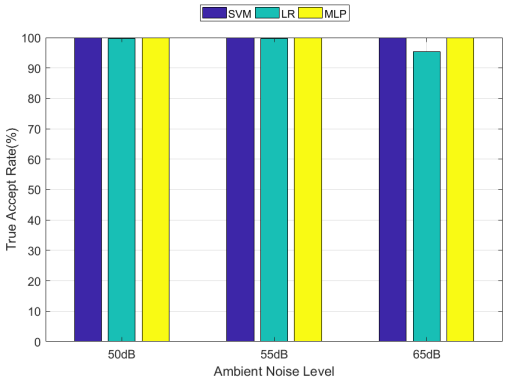

Fig. 18. The true accept rate of three volunteers under different ambient noise levels.

walking and after running respectively. For a home setting, the true accept rate is 94.7%, 97.6%, 100%, 100%, 93.3% for SVM, 90.4%, 94.3%, 97.0%, 96.7%, 90.5% for SVM and 93.7%, 98.3%, 100%, 100%, 93.3% for MLP when reclining, sitting, standing, after walking and after running respectively. In most cases, our system achieves satisfactory accuracy. It is found that 'standing' achieves the best performance because sitting or reclining postures may limit the muscle movement while breathing deeply but when the user is standing, s/he can contract and relax the muscles freely. The results show TAR at home is lower than that in the office because the user may feel more relaxed at home. However, the difference is very small and does not disturb the daily usage. The results also indicate the accuracy after running is not as good as other cases. This is because after running, the heart beat louder which interferes with the collection of deep breathing sounds. This is acceptable since in a fitness scenario, the device will generally be accessed before running and will not conduct liveness detection until the device is taken off.

To test the impact of ambient noise, we play music using a laptop to make the average sound levels measured at the position of device at 50dB, 55dB and 65dB respectively. In general, 50dB, 55dB and 65dB are the average sound levels in the library, the office and road traffic noise heard at a distance of 25m respectively. We collect 10 samples of deep breathing for each of the above settings. We adopt the classification model from Sec. 6.3 and train the system with data collected in this Sec. 6.3.

Fig. 18 shows the true accept rate is 100%, 100%, 100% for SVM, 99.8%, 99.8% and 95.4% for LR, and 100%, 100%, 100% for MLP under 50dB, 55dB and 65dB respectively. The results indicate that our system is robust under different levels of ambient noise.

To sum up, the results above show that SVM and MLP are more robust than LR in different contexts because they separate the features in a nonlinear way. Besides, the system can achieve a robust performance with only three volunteers' data included in the training set.

7 DISCUSSION

In this section, we discuss some limitations and potential solutions of our system.

7.1 Impact from Other Activities

To achieve better performance, the users are required to stay stable when performing the deep breathing. The larger body movement like walking, or running during the liveness detection can disturb the gyroscopic data.

However, this problem can be solved by wearable's paired smartphone that helps to eliminate the effect induced by other activities, which is one of our future works. Also our system does not require the user to stay stationary before or after liveness detection. The evaluation shows BreathLive is robust when user are under different contexts.

7.2 Different Classifiers and Potential Improvement

In this paper, we use three different machine learning classifiers: SVM, LR and MLP. It is obviously that the LR has the lowest accuracy comparing with other two methods. This is because SVM and MLP are nonlinear classifiers that fit for separating nonlinear features. When only considering SVM and MLP, MLP has a slightly better performance in all the attack scenarios and a slightly more robust than SVM. Besides, MLP has an advantage: online training for MLP is much easier than SVM. MLP model can be simply updated by Stochastic Gradient Descent or Adam. On the contrary, online training of SVM with RBF kernel is not trivial. However, MLP also has its own disadvantages: the training time for MLP is longer than SVM due to the more parameters.

Although the performance is satisfying, there is room for improvement in defending attacks more effectively. MLP achieves the lowest EER than the other two classifiers, thus, it inspires us that nonlinear classifiers with multiple layers such as deep neural network may enhance the system performance. However, the dataset we collected is not large enough for deep learning, which may result in over-fitting. Collecting more data to train a deep neural network is one of our future works.

7.3 High-speed Cameras

High-speed cameras can be a potential threat to our system. Assume there exists an attacker who is expert on video processing and has a deep understanding of our system. S/he steals the deep breathing sounds of a victim and meanwhile captures video clip of the victim with a high-speed camera. Then the attacker fully recovery the gyroscopic data based on the video. By using a machine such as a robotic chest to mimic the behaviour based on the gyroscopic data and feed the system with the stolen deep breathing sound, the attacker can access the device of the victim. However, this type of attack requires multiple expert skills. In addition, recalling that there exists randomness in deep breathing behaviour, sound and gyroscopic data stolen from different times cannot access our system from our evaluation. The requirement of stealing both signals from the same deep breathing of victim makes it more difficult.

7.4 Liveness Detection for ECG Authentication

Although we are aiming at heart sound authentication, the same principal can be implemented for ECG authentication. Instead of using a microphone, ECG authentication uses electrodes on a human body to capture the electric signal generated by heartbeat. From the literature, the principle of human movement is that the motor neurons transmit electrical signals, which is called electromyography(EMG), to cause relevant muscles to contract. The EMG signals can be captured by the electrodes, so there exists the potential that the EMG signal of the human chest movement is highly correlated to the gyroscopic data during deep breathing.

8 RELATED WORK

There are many works related to our work, we mainly focus on the following few areas:

(1) non-speech body sound detection.
(2) breathing monitor for different purposes.
(3) liveness detection of sound authentication.

8.1 Non-speech Body Sounds

Non-speech inner body sounds such as breathing sounds, heart sounds, eating sounds etc. can provide much useful information for detecting human behaviour and monitoring health status. There are many researchers trying to build a system to record and analyze these body sounds to derive such useful information.

Amft et al.[12] have detected whether the owner is eating and are able to recognize the category of food by analyzing the chewing sound captured from a microphone in the human ear. Their system achieves up to 99% accuracy in eating recognition and between 80% to 100% on food type classification. Eric et al.[36] proposed a privacy preserving coughing sensing system by putting a mobile phone in front of the user. PCA has been adopted to extract the coughing sound from ambient sound, thus preserving the privacy of the owner. Their system gets a 92% true positive rate and only 0.5% false positive rate. Bodyscope[54] is a wearable acoustic sensor for activity recognition which can capture sound from the human throat. It consists of a modified Bluetooth headset, an embedded microphone and the chestpiecce of a stethoscope. Their system can identify four different activities(eating, drinking, speaking and laughing) at about 71.5% accuracy. Bodybeat[43] is a mobile sensing system for capturing a wild range of non-speech body sounds and recognizing physiological reactions that generate these sounds in a real-life scenario. It is built with a diaphragm, a piezoelectric sensor and an acoustic isolator. In the paper, the authors test different materials of these components and find the optimum solution to record high quality body sound. They also develop a classification algorithm to distinguish different sounds of human behaviour(eating, drinking, deep breathing, coughing, laugh, silence, speech) at about 71.2% accuracy. SymDetector[21] is a wearable multi-sensor system for smart eating detection. Microphone, piezoelectric and accelerometer are positioned on the throat of the subject. Their system was tested on 7 participants with 14 eating episodes, resulting in 80.8% F-measure and 91.4% on eating and alone-time. All the work above motivated us to build a system that can recording breathing sounds from the inner body.

8.2 Breathing Monitor

Breathing occurs naturally for every live human and can reveal a multitude of useful information for fitness, medical care and even security purposes. Fitness has been investigated for a long time, many systems have been proposed to analyze the respiratory cycle[11, 14, 37, 38, 51], vital capacity[25, 35], and sleeping apnea[42]. Bates et al.[14] estimate the respiratory rate and flow waveform from a tril-axial accelerometer. SpiroSmart[35] measures the lung function with the help of the built-in microphone on a smart phone. The flow rate during exhalation is estimated by several signal processing algorithms, their method which was tested on 50 participants, shows a 5.1% mean error for common measures of lung function. SpiroCall[25], an advanced version of SpiroSmart, measures the lung function on any mobile phone without installing an app. The user can send the recording of the forced expiratory through the GSM voice channel and the lung function can be estimated on the service. Their system shows only 6.2% mean error on the four major lung function measures and can be used in many developing countries.ApneaApp[42] aims at monitoring the chest and abdomen movements to detect sleep apnea. Their system acts as an active sonar system which emits and receives the frequency-modulated sound signals. Thus different types of apnea can been distinguished at high accuracy. SymDetector[46] detected four types of respiratory symptoms in daily life.

The auscultation of the lung sound is one of the basic medical checks for human health. Some researchers focus on analyzing abnormal breathing sounds, for example, wheezing[28, 31, 47], crackling[17, 29], snoring[13, 53]. Styliani et al.[47] has proposed a time-frequency analysis algorithm to detect wheezing and shows a high accuracy and noise robustness. Charleston-Villalobos[17] adopts empirical mode decomposition to distinguish between different crackling sounds and normal breathing sound. Some researchers investigate disease related breathing sound behaviour, such as pneumonia[41], and sleep quality[30]. Raymond et al.[41] proposed a multi-channel

lung sound analyzer and found significant differences betwen lung sounds in patients with pneumonia and in asymptomatic controls.

However, our work is not aimed at detecting abnormal breathing sound but utilizing the deep breath for security purposes. The most relevant work is BreathPrint[18]. Chauhan et al. found different people have different breathing behaviour. They test three different breathing gestures: snoozing, normal and deep breathing to distinguish different people. They proposed a novel segmentation algorithm and extracted reliable features. The system tested under real life scenarios and shows high accuracy even under replay and impersonation attacks. However, the attacker in their scenario cannot capture the breathing sound while being too close to the victim. Their system may fail considering our scenario that the breathing sound print out may leaked from a fitness or medical device. And also Breathprint also utilizes stable features in breathing movements while we try to utilize the randomness in deep breathing.

8.3 Liveness Detection

Liveness detection has been widely used to defend against replay attacks in mobile authentication systems. Since biometrics, especially sound authentication, is vulnerable to reply attacks, liveness detection has been attracting a lot of interest. To defend against such attacks, many methods utilizing the sound features of the voice itself have been proposed[48, 49, 52]. However, the false positive rate is too high.

Combining microphone with other sensors is also a choice. Chetty et al.[20] detects lip movement by camera for liveness detection. VoiceLive[57] found different phonemes are generated from different positions in human body, which can be localized by multiple microphones. The entropy of the positions' possibilities ensure the security of their system and can be used in liveness detection. However, different from voice authentication, heart sound can only be generated from four positions in heart which means the entropy is too low for security purposes. Chen et al.[19] found the loudspeaker can generate a magnetic field while playing sound. Their system needs the user to move mobile phone in the air. The distance between the voice and the microphone has been calculated by gyroscope and accelerometer. Then such distance will compare with the variance in a magnetometer to determine whether the sound was generated by a loudspeaker. However, this approach is unfit in our scenario. This is because heart sound is hard to capture if the microphone is not in contact with the human body which is contrary to the requirement of moving the microphone in the air.

9 CONCLUSION

In this paper, we have proposed a novel liveness detection system, BreathLive, for heart sound authentication on a chest-worn device to defend against replay attack. BreathLive relies on the black-box and randomness property of deep breathing. After the user takes a deep breath, the microphone and gyroscope will capture the signal simultaneously. The bandpass filter, short term energy calculation and synchronization is used to preprocess both signals. Then the $< begin, end >$ duration is segmented and reliable features are extracted. In the classifier training phase, the system will be trained with the data of the owner and other users. In the liveness detection phase, the system will determine whether it is a live user or a replay attacker. We also implement the whole system and conduct comprehensive experiments under different attacker scenarios. The equal error rate of BreathLive is 4.0%, 6.4% and 8.3% for random impersonation attacks, advanced impersonation attacks and advanced replay attacks respectively. Our extensive experiments prove the system can be robust to different contexts including user activities and ambient noise with a small training set.

ACKNOWLEDGMENTS

This work was supported in part by the RGC under Contract CERG 16212714, 16203215, Contract ITS/143/16FP-A, and in part by the Grant AWS14J011 and Guangdong Natural Science Foundation No. 2017A030312008.

REFERENCES

[1] 2011. Nymi | Convenient Authentication Anywhere. https://nymi.com/. (2011).
[2] 2016. APPLE WATCH SERIES 2. https://www.apple.com/hk/en/watch/. (2016).
[3] 2016. SEEQ Mobile Cardiac Telemetry System. http://www.medtronicdiagnostics.com/us/cardiac-monitors/seeq-mct-system/index.htm. (2016).
[4] 2017. 3M Littmann Electronic Stethoscope Model 3200. http://www.littmann.com/3M/en_US/littmann-stethoscopes/products/~/3M-Littmann-Electronic-Stethoscope-Model-3200?N=5932256+8711017+3293188392&rt=rud. (2017).
[5] 2017. Breathing- Wikipedia. https://en.wikipedia.org/wiki/Breathing. (2017).
[6] 2017. Diaphragmatic breathing- Wikipedia. https://en.wikipedia.org/wiki/Diaphragmatic_breathing. (2017).
[7] 2017. How safe is your fitness tracker? Hackers could steal your data and sell the information to health companies. http://www.dailymail.co.uk/sciencetech/article-4049154/How-safe-fitness-tracker-Hackers-steal-data-sell-information-health-companies.html. (2017).
[8] 2017. POLAR H10 HEART RATE SENSOR. https://www.polar.com/hk-en/products/accessories/h10_heart_rate_sensor. (2017).
[9] 2017. Samsung Galaxy S8 | S8+. http://www.samsung.com/global/galaxy/galaxy-s8/. (2017).
[10] 2017. Thinklabs One - Digital stethoscope. http://www.thinklabs.com. (2017).
[11] Heba Aly and Moustafa Youssef. 2016. Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones. In *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 1–9.
[12] Oliver Amft, Mathias Stager, Paul Lukowicz, and Gerhard Troster. 2005. Analysis of chewing sounds for dietary monitoring. In *UbiComp*, Vol. 5. Springer, 56–72.
[13] Ali Azarbarzin and Zahra MK Moussavi. 2011. Automatic and unsupervised snore sound extraction from respiratory sound signals. *IEEE Transactions on Biomedical Engineering* 58, 5 (2011), 1156–1162.
[14] Andrew Bates, Martin J Ling, Janek Mann, and DK Arvind. 2010. Respiratory rate and flow waveform estimation from tri-axial accelerometer data. In *Body Sensor Networks (BSN), 2010 International Conference on*. IEEE, 144–150.
[15] Francesco Beritelli and Andrea Spadaccini. 2011. Human identity verification based on heart sounds: recent advances and future directions. *arXiv preprint arXiv:1105.4058* (2011).
[16] Frans A Boiten, Nico H Frijda, and Cornelis JE Wientjes. 1994. Emotions and respiratory patterns: review and critical analysis. *International Journal of Psychophysiology* 17, 2 (1994), 103–128.
[17] Sonia Charleston-Villalobos, Ramón González-Camarena, Georgina Chi-Lem, and Tomás Aljama-Corrales. 2007. Crackle sounds analysis by empirical mode decomposition. *IEEE Engineering in medicine and biology magazine* 26, 1 (2007), 40.
[18] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing Acoustics-based User Authentication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 278–291.
[19] Si Chen, Kui Ren, Sixu Piao, Cong Wang, Qian Wang, Jian Weng, Lu Su, and Aziz Mohaisen. 2017. You Can Hear But You Cannot Steal: Defending against Voice Impersonation Attacks on Smartphones. In *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 183–195.
[20] Girija Chetty and Michael Wagner. 2004. Automated lip feature extraction for liveness verification in audio-video authentication. *Proc. Image and Vision Computing* (2004), 17–22.
[21] Eli Cohen, William Stogin, Haik Kalantarian, Angela F Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2016. SmartNecklace: designing a wearable multi-sensor system for smart eating detection. In *Proceedings of the 11th EAI International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 33–37.
[22] Phillip L De Leon, Michael Pucher, Junichi Yamagishi, Inma Hernaez, and Ibon Saratxaga. 2012. Evaluation of speaker verification security and detection of HMM-based synthetic speech. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 8 (2012), 2280–2290.
[23] Mohammad Omar Derawi, Bian Yang, and Christoph Busch. 2011. Fingerprint recognition with embedded cameras on mobile phones. In *International Conference on Security and Privacy in Mobile Information and Communication Systems*. Springer, 136–147.
[24] Simon Eberz, Nicola Paoletti, Marc Roeschlin, Marta Kwiatkowska, I Martinovic, and A Patané. 2017. Broken hearted: How to attack ECG biometrics. (2017).
[25] Mayank Goel, Elliot Saba, Maia Stiber, Eric Whitmire, Josh Fromm, Eric C Larson, Gaetano Borriello, and Shwetak N Patel. 2016. Spirocall: Measuring lung function over a phone call. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5675–5685.
[26] Volker Gross, Anke Dittmar, Thomas Penzel, Frank Schuttler, and Peter Von Wichert. 2000. The relationship between normal lung sounds, age, and gender. *American journal of respiratory and critical care medicine* 162, 3 (2000), 905–909.
[27] Joel Grus. 2015. *Data Science from Scratch: First Principles with Python* (1st ed.). O'Reilly Media, Inc.
[28] İnan Güler, Hüseyin Polat, and Uçman Ergün. 2005. Combining neural network and genetic algorithm for prediction of lung sounds. *Journal of Medical Systems* 29, 3 (2005), 217–231.

[29] Leontios J Hadjileontiadis and Ioannis T Rekanos. 2003. Detection of explosive lung and bowel sounds by means of fractal dimension. *IEEE Signal Processing Letters* 10, 10 (2003), 311–314.
[30] Tian Hao, Guoliang Xing, and Gang Zhou. 2013. iSleep: unobtrusive sleep quality monitoring using smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 4.
[31] Antoni Homs-Corbera, José Antonio Fiz, José Morera, and Raimon Jané. 2004. Time-frequency detection and analysis of wheezes during forced exhalation. *IEEE Transactions on Biomedical Engineering* 51, 1 (2004), 182–186.
[32] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression*. Vol. 398. John Wiley & Sons.
[33] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. 2003. A practical guide to support vector classification. (2003).
[34] SS Kraman. 1984. The relationship between airflow and lung sound amplitude in normal subjects. *Chest* 86, 2 (1984), 225–229.
[35] Eric C Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N Patel. 2012. SpiroSmart: using a microphone to measure lung function on a mobile phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 280–289.
[36] Eric C Larson, TienJui Lee, Sean Liu, Margaret Rosenfeld, and Shwetak N Patel. 2011. Accurate and privacy preserving cough sensing using a low-cost microphone. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 375–384.
[37] Xuefeng Liu, Jiannong Cao, Shaojie Tang, Jiaqi Wen, and Peng Guo. 2016. Contactless Respiration Monitoring Via Off-the-Shelf WiFi Devices. *IEEE Transactions on Mobile Computing* 15, 10 (2016), 2466–2479.
[38] Junyi Ma, Yuxiang Wang, Hao Wang, Yasha Wang, and Daqing Zhang. 2016. When can we detect human respiration with commodity wifi devices?. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 325–328.
[39] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. 2014. Gyrophone: Recognizing Speech from Gyroscope Signals.. In *USENIX Security Symposium*. 1053–1067.
[40] Ashok Mondal, Parthasarathi Bhattacharya, and Goutam Saha. 2013. An automated tool for localization of heart sound components S1, S2, S3 and S4 in pulmonary sounds using Hilbert transform and HeronâĂŹs formula. *SpringerPlus* 2, 1 (2013), 512.
[41] Raymond LH Murphy, Andrey Vyshedskiy, Verna-Ann Power-Charnitsky, Dhirendra S Bana, Patricia M Marinelli, Anna Wong-Tse, and Rozanne Paciej. 2004. Automated lung sound analysis in patients with pneumonia. *Respiratory Care* 49, 12 (2004), 1490–1497.
[42] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 45–57.
[43] Tauhidur Rahman, Alexander Travis Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. 2014. BodyBeat: a mobile system for sensing non-speech body sounds.. In *MobiSys*, Vol. 14. 2–13.
[44] John L Semmlow and Benjamin Griffel. 2014. *Biosignal and medical image processing*. CRC press.
[45] H Sumbul and A Hayrettin Yuzer. 2015. Measuring of diaphragm movements by using iMEMS acceleration sensor. In *Electrical and Electronics Engineering (ELECO), 2015 9th International Conference on*. IEEE, 166–170.
[46] Xiao Sun, Zongqing Lu, Wenjie Hu, and Guohong Cao. 2015. SymDetector: detecting sound-related respiratory symptoms using smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 97–108.
[47] Styliani A Taplidou and Leontios J Hadjileontiadis. 2007. Wheeze detection based on time-frequency analysis of breath sounds. *Computers in biology and medicine* 37, 8 (2007), 1073–1083.
[48] Jesús Villalba and Eduardo Lleida. 2011. Detecting replay attacks from far-field recordings on speaker verification systems. *Biometrics and ID Management* (2011), 274–285.
[49] Jesús Villalba and Eduardo Lleida. 2011. Preventing replay attacks on speaker verification systems. In *Security Technology (ICCST), 2011 IEEE International Carnahan Conference on*. IEEE, 1–8.
[50] Elke Vlemincx, James L Abelson, Paul M Lehrer, Paul W Davenport, Ilse Van Diest, and Omer Van den Bergh. 2013. Respiratory variability and sighing: a psychophysiological reset model. *Biological psychology* 93, 1 (2013), 24–32.
[51] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. 2016. Human respiration detection with commodity wifi devices: do user location and body orientation matter?. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 25–36.
[52] Zhi-Feng Wang, Gang Wei, and Qian-Hua He. 2011. Channel pattern noise based playback attack detection algorithm for speaker recognition. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, Vol. 4. IEEE, 1708–1713.
[53] Azadeh Yadollahi and Zahra Moussavi. 2010. Automatic breath and snore sounds classification from tracheal and ambient sounds recordings. *Medical engineering & physics* 32, 9 (2010), 985–990.
[54] Koji Yatani and Khai N Truong. 2012. BodyScope: a wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 341–350.
[55] Meirav Yosef, Ruben Langer, Shaul Lev, and Yael A Glickman. 2009. Effect of airflow rate on vibration response imaging in normal lungs. *The open respiratory medicine journal* 3 (2009), 116.
[56] Hansong Zeng. 2012. *Bio-Inspired Inertial Sensors for Human Body Motion Measurement*. The Ohio State University.

[57] Linghan Zhang, Sheng Tan, Jie Yang, and Yingying Chen. 2016. Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1080–1091.

Received August 2017; revised November 2017; accepted January 2018