Breathing Sound-based Exercise Intensity Monitoring via Smartphones

Yanzhi Ren[1], Zhourong Zheng[1], Hongbo Liu(Corresponding Author)[1], Yingying Chen[2], Hongwei Li[1], Chen Wang[3]

[1]Dept. of CS, University of Electronic Science and Technology of China, 611731, China

[2]WINLAB, Rutgers University, NJ, 08854, USA

[3]Louisiana State University, LA, 70803, USA

[1]{renyanzhi05, hongweili}@uestc.edu.cn, [1]hongbo830117@gmail.com

[2]yingche@scarletmail.rutgers.edu

[3]chenwang1@lsu.edu

Abstract—Exercise intensity monitoring of physical activities has drawn increasingly attention as the awareness of the exercise intensity is of great importance for a person to achieve optimal training outcomes. For example, over-training could lead to excessive fatigue and loss of motivation for exercise. Traditional exercise intensity monitoring systems utilize GPS data to track the user's intensity of cardio activities through his/her position and speed. Such systems however become invalid for indoor exercises on stationary fitness equipments such as the treadmill or exercise bike. Recent work in using body-worn sensors to track the user's heart rate for exercise intensity monitoring usually involves additional wearable sensors which are only available on some particular fitness equipments, and thus are hard to be used in all occasions. This work presents an exercise intensity monitoring system which is capable of detecting a person's exercise intensity via smartphones. Our system exploits the off-the-shelf smartphone and its headphone to capture the user's breathing sound. Given the captured acoustic data, our system performs data pre-processing to remove the environmental noise and identify the non-silent acoustic frames based on the signal energy. Our system then conducts breathing event detection for non-silent frames, and further calibrates the detection results by utilizing the high correlation between breathing cycles to improve the detection accuracy. Moreover, our system can estimate the person's exercise intensity based on features extracted from the frames which contain breathing sound. Our experiments involving 9 subjects over four-month time period demonstrate that our proposed exercise intensity monitoring system is robust and accurate in both indoor and outdoor environments.

I. INTRODUCTION

Due to the accelerated pace of life and growing work pressure, people are paying more attention to their physical health and actively participating in daily aerobic exercise, such as running, jogging, cycling and dancing [1]. Recent studies show that regular exercise participation can improve the function of the cardiovascular system, increase the muscle strength, help with weight loss and reduce the risk of chronic disease such as diabetes, depression and obesity [2]–[4]. However, improper exercise intensity may damage the health and even result in unnecessary injuries or death in some specific cases [5]. In addition, over-training can also lead to excessive fatigue and loss of motivation for exercise [6]. Thus, the awareness of the exercise intensity is of great importance to achieve optimal training outcomes [7]. In this work, we seek to accurately measure the aerobic exercise intensity to provide users with reliable fitness guidance and feedback.

However, monitoring the exercise intensity is not trival. The main challenge lies in solutions providing accurate exercise intensity estimation across different exercise scenarios without requiring dedicated devices. Global Positioning System (GPS) sensor data collected from smart devices is effective to infer the intensity of cardio activities such as running or cycling through the position and speed information [8]. However, it is hard to be applied to measuring the intensity of activities with indoor stationary fitness equipments such as the treadmill, exercise bike or elliptical machine. In addition, the exercise intensity is a subjective measure of how hard physical activity feels to people while they are doing it, and such perceived level of exertion may be different among people even though they are doing the same exercise [9]. Thus, it is also relatively difficult to measure exercise intensity accurately for different users by merely using the position and speed information. Some recent techniques measure exercise intensity based on heart rate [10], but they usually involve additional sensors (e.g., chest-worn sensors) which are not always available across different scenarios. Along this direction, some commercial wrist-worn products such as Apple watch [11] or Garmin watch [12] have been designed to monitor the user's exercise intensity through heart rate using their embedded Photoplethysmography (PPG) sensors. However, it is not an easy task to accurately measure heart rate with such wrist-worn products during exercise [13]. Furthermore, a time delay (e.g., dozens of seconds) also exists between the heart rate changes and the actual exercise intensity changes [14], and it makes such system unable to provide real-time feedbacks to users.

To overcome the aforementioned weakness of existing solutions, we propose an accurate exercise intensity monitoring system without the involvement of dedicated sensors. Studies [15], [16] also show that the breathing sounds, characterized by breathing volume and rate, are correlated with the exercise intensity in real time. Our system thus captures the breathing sound induced by the air flow for exercise intensity estimation via the off-the-shelf smartphone and its headphone. To the best of knowledge, our work is the first that utilizes breathing sound to measure the exercise intensity by exploiting

the smartphone. But there are several challenges that need to be addressed: First, due to the fact that the breathing sound is intermittent, the collected acoustic signal usually contains many silent segments. Our proposed system should be able to identify non-silent segments from the acoustic signal accurately. Second, the breathing sound is usually relatively weak and is also susceptible to environmental interferences, such as the music in the gym, the running footsteps, the sound of wind or the traffic. Such environmental noises are mixed together with the breathing sound and may pose significant difficulty in identifying the breathing sound accurately from the recording. Third, the breathing pattern varies significantly among users in terms of frequency range and loudness, so it requires the exercise intensity detection method to be adaptive to different users.

Our system consists of four components to address the above challenges for accurate exercise intensity monitoring: *Acoustic Data Pre-processing*, *Breathing Sound Detection*, *Exercise Intensity Estimation* and *User Profile Construction*. Given the recorded sound signals, we first perform Acoustic Data Pre-processing to remove the environmental noise and identify non-silent frames. In Breathing Sound Detection, we extract acoustic features from each frame to detect possible breathing events embedded in the recorded sound signal. A detection result calibration technique is then applied to further improve the accuracy of breathing sound detection by leveraging the breathing cycle correlation inherent in a user's acoustic signal. Our Exercise Intensity Estimation component further calculates the statistical features from the identified breathing frames to perform the exercise intensity estimation for the user. In the meanwhile, during User Profile Construction, both acoustic and statistical features are extracted from pre-collected training breathing sound to construct the user's breathing profile and exercise intensity profile for breathing event detection and exercise intensity estimation, respectively. We summarize our main contributions as follows:

- We design an exercise intensity monitoring system leveraging the smartphone and its headphones to capture the breathing sound of users.
- We develop a robust breathing sound detection scheme by exploiting the inherent correlation relationship between the user's breathing cycles.
- We show that our approach has the capability to achieve unobtrusive exercise intensity monitoring for different users by using the detected breathing sound.
- We evaluate our system with 9 subjects over four-month time period in both indoor and outdoor environments. The results show that our system is highly accurate and robust under various scenarios.

II. RELATED WORK

There have been active studies on physical exercise monitoring leveraging various devices. Accelerometers can be attached on the user's body [17] or workout glove [18] to recognize the characteristics of exercises and count how many repetitions he/she has done so far. In addition, the passive RFID tags [19] can also be attached on dumbbells to recognize the free-weight workout in the gym. However, some dedicated hardware is required by these techniques. Along this direction, some commercial products such as Gym watch [20] have been designed for fitness monitoring. However, it still requires users to wear some additional sensors during exercise. Furthermore, all the systems/products mentioned above mainly focus on recognizing the type of exercise being performed or counting the number of exercise repetitions, and do not have the capability to perform exercise intensity monitoring.

Furthermore, there is also some work dedicated for exercise intensity monitoring. The smart devices may use the Global Positioning System (GPS) sensor data to track the user's intensity of cardio activities such as running or cycling through his/her position and speed [8]. However, this scheme could only infer the user's overall exercise intensity through analyzing the statistical information of the GPS data, and cannot provide real-time feedback to the user. In addition, it also becomes invalid for indoor exercises on fitness equipments such as the treadmill, exercise bike or elliptical machine. The heart rate based exercise intensity monitoring method has also been proposed [10]. The basic idea of this technique is that there exists a relationship between the user's exercise intensity and his/her heart rate, and it usually involves either chest worn sensors or commercial wrist-worn products (e.g., Apple watch [11] or Garmin watch [12]). However, body-worn sensors are usually limited to fitness usages while wrist-worn products do not have the capability to perform accurate heart rate monitoring during exercise [13]. Furthermore, there usually exists a time delay which is usually longer than dozens of seconds [14], between the heart rate changes and the actual exercise intensity changes, making the system unable to provide real-time feedbacks to the user.

Our work focuses on the aspect of providing accurate exercise intensity estimation via monitoring the user's breathing. Several work has been proposed for breathing monitoring. The capnometer system [21] measures the carbon dioxide concentration in exhaled air using a gas analyzer. However, the cost of such system is relatively high, making it unaffordable for most users. Some recent studies show that the wireless signal can be utilized for breathing rate detection [22], [23]. In particular, these approaches leverage dedicated wireless devices (e.g., Doppler radar-based sensors or WiFi devices) to capture small changes of the received signals caused by the user's breathing behavior. However, they are usually vulnerable to ambient interference or environment changes. Furthermore, they can only monitor the user's breathing rate when he/she is at rest (i.e., sitting or sleeping), and are not suitable for exercise scenarios. Hao *et al.* [24] propose to use the smartphone and its earphone to detect users' breathing sound, and further involves the smartphone accelerometer to assist the breathing rate detection. However, this method does not have the capability of exercise intensity monitoring.

Unlike the aforementioned work, we aim to perform exercise intensity monitoring by leveraging the user's breathing

(a) Low intensity   (b) High intensity
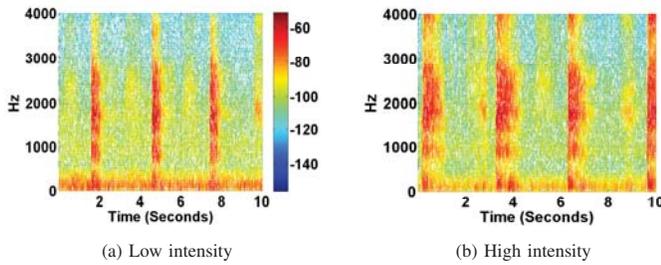
Fig. 1. Spectrograms of breathing sound under different exercise intensities for a specific user.

sound captured by the headphone, which is readily available with almost all the smartphones. Our proposed system does not need the active participation of users and is also easy to use without requiring any dedicated sensors or professional installations.

III. FRAMEWORK OVERVIEW

In this section, we discuss the preliminaries, system requirements and overview of our system design.

A. Preliminaries

In this work, to monitor the exercise intensity, we exploit the user's breathing sound captured by the off-the-shelf headphone. The benefit of using breathing sound is two-fold: (1) Universality: It is essential to have breath during exercise and the breathing sound is also correlated with the real-time exercise intensity of the user [15], [16]. It is thus feasible to provide an accurate and universal solution for exercise intensity monitoring under various scenarios. (2) Low cost and easy-to-use: we only rely on the microphone of the headphone to capture the breathing sound and facilitate intensity estimation without involving specialized hardware such as chest strap or wrist band. Thus, motivated by these aforementioned benefits, utilizing the characteristics within the breathing sound provides great opportunity for exercise intensity monitoring.

To validate the feasibility of exercise estimation leveraging breathing sound, Figure 1 presents an example showing the spectrograms of the breathing sound collected from a user under different exercise intensities. We can observe that the spectrograms of breathing sounds are significantly different between low and high exercise intensities. Specifically, the energy of the breathing sound under low exercise intensity lies mostly in the low frequency range (i.e., less than 3000 Hz). Whereas for high exercise intensity, most of the energy exists in the higher frequency range. In addition, when comparing with the low exercise intensity, we also find that more energy concentrates in the breathing phase when the user is in the high exercise intensity. These observations strongly suggest that we could perform exercise intensity monitoring by leveraging the user's breathing sound.
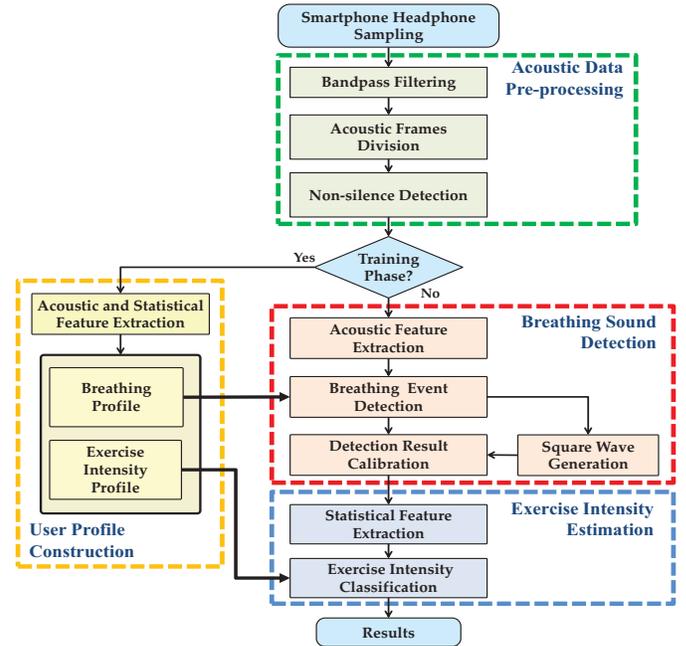
Fig. 2. System flow of the exercise intensity monitoring.

B. System Requirements

Our system aims to achieve noninvasive exercise intensity monitoring using smartphone and its headphone. Specifically, our system is designed to meet the following requirements:

Supporting Accurate Exercise Intensity Monitoring. Our system should be able to extract unique characteristics from a user's breathing sound for accurate exercise intensity monitoring. It could offer an important indicator for most aerobic training workouts to achieve the optimal training outcomes.

Easy to Use. Our system should be low cost and easy-to-use. Specifically, it should re-use existing devices in our daily life without dedicated sensors attached to the user's body since such sensors may affect the user's normal body movements during the exercise.

Robust Across Different Environments. As the background noises including the music in the gym, the sound of wind or the traffic outside are unavoidable during exercise, our system should be able to capture the weak breathing sound and provide accurate exercise intensity monitoring by mitigating the impact of noise.

Low Detection Latency. Our system should be able to estimate the user's exercise intensity with small number of acoustic measurements. In this way, the system can process the data on the fly and provide fast feedback to the user.

C. System Overview

The basic idea of our system is to utilize smartphone headphone to capture the breathing sound for exercise intensity monitoring. The smartphone headphone is used extensively during exercise. It is reported that over 60 percent of American runners use their headphone and smartphone to listen to music while running [25]. Furthermore, many people also tend to use headphone to watch videos, listen to news, or operate other apps from smart devices when they do aerobic exercises. It

is thus possible for us to explore utilizing the smartphone's headphone to capture the breathing sound for exercise intensity monitoring.

As illustrated in Figure 2, the system takes as input the acoustic data captured by the smartphone's headphone. In the acoustic data pre-processing phase, the acoustic data is first processed to remove the environmental noise using a band-pass filter and further divided into frames. An energy based technique is then adopted to accurately capture the non-silent frames from the acoustic signal. The next two components of our system are the breathing sound detection and exercise intensity estimation. Given the input non-silent sound frames, we first perform acoustic feature extraction (i.e, MFCCs) from each frame and identify the frames containing breathing events. A detection calibration technique is then developed to further improve the accuracy of breath detection by leveraging the inherent breathing cycle correlation in a user's acoustic signal. Finally, our system further calculates the statistical features from the sound signals of the identified breathing frames. Based on both statistical features and MFCCs, we use the random forest classifier to perform the exercise intensity estimation.

In the meanwhile, our system also conducts a one-time training process before the exercise intensity estimation. During the training process, each user needs to breath several times under different exercise intensities in a relatively quiet environment. The training sound captured by the headphone is fed into the same acoustic data pre-processing component. Similarly, both acoustic and statistical features are extracted from the non-silent frames and they are used as the user's breathing profile and the exercise intensity profile. At run-time, such profiles will then be utilized to train classifiers for breathing sound detection and exercise intensity estimation, respectively.

IV. EXERCISE INTENSITY ESTIMATION BASED ON BREATHING SOUND

In this section, we present the detailed system implementation of our breathing sound based exercise intensity monitoring system.

A. Acoustic Data Pre-processing

The basic idea underlying our monitoring system is based on the observation that the characteristics of user's breathing sound are relatively stable under certain exercise intensity and differ significantly between different exercise intensities. However, most people usually do exercise in a relatively noisy environment. The recording of breathing sound is susceptible to background noises, such as music in the gym, traffic noises or even the running footsteps. Furthermore, the inherent thermal noise of recording devices also affects the breathing sound recording. Both background and thermal noises could significantly degrade the performance of our exercise intensity monitoring system. In addition, due to the fact that the breathing is intermittent, the acoustic signal usually contains many silent segments. Thus, to build a robust system, we first perform the acoustic data pre-processing to reduce the impact of noises and then apply a non-silence detection method to detect the "clean" breathing sound segments.

1) Noise Reduction: Noise reduction aims to clean the recorded breathing sound by removing the noise components. In our work, we adopt a bandpass filter to remove the sound components with high or low frequency that are irrelevant to breath events. Specifically, given the sampling frequency as 8 kHz, the recorded sound signal is segmented into multiple frames with equivalent length $M = 128$. Then we apply the bandpass filter with lower and upper cutoff frequencies, 100 Hz and 3400 Hz, to each frame for noise reduction. The lower cutoff frequency of 100 Hz could filter the thermal noise at lower frequency band, while the upper cutoff frequency of 3400 Hz ensures that most breathing-related sound components are included [26].

2) Non-silence Detection: After noise reduction, our system performs non-silence detection to detect "clean" breathing segments. Assuming $\{r(l), 0 \leq l \leq N - 1\}$ are N acoustic samples and they have been equally divided into L frames, each frame is represented as $F_i = \{r(l), (i - 1)M \leq l < iM - 1\}, i = 1, ..., L$ with $LM = N$. Intuitively, the frames containing sound have higher energy than other frames, so we first attempt to identify the non-silence frames by calculating the energy for each frame. Specifically, the signal energy E_i of the frame F_i is derived as:

$$E_i = \frac{1}{M} \sum_{l=(i-1)M}^{iM-1} |r(l)|^2, i = 1, \cdots, L. \tag{1}$$

Next, a threshold-based method is developed to perform non-silence detection based on the signal energy $\{E_i, i = 1, \cdots, L\}$ across different frames. If E_i is larger than a pre-defined threshold Th, a non-silence frame F_i is identified. However, it is challenging to determine an appropriate threshold for non-silence detection as it may vary under different noise profiles. Therefore, we develop a dynamic threshold selection scheme based on the criteria introduced in [27]. Specifically, we first compute the histogram of the energy sequence $\{E_i, i = 1, ..., L\}$ and then apply a smoothing filter to this histogram. If K_1 and K_2 are the positions of the first and the second local maxima points in the histogram, the threshold Th is given as:

$$Th = \frac{\alpha K_1 + K_2}{\alpha + 1} \tag{2}$$

where the α is a weight parameter defined by the user. In this work, we empirically choose $\alpha = 5$ for the best performance.

Example. Figure 3 shows an example of non-silence detection based on a 20-second sound signal collected from a specific exerciser. The breathing sound signal is depicted in Figure 3(a), and the corresponding energy across different frames is shown as the green line in Figure 3(b). The threshold obtained from Equation 2 is also displayed as the black line in this figure. The non-silence frames could be accurately identified with the threshold, and represented in blue color in

(a) Non-silence segments identification.

(b) Energy sequence extraction.
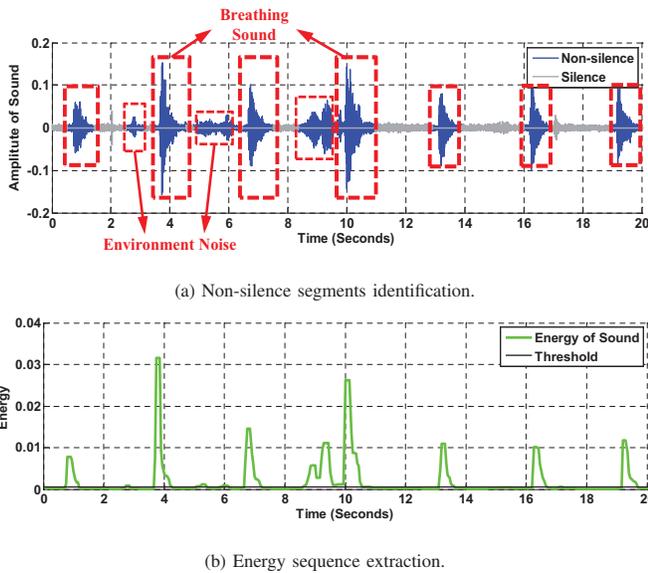
Fig. 3. An example of non-silence detection.

Figure 3 (a). This encouraging result confirms the feasibility of our approach to detect non-silence segments in the sound signal.

B. Breathing Sound Detection

Non-silence segments cannot be directly used to infer exercise intensity, so we need to extract representative breath-related features from the sound signals, and then adopt a Gaussian mixture model (GMM) as classifier to identify breathing events. To further improve the accuracy of the breath detection, we also propose a calibration technique by utilizing a square wave generated by the most likely breathing rhythm to reduce the number of incorrectly detected frames.

1) Acoustic Feature Extraction: The Mel-frequency cepstral coefficients (MFCCs), which are widely used in audio signal processing, could capture the unique characteristics of sound signal and are also insensitive to varying acoustic profiles. Thus, it can be used to distinguish different sound events (i.e., breathing or non-breathing events) embedded in the sound signals. Specifically, we compute the MFCCs for each frame F_i and take the first 12 coefficients as representative features (i.e., we have 12 features for each frame in total) for breathing event detection.

2) Breathing Event Detection: To detect breathing event in the non-silence frames, we first need to build the breathing sound profile for each individual user due to the significant differences of breathing pattern among different users in terms of strength, frequency range and acoustic characteristics. It is therefore necessary to design a training process to facilitate breathing event detection for different users. Specifically, the user is asked to record his breath sounds for several times under different exercise intensities in a relatively quiet environment. The quiet environment ensures a "clean" breathing sound to be captured with as little noise as possible. In the meanwhile, our non-silence detection scheme could identify the frames with breathing sound automatically and save the efforts for manually labeling breathing sound in the raw

acoustic signal. These detected breathing frames are then used to construct the user's breathing profile represented by MFCCs.

Based on the breathing profiles of different users, we develop a GMM-based classifier to detect breathing events. GMM is a probabilistic classification model and assumes instances from a dataset are generated by a mixture of Gaussian distributions with different mean and variance [28]. To build a GMM classifier with limited training data and avoiding overfitting, we empirically set the number of Gaussian distributions in GMM as 4. Specifically, two GMM models are built for each user for the classifier training: The first model, denoted as GMM_b, is created based on MFCCs extracted from the frames including user's breathing sound; while the other model GMM_n is constructed using MFCCs extracted from the frames with non-breathing sound.

During the breathing event detection phase, the MFCCs extracted from a run-time acoustic frame F_i are fed into GMM_b and GMM_n to obtain two different likelihood values: pb_i and pn_i. If F_i is non-silent and $pb_i > pn_i$, the breathing sound is detected and we label the frame F_i as $R_i = 1$ (i.e., breathing). Otherwise, the label is set as $R_i = 0$ (i.e., non-breathing or silent), indicating the absence of breathing sound in frame F_i. After all the non-silence frames are visited, the detected breath sequence are given as: $R = \{R_i, i = 1, ..., L\}$.

3) Detection Result Calibration: Although the breath events are roughly identified, it is still not applicable for accurate exercise intensity estimation due to that environmental noise still exists even after the noise reduction. Such noises, which may have similar acoustic features as the breathing sound, may cause a large number of "false positives" (i.e., non-breathing frames are incorrectly identified as breathing frames) in the detected breath sequence. To improve the detection accuracy, we take advantage of the high correlation between consecutive breathing cycles to mitigate the impact of background noise. This is due to the fact that breathing interval is relatively uniform over a short period of time, so breathing sound should also follow a cyclic pattern and appear periodically. Specifically, our system first identifies the breath period from the detected breath sequence R to generate a square wave sequence, which is referred as the simulated breath sequence. Our system then computes the correlation between the detected breath sequence and the simulated breath sequence under different parameter settings. The simulated breath sequence with the highest correlation value could best represent the user's possible breathing rhythm and will be used to reduce the number of incorrectly detected frames.

To generate the appropriate square wave, it is critical to estimate the user's breathing period first. Specifically, we calculate the similarity between any two frame labels from the detected breath sequence $R = \{R_i, i = 1, ..., L\}$ as a function of delay between them. Intuitively, the frame labels in R should be highly similar to each other when such delay is equal to the breathing cycle. Such delay could then be utilized to identify the user's breathing period. Specifically, we assume the possible minimum and maximum interval between
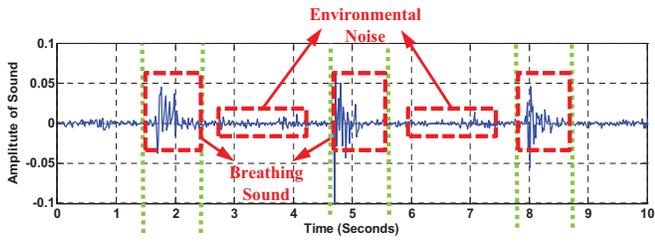
(a) Example of acoustic signal captured by the headphone.
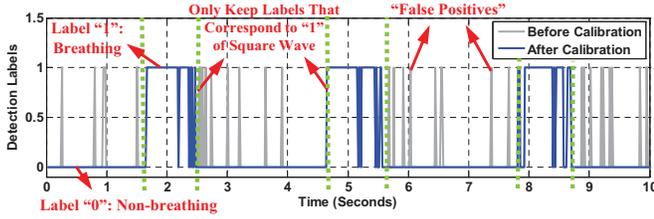
(b) Detection result calibration by utilizing the simulated breath sequence.
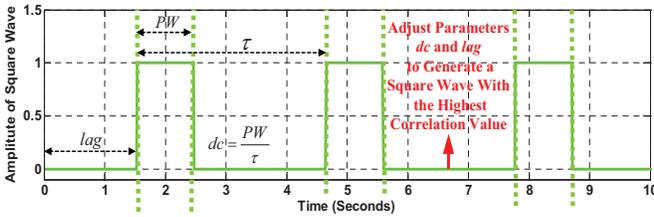
(c) The simulated breath sequence (i.e., square wave) generation.

Fig. 4. Illustration of detection result calibration.

the human's breathing cycles during exercise have T_{min} and T_{max} frames, respectively. Given the range of adult's normal breathing rate during exercise, we can empirically determine T_{min} and T_{max} as 75 and 375 frames, respectively. We then define a function $f(j)$ to measure the similarity as a function of delay j between them:

$$f(j) = \frac{\sum_{i=0}^{L-T_{\max}-1}(R_i - R_{i+j})^2}{N - T_{\max}}, T_{\min} \leq j < T_{\max}. \quad (3)$$

A larger value of $f(j)$ represents the similarity is lower and a smaller value represents the similarity is higher. Thus, we use the delay value τ that corresponds to the minimum value of $f(j)$ as the breathing period:

$$\tau = \arg \min_j f(j), \quad s.t., T_{\min} \leq j < T_{\max}. \quad (4)$$

We next generate a simulated breath sequence which could best represent the user's possible breathing rhythm to perform the detection result calibration. In particular, we generate a square wave sequence with a period of τ and refer it as simulated breath sequence. Our system then adjusts the duty cycle and time lag of the simulated breath sequence, and calculates the correlation between the simulated breath sequence and the detected breath sequence (i.e., $R = \{R_i, i = 1, ..., L\}$). The simulated breath sequence that gives the highest correlation value could represent the user's mostly likely breathing rhythm and will be utilized for calibration.

Specifically, we let $S = S(lag, dc) = \{S_i, i = 1, ..., L\}$ be a simulated breath sequence with a pre-determined period τ, where lag and dc denote its time lag and duty cycle, respectively. We then empirically set the search range of lag and dc as $[0.2, 0.6]$ and $[0, \tau]$, respectively. The correlation value is thus defined as:

$$c = \frac{Corr\{R, S(lag, dc)\}}{dc}, 0.2 \leq dc \leq 0.6, 0 \leq lag \leq \tau. \quad (5)$$

where the $Corr\{R, S(lag, dc)\}$ denotes the cross correlation between R and S. We next search for appropriate duty cycle dc and time lag lag to maximize the correlation value c between R and S. Finally, the corresponding dc and lag are chosen as the duty cycle and the time lag for the simulated breath sequence S.

In the calibration phase, we only keep the labels in the detected breath sequence which correspond to "1" of the simulated breath sequence and set other labels as "0". Specifically, for each R_i of R, the calibrated detection results are as follows:

$$R_i' = \begin{cases} R_i & \text{if } S_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $R' = \{R_i', i = 1, ..., L\}$. If $R_i' = 1$ indicates F_i does include the breathing sound we are interested, otherwise $R_i' = 0$.

Example. Figure 4 shows an example of detection result calibration. In this example, we collect the breathing sound from a user in a noisy environment and display it in Figure 4 (a). We then equally divide them into acoustic frames and perform breathing event detection to obtain the detected breath sequence as shown by both the blue and grey lines in Figure 4 (b). We can observe that environmental noises could cause some "false positives" in breathing event detection.

To improve the detection accuracy, we extract the breathing period τ and use it to generate a square wave, which is referred as the simulated breath sequence. We then adjust its parameters (i.e., duty cycle dc and time lag lag) and compute the correlation value c between such simulated breath sequence and the detected breath sequence. The simulated breath sequence that gives the highest correlation value, which is shown in Figure 4 (c), will be utilized for detection result calibration. Specifically, as illustrated by the blue lines in Figure 4 (b), only the detection labels which corresponds to "1" of the simulated breath sequence will be kept and other labels will be set as "0" (i.e., non-breathing) after the calibration. Thus, from Figure 4 we can observe that most "false positives" caused by environmental noise have been removed by our calibration algorithm. This result is encouraging as it indicates that our scheme is effective to identify breathing sound accurately even under noisy environments.

C. Exercise Intensity Estimation

To estimate the exercise intensity, we further extract the statistical features, which capture inherent characteristics of breathing sounds (e.g., the energy or variance), from the
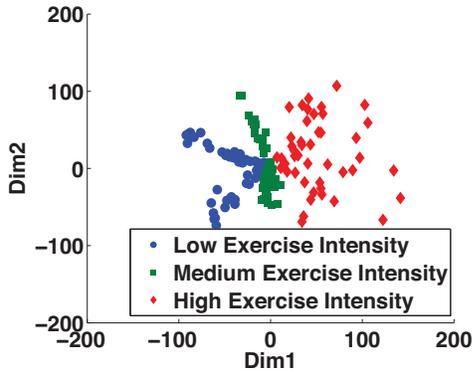
Fig. 5. An illustration of multidimensional scaling applied to features collected from a user's breathing sound under different exercise intensities.

frames that are identified as breathing events (i.e., the corresponding R'_i is detected as "1"). In particular, for each frame F_i, we derive 8 descriptive statistical features, including min, max, range, mean, standard deviation, root mean square, skewness and kurtosis. Based on such statistical features and MFCCs obtained from previous breathing sound detection component, we utilize the random forest [29] as classifier to estimate the user's exercise intensity (i.e., low, medium and high). The random forest is chosen due to its high accuracy and efficiency and it has been widely used for a large number of classification problems [29].

Specifically, in our classification model, we first label the training breathing frames corresponding to low, medium and high exercise intensities as "L", "M" and "H", respectively. We then select a subset of these frames to construct the user's exercise intensity profile, and train a random forest classifier with 50 bagged decision trees. In the exercise intensity estimation phase, the extracted statistical features and MFCCs obtained from the run-time breathing frames are fed into the random forest model and then the classifier outputs a prediction label (i.e., "L", "M" or "H") for each frame. To further improve the detection accuracy, we adopt a plurality vote based criteria, which chooses the most frequently occurring label among W consecutive breathing frames, to determine user's current exercise intensity. In this paper, we empirically choose $W = 50$ frames (i.e., 0.8 seconds) to achieve the best performance.

Feasibility Study. We next provide a feasibility study on how features (i.e., statistical features and MFCCs) derived for exercise intensity estimation change when they are extracted from acoustic frames which correspond to different exercise intensities. However, these feature vectors consist of 20 features (i.e., 8 statistical features and 12 MFCCs) and it is relatively difficult to visualize them in an intuitive way. To deal with this problem, we normalize each feature and compute features' dissimilarity matrix using Manhattan distance. The Multidimensional Scaling (MDS) [29] is then performed on such dissimilarity matrix. The MDS is an effective technique which could display the relative position of a number of multiple-dimensional objects in a two-dimensional figure, only given the distances among them. Intuitively, feature vectors

| (a) Running | (b) Cycling | (c) Elliptical Workout |
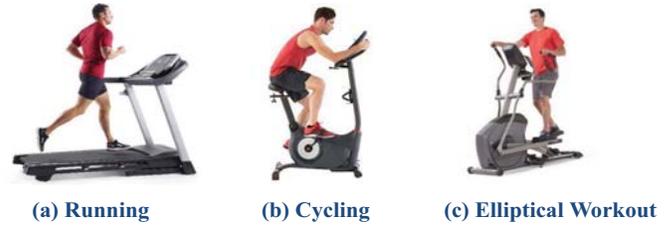
Fig. 6. Illustration of 3 types of exercises.

which are similar are placed close to each other on the map, whereas those vectors that are perceived to be very different are separated far away from each other.

Specifically, we collect one set of acoustic frames under each exercise intensity from a specific user with 50 frames per set. We compute statistical features and MFCCs for each frame, and then perform MDS on them. In particular, we display 20-dimensional features in a 2-dimensional figure as shown in Figure 5. In the example depicted in this figure, the nodes depicted by blue circle, green square and red diamond represent features extracted from low, medium and high exercise intensities, respectively. We can observe that a clear separation exists between features derived from different exercise intensities. This demonstrates that our proposed acoustic features can discriminate the user's breathing sound collected from different exercise intensities. These observations strongly confirm the feasibility of using our proposed features for exercise intensity estimation.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed exercise intensity monitoring system with 9 subjects over a period of four months. The following subsections detail our experimental methodology and results.

A. Experimental Methodology

We use two smartphones (i.e., Huawei Honor 10 and Huawei Mate 10) together with two Bluetooth headphones that record sounds under 8 kHz sampling rate. Each smartphone runs Android 9.0 operation system with 6 GB RAM and a 2.4 GHz Kirin 970 processor. The acoustic readings are collected and then written into a sound file on the smartphone

TABLE I
DEMOGRAPHICS OF VOLUNTEERS IN THE EXPERIMENT.

User ID	Gender	Age	Category
1	Female	10-20	Regular
2	Male	20-30	Rare
3	Male	20-30	Occasional
4	Male	20-30	Occasional
5	Male	20-30	Rare
6	Male	20-30	Rare
7	Female	20-30	Rare
8	Male	20-30	Rare
9	Male	30-40	Regular
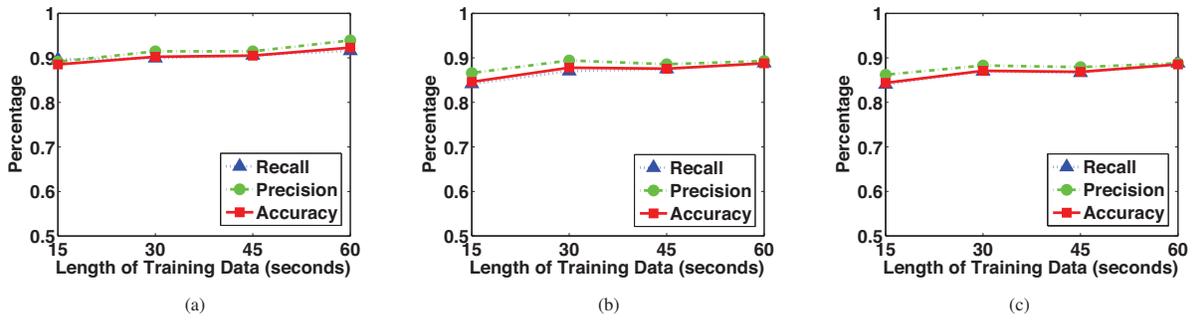
Fig. 7. Exercise intensity estimation for running exercise under different environments: (a) Quiet gym. (b) Noisy gym with music on. (c) Noisy playground with the traffic noise.
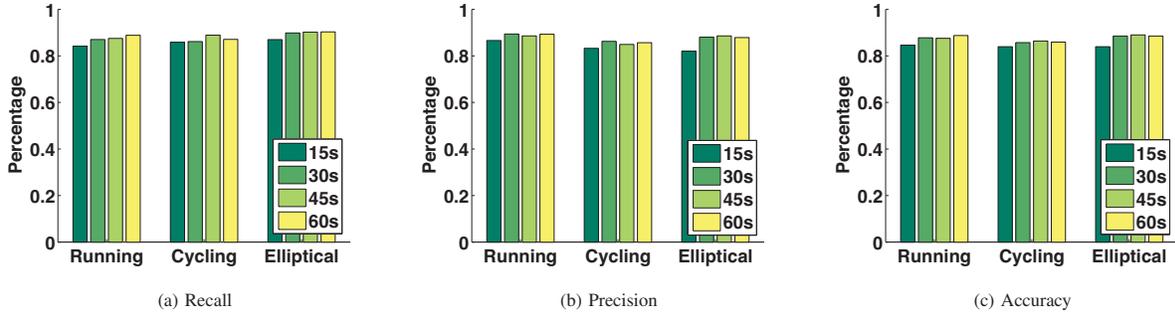
Fig. 8. Exercise intensity estimation for different types of exercise.

during the users' exercises. During the experiments, we let users connect the Bluetooth headset to smartphones, and then place the microphone of headphones near the user's nose. Such positions are quite natural since many users like to wear headset in similar positions during exercises. To make our system robust, users are also asked to adjust the position of the microphone so that the distance between the user and microphone is consistent during the experiment. We conduct experiments using three popular cardio exercises (i.e., running, cycling and elliptical workout as shown in Figure 6) under three different environments: *the quiet gym*, *the noisy gym with the music on*, *the noisy playground with the traffic noise*.

We conduct the experiments with 9 volunteers (ranging from 19 to 35 years old) over 4 months to evaluate the effectiveness of our system in exercise intensity monitoring. The demographics of subjects are detailed in Table I. A size of 9 volunteers is also typical for exercise monitoring studies [7], [24]. In addition, we also divide the subjects into 3 categories according to their self-report information: **Rare**: volunteers who rarely take any exercise; **Occasional**: volunteers who occasionally take short exercise; **Regular**: volunteers who usually take more than 60 minutes of exercise for each week. To obtain the ground truth of the exercise intensity, each volunteer is required to self-evaluate his/her exercise intensity (i.e., low or high) using the Borg's rating of perceived exertion (RPE) [30] after each exercise. Another student also accompanies with volunteers to record the exercise intensities according to their feedbacks. We use a subjective measure rather than an objective one (e.g., running speed) since the exercise intensity is a subjective measure of how hard physical activity feels to people and it may differ among people even

though they are doing the same exercise [9].

We use recall, precision and accuracy to evaluate the effectiveness of our system for exercise intensity monitoring. They are defined as follows:

Recall: the recall score for a certain exercise intensity is defined as the ratio of the number of the correctly detected instances associated with this exercise intensity to the total number of instances associated with this exercise intensity. The recall is thus the un-weighted mean of recall scores over all exercise intensities.

Precision: the precision score for a certain exercise intensity is defined as the ratio of the number of the correctly detected instances associated with this exercise intensity to the total number of instances which are detected as such exercise intensity. The precision is thus the un-weighted mean of precision scores over all exercise intensities.

Accuracy: the ratio of the number of the correctly detected instances to the total number of instances.

B. Impact of Environmental Noise

In the first set of experiments, we evaluate the performance of our proposed exercise intensity monitoring system for running exercise in Figure 7 with different lengths of training acoustic data under different environments (i.e., the quiet gym, the noisy gym with the music on, the noisy playground with the traffic noise). The legends "recall", "precision" and "accuracy" in Figure 7 denote the recall, precision and accuracy of our system, respectively.

Figure 7 (a) to (c) present the recall, precision and accuracy under different environments when the length of training acoustic data varies. We observe that the overall value of

recall, precision and accuracy remain more than 85 % across all environments when the length of training data is longer than 30 seconds. This demonstrates that our system can achieve a satisfactory performance with longer length of training data even under noisy environments. In addition, it also indicates that a training length of 30 seconds is sufficient for our scheme to achieve a high accuracy of exercise intensity estimation.

Further, this figure also shows that better detection recall, precision and accuracy can be achieved in the quiet environment, which indicates that the exercise intensities are more likely to be misclassified under a noisy environments. This is natural because it is relatively easier to identify the breathing sound from the noise in a relatively quiet environments. Finally, the figure clearly demonstrates that similar recall, precision and accuracy are achieved in the noisy gym and the noisy playground. This demonstrates that our system is also robust to both indoor and outdoor environments.

C. Robustness to types of exercise

We next study the robustness of our system when the run-time acoustic data is collected from different exercise. Figure 8 (a) to (c) present the recall, precision and accuracy under different lengths of training acoustic data when the run-time measurements are collected from running, cycling and elliptical workout respectively in a noisy gym with the music on.

We observe that the overall recall, precision and accuracy are higher than 80 % under different lengths of training data across all types of exercise. This result is encouraging as it indicates that our system is able to provide an accurate and universal solution for exercise intensity monitoring under various exercises. Further, we find that the recall, precision and accuracy decrease when the user is exercising on the cycling machine. This is because the noises generated by the exercise bike is relatively higher than the elliptical machine and the treadmill in our experiments, and such noise could degrade the performance of our system on both breathing sound detection and exercise intensity estimation.

Again, this figure also clearly shows that better performance could be achieved when the length of training acoustic data is longer than 30 seconds and it stabilizes when the training length is longer than 45 seconds. This is due to the fact that more breathing cycles which exist in the training data could capture the user's breathing characteristics more accurately. We also find that a length of 30 seconds is enough for our system to achieve a high accuracy of exercise intensity monitoring.

D. Effectiveness of breathing sound detection

Finally, we evaluate the effectiveness of our proposed breathing sound detection scheme. Specifically, we collect 3 acoustic recordings from each of 5 users with two-minute length for each recording. To get the ground truth of the breathing event, volunteers are required to listen to the acoustic recording and manually label the breathing sound from the raw data. We then use the breath cycle detection rate, which

(a) The quiet gym   (b) The noisy gym
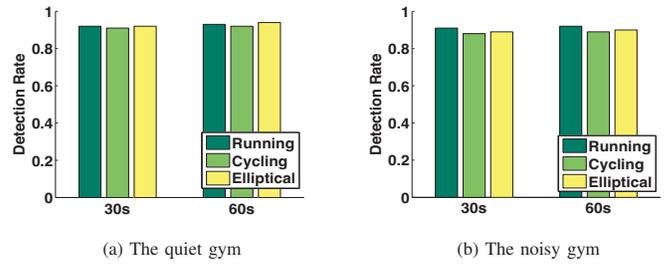
Fig. 9. Breath cycle detection rate under different environments.

is defined as the percentage of breath cycles that are accurately identified to evaluate the effectiveness of our proposed breath sound detection scheme.

Figure 9 (a) and (b) depict the detection rate derived from 3 exercises under different environments when the length of training acoustic data changes. We observe that the overall performance of our breathing sound detection scheme can achieve over 85 % detection rate across all scenarios. This indicates that our method is robust to different environmental noises and exercises. Further, we can observe that the detection rate increases when we conduct the experiments in a quiet gym. This is consistent with our expectations: it would be easier to identify the breathing sound accurately from the background noise in a relatively quiet environment. In addition, we can find that the detection rate becomes relatively lower for the elliptical workout. This is due to that the noise generated by the elliptical machine is relatively higher in our experiments and it can impact the performance of our proposed breath sound detection scheme. Overall, these observations indicate that our proposed breathing sound detection scheme can identify breathing events accurately.

VI. Conclusion

In this paper, we propose a practical system which can monitor an individual's exercise intensity utilizing the off-the-shelf smartphones. In particular, our system uses the smartphone's headphone to capture the user's breathing sound and estimate his/her exercise intensity. Our data pre-processing scheme could remove the environmental noise and detect the non-silent acoustic frames based on the signal energy. Our system then performs breathing event detection and calibrates the detection results by exploiting the high correlation between breathing cycles to improve the detection accuracy. Furthermore, our system estimates the person's exercise intensity based on features extracted from the frames which contain breathing sound. Through extensive experiments involving 9 subjects over four months time period, we show that our proposed exercise intensity monitoring system is accurate and robust in both indoor and outdoor environments. This strongly confirms the feasibility of using the smartphone and its headphone for exercise intensity monitoring.

REFERENCES

[1] "Physical Activity," https://www.who.int/dietphysicalactivity/pa/en/, 2019.

[2] E.T.Howley, "Type of activity: resistance, aerobic and leisure versus occupational physical activity," *Med Sci Sports Exerc.*, 2001.

[3] B. Pedersen and B. Saltin, "Evidence for prescribing exercise as therapy in chronic disease," *Scand J Med Sci Sports*, 2006.

[4] A. Semeco, "The top 10 benefits of regular exercise," 2017. [Online]. Available: https://www.healthline.com/nutrition/10-benefits-of-exercise

[5] B. Celli and et al, "Standards for the diagnosis and treatment of patients with copd: a summary of the ats/ers position paper," *European Respiratory Journal*, 2004.

[6] H. Kuipers and H. Keizer, *Overtraining in elite athletes. Review and directions for the future*. Sports Med., 1988.

[7] I. Pernek, G. Kurillo, G. Stiglic, and R. Bajcsy, "Recognizing the intensity of strength training exercises with wearable sensors," *Journal of Biomedical Informatics*, 2015.

[8] "MapMyRun," https://www.mapmyrun.com/, 2019.

[9] "How to measure exercise intensity," http://diet.mayoclinic.org/diet/move/how-to-measure-exercise-intensity, 2020.

[10] "Monitoring Excercise Intensity Using Heart Rate," https://www.nifc.gov/FireFit/documents/heartrate.pdf, 2001.

[11] "Apple Watch," https://www.apple.com/watch/, 2018.

[12] "Garmin watch," http://www.garmin.com/en-US/, 2018.

[13] "Wrist-worn Heart Rate Monitors Less Accurate Than Standard Chest Strap," https://www.acc.org/about-acc/press-releases/2017/03/08/14/02/wrist-worn-heart-rate-monitors-less-accurate-than-standard-chest-strap, 2017.

[14] L. H. Epstein, R. A. Paluch, L. E. Kalakanis, G. S. Goldfield, F. J. Cerny, and J. N. Roemmich, "How much activity do youth get? a quantitative review of heart-rate measured activity," *Pediatrics*, 2001.

[15] "Respiratory Responses to Exercise," https://www.ptdirect.com/training-design/anatomy-and-physiology/acute-respiratory-responses, 2018.

[16] A. Yadollahi and Z. Moussavi, "Comparison of flow-sound relationship for different features of tracheal sound," in *Proceedings of IEEE EMBC*, 2008.

[17] F. Mokaya, R. Lucas, H. Y. Noh, and P. Zhang, "Myovibe: Vibration based wearable muscle activation detection in high mobility exercises," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015.

[18] K. Chang, M. Y. Chen, and J. Canny, "Tracking free-weight exercises," in *Proceedings of the 9th international conference on Ubiquitous computing (ACM UbiComp)*, 2007.

[19] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "Femo: A platform for free-weight exercise monitoring with rfids," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, 2015.

[20] "Gymwatch," https://www.gymwatch.com/, 2018.

[21] F. Takatori, S. Yamamori, M. Inoue, S. Abe, and K. Miyasaka, "A novel mainstream capnometer system for non-intubated pediatric patients requiring oxygen administration," in *IEEE Eng Med Biol Soc*, 2011.

[22] X. Wang, C. Yang, and S. Mao, "Tensorbeat: Tensor decomposition for monitoring multiperson breathing beats with commodity wifi," *ACM Transactions on Intelligent Systems and Technology*, 2017.

[23] F. Lin, Y. Zhuang, C. Song, A. Wang, Y. Li, C. Gu, C. Li, and W. Xu, "Sleepsense: A noncontact and cost-effective sleep monitoring system," *IEEE Transactions on Biomedical Circuits and Systems*, 2017.

[24] T. Hao, G. Xing, and G. Zhou, "Runbuddy: A smartphone system for running rhythm monitoring," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015.

[25] "The State of the American Runner," https://www.runnersworld.com/runners-stories/a20825842/the-state-of-the-american-runner-2016/, 2016.

[26] N. Gavriely, M. Nissan, A. H. Rubin, and D. W. Cugell, "Spectral characteristics of normal breath sounds," *J Appl Physiol*, 1981.

[27] T. Giannakopoulos, "A method for silence removal and segmentation of speech signals, implemented in matlab," *University of Athens, Athens*, 2009.

[28] D.Titterington, A.Smith, and U.Makov, *Statistical Analysis of Finite Mixture Distributions*. Wiley, 1985.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining Inference, and Prediction*. Springer, 2001.

[30] G. Borg, "Psychophysical scaling with applications in physical work and the perception of exertion," *Scand J Work Environ Health*, 1990.