

**Enabling Smart Health Applications via Active Acoustic Sensing on
Commodity Mobile Devices**

by

Xingzhe Song

Bachelor of Engineering, Harbin Institute of Technology, China, 2015

Master of Engineering, Harbin Institute of Technology, China, 2017

Submitted to the Graduate Faculty of
the Swanson School of Engineering in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH
SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Xingzhe Song

It was defended on

July 12, 2022

and approved by

Dr. Amro El-Jaroudi, PhD, Associate Professor, Department of Electrical and Computer
Engineering

Dr. Zhi-hong Mao, PhD, Professor, Department of Electrical and Computer Engineering

Dr. Murat Akcakaya, PhD, Associate Professor, Department of Electrical and Computer
Engineering

Dr. Theodore Huppert, PhD, Associate Professor, Department of Electrical and Computer
Engineering

Dr. Hongshuai Li, MD, PhD, Assistant Professor, Department of Orthopedics and
Rehabilitation, University of Iowa

Dissertation Director: Dr. Wei Gao, PhD, Associate Professor, Department of Electrical
and Computer Engineering

Copyright © by Xingzhe Song
2022

Enabling Smart Health Applications via Active Acoustic Sensing on Commodity Mobile Devices

Xingzhe Song, PhD

University of Pittsburgh, 2022

Past decades have witnessed the development and prosperity of mobile and wearable devices with their merits of portability, energy efficiency and computational capability. Among applications implemented on these devices, smart health is an emerging field that deploys mobile applications to monitor vital signals, manage healthcare records and conduct disease diagnosis. However, existing techniques are limited to measure health metrics with evident biomarkers such as induced sound or visual change unless dedicated medical sensors are attached. To improve the practicality and feasibility of smart health applications, this thesis aims to utilize active acoustic sensing on subtle biomarkers that are not audible or observable.

This thesis has three major aspects to establish the active acoustic sensing framework using speaker-microphone pair that are widely available on mobile devices. First, a new system design is proposed to support complete, accurate yet reliable spirometry tests in regular home settings. To achieve this, it measures the chest wall motion based on sonar system and interprets such motion into lung function indices. Secondly, when the biomarker cannot be intuitively captured as object motion, a channel estimation approach is adopted to quantify muscle tremor induced by muscle fatigue. Lastly, with additional sensing attachments and physiological correlation, active acoustic sensing functionality is further broadened to facial expressions recognition, an intrinsic indicator of mental well-being.

Table of Contents

1.0 Introduction	1
2.0 Literature Review	3
2.1 Mobile Health	3
2.2 Acoustic Motion Tracking and Gesture Recognition	3
2.3 Smart Health via Acoustic sensing	4
3.0 Background & Theory	5
3.1 Clinical Lung Function Tests	5
3.2 Muscle Fatigability	9
3.3 Human Facial Expression	13
4.0 Methodology	17
4.1 Spirometry with Smartphones	17
4.1.1 Chest Motion Tracker	19
4.1.2 Irrelevant Motion Remover	20
4.1.3 Lung Function Estimator	29
4.2 Monitoring Muscle Fatigue with Smartphones	30
4.2.1 Channel Estimation	31
4.2.2 Fatigue Interpretation	34
4.3 Facial Expression Recognition with Headphones	36
4.3.1 Headphone Acoustic Sensor	38
4.3.2 Acoustic Channel Estimation	42
4.3.3 Recognizing Expressions from Acoustic Signals via Machine Learning	44
5.0 Evaluations	50
5.1 Evaluation Setup	50
5.1.1 Lung Function Tests	50
5.1.2 Muscle Fatigability	51
5.1.3 Facial Expression Recognition	54

5.2 Results	55
5.2.1 Lung Function Indices Error	55
5.2.2 Muscle Fatigability Consistency	64
5.2.3 Expression Recognition Accuracy	70
6.0 Conclusions	81
Bibliography	83

List of Tables

1	Acoustic impedance of different media	40
2	Classification accuracy with different selections of facial keypoints . . .	47
3	Patients' Information	60
4	Pearson correlation between chest wall motion features and lung function indices.	60
5	Linear correlation between chest motion features and lung function in- dices (%pred)	61
6	Statistical correlation among overweight patients	61
7	The speed of fatigue accumulation (slope value) over different subjects .	66
8	Channel estimation variance of different categories of facial expressions	74

List of Figures

1	Smart devices	1
2	Spirometry protocol	6
3	Flow-volume loop graph	7
4	Illustration of chest wall motion	7
5	Monitoring muscle fatigue	10
6	The basic component of human muscular system: motor unit (MU) [108]	10
7	The process of how force is generated	11
8	Fatigue-induced muscle tremor cause	12
9	Clinical evidence of the tremor and fatigue [32]	12
10	Facial expressions are created by different movements of facial muscles.	14
11	Spirometry using different equipment and devices	18
12	The spirometer design based on smartphones	18
13	Design overview	19
14	Stationary smartphone	20
15	Hand-held smartphone	21
16	Motion with body movements	21
17	Estimating lung function from chest motion	22
18	Uneven distribution of samples	23
19	I/Q trace segmentation	24
20	Approximation to an arc	25
21	Chest motion from low-quality spirometry tests	25
22	Without irrelevant body motions	26
23	With backward body motion	27
24	With forward body motion	27
25	Multiple choices of p_{start} and P_{end}	29
26	Proposed neural network	30

27	Design overview	31
28	Sub-channels division based on different delays of the training sequence	33
29	The waveform difference between isometric or isotonic contractions . . .	34
30	The concentration of a complex time series interpreted as fatigue level .	35
31	The accumulation speed of fatigue	35
32	The system recognizes facial expressions from the acoustic signal trans- mitted above the face skin surface.	37
33	System Design	38
34	Three possible pathways of ultrasound signal propagation	39
35	Leakage measurement of different headphone model	40
36	Diagrams of the speaker and microphone	41
37	Rewiring the 3.5mm audio jack to implement the required sensing setup	42
38	Raw signal collected with regular and re-wired microphones	43
39	High correlation coefficients on the encoded signal	44
40	Model performance using a single network based on acoustic signal . . .	45
41	Keypoints on human faces	47
42	Student network design	48
43	Main screens of smartphone app	50
44	Evaluation setup	52
45	Protocol procedures and instructions of the smartphone app	52
46	Experimental setup for bicep muscle	53
47	Evaluation setup	54
48	Trade-off between accuracy and overhead of neural network inference .	56
49	Lung function estimation error	56
50	Different distances	57
51	Different tilting angles	58
52	Impact of wearing different clothes	59
53	Average errors of estimating lung function indices (all in %pred)	62
54	Different age subgroups	64
55	Different gender subgroups	65

56	Different disease subgroups	66
57	Experimental setup with EMG device	67
58	The integration of EMG signal (IEMG)	67
59	Slope values of 3 different weights measured by EMG and this technique	68
60	The data of subject #4 with weaker muscle	68
61	Consistent sessions (%) of each individual	69
62	Cases of inconsistent sessions	70
63	Average accuracy over all users	71
64	Accuracies of individual users	71
65	Recognition accuracy without consecutive frame dependency	72
66	Leave-one-user-out accuracy of different facial expressions	73
67	Confusion matrix of facial expression recognition	73
68	Over-ear headphone (The impact of signal strength)	74
69	In-ear headphone (The impact of signal strength)	75
70	Over-ear headphone (The impact of signal frequency band)	76
71	In-ear headphone (The impact of signal frequency band)	76
72	Over-ear headphone (The impact of signal propagation direction)	77
73	In-ear headphone (The impact of signal propagation direction)	77
74	Over-ear headphone (Impact of different user activities)	78
75	In-ear headphone (Impact of different user activities)	79
76	Impact of different skin conditions	79
77	Impact of head accessories and clothes	80

1.0 Introduction

Past decades have witnessed the development and prosperity of mobile and wearable devices (Figure 1) such as smartphones, smart watches and virtual/augmented reality (VR/AR) devices [90, 63]. Statistical data shows that global smartphone units increases from 122 million to 1.4 billion in 2015 with more than \$400 billion global revenue [30]. Nowadays, smartphone ownership is growing rapidly so that around 80% of the U.S. owns a smartphone, which is a 33% increase compared to 2011 [105]. Moreover, from a statistical research conducted in China with the largest growing market of wearable devices, the market size enlarges from 64.9 million dollars in 2012, 183.8 million dollars in 2015 and a huge leap to 736.9 million dollars in 2017 [125]. With such a huge market, a variety of applications have been implemented to facilitate users of smart devices to improve life style, explore more ways for entertainment, and monitor health status routinely, etc [83, 7, 103].



Figure 1: Smart devices

Among applications implemented on these smart devices, smart (mobile) health is an emerging field that uses medical sensors and mobile applications to monitor health status, conduct primary disease diagnosis and manage and healthcare records [109, 106]. As of 2017, over 318,000 health apps were available on app market around the world, with more than 200 being introduced each day [52]. Besides the applications that simply use smart devices as the storage healthcare data for better health status tracking and disease management [1], these mobile systems utilizes built-in sensors on smart devices (e.g., camera, microphone, accelerometer, gyroscope, magnetometer and ambient light sensors, etc) to measure certain

biomarkers that are related to human physiological characteristics. Smart health techniques implemented on smart devices have essential advantages over dedicated medical hardware from the aspect of low cost, easy distribution and simple operability.

The camera on smart devices can capture more fine-grained and comprehensible information. Meanwhile the competition between manufacturers on image quality also greatly benefits applications based on cameras [39]. However, as the unavoidable weakness of vision-based sensor, the power consumption hinders the camera for continuous functioning [104], which can be a crucial requirement for health monitoring. On the other hand, inertial measurement unit (IMU) sensors are usually used for motion sensing such as step counting and exercise monitoring due to their energy efficiency. But as the intrinsic disadvantage of low signal-to-noise ratio (SNR) that applies to every IMU sensor [119], their applications are limited to recognizing repetitive patterns instead of any fine-grained healthcare biomarkers.

As the widely available sensor on any smart devices, microphones working at a sampling rate of 48kHz can record human voices and some narrow band at ultrasonic frequencies. With the ultrasound, various applications based upon acoustic sensing can be explored. With proper system and algorithm design, the microphone can capture the sound passively that is related to wellness (e.g., breath, heartbeat and exercising activities, etc.). Alternatively, if the speaker is also integrated into the system design as well, the healthcare functionality of modern smart devices can be further extended to biomarkers that are inaudible. The rationale for active acoustic sensing is that any externally observable variations can affect the line-of-sight propagation pathway of the acoustic signal. After reflection, the reflected signal carries information of the measured object so that the system is capable of utilizing such information as biomarkers correlated with health monitoring.

This thesis aims to prove the feasibility and effectiveness of adopting active acoustic sensing techniques for healthcare applications on smart devices. Specifically, the thesis includes several applications that are lung function evaluation based on acoustic sonar, muscle fatigue monitoring based on acoustic channel estimation. Lastly, active acoustic sensing functionality is further broadened to facial expression recognition, which is so far impractical on any smart devices.

2.0 Literature Review

This chapter introduces the current work and progress about smart health applications. Based on different sensing modalities, these applications have various setups and designs.

2.1 Mobile Health

In the past years, many mobile health systems have been built based on smartphones. For example, by using the smartphone camera, [14] measures the blood oxygen level by analyzing the flashlight reflections from a fingertip. [15] estimates the blood pressure level from human ear and realizes frequent blood pressure monitoring without affecting normal activities. To assess the health status, [8] analyzes the pH of sweat using pulse oximeter and can be integrated into smartwatches. Research efforts have been made to monitor human bodies from remote. Conventional techniques use infrared devices [40] or depth cameras [54] for imaging analysis, and later research achieved accurate sensing with RF signals for as fall detection [114], gesture recognition [132] and motion tracking [67]. In particular, these techniques have been adopted for contactless breath monitoring [64, 126, 48, 85, 71, 127].

2.2 Acoustic Motion Tracking and Gesture Recognition

Since acoustic signals could be produced by commodity smartphones, they have been widely used to track humans' hand and body motions. Earlier techniques require humans to hand-hold the smartphone and track the smartphone's motion through signal propagation delay [89], Doppler shift [5, 50, 128] and frequency difference [75, 120]. Later schemes, instead, focus on device-free motion tracking that analyzes the reflected signal from the target being tracked [84, 123, 129, 111]. These device-free systems, however, require the smartphone to be always stationary.

2.3 Smart Health via Acoustic sensing

Acoustic sensing has been widely used in smart and wearable devices for health monitoring applications. Such sensing techniques can mainly be categorized into two kinds as passive and active acoustic sensing. Passive sensing analyzes the unique sound produced during certain activities. For example, someone’s tooth brushing manner can be evaluated from the brushing sound by estimating how much time he (she) spends on cleaning each mouth region, which can help maintain tooth hygiene [88]. Furthermore, PDVocal [131] can primarily diagnose Parkinson diseases based upon the abnormal sound a patient may produce during talking and walking.

On the contrary to capturing any existing sound, systems using active acoustic sensing transmit a signal and extract useful information from the received signal. With the prevalence of smart speakers and smartphones, human heart rhythms [121] and lung functions [107] can be estimated by tracking chest motions via active sonar techniques. In addition, active sensing can utilize sound properties like Doppler shift to detect sudden motions like falling [68], which can lead to significant spectrogram variations on the received signal. This work is also based on active acoustic sensing but has orthogonal target with above techniques. Moreover, the system focuses on more subtle changes as the facial skin deformation and propose new measurement metrics as the estimation of acoustic signal propagation channel.

3.0 Background & Theory

This chapter describes healthcare problems this thesis aims to solve. By adopting different acoustic sensing rationales, three widely known applications are introduced.

3.1 Clinical Lung Function Tests

Respiratory diseases, such as asthma, chronic pulmonary disease (COPD) and acute respiratory distress syndrome (ARDS), constitute a significant public health challenge [6]. Over 330 million people worldwide have asthma, including 8.4% of children and 7.7% of adults [31]. These diseases are characterized as various types of airway obstruction. *Spirometry*, as the most commonly used pulmonary function testing (PFT), assesses such obstruction by measuring the volume and velocity of breathing airflow [116], and is crucial in disease evaluation and monitoring [78]. It is also used to judge shortness of breath and airway inflammation, both of which are important symptoms of the coronavirus disease (COVID-19).

Spirometry measures how fast and how much air the patient can breathe out. Before a test starts, the patient exhales all air from the lung. Then, as shown in Figure 2, a spirometry test consists of two stages: the patient first takes a full inhalation and then exhales as hard as possible, until no more air can be breathed out [81]. As shown in Figure 3, measurements from spirometry are represented by a flow-volume loop graph that depicts the correlation between the volume and velocity of airflow, and the graphs of patients with pulmonary diseases are significantly different from those of healthy people. For example, patients with upper airway obstruction (UAO) exhibit apparent plateaus in the graph [12], and asthma patients exhibit “scooped” curves in the exhaling part of graph [76]. In practice, clinicians usually extract certain features from the graph as lung function indices, for more convenient disease evaluation and monitoring. These indices include:

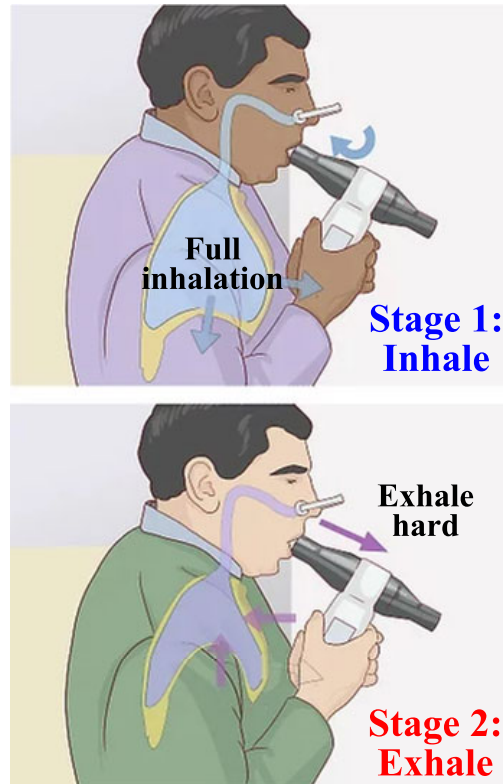


Figure 2: Spirometry protocol

- **Peak expiratory flow (PEF)** is the maximum airflow velocity in exhalation. The average PEF of healthy males and females is around 10L/sec and 8L/sec, respectively. The PEF of asthma patients is as low as 5L/sec [37, 95].
- **Forced expiratory volume in 1 second (FEV1)** is the exhaled air volume in the first second of exhalation, and indicates the airway's resistance against breath [11]. The average FEV1 for healthy people is 3.75L, but that of COPD patients could be as low as 2L [62].
- **Forced vital capacity (FVC)** is the total air volume exhaled. Decline of FVC indicates disease deterioration, and can reach 2.5L for COPD patients [23, 9]. The average FVC of healthy males and females is 5.25L and 3.75L, respectively [61].
- **FEV1/FVC** is the ratio of FEV1 and FVC. This ratio should be >80% among healthy people, but could be as low as 50-60% among asthma patients [17].

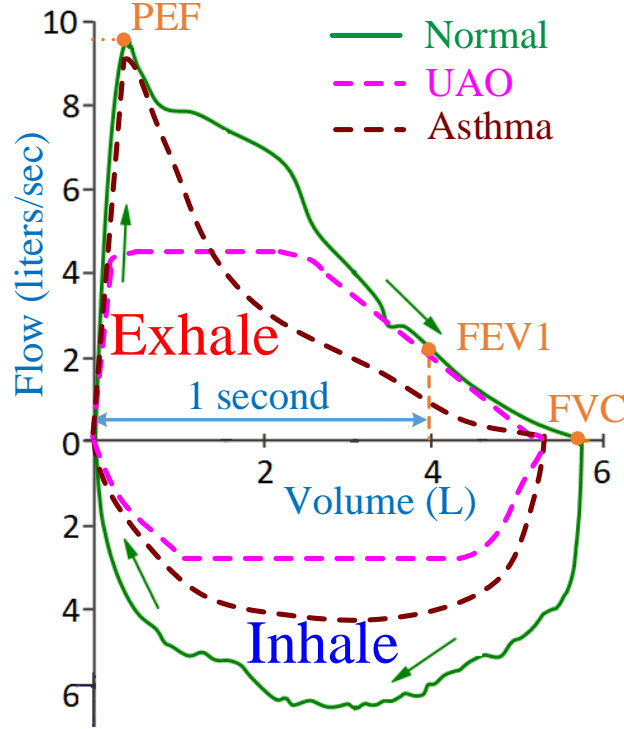


Figure 3: Flow-volume loop graph

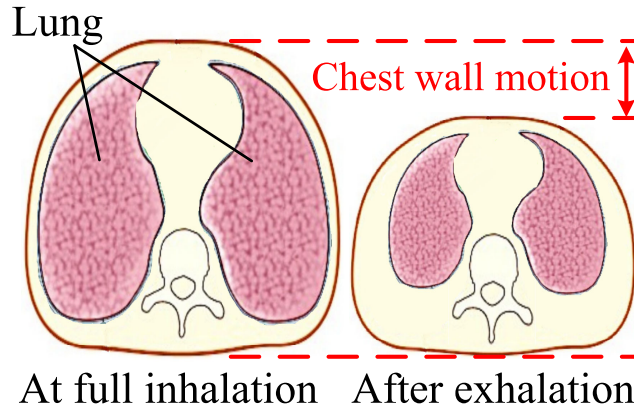


Figure 4: Illustration of chest wall motion

In clinical practice, PEF measurements are highly variable [43, 13], and clinicians mainly use other three indices to evaluate lung function [16]. To ensure accuracy, a patient usually completes multiple (3-8) spirometry tests [28, 27], and the maximum difference of FEV1 and FVC readings in these tests should be $<0.15\text{L}$ [81]. In this way, measurement error of in-clinic spirometry is around 5% [23, 61], which is the baseline to evaluate the performance.

Since lung function greatly varies over individuals, the raw values of lung function indices are seldomly used in clinic. Instead, clinicians usually categorize patients into subgroups according to their demographics (e.g., age, gender, race, etc), and then convert the raw values of lung function indices into percentiles (%pred) over healthy people’s data in the subgroup, provided by Global Lung Function Initiative (GLI) [93]. Typically, a percentile lower than 70% indices high risks of pulmonary diseases. The system will use such percentiles as indicators to measure human lung function.

Correlation between lung function and chest wall motion has been clinically validated. Such correlation, as shown in Figure 4, originates from humans’ ribcage expansion and contraction when breathing [24]. These ribcage movements, when measured by a pneumotrace chest band, are consistent with the fluctuation of lung volume. Clinical studies also showed that humans’ lung volume is proportionate to the ribcage motion [94]. When this work is used in practice, since lung expansion and contraction cause deformation over the entire upper body, individual body shape differences can produce large variations over estimation results. For example, the belly area also reflects the respiration motion but can be less reliable because of fat accumulation. As a consequence, this system requires the smartphone point towards the central part of upper chest area to ensure best lung function estimation accuracy.

Such correlation motivates us to measure the volume of breathing airflow from external, through the displacement of chest wall in spirometry tests. Similarly, the velocity of airflow can be measured from the speed of chest motion. In particular, clinical studies showed that asthma and COPD patients have significantly reduced chest wall motion [59, 96], due to the abnormal changes of chest dimensions and the subsequent lateral ribcage indrawing. For example, patients and healthy people could have 20mm mean difference on the ribcage anteroposterior motion, as well as 10mm mean difference on the upper lateral motion [34].

Such correlation has also been clinically validated to be significant and consistent across different human groups, such as different age groups, males and females with different chest structures and conditions [25], obese people with high BMI [96], etc. Hence, by measuring the patients’ chest wall motion, the proposed system could potentially serve as a useful tool for pulmonary disease evaluation and tracking out of clinic.

3.2 Muscle Fatigability

Muscles are essential to humans since they are responsible for body mobility and posture maintenance. The overuse of muscle may cause feelings of soreness, trembling, and cramping, which are major symptoms of muscle fatigue. Muscle fatigability measures how fast and easily the muscle gets fatigued, and is crucial for diagnosing many muscular diseases such as muscular disorders and dystrophy that influence 1 in every 3500 males but has no cure so far [99, 4]. During clinical visits, the muscle fatigability can be evaluated by lactic acid accumulation using a blood test [102] or myoelectric signals using electromyography (EMG) [22], but frequent daily evaluation of muscle fatigue out of clinic is usually preferred in most of muscular disease diagnosis for long-term tracking of disease progress and medication efficacy [21]. Unfortunately, effective methods for such out-of-clinic evaluation are currently missing, and patients instead are only asked to self report their feelings of fatigue by completing certain questionnaires [20]. This method is quite inaccurate because different individuals may have subjective definitions of fatigue feeling, which makes the evaluation of muscle fatigue unreliable. The key to accurate evaluation, on the other hand, calls for an objective biomarker that reliably indicates the fatigue level under normal daily circumstances out of clinic.

To address this challenge, a new technique is invented that uses commodity smartphones for objective muscle evaluation out of clinic. This technique is built on the physiological fact that human muscles will unconsciously tremble or shake when getting fatigued and such tremor is proportionate to the level of muscle fatigue [70]. Since this muscle tremor can produce shape changes on the muscle surface, the system could measure the muscle fatigue from such changes which affect the acoustic signal being transmitted between the smartphone’s built-in speaker and microphone, as shown in Figure 5. More specifically, the smartphone transmits acoustic signals towards the muscle piece being monitored and receives reflected signals by its microphone. This signal flight path forms a propagation channel, in which any physical change of the nearby muscle objects disturbs the channel and can hence be captured from the channel estimation.

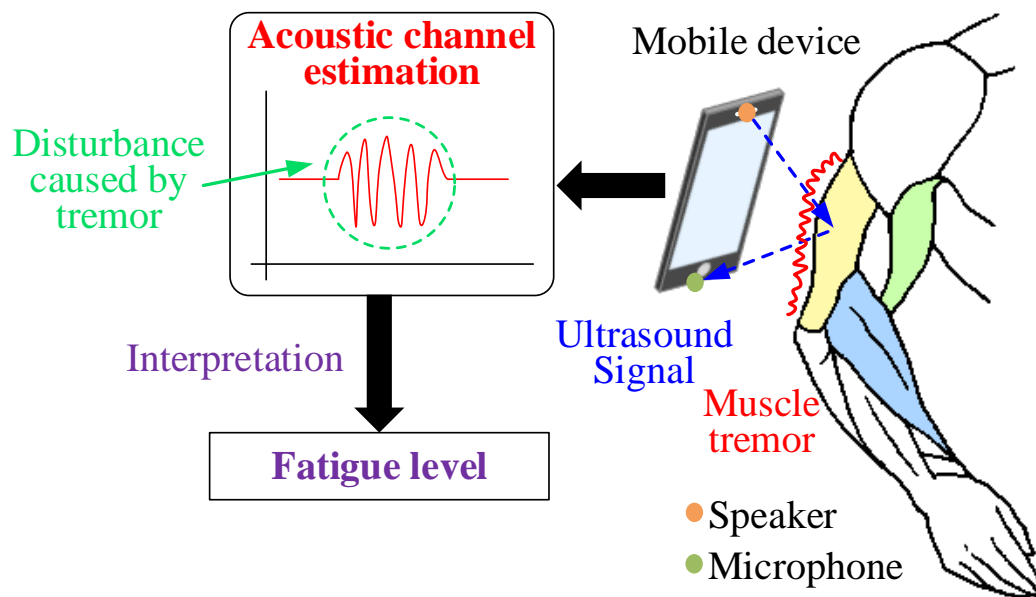


Figure 5: Monitoring muscle fatigue

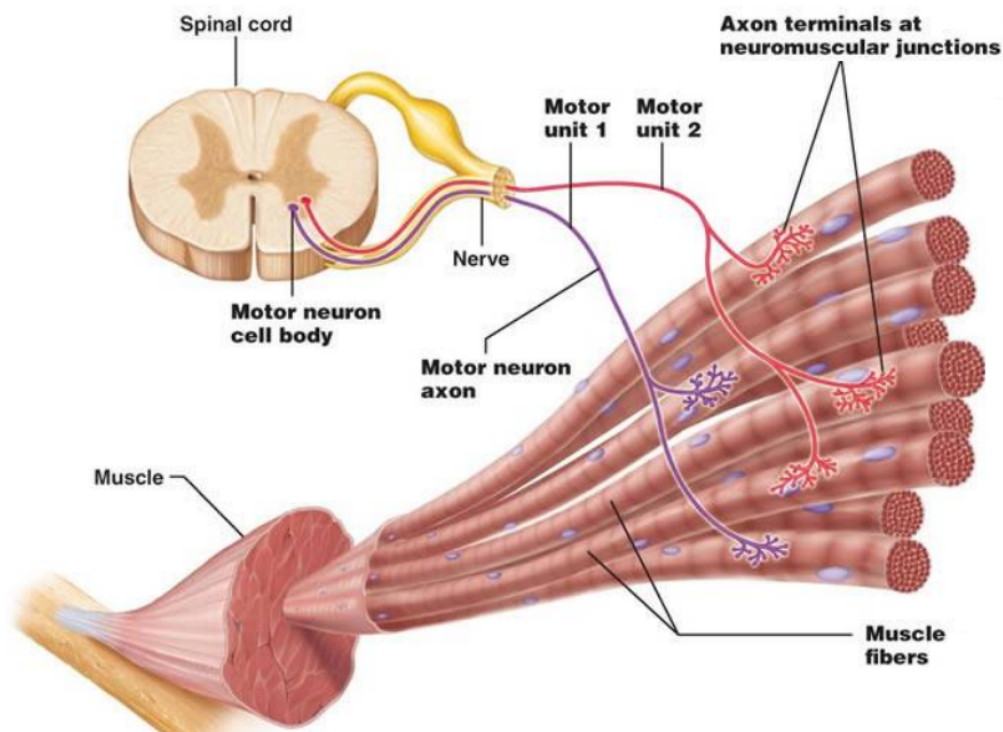


Figure 6: The basic component of human muscular system: motor unit (MU) [108]

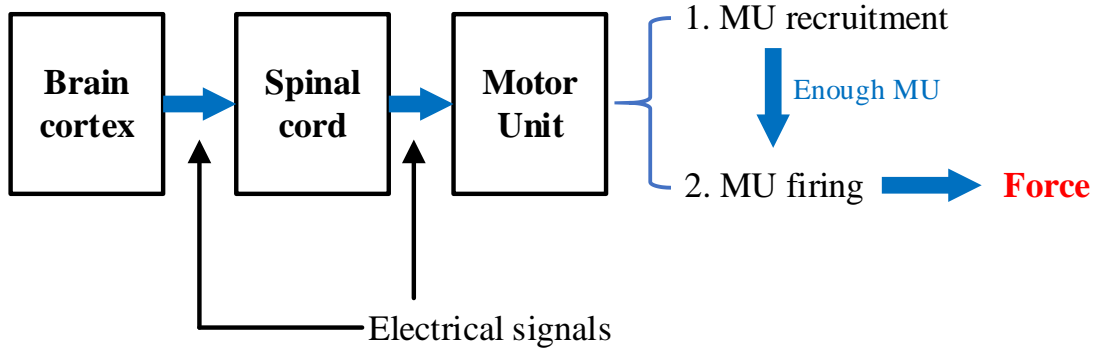


Figure 7: The process of how force is generated

The basic component of human muscular system is the motor unit (MU), which consists of motor neuron and muscle fibers. As shown in Figure 6 [101], the motor neurons lie in the neuron pool located at spinal cord and each of them controls and innervates a set of muscle fibers. The muscle fibers constitute different types of muscle all over the human body and they are connected to a specific motor neuron by muscle nerve. The communication between the motor neuron and muscle fibers is achieved by myoelectric signals, which is also the principle of clinical EMG diagnosis [33]. The number of MUs per muscle varies from 100 (e.g., hand muscle) to 1000 or more (e.g., limb muscle), but the relationship is one-to-one that each MU is only responsible for one particular muscle [124].

Figure 7 describes how the force is generated under the collaboration between human brain cortex and MUs via electrical signal. The signal originates from the brain and transmits through spinal cord to specific motor units [79]. Then based on how much force is needed, the MU recruitment starts. The more force being generated, the more muscle fibers are required to get involved, so the more MUs need to be recruited. When enough MUs are ready to trigger the contraction of muscle, a new process begins as the so-called MU firing. The firing makes every motor neuron innervate their corresponding muscle fibers and generates the final force as what humans feel during muscle exercises [108].

Tremor is one of common consequences of muscle fatigue and the cause of tremor can be explained from the physiological aspect. During MU firing, each MU triggers the muscle with an impulse signal [2]. As shown in Figure 8, when the muscle has not been fatigued,

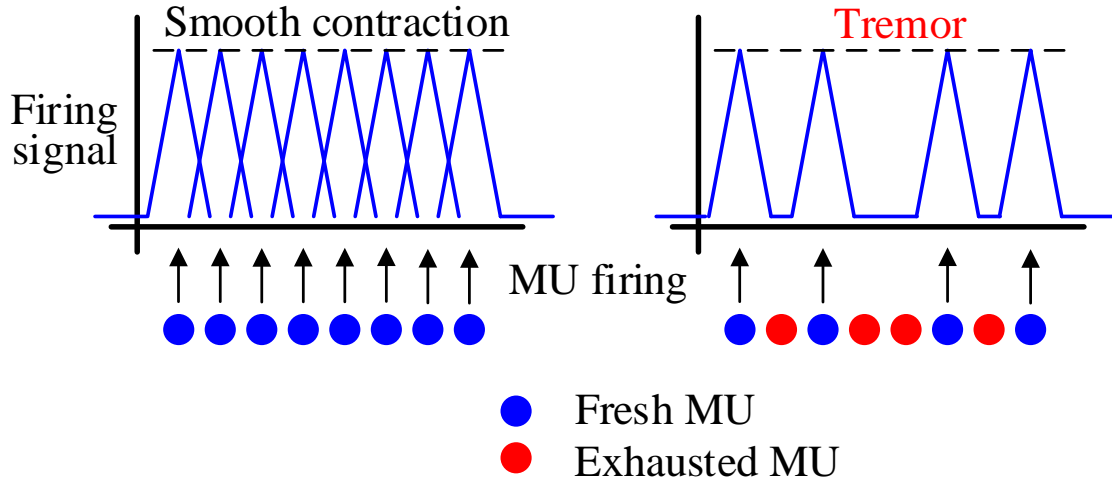


Figure 8: Fatigue-induced muscle tremor cause

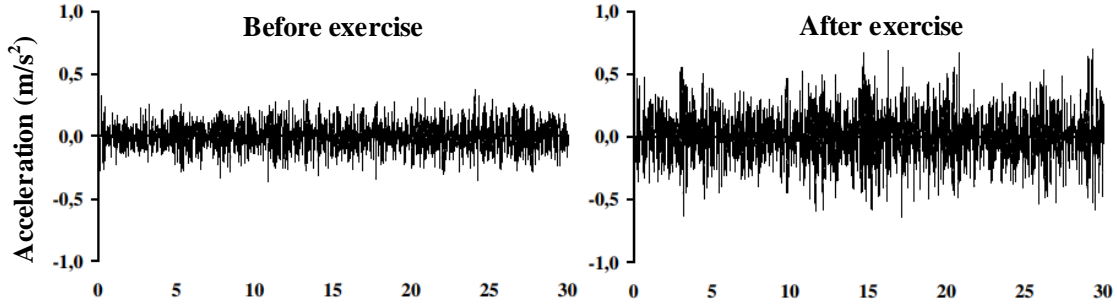


Figure 9: Clinical evidence of the tremor and fatigue [32]

all the MUs functions normally so that the combinational results of many MUs lead to a smooth contraction of the muscle. However, after overusing the muscle, the lactic acid starts accumulating in the muscle and reduces the intracellular pH. This phenomenon will decrease the conduction velocity of muscle fibers so that some of the MUs exhaust to maintain the force [82]. The exhaustion of these MUs will cause the fatigue that humans feel, and in the meantime, the contraction becomes less consistent and produces the tremor that can be observed [33].

Clinical study also showed that the amplitude of muscle tremor is proportionate to the level of muscle fatigue [32]. As shown in Figure 9, the subjects in this research were asked to contract their muscle before and after exercises with great intensity. From the readings reported by accelerometers placed on the muscle, it is observed that the average amplitude of tremor increases when the muscle gets fatigued. Motivated by such correlation between muscle tremor and fatigue, the technique measures the muscle tremor externally as the physical disturbance caused in an acoustic signal channel and further uses such tremor to evaluate the actual muscle fatigue.

Muscle contraction can be categorized as isometric and isotonic contractions [56]. Though both phases can generate force, isometric contractions do not change the length of muscle while isotonic contractions require lengthening or shortening the muscle. Taking common bicep exercises as an example, isometric contraction happens when the subject attempts to hold the dumbbell with no arm movements and the isotonic contraction happens when the subject lifts up (down) the dumbbell by moving the lower arm. Moreover, the upper arm shows more evident geometrical change (bigger) during isotonic than isometric contractions. Fatigue induced tremor exhibits over both isometric and isotonic phases [113], but based on the rationale of this technique that monitoring channel disturbances caused by physical change, tremor can be monitored with more confidence if no body motion or muscle geometrical change exhibits during isometric contractions. Since the work targets on bicep muscles in this thesis, differentiating different phases of muscle contraction is also essential in this design.

3.3 Human Facial Expression

In human communication and interaction, facial expressions serve as natural and effortless signals to convey ourselves besides verbal languages [66]. With recognition and analysis of facial expressions, variety of applications can be deployed [53]. For example, mental well-being, which is considered as an underlying challenge during COVID-19 pandemic [46], can be identified and tracked through recognizing the appearance of positive expressions [44, 117].

In addition, facial expressions recognition also provide important information and feedbacks for human-computer interaction (HCI) such as virtual reality [47], cognitive robotics [73] and customer analytics [51], etc.

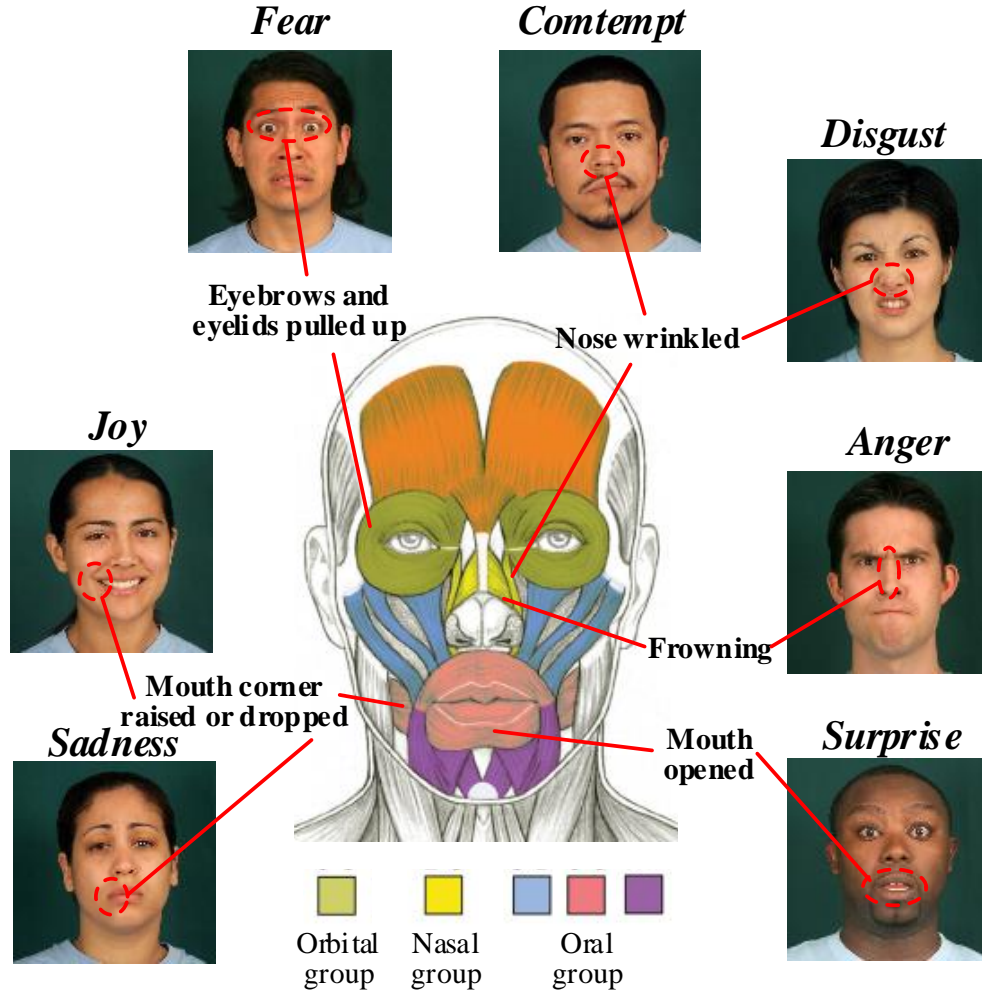


Figure 10: Facial expressions are created by different movements of facial muscles.

Most of current computing approaches to recognizing human facial expressions use images of human faces as the input [57], and can accept such images being taken from either front [72] or side [19] of the user. However, their accuracy and reliability significantly drop when the ambient light condition degrades [26, 19]. Continuous camera use will also incur very high power consumption that is not affordable on most battery-powered mobile devices. Moreover, in certain scenarios such as virtual/augmented reality, the user's face is always occluded by the headset. Using camera cannot ensure accurate facial expression recognition, which hinders the avatar creation and immersive communication in future metaverse. Lastly,

camera-based facial recognition requires clear pictures of subject's facial area. As there are lots of system adopting facial information to perform user authentication, leakage of these bio-metric information will inevitably raise privacy concerns.

Instead, some recent research seeks to exploit the correlation between facial expressions and facial muscle movements, which can be detected via either surface electromyography (EMG) [38, 77], on-head strain sensor [65] or in-ear air pressure sensor [3]. Facial expressions have also been considered as the outcome of human emotions, which can be monitored from different physiological signals [135, 60, 36] and human activities [134, 74]. However, all these techniques require custom sensing hardware, which are expensive and difficult to be integrated into commodity systems in practical use. The use of such custom hardware, on the other hand, also make these systems highly sensitive to the random variations of the human body, user mobility, and the surrounding environment. For example, sensing electrodes may be misplaced due to human body movements and hence result in substantial recognition errors [100, 41].

Facial expressions are induced by the contraction of different skeletal muscles laying underneath human facial skin. Compared with other skeletal muscles on human body, since facial muscles insert directly into the skin, much evident external changes can be observed during their contractions [35]. These changes lead to various facial expressions that reflect one's different emotions and feelings. As shown in Figure 10, muscles related to facial expressions are commonly divided into 3 categories: orbital, nasal and oral.

Orbital group: it contains muscles of orbicularis oculi and corrugator supercilii, which surround the eye socket and control the movements of eyelids and eyebrows. For example, under the emotion of sadness, the inner corner of eyebrows raises and the eyelids become loose. Moreover, the entire orbital group pull up and together when someone has a fear feeling.

Nasal group: it contains the muscles of nasalis and procerus that are associated with the movements of nose and the skin around it. At expressions like disgust and contempt, nose wrinkles appear under the control of nasalis. Meanwhile, procerus muscle also contributes to frowning expressions such as anger or sadness. Other than making expressions, nasal muscle group is also essential to respiration.

Oral group: it contains orbicularis oris, buccinator and lots of smaller muscles. These muscles controlling the movements of the mouth and lips are not only required during vocal communications, but can produce various facial expressions by controlling the position of mouth corners and the opening or closure of the mouth. For example, joy and sadness can be easily differentiated by the raised or dropped lip corner. Unconscious opening of mouth as well as the extent of such opening are natural indicators of emotions like surprise and fear.

In practice, different facial expressions are the synthesized outcome of all facial muscle movements. Identifying the change of specific facial muscle cannot determine an expression confidently because certain muscles are involved in different facial expressions [133], thus the system captures the combined effect of facial muscle movements reflected on ultrasound channel variations.

4.0 Methodology

This chapter describe methods used to implement the sensing systems with respect to known problems.

4.1 Spirometry with Smartphones

Ideally, spirometry should be daily conducted out of clinic, to timely detect and avoid frequent disease exacerbations that cause emergency department visits or hospitalizations [10, 55]. However, current spirometers in clinic, as shown in Figure 11(a), are too bulky for daily home use. Recent efforts, as shown in Figure 11(b), reduce the size of spirometers but their costs ($> \$2,000$) are still too high for home use. Low-cost spirometers priced at $< \$100$ [98, 80] are mostly inaccurate and could produce $> 20\%$ error [69, 42]. In this work, a novel system is designed that uses commodity smartphones to support complete, accurate yet reliable spirometry tests out of clinic, with various environmental and human factors. As shown in Figure 12, the system design builds on the close correlation between lung function and chest wall motion of humans, which has been widely validated in clinical practice [34, 59, 96, 94]. The system measures chest wall motion as an externally observable biomarker, and interprets such motion into lung function indices. To measure such motion, the system transmits ultrasound signal with the smartphone’s speaker, and analyzes the signal being reflected by the patient’s chest wall and received by the smartphone’s microphone.

As shown in Figure 13, to use the system, the patient holds the smartphone and points the phone’s bottom speaker and microphone to the chest. Then, the patient follows the spirometry protocol to inhale and exhale. Being similar with in-clinic spirometry, the patient should maintain an upright posture by leaning the back against a chair backrest [81]. In this way, the body trunk remains steady during the spirometry test, and the measured chest wall motion is only caused by inhalation and exhalation.

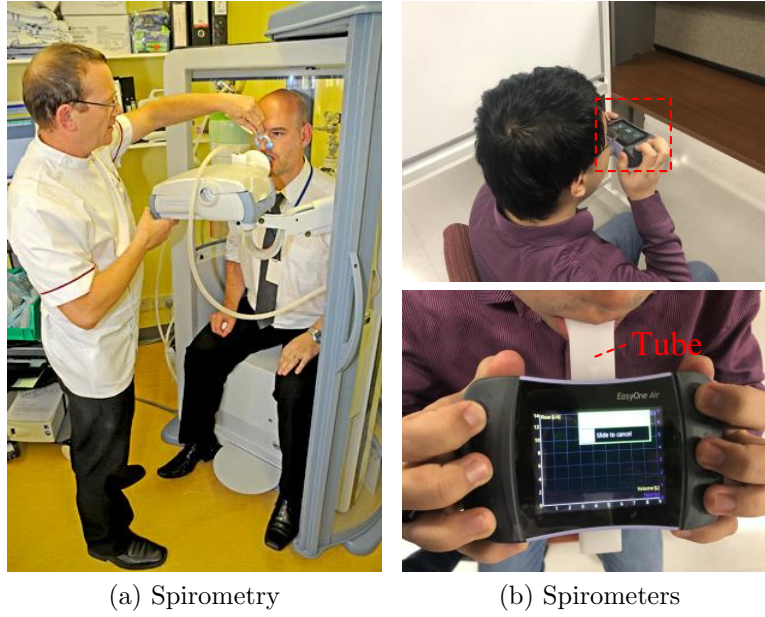


Figure 11: Spirometry using different equipment and devices

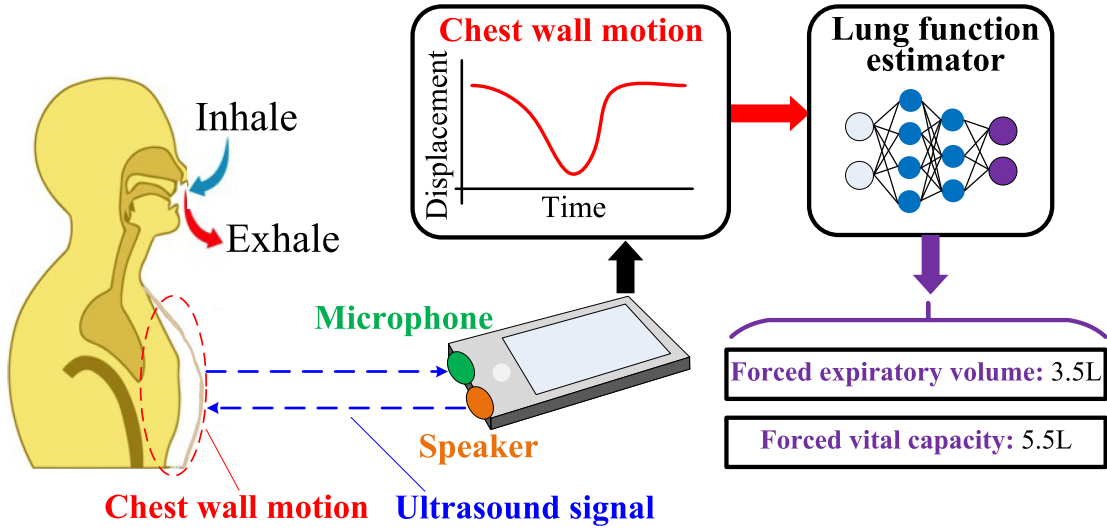


Figure 12: The spirometer design based on smartphones

The system tracks the patient's chest wall motion in both inhalation and exhalation stages of spirometry tests. To ensure accuracy, such chest wall motion will be first examined and corrected to remove any impact of irrelevant motions. Afterwards, the corrected chest

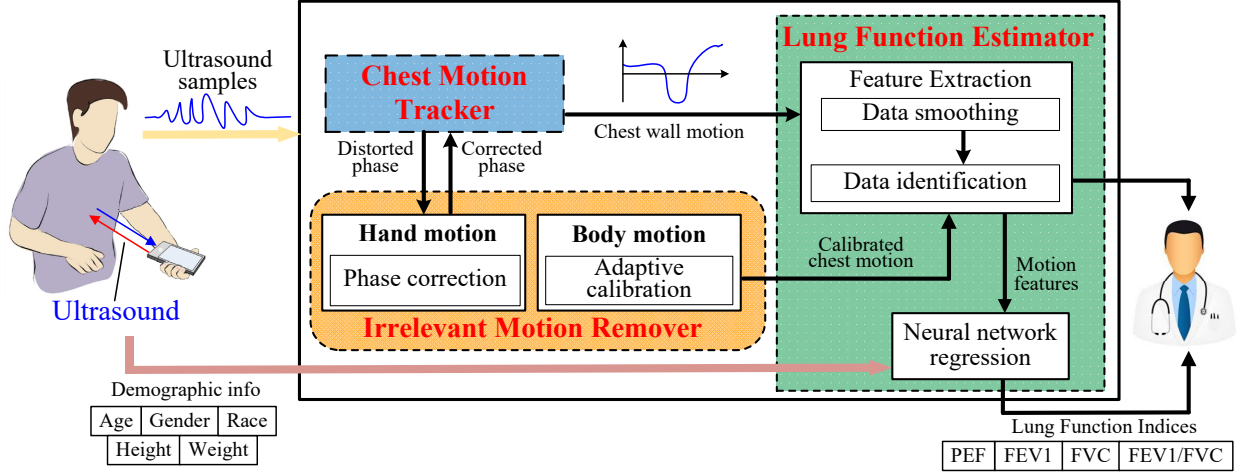


Figure 13: Design overview

motion will be used as the input to neural network regression, which computes lung function indices based on the nonlinear correlation between chest wall motion and human lung function. These lung function indices will be converted into prediction values (%) based on the patient's demographic information, and then reported to pulmonary doctors for remote disease evaluation and monitoring.

4.1.1 Chest Motion Tracker

Since humans' chest wall displacement in spirometry is usually lower than 60mm [58], the error of chest motion tracking should be at most 3-4 millimeters, so that the error of lung function estimation could be within 5%. To achieve this accuracy, the system measures the phase change between the transmitted and received ultrasound signal. Considering the transmitted signal as $A \cos(2\pi ft)$, the phase of the received signal, after being reflected by the chest wall, is $\varphi(t) = 4\pi fd(t)/c$, where c is sound speed and $d(t)$ is the distance between chest wall and smartphone at time t . When the chest wall moves during a time period $[t_0, t_1]$, its displacement is

$$\Delta d = d(t_1) - d(t_0) = -c/(4\pi f) \cdot (\varphi(t_1) - \varphi(t_0)). \quad (4-1)$$

When the ultrasound signal's frequency ranges between 17kHz and 24kHz, a 2mm displacement causes the signal path length to change by 4mm and corresponds to a phase change between 0.4π and 0.56π , large enough to be detected. Such detectability also allows us to use multiple signals with different frequencies in this range to further improve the tracking accuracy.

4.1.2 Irrelevant Motion Remover

Since the chest wall's motion is measured as its relative displacement from the smartphone, it could be easily affected by the smartphone's random movements: the patient cannot keep the hand-held smartphone to be 100% stationary, and his/her body may unconsciously lean forward or backward when exhaling hard. To remove such irrelevant movements from the measured chest wall motion, an intuitive method is to measure the smartphone's movements using its built-in accelerometers. This approach, however, is inaccurate due to the error accumulation, when converting accelerometer readings into displacement via double integration [128].

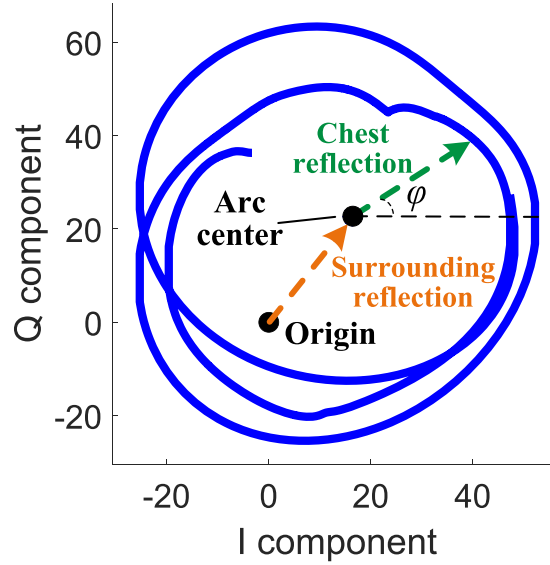


Figure 14: Stationary smartphone

Instead, the system investigates the abnormal characteristics from the measured chest wall motion itself. As shown in Figure 14, if the smartphone is stationary, I/Q trace of the received signal should be a regular collection of concentric arcs. The I/Q trace can form

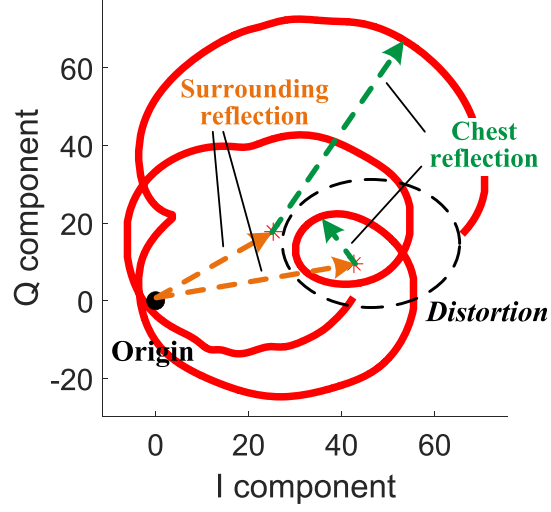


Figure 15: Hand-held smartphone

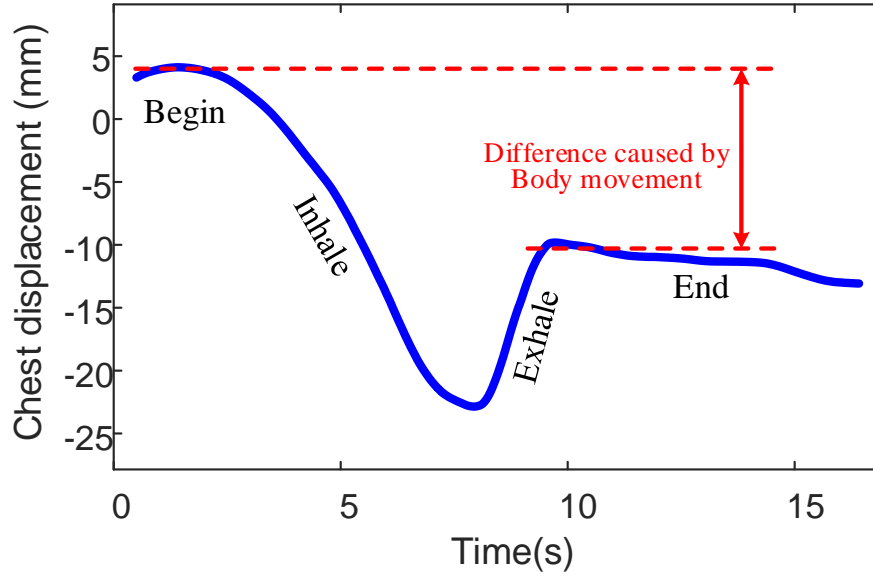


Figure 16: Motion with body movements

either a circle or ellipse because the radius can vary based on different characteristics of the chest wall motion. But as long as the origin of I/Q trace determined by all surrounding reflections does not vary much, phase change can always be intuitively calculated. Otherwise, I/Q trace produced from a hand-held smartphone with random movements will be arbitrarily

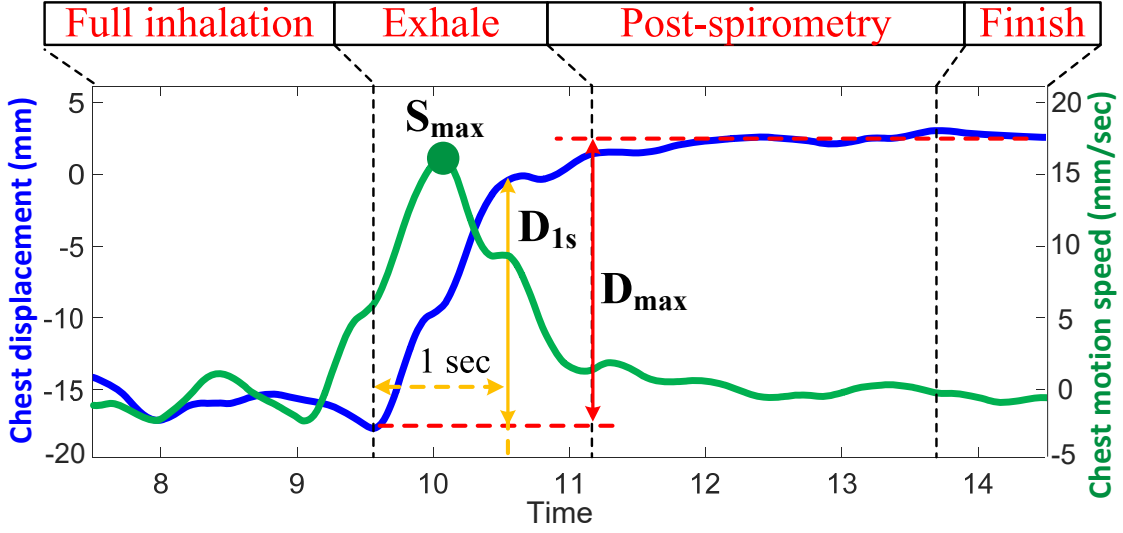


Figure 17: Estimating lung function from chest motion

distorted. The reason, as shown in Figure 15, is that the received signal always contains reflections from both the chest wall and surrounding objects. When the smartphone is hand-held and randomly moving, the surrounding reflection varies and distorts the cumulatively received signal.

Motions of a hand-held smartphone distort the I/Q trace of the received ultrasound signal and result in irregular phase variation. To address such variation, it assumes that the hand motion speed is slower than the chest wall motion speed during a spirometry test. The assumption is valid if the patient follows the spirometry protocol correctly by maintaining a steady upper body posture when exhaling. In this way, phase changes of I/Q trace is mainly caused by the moving chest wall. The system first divides the I/Q traces into short segments, during each of which the smartphone can be assumed as motionless. Then, it approximates each segment back to the closest circular arc on the complex plane. The phase variation can then be corrected by normalizing the centers of all the arcs back to the origin on the complex plane, making the I/Q trace to a collection of concentric arcs as shown in Figure 14.

Segmentation: One intuitive method is to divide the I/Q trace into segments with the equal number of signal samples, but is ineffective when chest motion is measured as phase

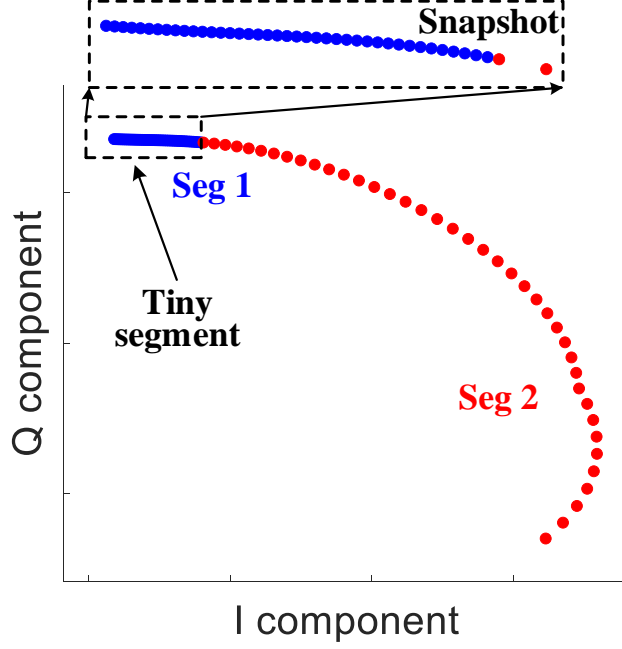


Figure 18: Uneven distribution of samples

change: as shown in Figure 18, since each sample $[I(t), Q(t)]$ has a phase of $\tan^{-1}(Q(t)/I(t))$, smaller chest motion results in many consecutive samples with similar phases, and creates many unwanted tiny segments.

Instead, the system segments the I/Q trace based on its specific phase change over time. As shown in Figure 19, for every two consecutive samples at time t_1 and t_2 , the system computes the phase change as $(Q(t_2) - Q(t_1))/(I(t_2) - I(t_1))$, and produce a new segment once the cumulative phase change exceeds a threshold. For example, when this threshold is $\pi/2$, the I/Q trace in Figure 19 is divided into 6 segments. In practice, it adaptively adjusts this threshold, to make sure that each segment contains a sufficient number of signal samples for correcting the signal distortions. The segment 2 in Figure 19, as an instance, corresponds to fast chest motion and hence a larger threshold of π is being applied.

Random signal noise may be produced by the hardware imperfection of smartphones or surrounding signal sources (e.g., spinning fans), and temporarily fluctuates the signal phase as shown in Figure 19. Such phase fluctuation may result in small unwanted segments, but

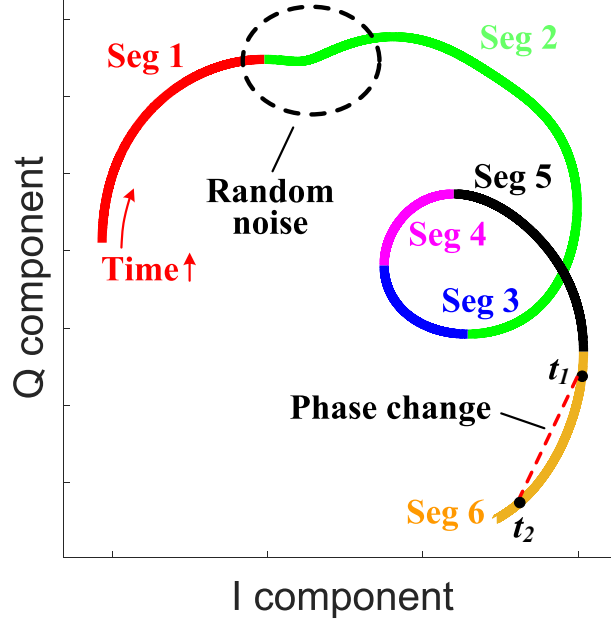


Figure 19: I/Q trace segmentation

unfortunately cannot be removed by smoothing the I/Q trace with a sliding window, due to the uneven distribution of signal samples shown in Figure 18. Instead, such unwanted segments are avoided by setting up a threshold on the minimum segment length. This threshold is empirically set as three times of the standard deviation of I/Q samples during full inhalation in spirometry, where the chest wall motion is considered as minimum.

Correction: As shown in Figure 20, the system approximates each segment to the closet circular arc, indicated by the best linear unbiased estimates (BLUE) of arc center (I_c, Q_c) and radius r_c . This arc¹ is estimated as $[\hat{I}_c, \hat{Q}_c, \hat{\theta}]^T = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y}$, where

$$\mathbf{H} = \begin{bmatrix} 2I(1) & 2Q(1) & 1 \\ 2I(2) & 2Q(2) & 1 \\ \vdots & \vdots & \vdots \\ 2I(N) & 2Q(N) & 1 \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} I(1)^2 + Q(1)^2 \\ I(2)^2 + Q(2)^2 \\ \vdots \\ I(N)^2 + Q(N)^2 \end{bmatrix}, \quad (4-2)$$

N is the segment's number of samples, and $\hat{\theta} = \hat{r}_c^2 - \hat{I}_c^2 - \hat{Q}_c^2$. Each sample is then mapped to the arc individually.

¹The corresponding circular arc is represented as $(I - I_c)^2 + (Q - Q_c)^2 = r_c^2$.

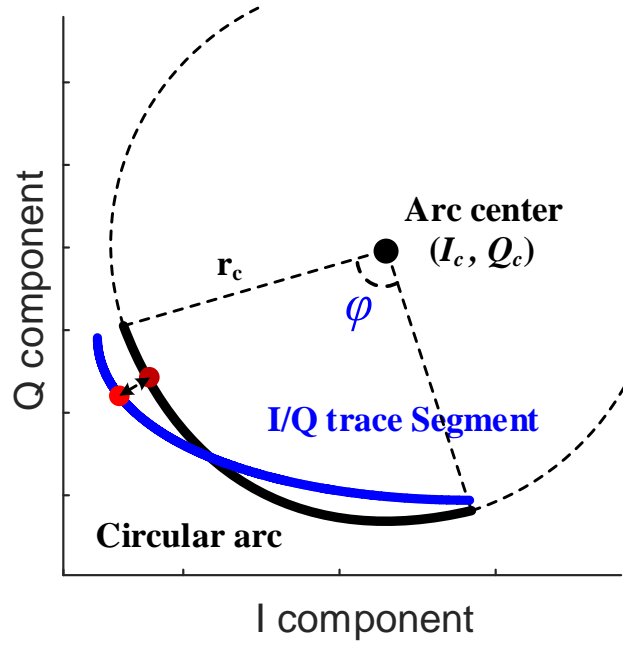


Figure 20: Approximation to an arc

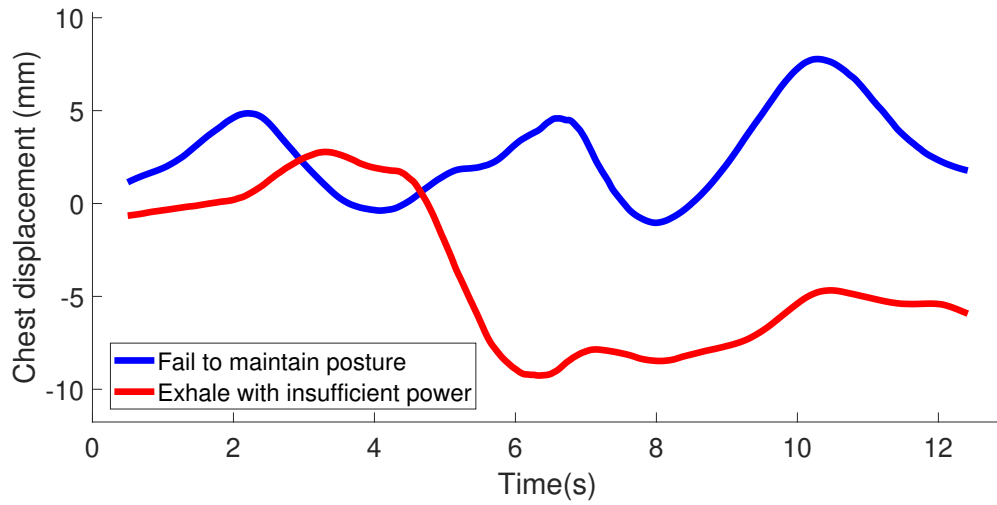


Figure 21: Chest motion from low-quality spirometry tests

Another challenge of such lung function estimation is the heterogeneous human factors, which may impair the data quality in spirometry tests and make it difficult to correctly identify the exhalation stage. Ideally, the measured chest wall motion, as shown in Figure

17, should exhibit a sole rapid change of chest displacement of at least 10-15mm, as a result of hard exhalation. However in practice, patients may not fully follow the spirometry protocol, due to lack of clinician's guidance or weak body conditions. As shown in Figure 21, they may exhale with insufficient power and result in inadequate chest displacement; they may fail to keep the upright posture and produce abnormal chest motions.

According to the spirometry protocol, the patient's chest wall should be at the same position before and after a spirometry test, if the patient's body does not move during the test. The impact of body movements on the measured chest wall motion, as shown in Figure 16, can hence be identified as the difference of chest wall position before and after a spirometry test.

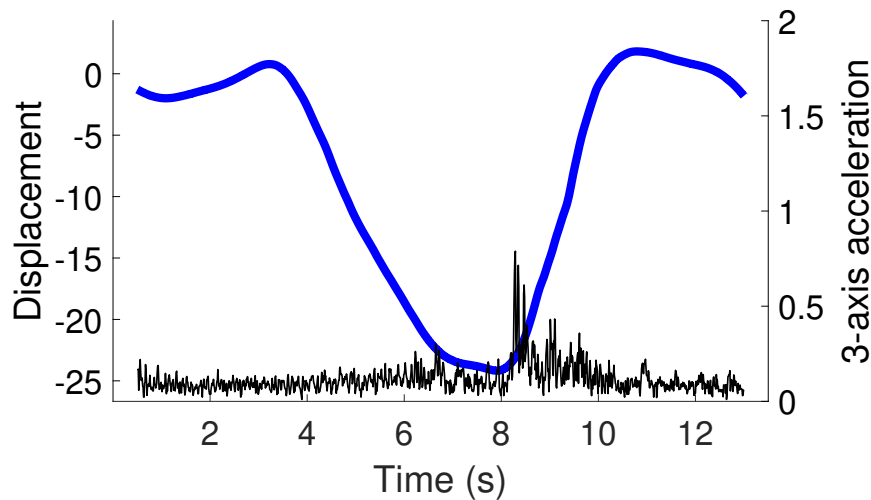


Figure 22: Without irrelevant body motions

The proposed calibration builds on the fact that, the patient's chest wall should be at the same position before and after a spirometry test if there is no extra body motions during the test, because the patient is supposed to have full inhalation and exhalation in spirometry. To verify this, a digital accelerometer is attached to the patients' upper body to measure their body motions in spirometry, and results in Figure 22 match the expectation. In contrast, with noticeable body motions being indicated by the high accelerometer readings during exhalation, the patients' chest wall position will be largely changed after spirometry tests, as shown in Figure 23 and 24.

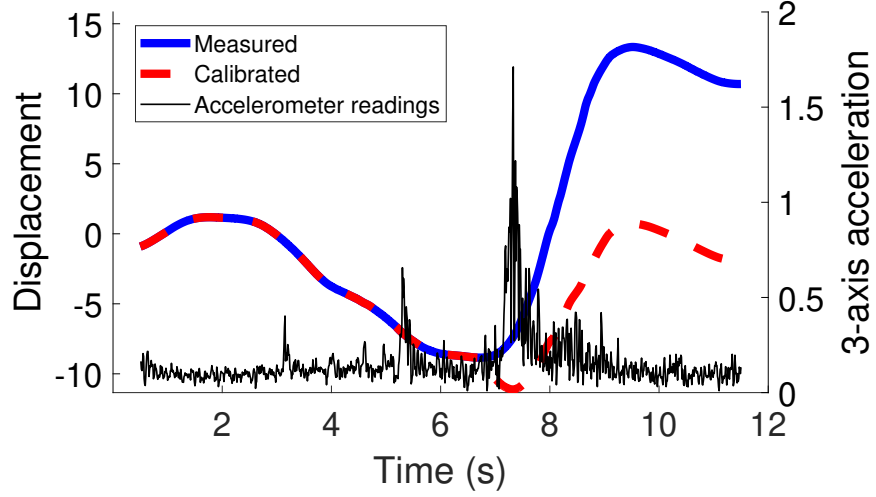


Figure 23: With backward body motion

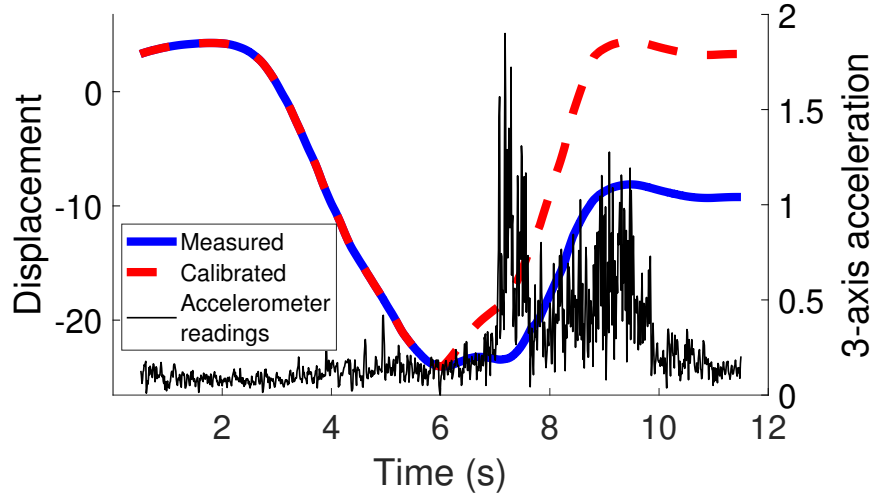


Figure 24: With forward body motion

In practice, such body motions could be either unidirectional or bidirectional during exhalation. For unidirectional (i.e., either moving forward or backward) motions, this approach to removing its impact uses the chest wall position before spirometry as the baseline, to proportionally calibrate every signal sample after the exhalation starts. The outcomes, as shown in Figure 23 and 24, effectively remove the difference of chest wall positions before and after spirometry.

In some other cases, the patient's body may move both back and forth during exhalation. When such bidirectional body motion is small, the system removes this motion through adaptive smoothing: it adapts the smoothing window (W) to the momentary chest motion speed (S) as $W = (1 - |S/S_{\max}|) \cdot f_s$, where S_{\max} is the maximum chest motion speed and f_s is the ultrasound signal's sampling rate. In this way, slower motion leads to a larger window that produces a more smooth motion curve. Rapid motion results in a smaller window to avoid missing details in the motion pattern.

Big bidirectional body motions, on the other hand, indicate that the patient does not follow the spirometry protocol and the system will instead judge the corresponding spirometry test as invalid.

The exhalation stage in a spirometry test is indicated by a starting point (p_{start}) and an ending plateau (P_{end}). A valid p_{start} should be a local minimum on the curve of chest wall displacement, and a valid P_{end} should correspond to a period of sufficiently small chest motion². However in practice, as shown in Figure 25, there may be multiple possibilities of such starts and ending plateaus due to the heterogeneous human factors. Three following criteria are used to decide the best choices of p_{start} and P_{end} :

- p_{start} and P_{end} are always decided in pairs, and the p_{start} is always located before P_{end} .
- The average chest wall displacement within P_{end} should be higher than 90% of the maximum displacement between p_{start} and P_{end} .
- If multiple pairs are available, the pair that corresponds to the maximum chest displacement in exhalation is selected.

Based on such decision, the motion features (S_{\max} , D_{1s} and D_{\max}) can be calculated as shown in Figure 17. If a valid pair of p_{start} and P_{end} cannot be found, it is considered that the patient did not fully follow the spirometry protocol (e.g., the body moves back and forth during exhalation), and data in this spirometry test has low quality. The system will exclude such data from lung function estimation.

²According to the pulmonary clinicians, if the variance of chest wall displacement within 1.5 seconds is less than 15% of the maximum displacement in a test, it is considered that 1.5 seconds of chest wall motion as a valid P_{end} .

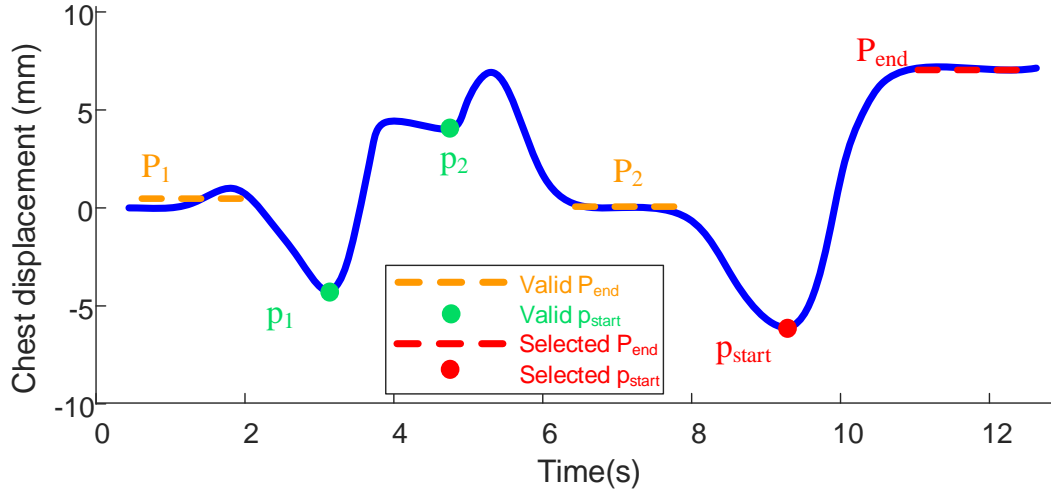


Figure 25: Multiple choices of p_{start} and P_{end}

4.1.3 Lung Function Estimator

As shown in Figure 17, the system uses the following features of chest wall motion during the exhalation stage in spirometry to estimate lung function indices:

- The maximum speed of chest wall motion (S_{max}), which corresponds to PEF.
- The chest wall displacement in the first second of exhalation (D_{1s}), which corresponds to FEV1.
- The maximum chest wall displacement (D_{max}), which corresponds to FVC.

The system quantifies such correlation with a neural network regression model, which is built with data from clinical studies. Since such clinical data from patients is usually low in volume and could hence result in overfitting when training the model [18], a Bayesian regularized neural network is used, which has good capability of generalization that avoids overfitting [86, 115]. In the clinical study, patients do spirometry tests using the proposed system and clinical-grade spirometers at the same time. As shown in Figure 26, the outcomes from the motion tracking and spirometers' measurements are then used as the inputs and outputs, respectively, to train the neural network³. Afterwards, the trained neural network is loaded to patients' smartphones for out-of-clinic use.

³Lung function indices measured by spirometers are converted into %pred before being used to train the neural network.

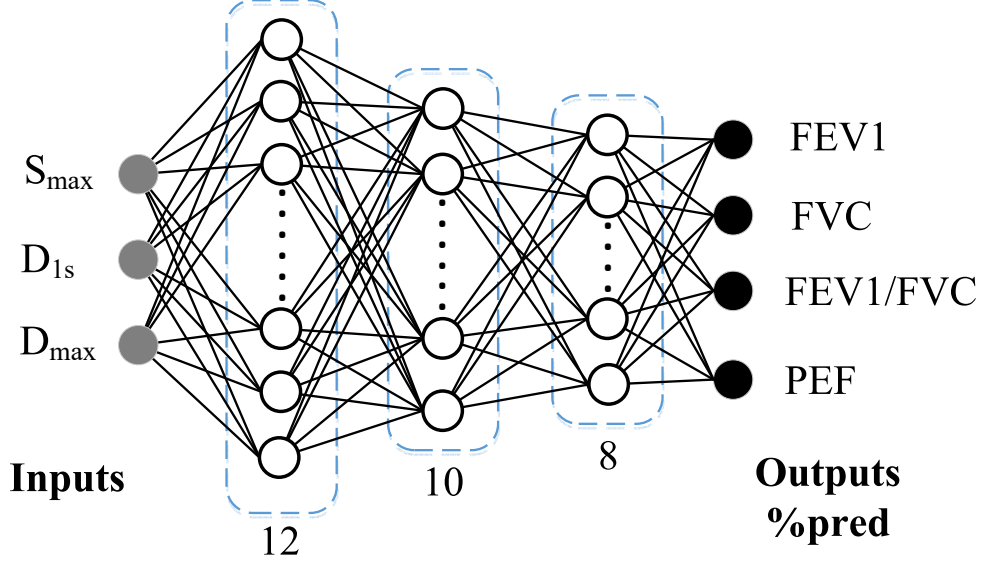


Figure 26: Proposed neural network

Both the accuracy and overhead of such inference depend on the complexity of neural network. A neural network with more hidden layers and numbers of neurons improves the inference accuracy, but also increases its computation overhead. The system empirically uses three hidden layers in the neural network, and then balance between these two aspects by tuning the numbers of neurons in each layer.

4.2 Monitoring Muscle Fatigue with Smartphones

As shown in Figure 27, the technique of muscle fatigue evaluation consists of three consecutive modules. First, the channel estimation module converts the received ultrasound signal samples into an estimation matrix that represents the acoustic channel variation over time. To exclude irrelevant signal reflections, the transmitted signal is specifically modulated with a known codeword sequence. Such operation can divide the acoustic channel into different distance ranges (sub-channels), and the system only needs to extract the channel information corresponding to the range where muscle is placed.

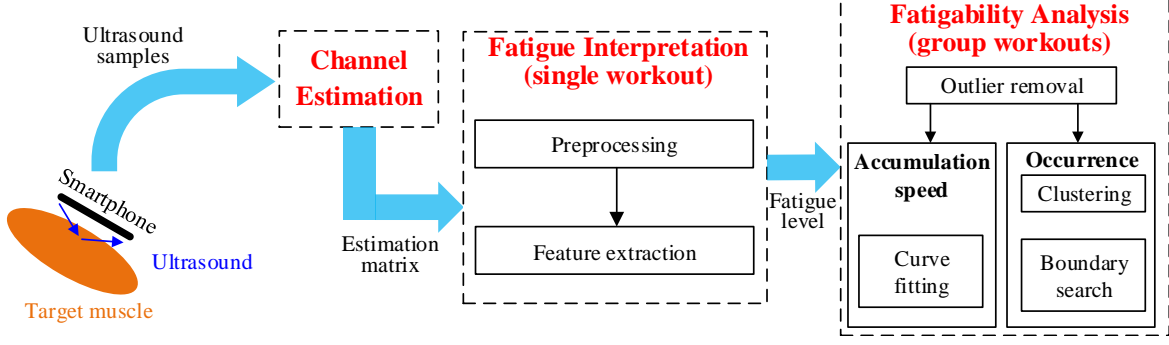


Figure 27: Design overview

Second, in every single workout, the system interprets the result of the above channel estimation into the corresponding fatigue level using the fatigue interpretation module. For the bicep muscle, this module identifies the muscle contraction phases and keeps only data during isometric contraction because it does not involve any body motion or geometrical change of muscle that creates extra disturbances on the acoustic channel. Then, the system extracts specific features from the channel estimation matrix to quantify the fatigue level. These features will be the raw inputs for further fatigability analysis.

Lastly, the fatigability analysis module evaluates the muscle condition using all fatigue level quantifications in a group of workouts. Since muscle fatigue should monotonically increase over consecutive workouts, certain outlier workouts that exhibit abnormal fatigue values are removed firstly. Then, this work investigates the fatigability by analyzing the speed of fatigue accumulation and the actual occurrence of fatigue in a group of workouts. For the accumulation speed, the system uses a single numerical feature to represent the fatigue level via linear curve fitting. For the occurrence of fatigue, this work defines a feature set and cluster them into two categories as non-fatigue and fatigue.

4.2.1 Channel Estimation

To obtain the estimation matrix that carries information of muscle tremor, this technique contains two major steps of signal design and channel estimation.

Signal design: The signal transmitted by smartphone speaker contains single carrier frequency and the system encodes the carrier wave with a known sequence through BPSK modulation. To achieve efficient computation of channel estimation [92], the sequence is

chosen as a 26-bit GSM training sequence that is originally used in GSM communication protocol for channel equalization. Such pre-known sequence, though primarily estimates the characteristics of physical path in RF systems, is also applicable to acoustic channel due to the similarity between signal processing methods of both systems. Besides, since commodity smartphones always have co-located speaker and microphone (e.g., at the bottom side of the smartphone), their direct transmission, if not properly tackled, will have significant impact on the channel estimation. Therefore, a fixed idle period is appended after each modulated sequence to make the transmitted signal perform like an impulse signal. With only reflections from external objects, the received sequence can be correctly demodulated to ensure the correctness of the channel estimation.

Channel estimation: The received signal sequence \mathbf{y} , after demodulation, is expressed as $\mathbf{y} = \mathbf{M}\mathbf{h} + \mathbf{n}$, where \mathbf{h} is the complex channel impulse response, \mathbf{n} is white Gaussian noise and \mathbf{M} is the circulant training sequence derived from the known sequence mentioned in the signal design [92]. Let $\mathbf{m} = \{m_0, m_1, \dots, m_{L+P-1}\}$ denote the known signal sequence being transmitted and $L + P = 26$ denote the matrix size (L and P are estimation parameters), then \mathbf{M} is formed as:

$$\mathbf{M} = \begin{bmatrix} m_L & \cdots & m_1 & m_0 \\ m_{L+1} & \cdots & m_2 & m_1 \\ \vdots & & \vdots & \vdots \\ m_{L+P-1} & \cdots & m_P & m_{P-1} \end{bmatrix}_{P \times (L+1)}. \quad (4-3)$$

The best linear unbiased estimate (BLUE) for $\mathbf{h}_{(L+1) \times 1}$ is then computed as

$$\hat{\mathbf{h}} = (\mathbf{M}^H \mathbf{M})^{-1} \mathbf{M}^H \mathbf{y}, \quad (4-4)$$

where this work uses partial decoded sequence $\mathbf{y}_{P \times 1}$. For the GSM training sequences, the given solution to $\hat{\mathbf{h}}$ provides the periodic autocorrelation function of the codeword sequence in the received signal. Since objects at different distances produce delayed reflections of the same transmitted ultrasound signal, the system could extract the channel estimation corresponding to a certain distance range from the smartphone by shifting the original training

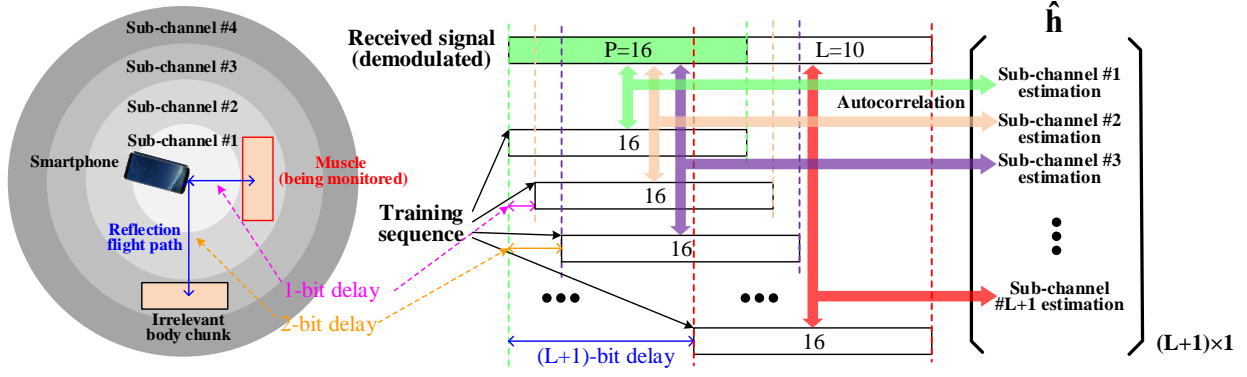


Figure 28: Sub-channels division based on different delays of the training sequence

sequence and autocorrelating it with different segments of the received signal after demodulation. Such operation, as shown in Figure 28, divides the entire surrounding area into smaller ranges (sub-channels), and the aforementioned parameter L determines the number of such sub-channels. Note that, larger L produces more sub-channels with individual estimations, but the correspondingly smaller P will make the estimation less reliable as the length of autocorrelation is reduced [129]. Empirically, this work selects $P = 16$ and $L = 10$ to have the best estimation results.

In summary, the preliminary metric $\hat{\mathbf{h}}$ is a complex vector, each element of which represents the channel estimation value at a specific distance range. Motion artifacts (e.g., body and hand motions) from other body chunks happen in distance ranges that are irrelevant to target muscle so that they will not affect the monitoring result. For those sub-channels farther from the phone, although the signal can still capture the variation within them, the interference and signal attenuation become more significant compared with closer ones. To improve the system reliability, this work stipulates that the system should be used by placing the muscle being monitored close to the phone ($<10\text{cm}$). Then, under the training sequence bit rate of 5k/s , the system will only search for tremor caused channel variation in the first three sub-channels (range of 10.2cm^4).

⁴Each sub-channel corresponds to a distance of $V_s/2B$, where V_s is the speed of sound (340m/s) and B is the bit rate.

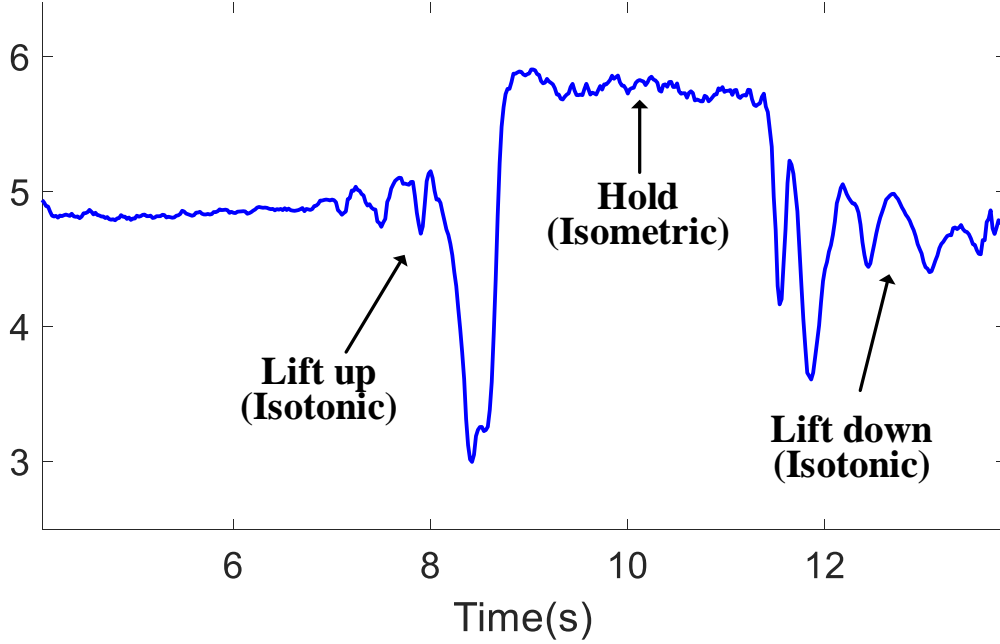


Figure 29: The waveform difference between isometric or isotonic contractions

4.2.2 Fatigue Interpretation

Next, the system interprets the channel estimation values into the level of muscle fatigue. When the human subject contracts the muscle to produce force, the system records every changing value of $\hat{\mathbf{h}}$ during the process. At the end of each workout, the system will have a time series of measurements on each sub-channel.

Depending on the type of exercise and muscle being used, the system further processes the received signal to locate the period of time that only the muscle tremor is the main cause of channel variation. For example, when a bicep workout is the target, the key of obtaining such period of time is to differentiate isometric and isotonic contractions because arm motion and the geometrical change of muscle may impact the tremor monitoring during an isotonic contraction (e.g., lifting up the forearm). Figure 29 shows the time series from one sub-channel during a complete bicep workout and this work demonstrates such impact by taking the moduli of those complex metrics.

From Figure 29, it is observed that both lift-up and lift-down phases lead to more evident changes of channel estimation due to the change of muscle geometry. As the physical change caused by tremor can be minor, it is very likely to be covered when being measured from

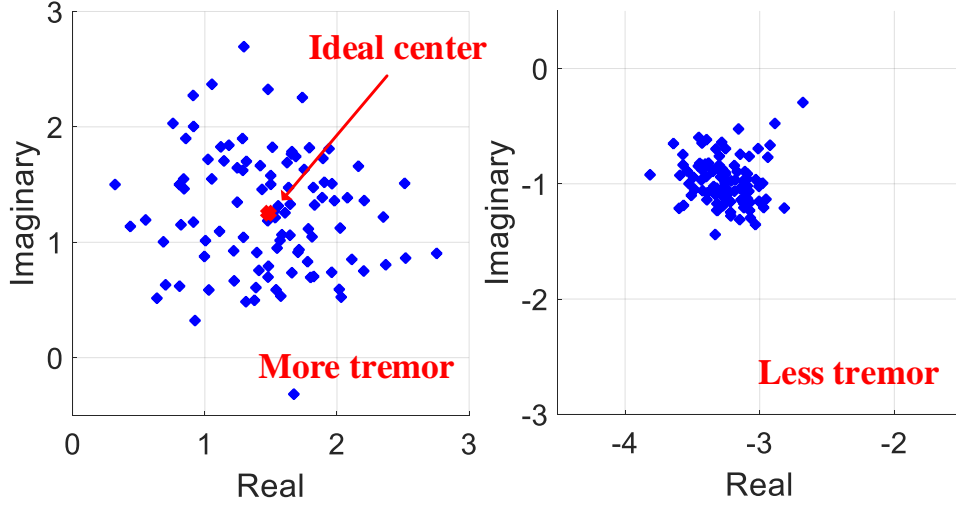


Figure 30: The concentration of a complex time series interpreted as fatigue level

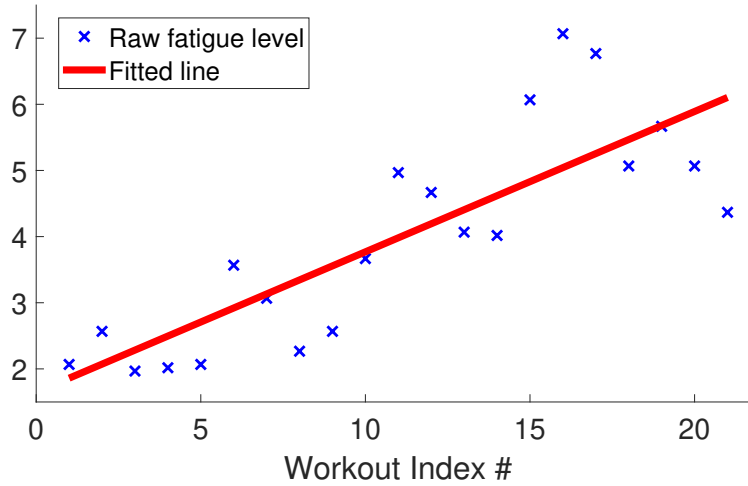


Figure 31: The accumulation speed of fatigue

isotonic contractions. To build solely the connection between tremor and channel variation, the system identifies the isometric phase by comparing the value range of a 3-second (stipulated in the test protocol) sliding window with a threshold. The threshold is set as 10% of the measured value over the 1st second when a subject remains static and prepares to begin the workout. Since this threshold is determined every time when the workout starts, it serves as a generic calibration process.

After addressing all significant factors that may affect the channel estimation, an approach is proposed to extract the channel variation led by muscle tremor and interpret it to fatigue level. This work exploits the channel estimation metric in the complex plane by assuming that any complex time series over time, if not being disturbed, would concentrate all at a single point on the complex plane. As shown in Figure 30, the more disperse the time series in the complex plane is, the more muscle tremor is exhibited. To measure the degree of dispersion, the system first finds the average point of all complex signal samples on the complex plane. Then, a new vector is formed by computing the distance between every signal sample and the average point. For a single workout, the quantitative fatigue level can be formulated as the mean of these distances.

4.3 Facial Expression Recognition with Headphones

This work aims to address the aforementioned limitations of human facial expressions recognition and instead propose a new sensing system that achieves precise and reliable recognition of human facial expressions by only using commodity headphones. As shown in Figure 32, the proposed system transforms the commodity headphone into an acoustic sensing device through simple audio jack re-wiring, which changes one of the headphone speakers into a microphone. With the inaudible ultrasound signal being transmitted from the headphone’s speaker to microphone, the system design builds on the fact that different facial expressions correspond to different skin deformations on the human face surface, which significantly affect the propagation of ultrasound signal above skin surface. This correlation, then, allows us to precisely recognize human facial expressions from the received ultrasound signal. Furthermore, since headphones are usually attached to the user’s ear, they are less likely to move and hence provide better reliability of continuous recognition over long time. Besides, since facial expressions are naturally caused by different emotions, this technique can work together with more specialized physiological sensors that record galvanic skin response (GSR) or electromyography (EMG) to further extend existing mobile devices’ functionalities.

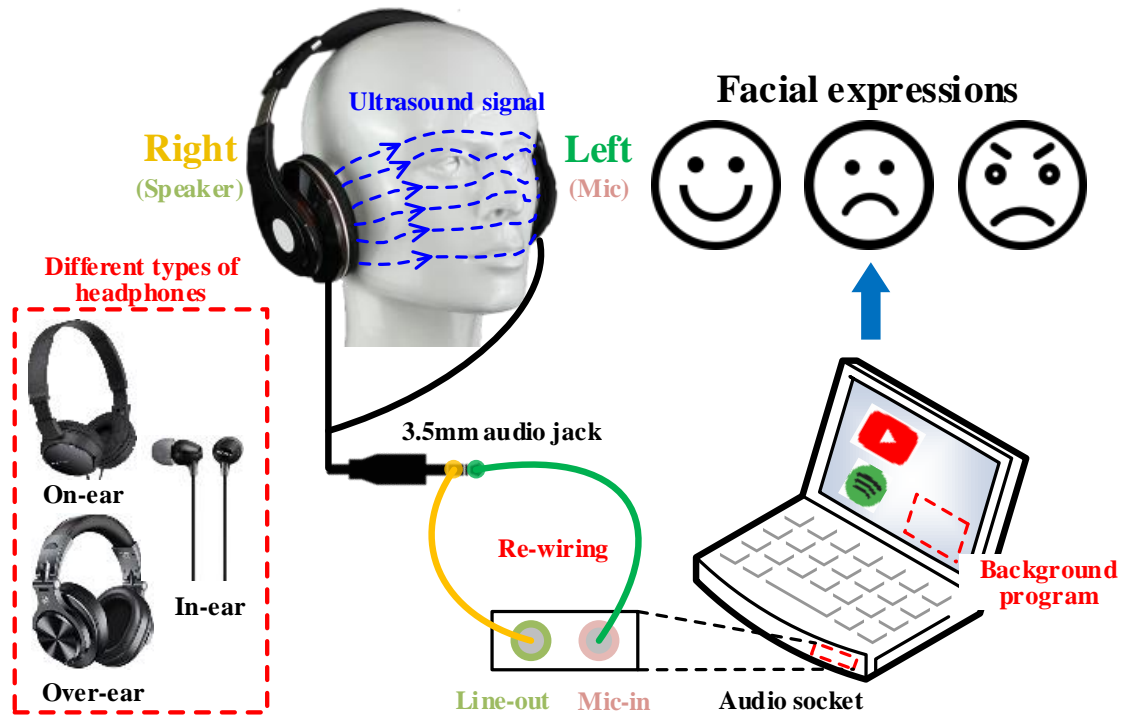


Figure 32: The system recognizes facial expressions from the acoustic signal transmitted above the face skin surface.

As shown in Figure 33, the system recognizes facial expressions when a wired headphone is worn regularly but in a way that one of the headphone speaker channels is converted into a microphone. From the acoustic signal, the system employs channel estimation technique to captures information related to human facial expressions. The recognition model consists of training and testing phases. During the training phase, the model needs to be assisted by an additional camera placed in front of user face. The headphone plays and records the ultrasound while the camera captures a video of facial images. After the images and sounds are collaboratively trained with the teacher-student network, the student network using acoustic only can be individually used in the testing phase to recognize the user's facial expressions.

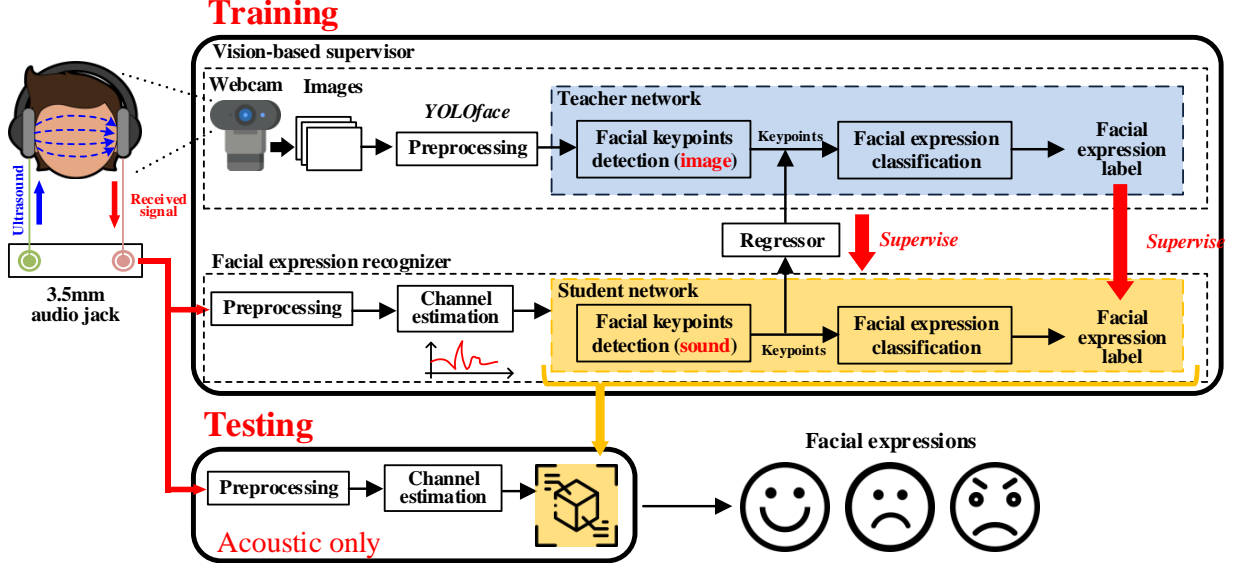


Figure 33: System Design

4.3.1 Headphone Acoustic Sensor

In practice, the ultrasound signal could be possibly propagated along three different pathways as shown in Figure 34 and listed below:

1. The signal propagates through the facial skin tissue.
2. The signal propagates first into the ear canal of the speaker side, and then reaches the other ear canal of the microphone side through the human head.
3. The signal propagates through the airborne pathway above the skin surface.

In the first pathway, the ultrasound signal will be greatly attenuated whenever it passes through the boundary between air and human tissue, in which the acoustic impedance decides such attenuation. As listed in Table 1, since the impedance in air is much smaller than that in human tissue, a significant amount of signal energy ($>99\%$) will be absorbed by the human tissue itself and the remainder will hence have negligible impact on the microphone's received signal. Similarly, in the second pathway through the ear canals, the majority of signal energy will be absorbed by the ear drum that captures the air vibration.

To this end, it is considered that the majority of signal being received at the transformed microphone is transmitted through the airborne pathway above the facial skin surface. In practice, when the headphone is being worn by the user, such propagated signal is the

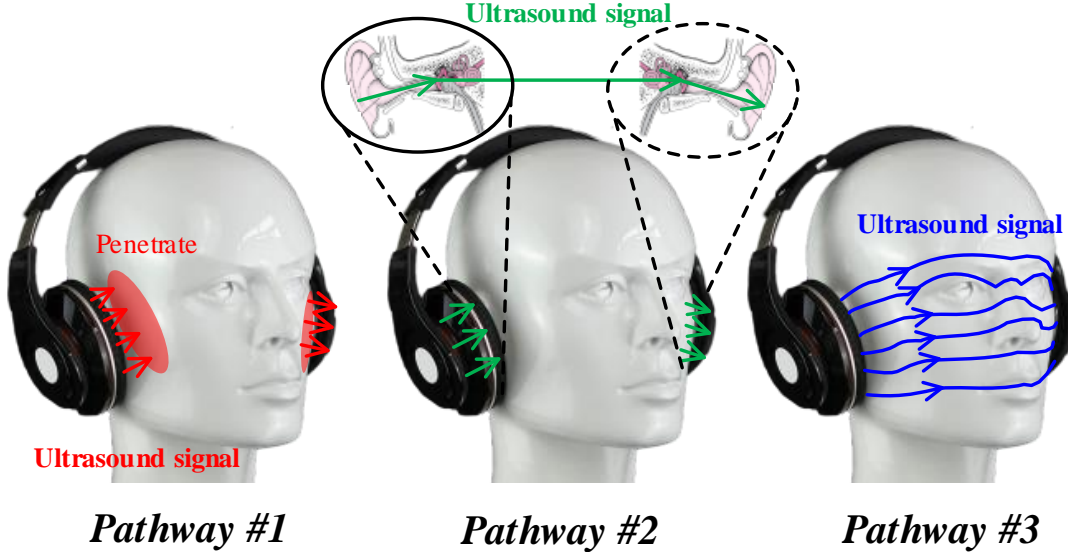


Figure 34: Three possible pathways of ultrasound signal propagation

outcome of the headphone sound leakage, which widely exists on almost every commodity headphone [87]. To verify such leakage, preliminary experiments are conducted on different types of headphone models⁵ by transmitting a sweeping sine wave from 20 Hz to 24 kHz with the maximum volume. Results in Figure 35 show that these commodity headphones are able to greatly suppress signal leakage within the audible band, but produce much higher signal leakage in the ultrasound band. In particular, on all the headphone models, Figure 35 shows that the sound pressure level (SPL) above 15 kHz is always above 30 dB and at least 15 dB higher than that in the audible band. These results verify that the ultrasound signal transmitted by commodity headphones, when being received by the transformed microphone on the headphone, is strong enough for precise recognition of human facial expressions.

Since sound is produced by vibrations and propagates through air as pressure waves, speakers and microphones are devices that essentially generate or capture such air vibrations. Drivers of a speaker or a microphone, as shown in Figure 36, share the same mechanism in fact. For the speaker, electric signal triggers the diaphragm to move through electromagnetic induction. Then, the diaphragm movements lead the surrounding air to vibrate thus create the sound. On the other hand, a microphone also has these driver components like diaphragm, magnet and coils, etc. But the microphone functions in a reverse way that

⁵On-ear: Sony MDR-ZX110, Over-ear: Audio-Technica ATH-M30x, In-ear: Samsung EO-EG920BW.

Table 1: Acoustic impedance of different media

Media	Density (kg/m^3)	Velocity (m/s)	Impedance ($10^6 kg/m^2/s$)
Air	1.16	344	0.0004
Water	998	1482	1.48
Muscle	1041	1580	1.65
Tissue	928	1430	1.33

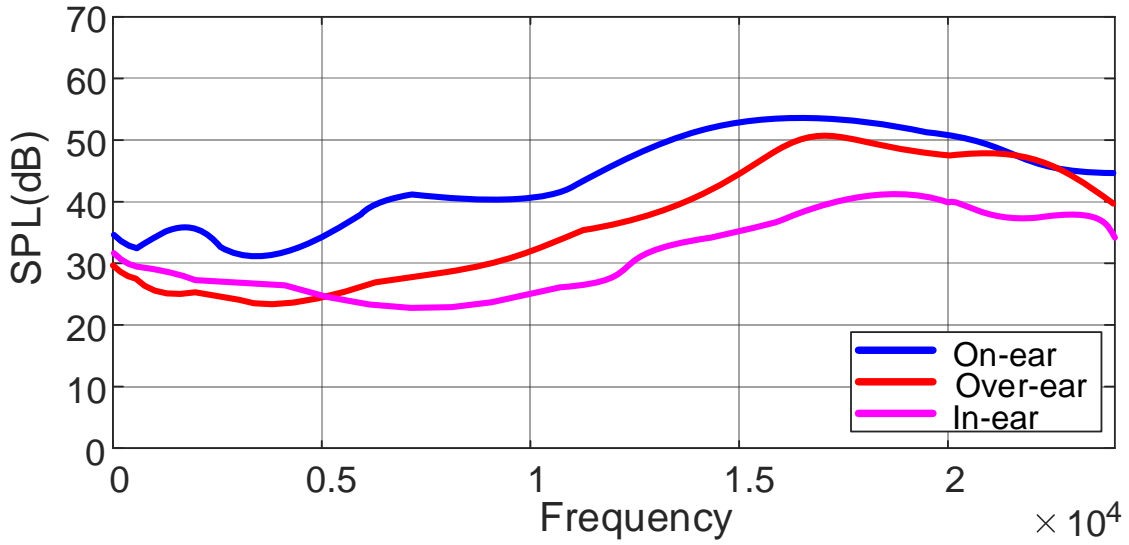


Figure 35: Leakage measurement of different headphone model

sound wave actuates the diaphragm first. With the similar electromanetic coupling effect of the magnet and coils, acoustic signal is then converted into electrical signal, which can be processed by hardware like desktop sound cards.

Based on such similarity, the hardware of a commodity speaker can be used as a microphone, by simply using the speaker's output interface as input. In practice, existing work showed that the 3.5mm socket of audio output can be converted to input via software, because any audio port of the sound processing hardware (e.g., the PC sound card) is designed to be configurable as either output or input [91]. However, this software approach can only

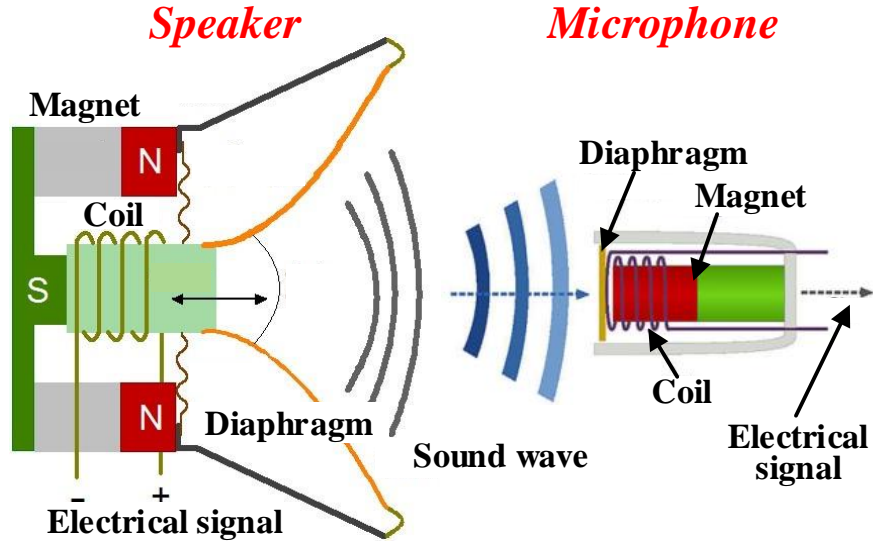


Figure 36: Diagrams of the speaker and microphone

configure both the left and right audio channels of a headphone as output or input at the same time, and hence cannot be used to transform a commodity headphone into an acoustic sensing device that consists of both speaker (output) and microphone (input).

Instead, the system transforms one speaker of the headphone into a microphone but retain another speaker unchanged, by re-wiring the audio jack to split the left and right audio channels as shown in Figure 37. More specifically, on a tip/ring/sleeve (TRS) or tip/ring/ring/sleeve (TRRS) audio jack, the system separates its tip and ring into two separate audio jacks, which are then plugged into the Line-out and Mic-in ports of the host system, separately. In this way, the sound will be played through the right speaker and then recorded by the left speaker⁶. In practice, unlike existing work that requires a custom IC board [29], this approach can be easily implemented with an audio jack splitter into separate 3.5mm audio jacks. Then, plugging either channel into the mic-in socket on a computer can turn one of the headphone speaker into an acoustic sensor (microphone). When the computer plays an ultrasound through a headphone configured in such a way, the acoustic signal will

⁶Similarly, the audio jacks can be re-wired in a reverse way to propagate sound from the left speaker to the right speaker.

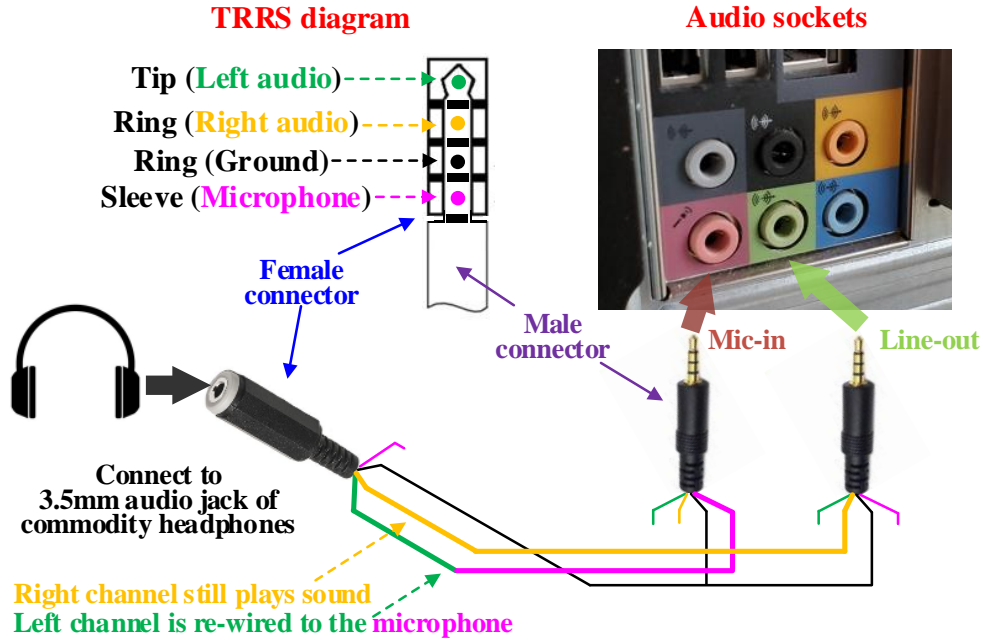


Figure 37: Rewiring the 3.5mm audio jack to implement the required sensing setup

propagate through the facial skin and be received on the other side. Since the propagation pathway captures the skin deformation caused by any facial expression, the system estimates the acoustic channel to extract the useful information.

4.3.2 Acoustic Channel Estimation

A major concern of connecting a commodity headphone speaker into a microphone socket is the audio quality, because the speaker specification is not designed for that purpose. Since regular speaker has a larger diaphragm, sound with similar energy will cause less vibrations so that the electrical signal produced is minute to be captured without a filter circuit and amplifiers. this work investigates how the sound quality varies between a regular microphone and a re-wired headphone speaker. Figure 38 shows recordings of the same sound being recorded at the same distance by these two microphones.

Compared with a regular microphone, the sound recorded by a re-wired microphone exhibits weaker signal strength on useful content but has a similar noise level. Such a sound composition makes the content less recognizable after being buried into the noise. To

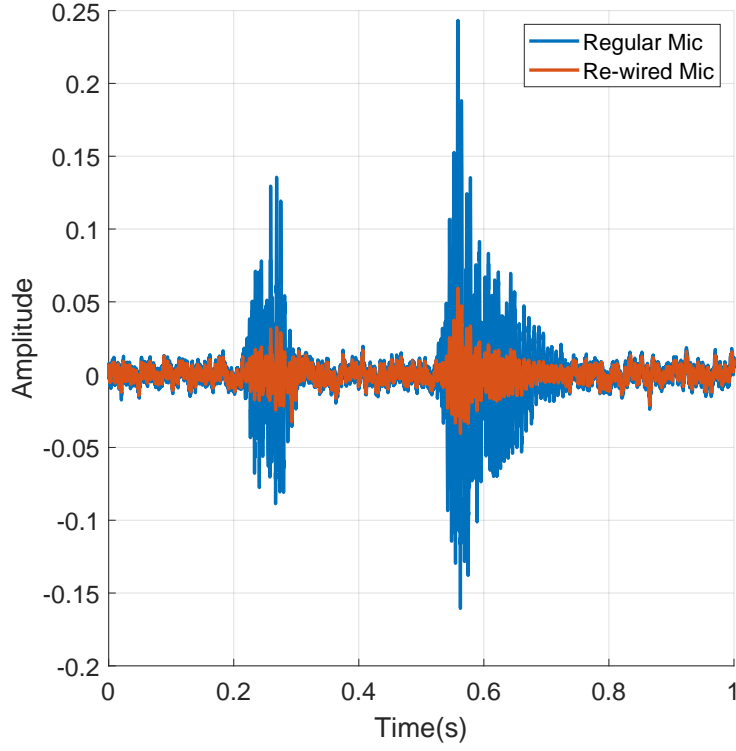


Figure 38: Raw signal collected with regular and re-wired microphones

overcome such difficulty and unveil the transmitted signal to facilitate the acoustic channel estimation, this work designs single-tone signals to carry a known sequence (60-bit generated randomly) by binary phase shifting keying (BPSK) modulation. At the receiver end, though the sound is distorted, it still shows great detection performance (high correlation coefficients) of these sequences as shown in Figure 39.

In details, after the signal is preprocessed by a high-pass filter to exclude audible sounds that may potentially impact the sensing accuracy, the system first synchronizes the received signal with known transmitted sequences at samples with highest correlation coefficients. Then, the signal is demodulated to find original symbols of the known sequences. Assuming the channel as $Y = HX + noise$, where $Y = [y_1, y_2, \dots, y_n]$ is the received signal, $X = [x_1, x_2, \dots, x_n]$ is the transmitted signal and x_n and y_n are known symbols, respectively, the acoustic channel H can be estimated from Equation 4–5.

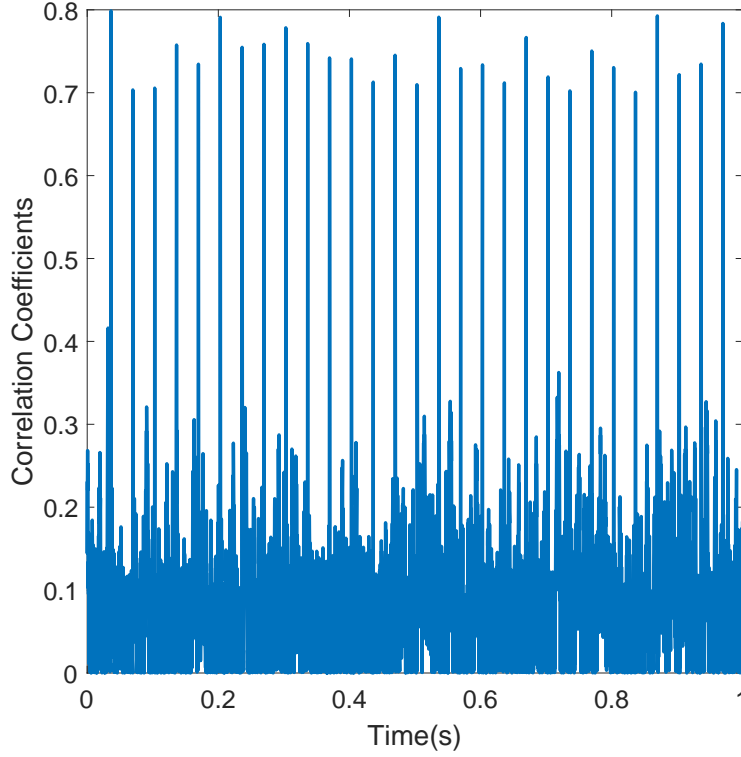


Figure 39: High correlation coefficients on the encoded signal

$$H = YX^T(XX^T)^{-1} \quad (4-5)$$

To better improve the estimation reliability, this system adopts frequency-division multiplexing that simultaneously transmits encoded signals at different frequency bands. At a time, there are multiple measurements of the channel estimation, and this work will treat all of them as raw features for the following neural networks.

4.3.3 Recognizing Expressions from Acoustic Signals via Machine Learning

With these acoustic channel estimations, an intuitive approach to recognizing facial expressions is to apply these channel estimations and the corresponding facial expression labels to train a neural network. For example, a long short-term memory (LSTM) classification model can be trained to learn from the time series data of channel estimations. However,

preliminary experiments using such as simple classification model lead to very low recognition accuracy as shown in Figure 40. The main reason of such low accuracy is that the correlation between the acoustic channel and facial expressions is subtle, and the acoustic channel is hence highly sensitive to the small changes of face skin deformation, which could be highly heterogeneous among different samples of the same type of facial expression. The capacity of a simple classification model, in this case, is insufficient to fully represent such subtle correlation.

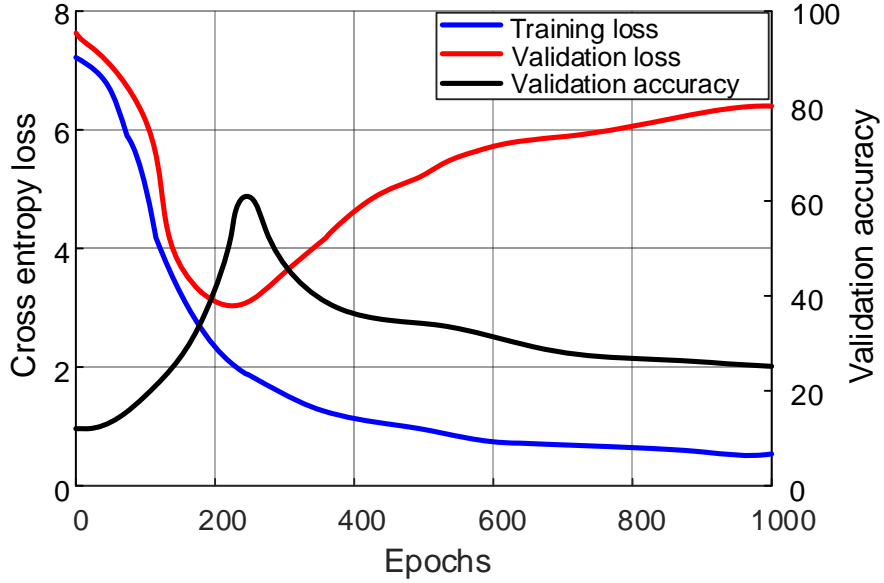


Figure 40: Model performance using a single network based on acoustic signal

On the other hand, existing work has demonstrated that using neural networks can precisely recognize humans’ facial expressions from their facial images. When using images of front faces, the accuracy of recognition using a single deep neural network (DNN) can achieve 95% [49, 122]. Based on such high accuracy, the system intends to transfer the latent knowledge about facial expressions being learned from such vision-based teacher network to the acoustic-based student network, to supervise the training of student network and hence increase its capacity.

Traditionally, such supervision is conducted by taking the teacher network’s output as the target label for the student network during the training procedure. This method, however, is ineffective in the proposed system because many small changes on facial deformation may not be captured by the teacher network from facial images but could result in large variation in the acoustic channel.

To address this challenge and ensure appropriate supervision, as shown in Figure 33, system divides both the teacher network and the student network into two sub-networks: the *first* sub-network detects the positions of a pre-defined set of facial keypoints (e.g., eyebrow ends, eye corners, nose and mouth corners, etc), and the *second* sub-network further recognizes the facial expression from the positions of these keypoints with respect to the human face contour.

As a result, the teacher network can supervise the training of both sub-networks in the student network. In particular, the teacher network’s learned knowledge on individual facial keypoints can be effectively transferred to the corresponding sub-network in the student network, hence ensuring that any small changes on facial deformation can be correspondingly mapped to the right keypoint. In facial keypoint detection of the teacher network, the system first preprocesses each facial image by resizing and cropping it⁷, to make sure that the coordinates of facial keypoints in all facial images are consistent. This consistency, then, allows us to universally compute the coordinates of facial keypoints in each image as their relative positions from the nasal tip in the image.

Afterwards, the system needs to decide the most appropriate set of keypoints being used, which should be the most important ones that represent facial expressions. For example, the positions of nose and mouth corners are important indicators of many facial expressions including joy, sadness, anger, disgust, etc. The system starts from a face mesh model that defines 468 facial keypoints as shown in Figure 41(a), and then select 30 keypoints with the highest importance in facial expression recognition to be used in the system for timely online inference.

In the system, such selection procedure is done by training a regular fully-connected neural network over the collected facial images. More specifically, the system uses a well trained model to detect the 468 keypoints in each facial image and use their coordinates as the neural network input. The facial expression labels of the images are used as the neural network output. After that, this work adopts the explainable AI technique [112] on this trained network to compute the integrated gradient (IG) of each keypoint, which quantita-

⁷Such preprocessing can be done by simply using existing image processing tools such as the YOLOface model [97].

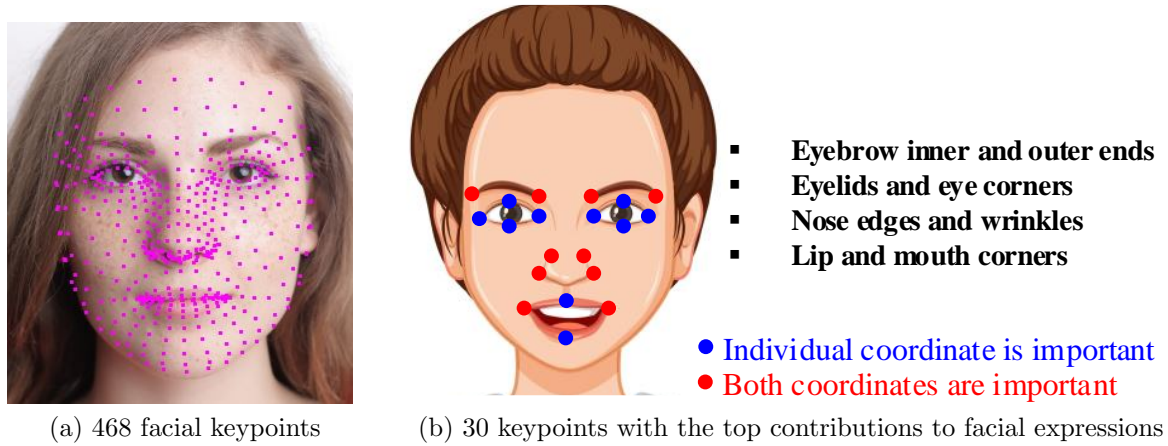


Figure 41: Keypoints on human faces

tively measures how this keypoint contributes to correct recognition of facial expressions. The top 30 keypoints with the highest contributions are then selected to be used in the following expression recognition task.

These selected 30 keypoints, as shown in Figure 41(b), mainly cover the edges and corners of eyebrow, eye, nose and mouth, hence depicting the contours of these important elements on human face. Note that, the contributions of the x and y coordinates of these keypoints are separately measured, and hence some keypoints may only have one coordinate that is important to facial expression recognition. Based on such selection, Table 2 shows⁸ that, when being applied to the same set of facial images being used in training, these 30 keypoints result in similar accuracy of facial expression recognition, compared to that using all the 468 keypoints.

Table 2: Classification accuracy with different selections of facial keypoints

Expressions	N	J	Sa	Su	A	F	D
468 keypoints(%)	96.5	95.3	95.9	95.2	94.9	95.3	95.1
30 keypoints(%)	97.1	95.5	96.6	96.1	94.7	95.9	96.5

The second sub-network, as the classifier of facial expressions, is being trained with the coordinates of the 30 selected facial keypoints on each facial image as input and the corresponding facial expression label as the output. In particular, to ensure its reliability

⁸(N=neutral, J=joy, Sa=sadness, Su=surprise, A=anger, F=fear, D=Disgust)

and accuracy, this work first pre-trains this network using a public dataset of facial images with known facial expression labels. Afterwards, when being used on a specific user, this classifier will be further improved with this user’s own facial images.

In order to facilitate the supervision from the teacher network, the student network in the system is also designed to be consisting of two sub-networks: the first sub-network decides the positions of the 30 selected facial keypoints from the input data of acoustic channel estimation and then the second sub-network similarly recognizes facial expressions from the detected facial keypoints. Since acoustic channel estimations are produced as time series data, this work uses the LSTM model as the basic processing kernel for facial keypoint detection. When multiple acoustic channel estimations from different frequency bands are simultaneously used, they are being applied to separate LSTM encoders and then merged together.

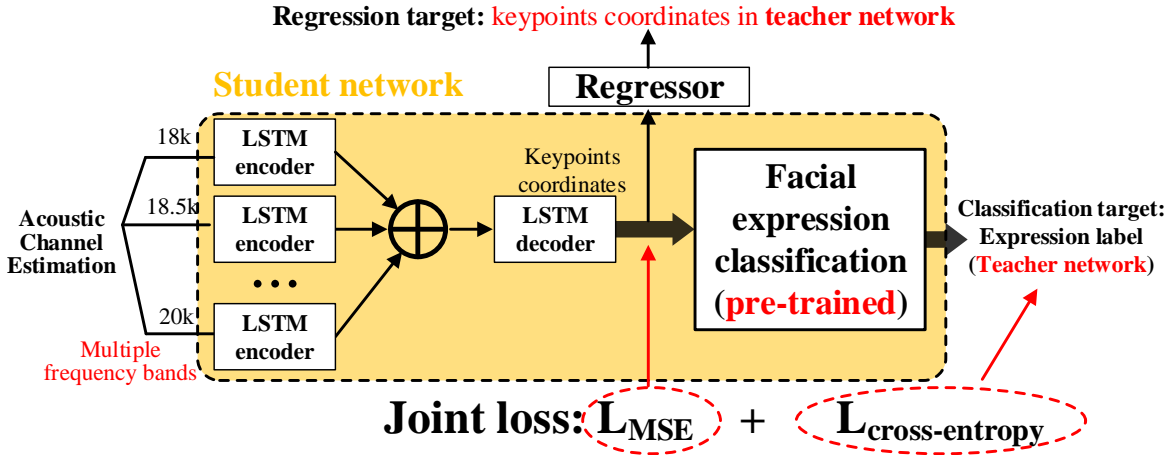


Figure 42: Student network design

As shown in Figure 42, the supervision on the student network is defined as a joint loss function that covers both facial keypoint detection and facial expression classification. The supervision on facial keypoint detection is designed as a regression task, which aims to minimize the mean square difference (MSD) between the teacher network’s and student network’s facial keypoint coordinates being detected. The supervision on facial expression classification, on the other hand, is done through the categorical cross entropy loss, which ensures that the student network produces similar results of facial expression recognition as

the teacher network does. The detailed model design is described as follows: (1) during the training phase, both LSTM network that consists multiple LSTM encoders and the regressor are trained by feeding acoustic channel estimations as model inputs. The first training goal is that the output of regressor should be as close as facial keypoint coordinates recognized by the vision-based teacher network. The second training goal is the output of expression classification model should predict the correct expression label. During the process, such collective training is controlled by minimizing the joint loss function. (2) In the testing phase after sufficient training data is provided, the LSTM network can not only extract channel estimation features to predict specific facial expression, but also ensure that these features are constrained in the searching space of accurately detecting the facial keypoint coordinates.

5.0 Evaluations

This chapter describes the evaluation setups and results of applications being proposed in the thesis.

5.1 Evaluation Setup

5.1.1 Lung Function Tests

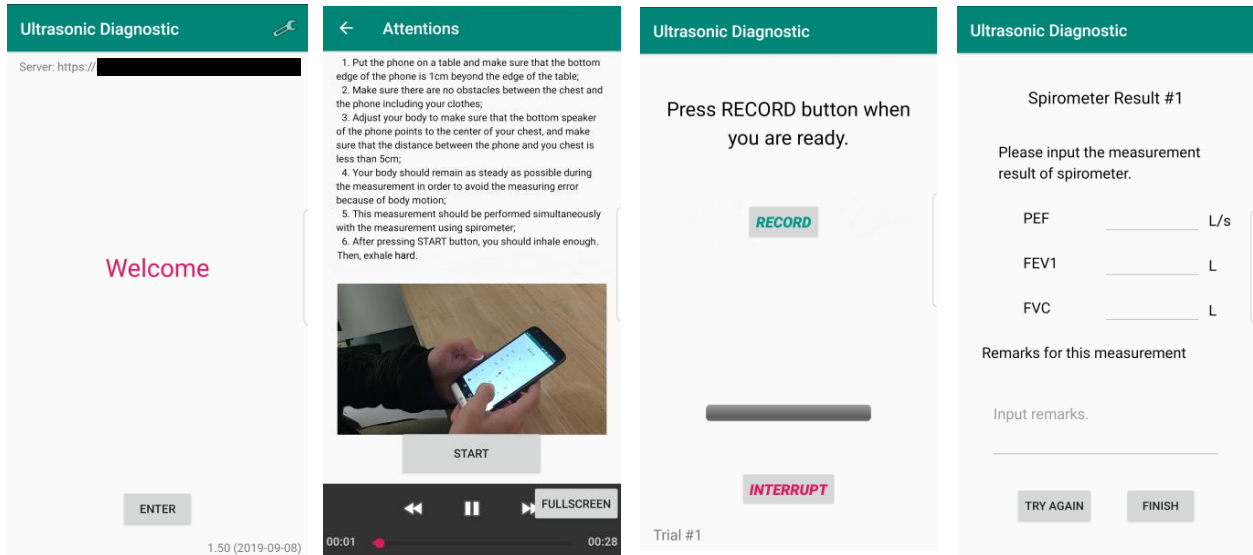


Figure 43: Main screens of smartphone app

As shown in Figure 43, an Android smartphone app that transmits multi-tone ultrasound signals is implemented. It uses 12 ultrasound frequencies ranging from 17kHz to 22.5kHz with 0.5kHz interval, which have been proved to have good frequency responses on commodity smartphones [118]. Being similar with previous work [118], the system disables the smartphone's Automatic Gain Control (AGC) to avoid unwanted fluctuations of the received signal amplitude when the ambient noise level varies¹. The measured chest wall motion is also adaptively smoothed before any feature extraction, using a flexible sliding window.

¹Since the is designed to receive ultrasound signal within a short range ($<20\text{cm}$), it has good SNR ($>40\text{dB}$) even when disabling the AGC.

The app is designed following instructions from pulmonary clinicians, to minimize the patients' cognitive and operational barriers when using the system out of clinic. Before a spirometry test starts, a tutorial with both texts and video is provided to demonstrate the protocol. A spirometry test using the system is: 1) fully automated that no manual inputs (e.g., indicating the start and end of inhalation and exhalation stages) are needed from the patient; 2) fully customizable that the patient can opt to pause or resume the ultrasound recording at any time. The application also allows uploading the test results using clinical spirometers, so that this app can also be used in clinical studies.

This work evaluate the accuracy of measuring humans' lung function over five healthy student volunteers with different body conditions². All experiments are conducted in a 10m×10m office with regular furniture and facilities, and student volunteers are instructed to strictly follow the spirometry protocol. Every student volunteer, as shown in Figure 44, conducts 50 spirometry tests using a portable spirometer, and simultaneously holds a smartphone running the system. The distance between the smartphone and chest wall varies between 5cm to 20cm.

For each student volunteer, the experiment uses the other four volunteers' data (spirometer readings and the measured chest wall motion) to train the neural network regression model. The trained model is used to convert this student volunteer's chest wall motion into lung function indices, which are then compared with the corresponding spirometer results to evaluate the measurement accuracy. Finally, the experiment result is averaged over all the five volunteers.

5.1.2 Muscle Fatigability

This work implements the technique as an Android smartphone app, which uses 20kHz as the carrier wave of the encoded signal. Though commodity smartphones usually have two pairs of speaker and microphone for music playback or voice call, the system chooses top speaker along with bottom microphone because this combination can ensure the best results of channel estimation from experiments. To minimize interaction needs, the subject only

²Ages: 21-28, Height: 173-188cm, Weight: 55-85kg.



Figure 44: Evaluation setup

needs to signal the start and end of each test by pressing a phone volume button. The test protocol is suggested by clinicians based on a simple bicep workout and entire procedures are automated by voice and screen instructions as shown in Figure 45.

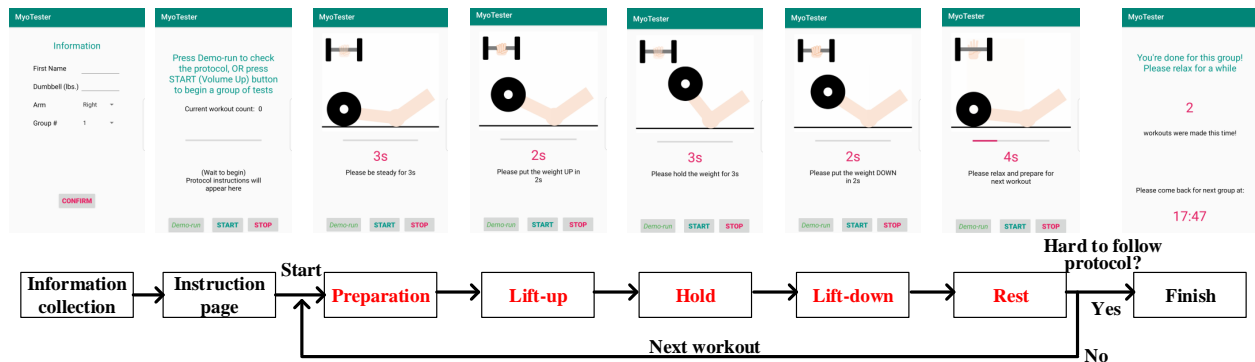


Figure 45: Protocol procedures and instructions of the smartphone app



Figure 46: Experimental setup for bicep muscle

The protocol stipulates that the basic unit of the test is a group, in which it requires the subject to complete as many biceps workouts as possible until he/she feels hard to follow the speed of voice instructions. Each workout includes steps of preparation, lift-up, hold, lift-down and a resting interval for the following workouts. By explicitly defining the duration of these steps and the elevation that the dumbbell should be lifted, this work standardizes the protocol so that people with different muscle condition should have similar feeling of fatigue after enough times of workout.

This work evaluates the technique over 5 student volunteers. All volunteers have different muscle strength and exercise habits, but none of them has ever experienced any muscle disorders. All experiments are done in a 10m×10m lab office with furniture and electronic equipment. Based on the evaluation goals, the protocol may ask the subjects to use dumbbells with different weights or complete multiple groups of test. During the test, the smartphone should be placed closely (5 to 10cm) with its screen facing towards the muscle being monitored. For bicep muscle, a handmade holder is used to fix the phone screen pointing to the inner area of upper arm as shown in Figure 46.

5.1.3 Facial Expression Recognition

The system is implemented using different types of commodity headphones, balun audio connectors and audio jack splitters, and transmit ultrasound signals in multiple frequency bands between 18 kHz and 21 kHz with an interval of 0.5 kHz. In the implementation, this work allows the system to monitor the user’s facial expressions during normal uses of the headphones (e.g., playing music or watching videos). To allow simultaneous transmission of ultrasound signal along with audible sound playback in these cases, the system uses built-in audio mixers on desktop OSes such as Windows and Linux.

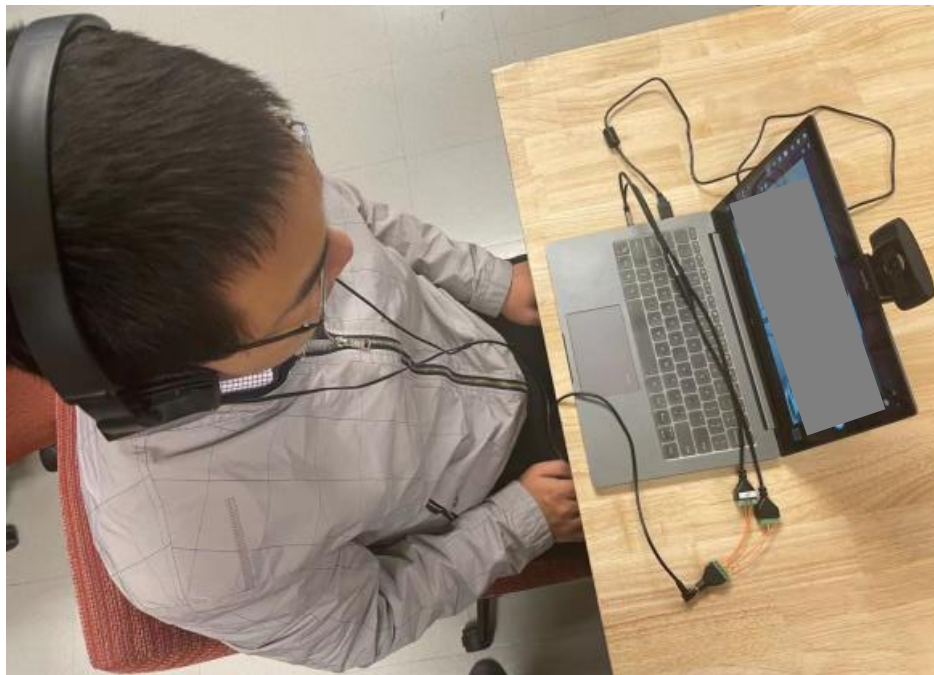


Figure 47: Evaluation setup

Based on this implementation, this work evaluates the accuracy of recognizing the facial expressions over 5 student volunteers, in a 10m×10m lab office with regular furniture. Due to the large variation among different human faces, the neural network models are separately trained and tested for each student volunteer, and the results are averaged over the 5 student volunteers. As shown in Figure 47, in these experiments, a volunteer sits in front of a camera and wears the re-wired headphone that connects to the host system. The volunteer then follows the host system screen’s instructions to make facial expressions in 7 different categories (neutral, joy, sadness, surprise, anger, fear and disgust) and retain each facial

expression for 5 seconds before switching to the next. Each experiment lasts for 10 minutes, during which the camera takes facial images at 30 FPS and acoustic sensing is conducted simultaneously. Since the acoustic signal has much higher sampling rate compared to the video frame rate, each video frame corresponds to one segment of sound recording. In total, each experiments result in 18k training samples of video frames and acoustic recordings respectively for both teacher and student network.

5.2 Results

5.2.1 Lung Function Indices Error

The accuracy of lung function estimation depends on the complexity of neural network being used. Such complexity also decides the computation overhead of online inference. This work investigates these two aspects with different numbers of neurons in each hidden layer of the neural network. In all experiments, the system uses a Samsung Galaxy S8 phone at 5cm from the chest for measurement. Results in Figure 48 show that errors of estimating all the lung function indices can be effectively constrained below 2.5%, when the average number of neurons in each layer is higher than 8. Otherwise, when this average number is reduced to 6, the estimation error could approach to 3%. On the other hand, the computing time of neural network inference is proportional to the neural network complexity. Based on these results, this work sets the numbers of neurons in the three hidden layers as [12,10,8], for all evaluations in this work.

This work first examines the measurement accuracy with the raw chest wall motion produced by the motion tracker with lower accuracy. As shown in Figure 49, the average errors of estimating FEV1, FVC and FEV1/FVC are 7.3%, 6.0% and 4.6%, respectively. In particular, it is generally harder to precisely estimate FEV1, because the uncertainty in human behaviors makes it difficult to precisely locate p_{start} . On the other hand, when the proposed techniques are used to remove irrelevant hand and body motions, the system can reduce the errors to 2.5% and hence reach the same level of accuracy as in-clinic spirometry.

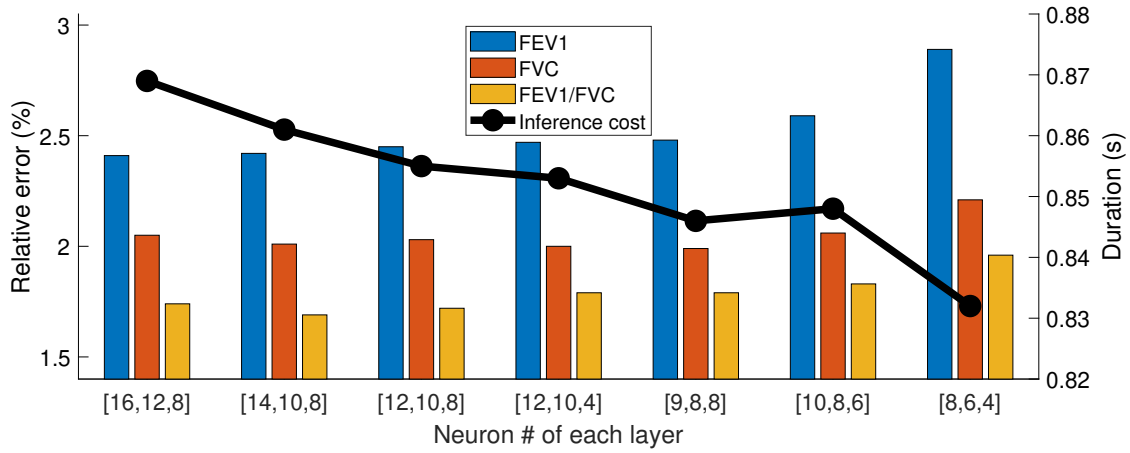


Figure 48: Trade-off between accuracy and overhead of neural network inference

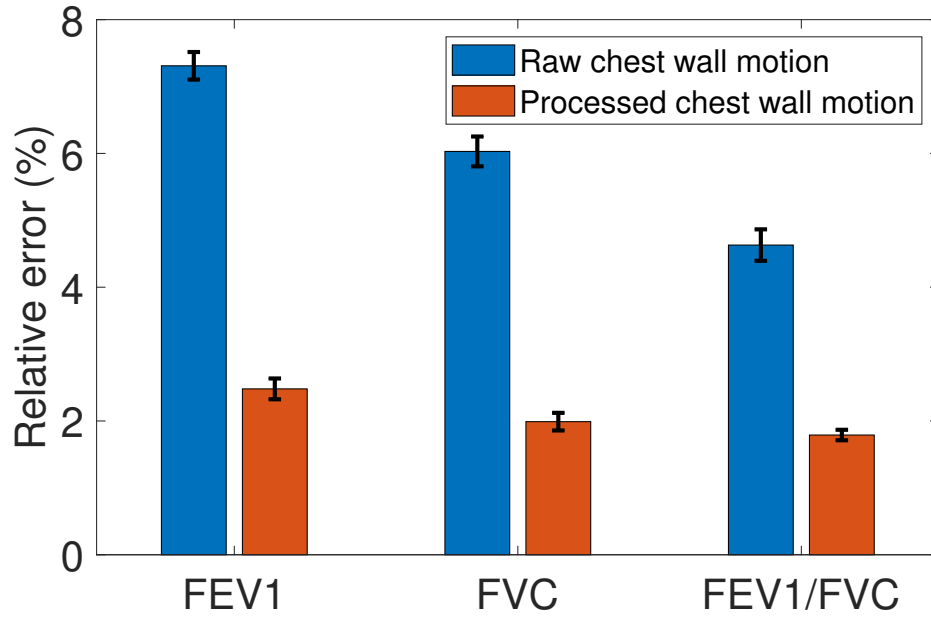


Figure 49: Lung function estimation error

When the distance between the smartphone and chest wall increases, the ultrasound signal strength attenuates and more multi-path interferences are also involved. As shown in Figure 50, when such distance increases from 5cm to 20cm, the error when measuring FEV1 slightly increases to 2.5%, because the extra signal distortions make it harder to correctly identify the start of exhalation stage. At the same time, this work is able to retain the

accuracy of measuring FVC and FEV1/FVC with negligible difference over these distances. These results enhance the usability of the system out of clinic, as users do not need to intentionally fix the distance between the smartphone and chest wall.

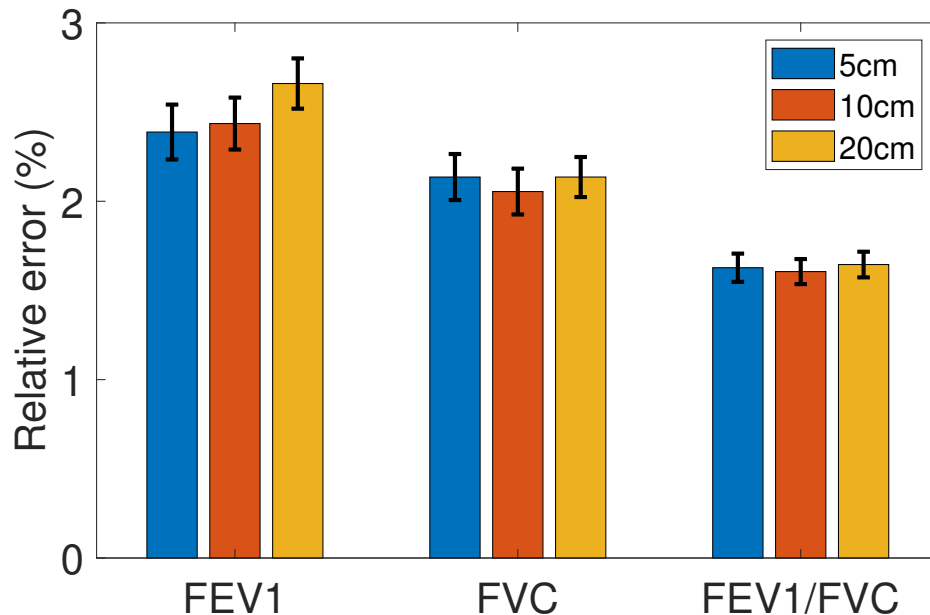


Figure 50: Different distances

When being hand-held, the smartphone may be randomly tilted from horizontal, especially during the exhalation stage. Such tilted positions change the reflection path of ultrasound signal and hence may also affect the measurement of lung function indices. To examine such impact, the system keeps the measurement distance to be 5cm and tilt the smartphone up and down with different degrees. Experiment results in Figure 51 show that when the smartphone is tilted up or down by 15 degrees, the increase of measurement error could be effectively restrained within 0.2%. Even when the smartphone is tilted to 30 degrees, such error will not exceed 3.5%.

The ultrasound signal, when penetrating through clothes, may create extra reflections that affect the chest motion tracking. To investigate such impact of clothes, the protocol instructs the volunteers to wear clothes of different tightness, thickness and texture, as shown in Figure 52(a). Experiment results in Figure 52(b) show that the system provides reliable lung function estimation with most types of clothes, and could effectively restrain the

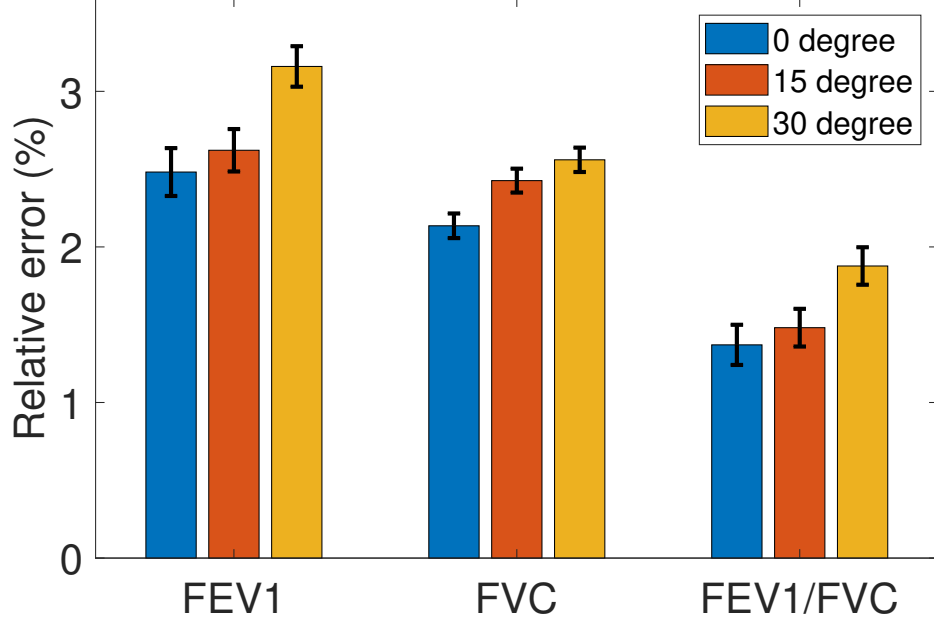


Figure 51: Different tilting angles

variance across different clothes within 0.5%. The only exception is when the patient wears a thick sweater, which results in significant signal attenuation and raises the measurement error to 7%. However, this thick sweater is for outdoor wear and the system does not expect any patient to wear it indoors.

Based on the high measurement accuracy over healthy humans, this work further conducts an observational clinical study at the Children’s Hospital of Pittsburgh, to investigate the accuracy of the system on measuring the lung function of pulmonary disease patients, particularly, pediatric patients with limited capabilities of cognition and cooperation. With the IRB approval³, 83 pediatric patients are recruited in 4 months, and requested each patient to use the system when they visited the hospital for in-clinic spirometry⁴.

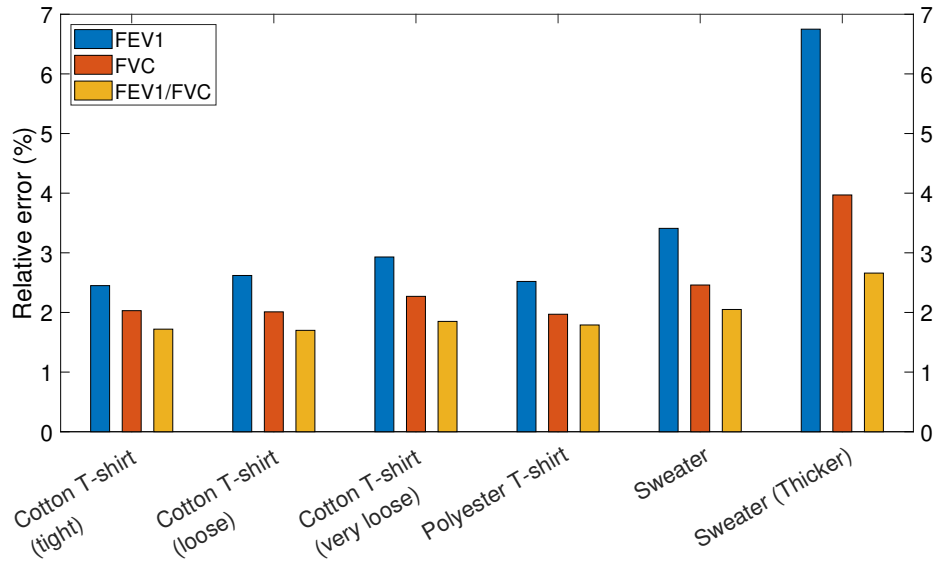
As shown in Table 3, the recruited patients cover a wide variety of different ages, genders, races, body conditions and diseases. A total number of 281 spirometry tests are measured, and 185 of them are considered as valid.

³University of Pittsburgh IRB approval No. STUDY19030114.

⁴Due to the hospital’s regulatory requirements to ensure clinical compliance and the pediatric patients’ body safety in the clinical room, the smartphone is fixed by a holder in front of chest rather than being hand-held.



(a) Different types of clothes being worn



(b) Measurement error with different clothes

Figure 52: Impact of wearing different clothes

This work statistically analyzes the correlation between patients' chest wall motion and lung function indices. Table 4 shows the Pearson correlation between chest wall motion features (S_{max} , D_{1s} , D_{max}) and lung function indices. It demonstrates that FEV1, FVC and FEV1/FVC have strong correlations ($P < 0.01$) with chest motion features, clinically validating the usefulness of the design. The correlation with PEF is weaker due to high variability of PEF measurements and uncertainty of pediatric patients' body movements.

Table 3: Patients' Information

Category	Characteristics	Number
Demographics	Tests per patient	2.2 ± 1.4
	Age (years)	11.5 ± 2.6
	Female (%)	40 (48.2)
Body conditions	Height (cm)	148.7 ± 15.3
	Weight (kg)	46.3 ± 19.9
	BMI (kg/m ²)	20.2 ± 5.3
	Underweight (%)	19 (22.9)
	Overweight (%)	14 (16.9)
	Obese (%)	1 (1.2)
Pulmonary diseases	Asthma (%)	62 (74.7)
	Cystic fibrosis (%)	21 (25.3)

Table 4: Pearson correlation between chest wall motion features and lung function indices.

Coefficients (P -value)	D_{1s}	D_{max}	S_{max}
FEV1	0.33 ($3.9e^{-6}$)	0.25 ($7.6e^{-4}$)	0.31 ($1.3e^{-5}$)
FVC	0.3 ($3.1e^{-5}$)	0.22 ($2.3e^{-3}$)	0.3 ($3.6e^{-5}$)
FEV1/FVC	0.17 ($1.8e^{-2}$)	0.15 ($4.2e^{-2}$)	0.16 ($3.1e^{-2}$)
PEF	0.12 ($9e^{-2}$)	-0.013 ($8.6e^{-1}$)	0.059 ($4.3e^{-1}$)

Furthermore, this work individually investigates the correlation between each lung function index and chest motion features, by using generalized linear regression. Results in Table 5 show that all three motion features are linearly correlated with FEV1 and FVC ($P < 0.01$). For example, each millimeter increment of D_{1s} is significantly associated with 1.26 increase of %pred FEV1 and 1.00 increase of %pred FVC. Such correlation with PEF, however, is weaker.

Table 5: Linear correlation between chest motion features and lung function indices (%pred)

β -Coefficients (P -value)	FEV1	FVC	FEV1/FVC	PEF
S_{max}	0.74 (< 0.01)	0.62 (< 0.01)	0.21 (0.03)	0.01 (0.16)
D_{1s}	1.26 (< 0.01)	1.00 (< 0.01)	0.37 (0.02)	0.04 (0.02)
D_{max}	0.54 (< 0.01)	0.43 (< 0.01)	0.19 (0.04)	0.01 (0.42)

Overweight and obese patients are more likely to exhibit abnormal chest wall motion. To investigate such impact on the system, the above statistical correlation is calculated over these patients. Results in Table 6 show that statistical correlations are generally weaker over these patients due to their heterogeneous external characteristics of the chest wall, and strong correlations are only found between chest motion features and FVC.

Table 6: Statistical correlation among overweight patients

β -Coefficients (P -value)	FEV1	FVC	FEV1/FVC	PEF
S_{max}	0.99 (0.06)	1.47 (< 0.01)	-0.31 (0.12)	-0.07 (0.02)
D_{1s}	1.81 (0.02)	2.46 (< 0.01)	-0.41 (0.17)	-0.09 (0.05)
D_{max}	0.44 (0.40)	1.12 (0.03)	-0.50 (< 0.01)	-0.08 (< 0.01)

Built on the statistical correlation, this work investigates the accuracy of lung function monitoring. Such accuracy is evaluated at both levels of individual spirometry tests and patients. Results in Figure 53 show that at test level, The system has an error of 15.45%

$\pm 1.82\%$ and $11.46\% \pm 1.23\%$ when estimating FEV1 and FVC, respectively. The main reason for such errors is that pediatric patients usually have difficulty in fully following the spirometry protocol and exhibit more diversity in chest wall motion. Besides, such estimation accuracy has been greatly improved from current low-cost spirometers ($>20\%$ in home use), and the accuracy in estimating FEV1/FVC is further reduced to $<10\%$.

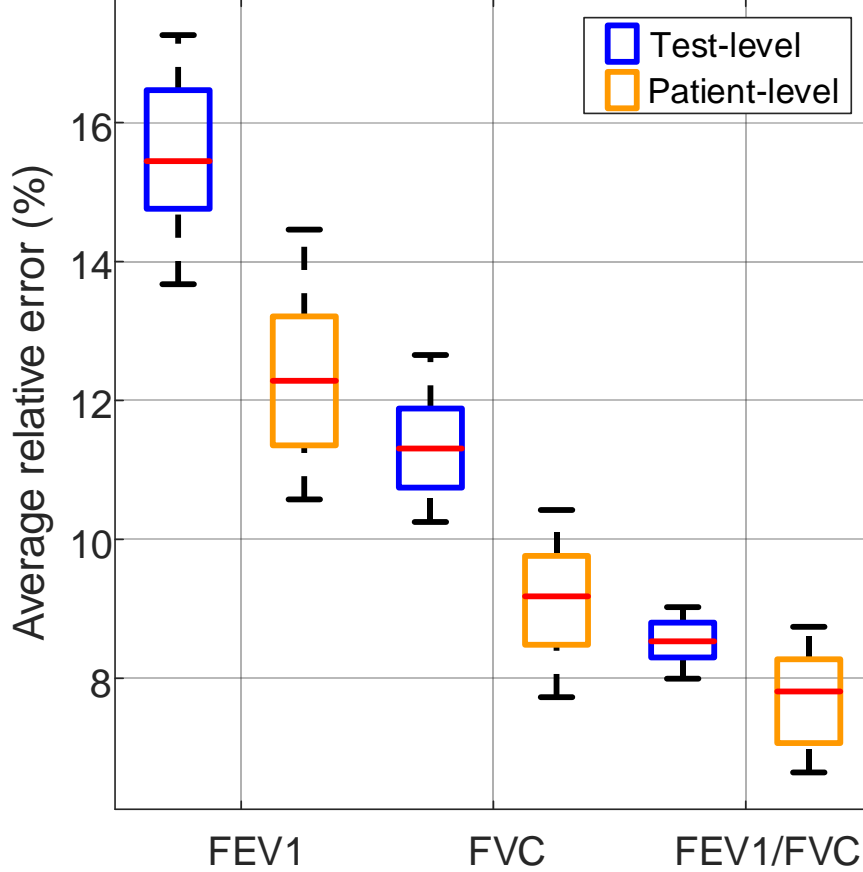


Figure 53: Average errors of estimating lung function indices (all in %pred)

The estimation error at patient level can be reduced to $12.52\% \pm 1.97\%$, $9.06\% \pm 1.39\%$, and $7.69\% \pm 1.69\%$ for FEV1, FVC and FEV1/FVC, respectively, by averaging the results from multiple spirometry tests of the same patient. Since a patient's spirometry tests are only valid if the variance of their results is smaller than 10% [27, 81], such patient-level estimation is more commonly used in practice. The accuracy of estimating FEV1/FVC, in particular, approximates that of in-clinic spirometry (around 5%). The system, hence, can be reliably used for disease tracking out of clinic.

In clinical practice, humans' lung function is highly relevant to their age, gender and the specific pulmonary disease they have. Hence, this work investigates the accuracy of lung function estimation over these different patient subgroups.

Age: This work divides patients into three age groups (8-10, 11-13 and 14-17), and numbers of patients in these groups are 34, 26 and 23, respectively. Results in Figure 54 show that the estimation is more accurate over older patients, and the error over age group of 14-17, in particular, has $>4\%$ less errors than other groups. The primary reason is that pediatric patients in low ages usually cannot maintain their bodies fully stationary during spirometry tests. In these cases, being different from in-clinic spirometry that measures patients' exhaling airflow through a mouthpiece, the system is more sensitive to patients' body motions and produces more errors. This is also exemplified by the ratio of valid spirometry tests in different age groups, which is 70.1% for the Age 14-17 group but only 61.9% for the Age 8-10 group. In older age groups, estimating FVC is more accurate compared to FEV1 estimation, because older patients have smaller variation of airway caliber with respect to lung size. It is hence easier to estimate their air volumes being exhaled [130].

Gender: Figure 55 shows that the system achieves higher estimation accuracy over females, and such difference could be up to 3%-4% for estimating FEV1/FVC. One possible reason is that girls' spirometry tests in the clinical study have a higher percentage (69.6%) to be valid, and this percentage is only 62.2% for boys' spirometry tests in the clinical study. Such difference is even more significant in low-age groups. Furthermore, from the clinical perspective, girls usually have wider but shorter airways in their childhood than boys, so that their momentary lung function indices (e.g., FEV1) could have less variation [45].

Disease: As shown in Figure 56, the system achieves better accuracy in estimating lung function of asthma patients, because asthma is predominantly a disease of airway obstruction. Estimation errors over cystic fibrosis (CF) patients, on the other hand, are higher and the difference could be $>5\%$. The main reason is that CF usually exhibits other symptoms such as infection and mucus accumulation, which affect lung function beyond the changes in airway mechanics. Besides, before hospitalization and treatment, CF patients tend to have markedly lower lung function on force exhaling. All these factors cause highly variant results on the FEV1 error, while FVC reflecting the entire breathing volume is less affected.

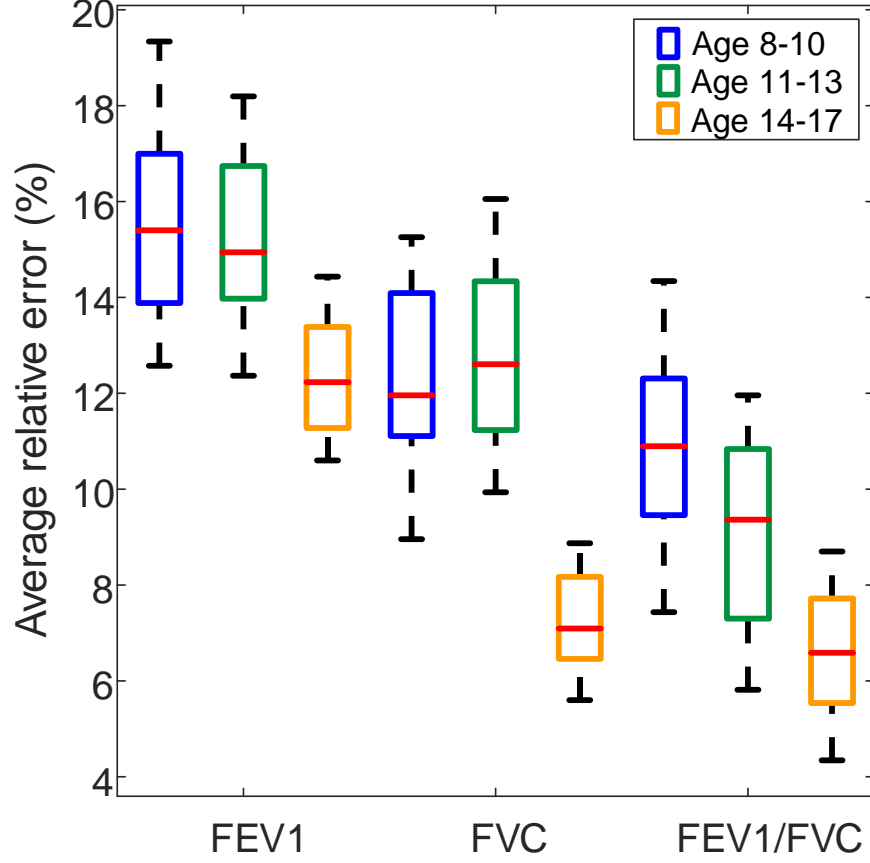


Figure 54: Different age subgroups

5.2.2 Muscle Fatigability Consistency

This work first evaluates the technique when being used to analyze the accumulation speed of muscle fatigue. The evaluation is based on a fact that bicep muscle would become fatigued faster when being exercised with heavier weights, so the system presets 3 weight options and ask the subject to complete at least 1 group of workouts for each weight option. The evaluation metric is whether the slope values become larger accordingly as the protocol increase the weights.

Meanwhile, the experiment also deploys an EMG device, as shown in Figure 57, to further validate such fact of accumulation speed and serve as the ground truth to evaluate this technique. In the scope of EMG analysis, muscle fatigue can be represented as the

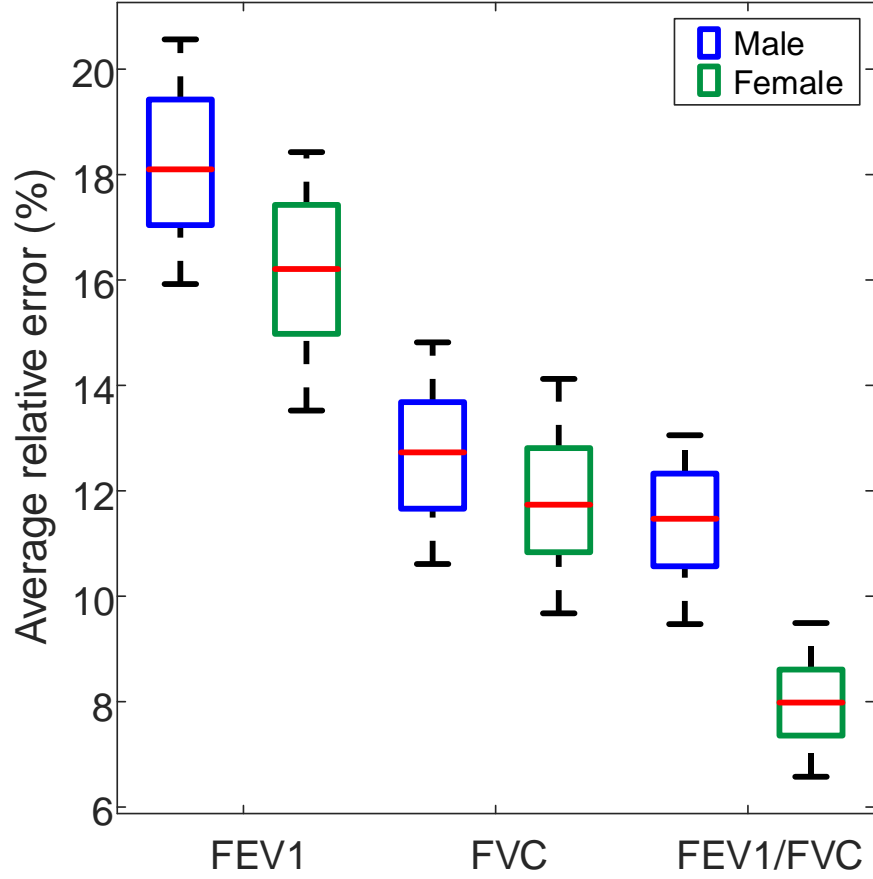


Figure 55: Different gender subgroups

integration of EMG signal (IEMG) while muscle contracts [110], namely the area within the EMG curve and time axis as shown in Figure 58. With similar insights of the system, this work performs the first order curve fitting over IEMG values measured in a group of workouts and use the slope value to represent the accumulation speed of fatigue.

Figure 59 shows the fitted lines of fatigue measurements for both EMG and this technique. The experiment includes 3 groups of workouts using weights of 9, 14 and 19 pounds respectively over the same subject. Because EMG has different measurement rationale with this technique, the scales of these values are different. However, from the correspondence between the slope value and actual weight being used, this work validates that fatigue accumulates faster as the protocol use heavier weights, which agrees with results measured by this technique as well.

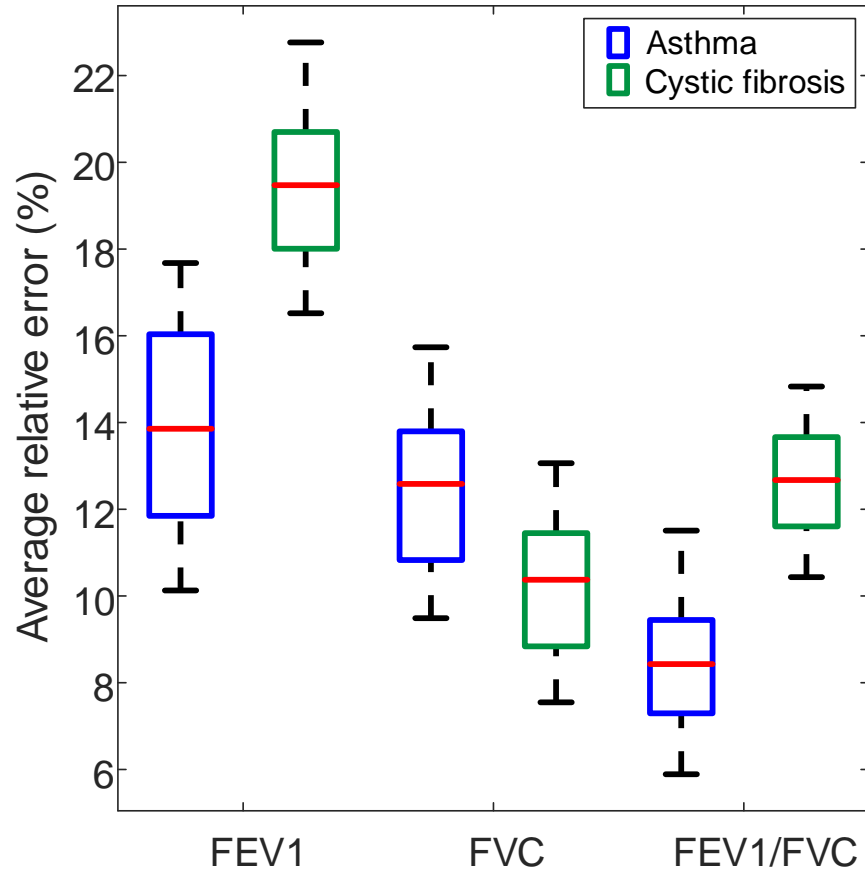


Figure 56: Different disease subgroups

Table 7: The speed of fatigue accumulation (slope value) over different subjects

Subject #	9 Pounds	14 Pounds	19 Pounds
1	1.28	1.51	3.64
2	0.42	0.95	2.34
3	1.02	1.62	3.22
4	0.52	0.46	0.65
5	1.09	1.37	2.92



Figure 57: Experimental setup with EMG device

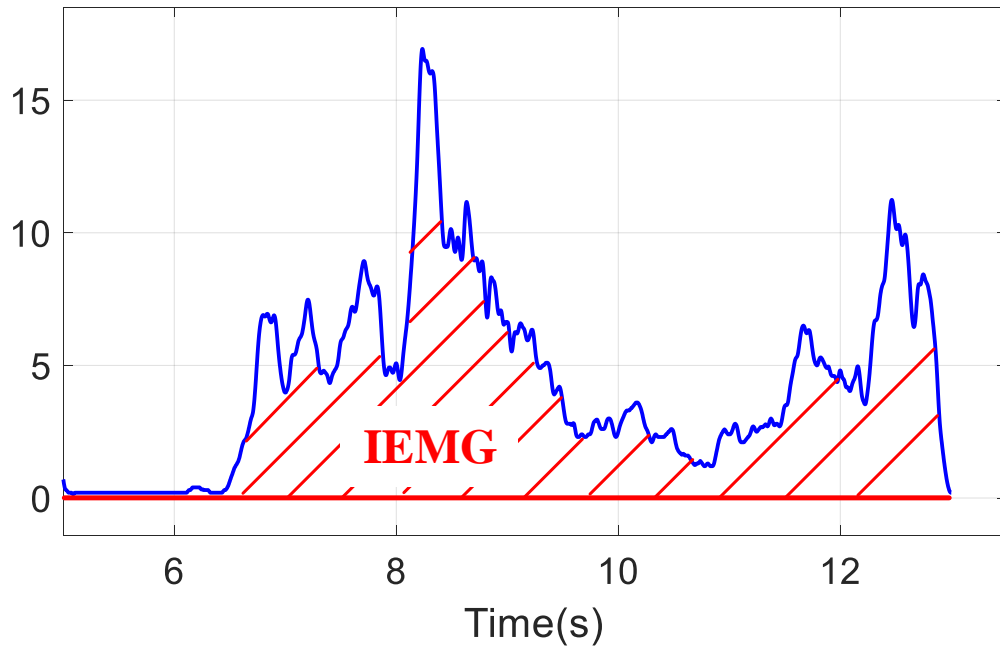


Figure 58: The integration of EMG signal (IEMG)

This work then evaluates the technique regarding its generality over more subjects with the same weight settings. Table 7 shows the slope values measured on all volunteers with different muscle strength, from which the same trend of increasing value can still be observed

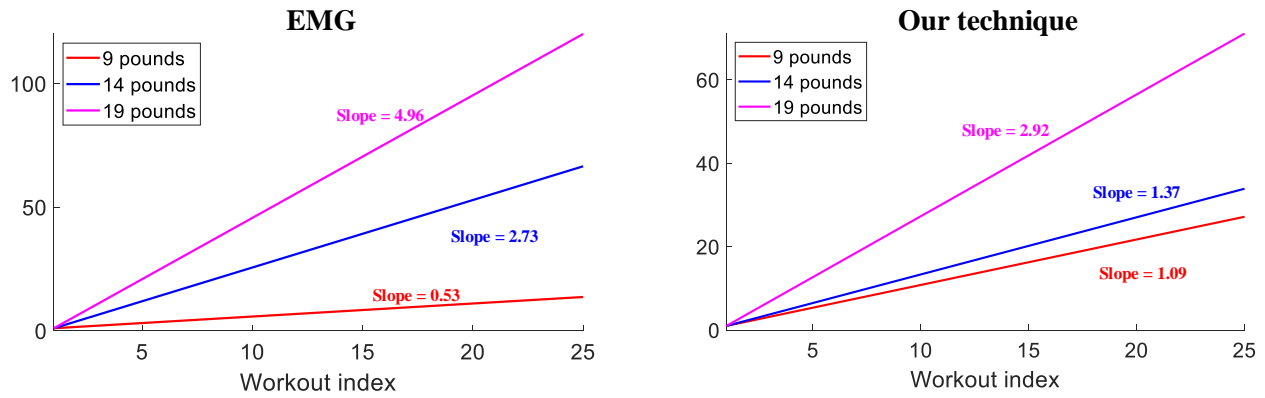


Figure 59: Slope values of 3 different weights measured by EMG and this technique

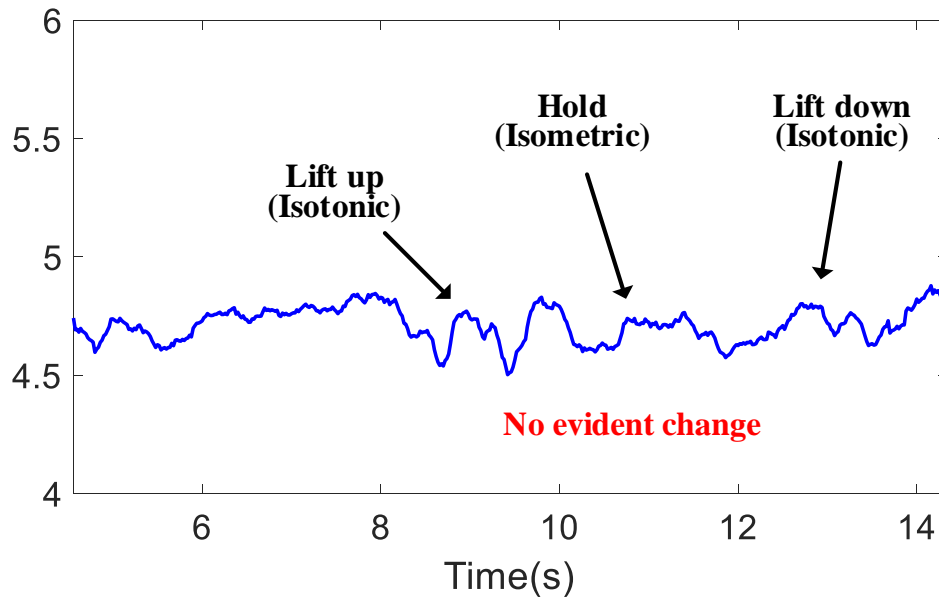


Figure 60: The data of subject #4 with weaker muscle

when the weight gets heavier. Noticeably, each individual tends to have a value distinct from all the others. It is considered that such results are caused by muscle development and strength of different individuals, especially for subject #4 whose measurements are quite similar with different weights. Due to less exercises and training compared with others, the bicep muscle of subject #4 is much weaker, which causes the monitored data exhibiting little difference between isometric and isotonic contractions, as shown in Figure 60. Unlike data collected from others that clearly shows the transition between these two phases such as the

waveform in Figure 29, subject #4's data hinders the estimation of muscle tremor as well as the fatigue monitoring. For the rest majority of people, this technique can effectively quantify the speed of fatigue accumulation.

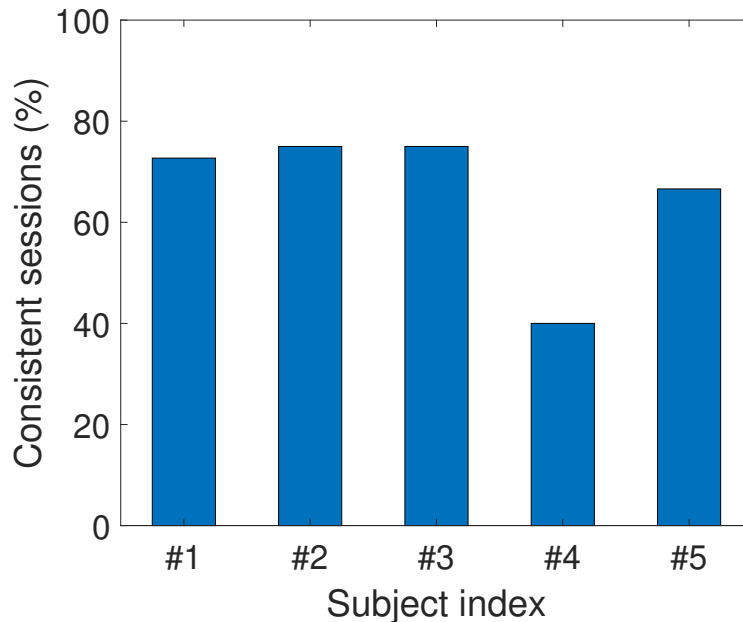


Figure 61: Consistent sessions (%) of each individual

To justify that the technique is effective and consistent with the nature of muscle fatigue, this work asks each subject to finish 3 groups (a session) of workouts using the same weight with an interval of one hour. Then, given that later groups are conducted based on initial tiredness led by previous group(s), this work considers that muscle fatigue, under the same weight, should occur earlier in later groups. As a result, the system defines a consistent test session as $G_1 > G_2 > G_3$, where G_i is the workout index of fatigue boundary in group i . Then, the evaluation metric is the percentage of consistent sessions among all test sessions being finished.

34 sessions of data are collected over all the five volunteers and 23 sessions (67.6%) achieve the aforementioned consistency after the identification of fatigue occurrence. Figure 61 presents the percentage of consistent sessions over 5 individuals, and the technique can achieve around 70% of consistency for the majority of participants. With the same reason for muscle weakness mentioned above, data of subject #4 produces only 40% of consistency. For all the other subjects, the ratio of consistency is generally higher than 75%.

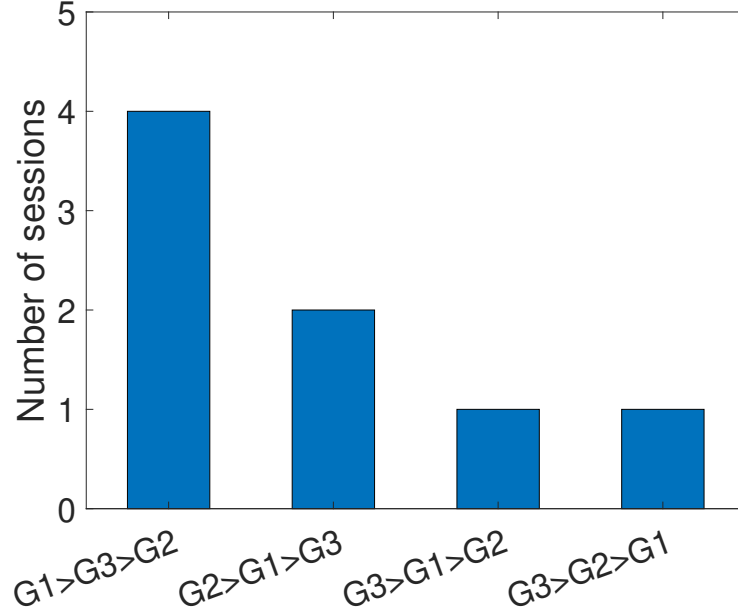


Figure 62: Cases of inconsistent sessions

For the 8 inconsistent sessions from all subjects except for subject #4, Figure 62 shows the orders of fatigue boundaries in detail. It is observed that only 1 session exhibits completely inverse order ($G3 > G2 > G1$), which is counter-intuitive for muscle fatigue. On the other hand, 4 out of 8 inconsistent sessions ranked the 1st group of workouts correctly but only mistakenly placed the 2nd and 3rd. Such inaccuracy between 2nd and 3rd groups is mainly caused by less evident cluster boundary after muscle exhaustion in the 1st group of that session. Since 1-hour rest interval may not be sufficient enough for fatigue recovery, during the 2nd or 3rd group of tests workouts are performed with exhausted muscle that gets tired quickly.

Overall, these results demonstrate that the workout index discovered by this technique is a potential indicator of actual fatigue occurrence, and the technique bears more confidence when being applied to muscle without initial tiredness.

5.2.3 Expression Recognition Accuracy

This work trains different models for each subject due to the heterogeneity of human facial muscle. The average accuracy of recognizing 7 categories of facial expressions over 5 student volunteers is shown in Figure 63, and such accuracy is above 80% for all 7 cate-

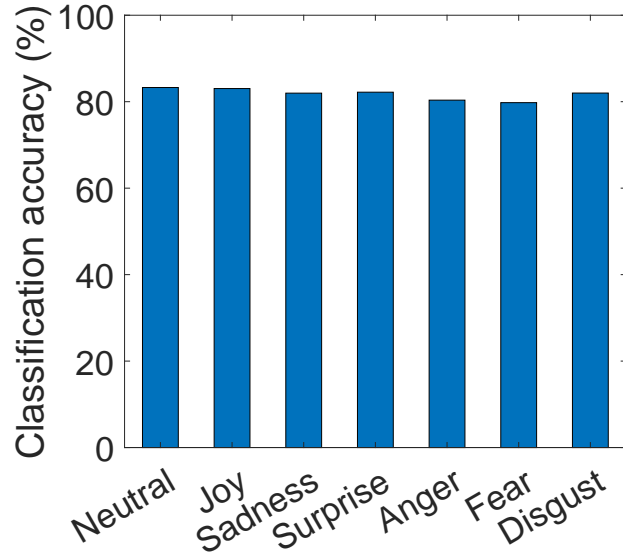


Figure 63: Average accuracy over all users

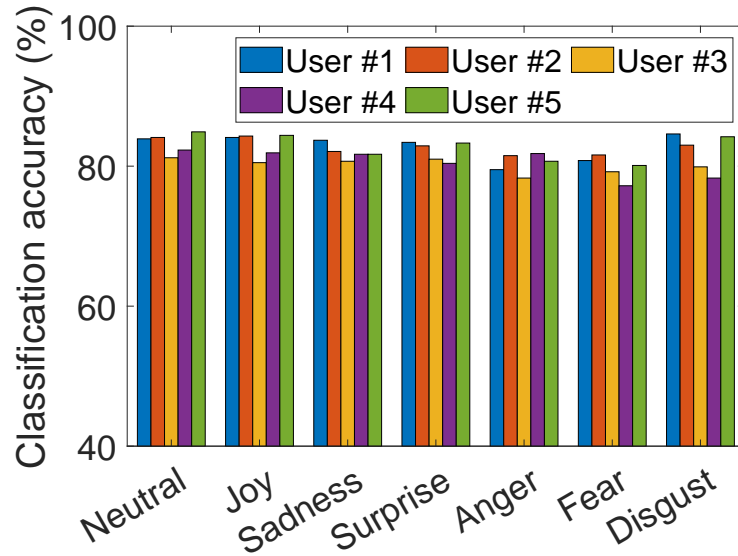


Figure 64: Accuracies of individual users

gories. Furthermore, Figure 64 shows the recognition accuracy over each individual student volunteer, and demonstrates that such accuracy varies little and retains at the same level over each student volunteer.

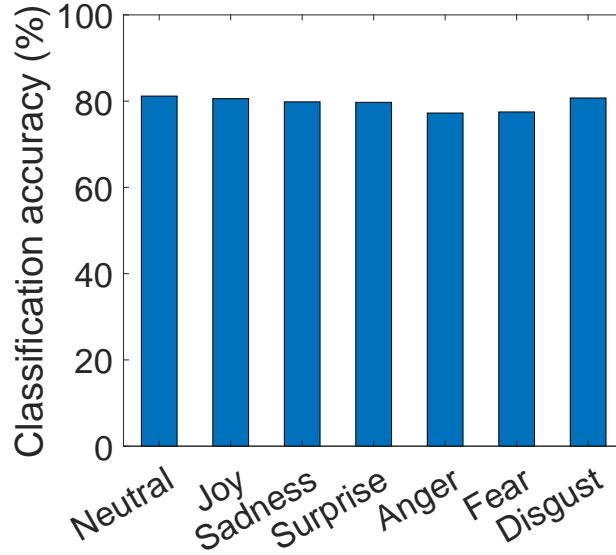


Figure 65: Recognition accuracy without consecutive frame dependency

Since the test above randomly selects data from a 10-minute session, there will be consecutive video frames or audio segments recorded from the same expression. Such dependency needs to be removed to evaluate the model consistency among different exhibitions of the same expression. To evaluate that, the 10-minute session is separated into beginning 8 minutes and ending 2 minutes. Data from the first 8-minute is used to train the model and last 2-minute is used to report the accuracy. With such a data selection, the test data will not come from the same time segment that is used to train the model. Figure 65 shows results of this data division approach, and the accuracy is comparable to that when the training sample is randomly selected.

Next, Figure 66 shows results of the leave-one-user-out model evaluation, in which this work trains the model using data collected from 4 users and report the recognition accuracy tested on the one left. The model still achieves 70% of accuracy when it is not tuned to specific user. This proves that, even without user-specific training, the teacher and student model can still work collaboratively to extract effective acoustic features correlated with facial expressions.

Furthermore, the confusion matrix of facial expression recognition is shown in Figure 67. Note that recognitions on the anger and fear expressions are less accurate when compared to those on other expressions. To investigate the reason of such inaccuracy, this work inspects

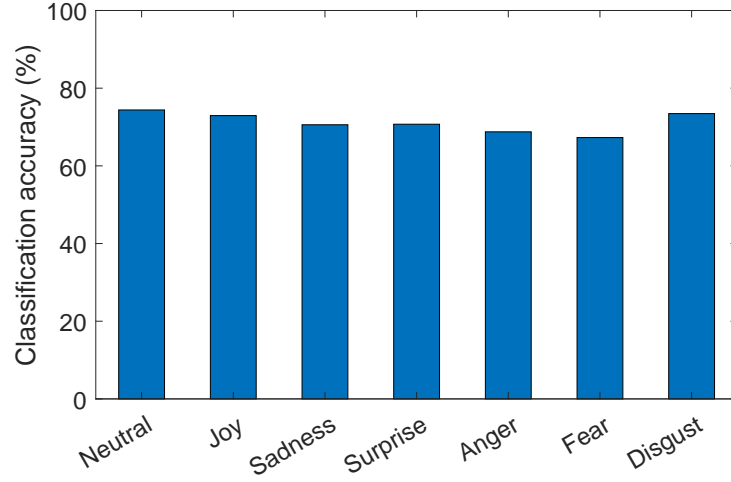


Figure 66: Leave-one-user-out accuracy of different facial expressions

Neutral	83.9	2.1	3.2	2.8	2.4	2.7	2.9
Joy	3.5	84.1	2.5	3.1	3.4	2.3	1.1
Sadness	2.1	2.6	83.7	2.8	2.9	3.1	2.8
Surprise	3.0	2.9	2.7	83.4	3.6	2.9	1.5
Anger	3.9	3.5	3.3	2.7	79.5	4.1	3.0
Fear	3.3	3.1	2.5	2.1	4.2	80.8	4.0
Disgust	2.5	2.7	2.8	3.3	2.0	2.1	84.6
	Neutral	Joy	Sadness	Surprise	Anger	Fear	Disgust

Predicted Class (%)

Figure 67: Confusion matrix of facial expression recognition

the acoustic channel estimations corresponding to each category of facial expressions, and found that these two categories of facial expressions result in larger variations of the acoustic channel, as shown in Table 8. The reason of such large variance is that maintaining the fear and anger expressions is more difficult for the users due to the involvement of more groups of facial muscles. As a result, different samples of these expressions may be largely heterogeneous and hence affect correct recognition of these expressions.

Table 8: Channel estimation variance of different categories of facial expressions

Expressions	N	J	Sa	Su	A	F	D
Variance	1.39	1.78	1.67	2.31	5.17	5.89	2.75

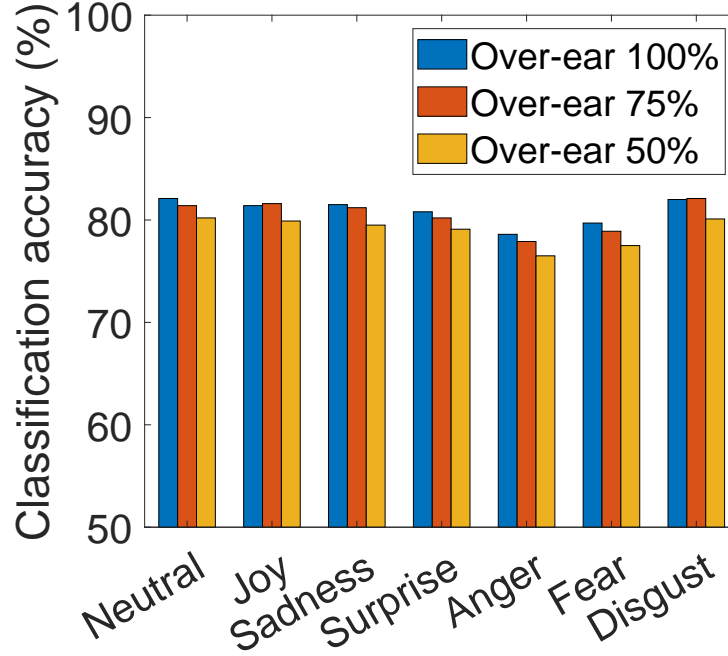


Figure 68: Over-ear headphone (The impact of signal strength)

The recognition accuracy depends on the strength of ultrasound signal being received by the transformed microphone. This work evaluates the impact of such signal strength with different volumes of the transmitted ultrasound, but the results in Figure 68 show that even when the sound volume at the transmitting speaker is reduced to 50% of the maximum, the recognition accuracy only slightly drops by 1.5%. In reality, as the transmit ultrasound is inaudible to most people, maximum volume can always be used in most cases.

The sound leakage from commodity headphones is the basic motivation of the system design. Since such leakage varies over different frequency bands, this work evaluates the system performance by transmitting acoustic signal at different frequency bands as shown in

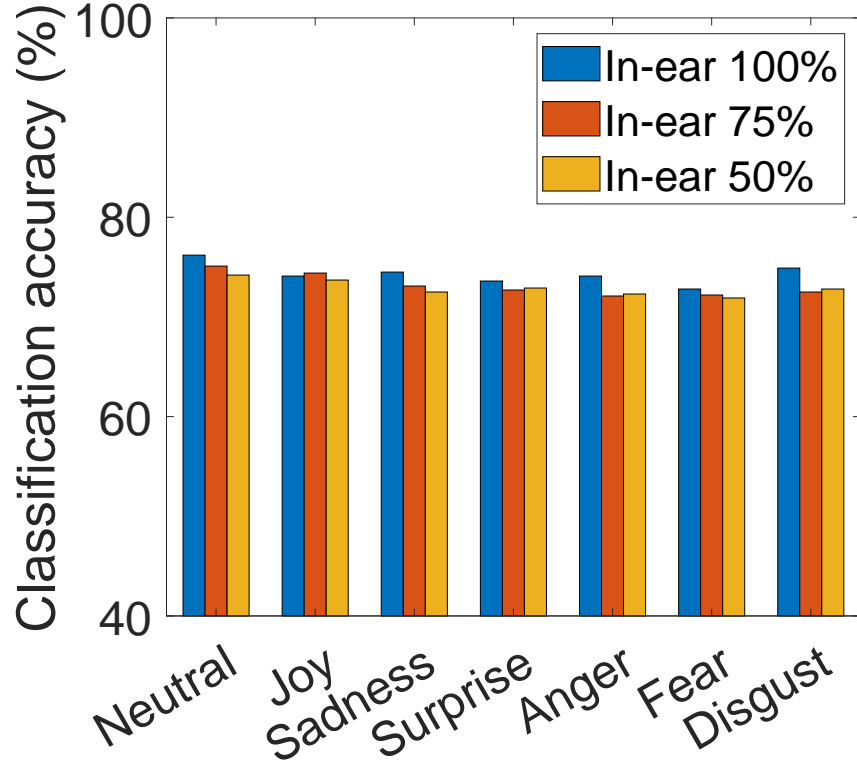


Figure 69: In-ear headphone (The impact of signal strength)

Figure 70. The results demonstrate that the system has the similar recognition accuracy over different ultrasound bands, but has slightly lower accuracy when the transmitted signal falls closer to the audible sound (e.g., around 15 kHz). Such accuracy drop is mainly because that commodity headphones are designed suppress the sound leakage within the audible bands.

The system transmits the ultrasound signal along an one-way direction because the re-wiring only transforms one speaker into a microphone. Given that some facial expressions may be asymmetric (e.g., contempt and fear as shown in Figure 10), the direction of ultrasound propagation may affect the accuracy of recognizing facial expressions. This work evaluates the impact of such direction of sound propagation. As shown in Figure 72, the impact of different propagation directions produce negligible impacts on the recognition accuracy. The main reason is that facial expressions are recognized based on estimations of the acoustic channel that covers the entire human face, and such a channel hence remains the same in both directions of signal propagation.

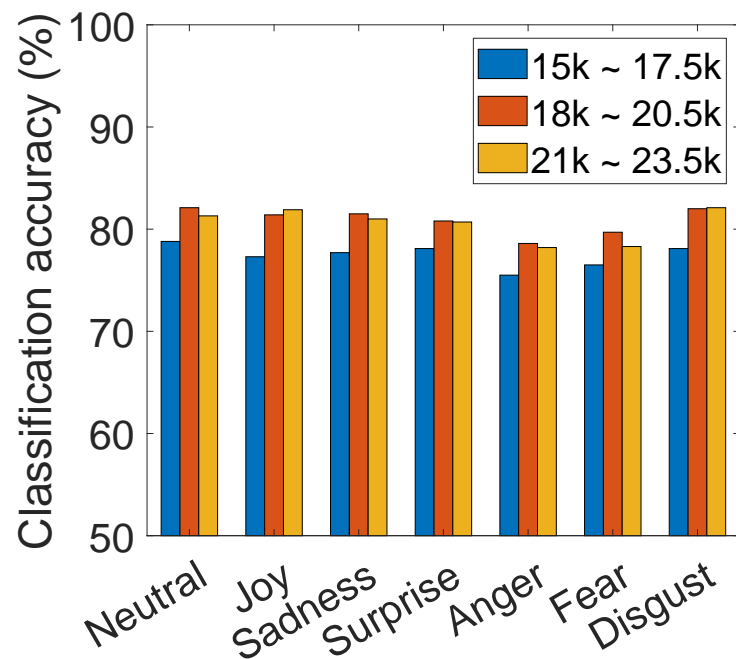


Figure 70: Over-ear headphone (The impact of signal frequency band)

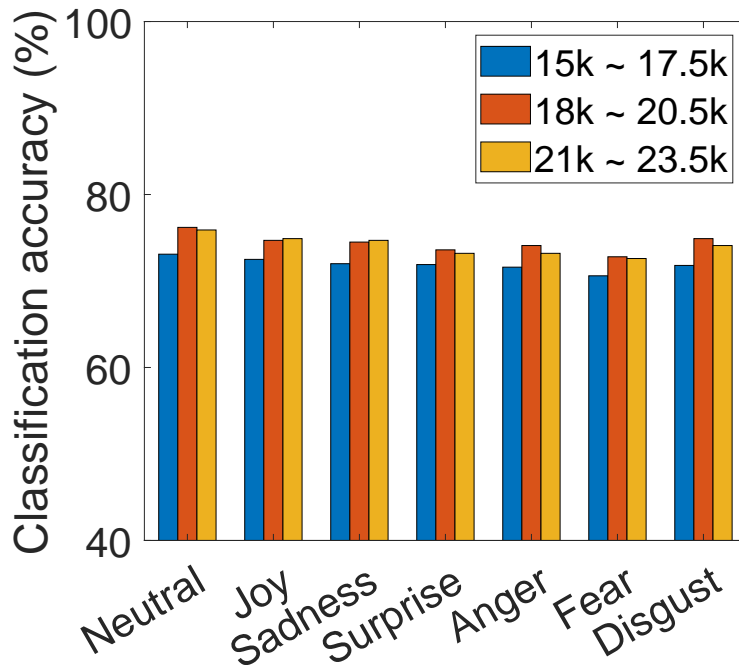


Figure 71: In-ear headphone (The impact of signal frequency band)

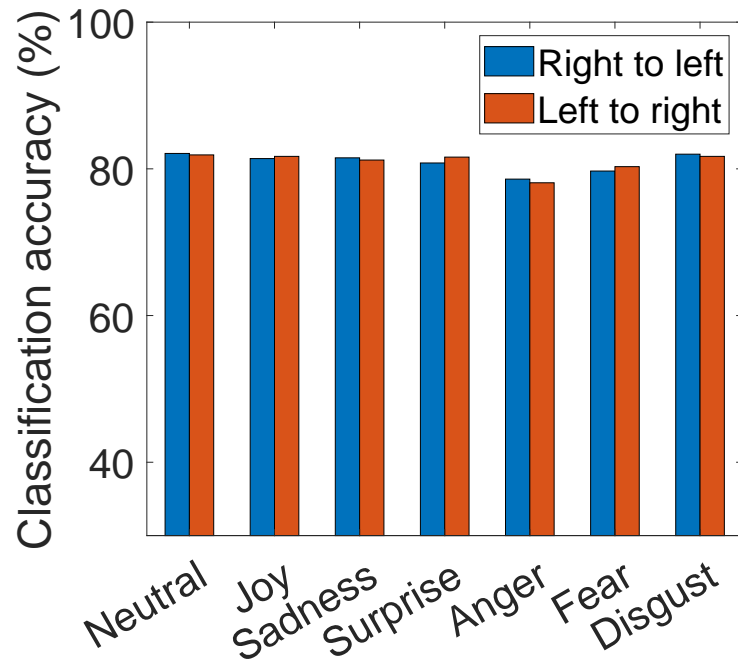


Figure 72: Over-ear headphone (The impact of signal propagation direction)

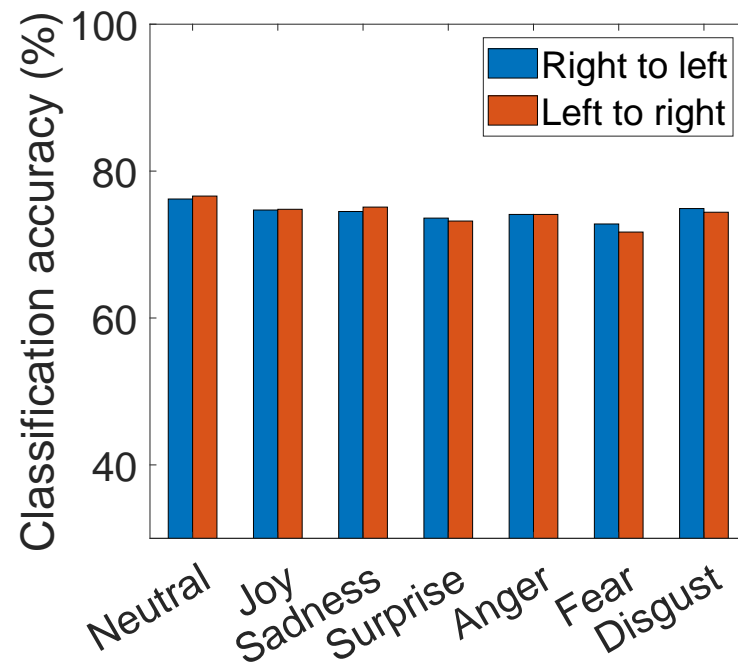


Figure 73: In-ear headphone (The impact of signal propagation direction)

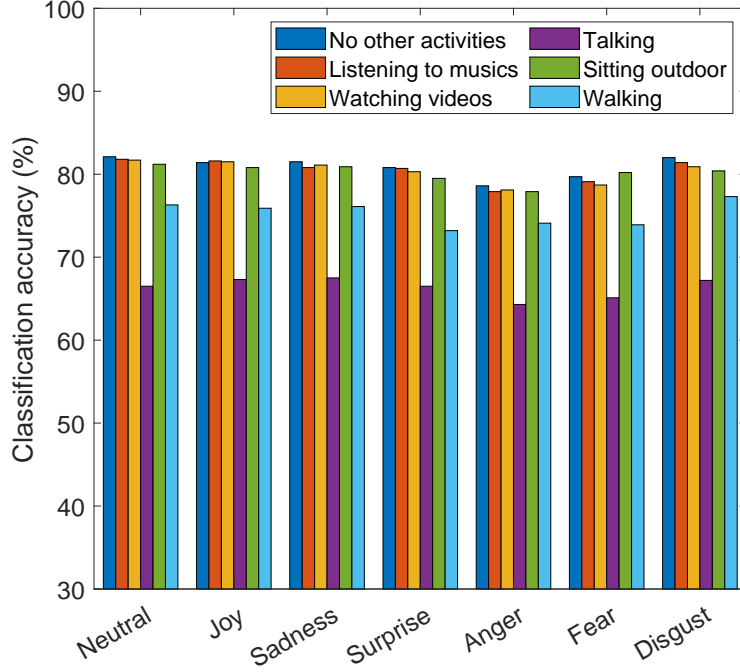


Figure 74: Over-ear headphone (Impact of different user activities)

As stated above, the system is designed to recognize the user’s facial expressions with other concurrent user activities. This work evaluates the impact of different user activities on the recognition accuracy, and evaluation results in Figure 74 show that the system retains high accuracy of recognizing facial expressions for most types of user activities, even when the user is using the headphone to listen to music or watch videos that mainly involve sounds in audible bands. To verify its practical application, the system also tests outdoor scenarios including sitting on a bench and walking. Because most outdoor noise can be filtered out during preprocessing, it can still achieve similar accuracy. Besides, a slight performance drop is noticed when the user is walking, such difference is caused by human walking motion may add movements on the receiver side of the headphone, thus impact the channel estimation results. However, the system will experience 15% accuracy drop if the user is talking while using the system, because of the extra movements of relevant facial muscles. Comparatively, it is also verified that vision-based recognition methods [122] will also experience 10-15% accuracy drop in similar situations, hence demonstrating the competitiveness to these existing methods.

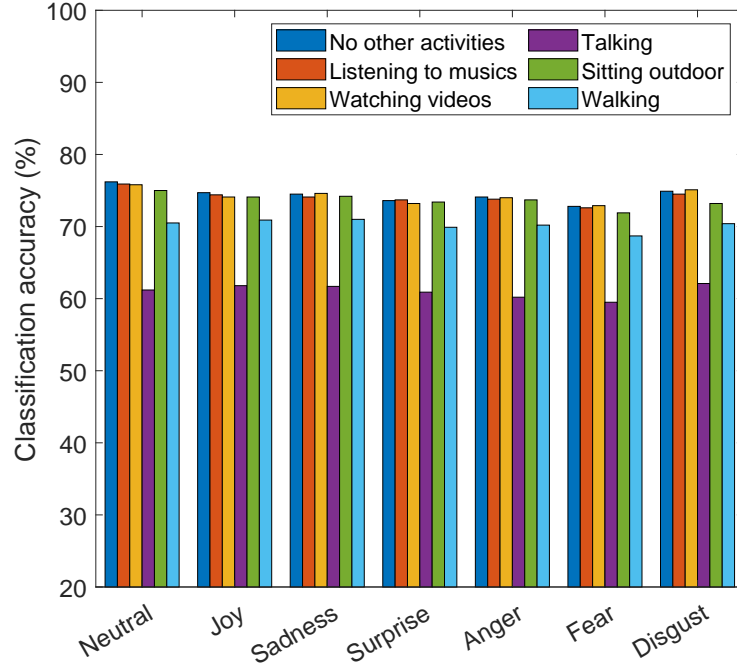


Figure 75: In-ear headphone (Impact of different user activities)

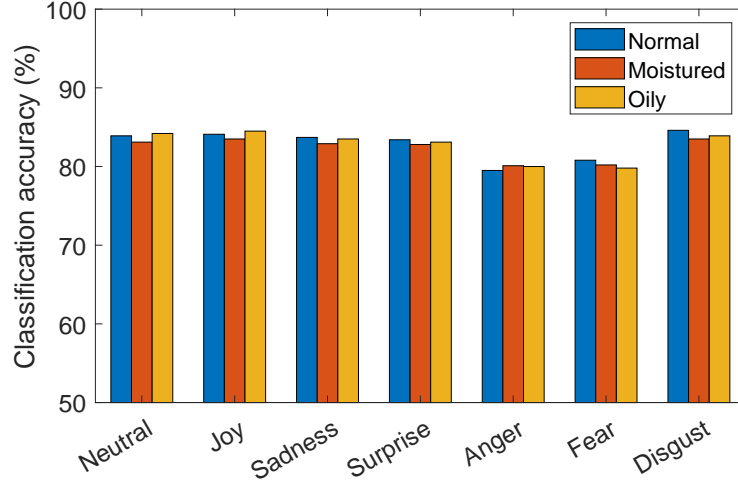


Figure 76: Impact of different skin conditions

Since the transmitted ultrasound signal propagates closely above the facial surface, its propagation may also be impacted by different skin conditions. To evaluate whether such factor can make an impact, this work trains the model with data only collected on normal

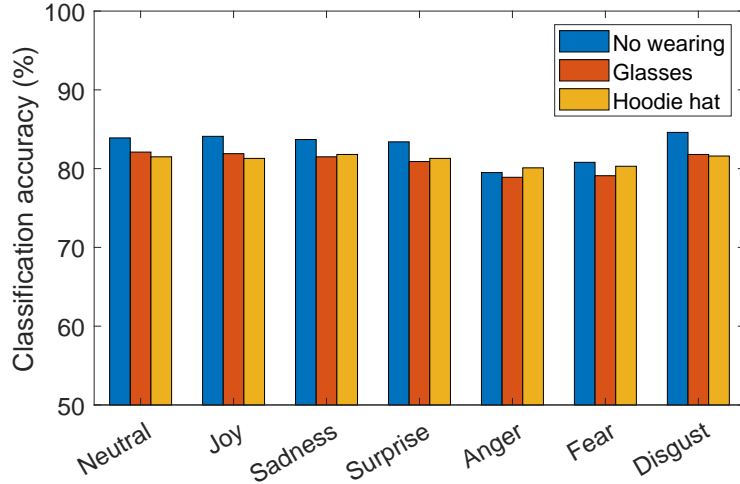


Figure 77: Impact of head accessories and clothes

faces, and test the system performance using data from other skin conditions (without any re-training or fine-tuning on these new data). The system emulates these skin conditions by applying cream and water onto the user face, and the accuracy of facial expression recognition, as shown in Figure 76, remains nearly the same under these different skin conditions.

Lastly, this work evaluates some external wearing objects that can impact the system’s practical usefulness. Figure 77 shows the recognition accuracy when a user wears glasses or a hoodie hat. Objects being worn may potentially change the acoustic channel characteristics by incorporating additional signal propagations. However, these results demonstrate that wearing accessories or clothes produces little impacts on the recognition accuracy. Essentially, as long as the accessory does not cover the facial skin (e.g., masks) to hide the face contour, the system can retain the recognition accuracy with little performance loss.

6.0 Conclusions

This dissertation proposes several smart health applications based on active acoustic sensing. In general, its research goals include spirometry test for human lung function estimation, the evaluation of bicep muscle fatigability and facial expressions recognition that can be used to infer human emotion well-beings.

Chapter 2 describes current research work related to the dissertation. Mobile health systems recently demonstrate their success out of the clinic based on the robustness of active motion tracking techniques. Compared to medical equipment with dedicated sensors, smart health applications using generic sensors not only require biomarkers that are not only correlated with certain health metrics, but also non-intrusively measurable to ensure practical usability.

Chapter 3 describes the medical background of all topics discussed in the thesis. It introduces spirometry and lung function indices with their clinical protocols, muscle contraction and force generation mechanisms and the correlations between facial expressions and facial muscle. These knowledge helps the development of methodologies to implement the smart health applications.

Chapter 4 describes the rationale of this work that is the active acoustic sensing techniques derived from motion tracking applications. Such active sensing is described as the speaker transmitting a pre-defined signal, then the signal propagates and will be received by the microphone. During the signal propagation, the physical variation of desired biomarker can affect signal characteristics due to impact of reflection and attenuation. From the received signal, these information can be uncovered to quantify the variation of these biomarkers. As a result, the received signal can be correlated with certain health evaluation metrics. In this chapter, three independent applications are presented to demonstrate how active sensing can be used on different biomarkers and how the system setup adapts to these applications. The first application uses the smartphone to conduct lung function spirometry test. Since it measures the biomarker of chest wall motion, bottom speaker and bottom mic are chosen to ensure better directional propagation pathway. The second application

quantifies muscle fatigue to monitor muscular disorders. It uses muscle tremor as an implicit biomarker, so bottom mic and top speaker are selected to ensure larger area to be captured by the acoustic signal. Finally, to recognize human facial expressions from the biomarker of facial muscle deformation, existing mobile devices cannot meet the sensing requirements of transmitting signals above the facial skin, so the setup is modified accordingly by rewiring commodity headphones.

Chapter 5 describes the evaluation setup for all smart health applications discussed in this thesis. In general, all applications work well on the recruited subject group. Specifically, the spirometry on smartphone has been clinically tested on pediatric asthma patients. To prove the practical usability of these systems, this chapter also includes test results with different influential factors that will impair the system performance. To summarize, all systems can work reliably and accurately in casual practice compared with medical equipment.

Bibliography

- [1] Robab Abdolkhani, Kathleen Gray, Ann Borda, and Ruth DeSouza. Patient-generated health data management and quality challenges in remote patient monitoring. *JAMIA open*, 2(4):471–478, 2019.
- [2] JH Allum, V Dietz, and HJ Freund. Neuronal mechanisms underlying physiological tremor. *Journal of Neurophysiology*, 41(3):557–571, 1978.
- [3] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 679–689, 2017.
- [4] Corrado Angelini and Elisabetta Tasca. Fatigue in muscular dystrophies. *Neuromuscular Disorders*, 22:S214–S220, 2012.
- [5] Md Tanvir Islam Aumi, Sidhant Gupta, Mayank Goel, Eric Larson, and Shwetak Patel. Doplink: using the doppler effect for multi-device interaction. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 583–586, 2013.
- [6] Katayoun Bahadori, Mary M Doyle-Waters, Carlo Marra, Larry Lynd, Kadria Alasaly, John Swiston, and Js Mark FitzGerald. Economic burden of asthma: a systematic review. *BMC pulmonary medicine*, 9(1):24, 2009.
- [7] Mirza Mansoor Baig, Hamid GholamHosseini, and Martin J Connolly. Mobile health-care applications: system design review, critical issues and challenges. *Australasian physical & engineering sciences in medicine*, 38(1):23–38, 2015.
- [8] Ananta Narayanan Balaji, Chen Yuan, Bo Wang, Li-Shiuan Peh, and Huilin Shao. ph watch-leveraging pulse oximeters in existing wearables for reusable, real-time monitoring of ph in sweat. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 262–274, 2019.
- [9] Surya P Bhatt, Pallavi P Balte, Joseph E Schwartz, Patricia A Cassano, David Couper, David R Jacobs, Ravi Kalhan, George T O’Connor, Sachin Yende, Jason L Sanders, et al. Discriminative accuracy of fev1: Fvc thresholds for copd-related hospitalization and mortality. *Jama*, 321(24):2438–2447, 2019.

- [10] Homer A Boushey, Christine A Sorkness, Tonya S King, Sean D Sullivan, John V Fahy, Stephen C Lazarus, Vernon M Chinchilli, Timothy J Craig, Emily A Dimango, Aaron Deykin, et al. Daily versus as-needed corticosteroids for mild persistent asthma. *New England Journal of Medicine*, 352(15):1519–1528, 2005.
- [11] PLP Brand and RJ Roorda. Usefulness of monitoring lung function in asthma. *Archives of disease in childhood*, 88(11):1021–1025, 2003.
- [12] GB Brookes and AJ Fairfax. Chronic upper airway obstruction: value of the flow volume loop examination in assessment and management. *Journal of the Royal Society of Medicine*, 75(6):425, 1982.
- [13] V Brusasco. Usefulness of peak expiratory flow measurements: is it just a matter of instrument accuracy? *Thorax*, 58(5):375–376, 2003.
- [14] Nam Bui, Anh Nguyen, Phuc Nguyen, Hoang Truong, Ashwin Ashok, Thang Dinh, Robin Deterding, and Tam Vu. Pho2: Smartphone based blood oxygen level measurement systems using near-ir and red wave-guided light. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pages 1–14, 2017.
- [15] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. ebp: A wearable system for frequent and comfortable blood pressure monitoring from user’s ear. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–17, 2019.
- [16] PMA Calverley. The clinical usefulness of spirometric information. *Breathe*, 5(3):214–220, 2009.
- [17] William D. Carey. *Current Clinical Medicine 2nd Edition*. Saunders, Cleveland, OH, 2010.
- [18] Po-Hsuan Cameron Chen, Yun Liu, and Lily Peng. How to develop machine learning models for healthcare. *Nature materials*, 18(5):410, 2019.
- [19] Tuochao Chen, Benjamin Steeper, Kinan Alsheikh, Songyun Tao, François Guimbretière, and Cheng Zhang. C-face: Continuously reconstructing facial expressions by deep learning contours of the face with ear-mounted miniature cameras. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 112–125, 2020.

- [20] Christopher Christodoulou. The assessment and measurement of fatigue. *Fatigue as a window to the brain*, pages 19–35, 2005.
- [21] Emma Ciafaloni, Deborah J Fox, Shree Pandya, Christina P Westfield, Soman Puzhankara, Paul A Romitti, Katherine D Mathews, Timothy M Miller, Dennis J Matthews, Lisa A Miller, et al. Delayed diagnosis in duchenne muscular dystrophy: data from the muscular dystrophy surveillance, tracking, and research network (md starnet). *The Journal of pediatrics*, 155(3):380–385, 2009.
- [22] Mario Cifrek, Vladimir Medved, Stanko Tonković, and Saša Ostojić. Surface emg based muscle fatigue evaluation in biomechanics. *Clinical biomechanics*, 24(4):327–340, 2009.
- [23] A Czaplinski, AA Yen, and Stanley H Appel. Forced vital capacity (fvc) as an indicator of survival and disease progression in an als clinic population. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(3):390–392, 2006.
- [24] Fernanda de Cordoba Lanza, Anderson Alves de Camargo, Lilian Rocha Ferraz Archija, Jessyca Pachi Rodrigues Selman, Carla Malaguti, and Simone Dal Corso. Chest wall mobility is related to respiratory muscle strength and lung volumes in healthy subjects. *Respiratory care*, 58(12):2107–2112, 2013.
- [25] Sophie Debouche, Laurent Pitance, Annie Robert, Giuseppe Liistro, and Gregory Rey-chler. Reliability and reproducibility of chest wall expansion measurement in young healthy adults. *Journal of manipulative and physiological therapeutics*, 39(6):443–449, 2016.
- [26] Shubhada Deshmukh, Manasi Patwardhan, and Anjali Mahajan. Survey on real-time facial expression recognition techniques. *Iet Biometrics*, 5(3):155–163, 2016.
- [27] Paul Enright, WM Vollmer, B Lamprecht, R Jensen, A Jithoo, W Tan, M Studnicka, P Burney, S Gillespie, and A Sonia Buist. Quality of spirometry tests performed by 9893 adults in 14 countries: the bold study. *Respiratory medicine*, 105(10):1507–1515, 2011.
- [28] Paul L Enright, Gwen S Skloot, Jean M Cox-Ganser, Iris G Udasin, and Robin Herbert. Quality of spirometry performed by 13,599 participants in the world trade center worker and volunteer medical screening program. *Respiratory Care*, 55(3):303–309, 2010.

- [29] Xiaoran Fan, Longfei Shangguan, Siddharth Rupavatharam, Yanyong Zhang, Jie Xiong, Yunfei Ma, and Richard E Howard. Headfi: bringing intelligence to all head-phones. In *MobiCom*, pages 147–159, 2021.
- [30] Ying Fan and Chenyu Yang. Competition, product proliferation, and welfare: A study of the us smartphone market. *American Economic Journal: Microeconomics*, 12(2):99–134, 2020.
- [31] National Center for Environmental Health. Most recent national asthma data, 2017.
- [32] Jan Gajewski. Fatigue-induced changes in tremor caused by physical efforts of different volume and intensity. *Acta of Bioengineering and Biomechanics*, 8(2):103, 2006.
- [33] Sc C Gandevia. Spinal and supraspinal factors in human muscle fatigue. *Physiological reviews*, 81(4):1725–1789, 2001.
- [34] JJ Gilmartin and GJ Gibson. Abnormalities of chest wall motion in patients with chronic airflow obstruction. *Thorax*, 39(4):264–271, 1984.
- [35] Anne M Gilroy. *Anatomy: an essential textbook*. 2013.
- [36] Atefeh Goshvarpour, Ataollah Abbasi, and Ateke Goshvarpour. An accurate emotion recognition system using ecg and gsr signals and matching pursuit method. *Biomedical journal*, 40(6):355–368, 2017.
- [37] Ian Gregg and AJ Nunn. Peak expiratory flow in normal subjects. *Br Med J*, 3(5874):282–284, 1973.
- [38] Anna Gruebler and Kenji Suzuki. Measurement of distal emg signals using a wearable device for reading facial expressions. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 4594–4597. IEEE, 2010.
- [39] Ngoc Ha. *Smartphone industry: The new era of competition and strategy*. 2016.
- [40] Ju Han and Bir Bhanu. Human activity recognition in thermal infrared imagery. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)-Workshops*, pages 17–17. IEEE, 2005.

- [41] Richard A Harrigan, Theodore C Chan, and William J Brady. Electrocardiographic electrode misplacement, misconnection, and artifact. *The Journal of emergency medicine*, 43(6):1038–1044, 2012.
- [42] Matthew J Hegewald, Heather M Gallo, and Emily L Wilson. Accuracy and quality of spirometry in primary care offices. *Annals of the American Thoracic Society*, 13(12):2119–2124, 2016.
- [43] Matthew J Hegewald, Michael J Lefor, Robert L Jensen, Robert O Crapo, Stephen B Kritchevsky, Catherine L Haggerty, Douglas C Bauer, Suzanne Satterfield, Tamara Harris, et al. Peak expiratory flow is not a quality indicator for spirometry: peak expiratory flow variability and fev1 are poorly correlated in an elderly population. *Chest*, 131(5):1494–1499, 2007.
- [44] Javier Hernandez, Mohammed Hoque, Will Drevo, and Rosalind W Picard. Mood meter: counting smiles in the wild. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 301–310, 2012.
- [45] M Hibbert, A Lannigan, J Raven, L Landau, and P Phelan. Gender differences in lung growth. *Pediatric pulmonology*, 19(2):129–134, 1995.
- [46] Md Mahbub Hossain, Samia Tasnim, Abida Sultana, Farah Faizah, Hoimonty Mazumder, Liye Zou, E Lisako J McKyer, Helal Uddin Ahmed, and Ping Ma. Epidemiology of mental health problems in covid-19: a review. *F1000Research*, 9, 2020.
- [47] Bitu Houshmand and Naimul Mefraz Khan. Facial expression recognition under partial occlusion from virtual reality headsets based on transfer learning. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pages 70–75. IEEE, 2020.
- [48] Chen-Yu Hsu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. Zero-effort in-home sleep and insomnia monitoring using radio signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–18, 2017.
- [49] Qionghao Huang, Changqin Huang, Xizhe Wang, and Fan Jiang. Facial expression recognition with grid-wise attention and visual transformer. *Information Sciences*, 580:35–54, 2021.

- [50] Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, Xufei Mao, Panlong Yang, and Yunhao Liu. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *IEEE INFOCOM*, pages 370–378, 2014.
- [51] Earnest Paul Ijjina, Goutham Kanahasabai, and Aniruddha Srinivas Joshi. Deep learning based approach to detect customer age, gender and expression in surveillance video. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6. IEEE, 2020.
- [52] IQVIA Institute. The growing value of digital health evidence and impact on human health and the healthcare system, 2017.
- [53] Anil K Jain and Stan Z Li. *Handbook of face recognition*, volume 1. Springer, 2011.
- [54] Ahmad Jalal, Md Zia Uddin, and T-S Kim. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE Transactions on Consumer Electronics*, 58(3):863–871, 2012.
- [55] Kerri A Johansson, Eric Vittinghoff, Julie Morisset, Joyce S Lee, John R Balmes, and Harold R Collard. Home monitoring improves endpoint efficiency in idiopathic pulmonary fibrosis. *European Respiratory Journal*, 50(1):1602406, 2017.
- [56] David A Jones and Joan M Round. *Skeletal muscle in health and disease: a textbook of muscle physiology*. Manchester University Press, 1990.
- [57] Samira Ebrahimi Kahou, Christopher Pal, Xavier Bouthillier, Pierre Froumenty, Çağlar Gülçehre, Roland Memisevic, Pascal Vincent, Aaron Courville, Yoshua Bengio, Raul Chandias Ferrari, et al. Combining modality specific deep neural networks for emotion recognition in video. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 543–550, 2013.
- [58] Hideo Kaneko and Jun Horie. Breathing movements of the chest and abdominal wall in healthy subjects. *Respiratory care*, 57(9):1442–1451, 2012.
- [59] Hideo Kaneko, Shuichi Shiranita, Jun Horie, and Shinichiro Hayashi. Reduced chest and abdominal wall mobility and their relationship to lung function, respiratory muscle strength, and exercise tolerance in subjects with copd. *Respiratory care*, 61(11):1472–1480, 2016.

- [60] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, 22(1):98–107, 2017.
- [61] HA Kerstjens, Bert Rijcken, Jan P Schouten, and Dirkje S Postma. Decline of fev1 by age and smoking status: facts, figures, and fallacies. *Thorax*, 52(9):820, 1997.
- [62] Peter Lange, Bartolome Celli, Alvar Agustí, Gorm Boje Jensen, Miguel Divo, Rosa Faner, Stefano Guerra, Jacob Louis Marott, Fernando D Martinez, Pablo Martinez-Camblor, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *New England Journal of Medicine*, 373(2):111–122, 2015.
- [63] Amanda Lazar, Christian Koehler, Theresa Jean Tanenbaum, and David H Nguyen. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 635–646, 2015.
- [64] Gregory F Lewis, Rodolfo G Gatto, and Stephen W Porges. A novel method for extracting respiration rate and relative tidal volume from infrared thermography. *Psychophysiology*, 48(7):877–887, 2011.
- [65] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. Facial performance sensing head-mounted display. *ACM Transactions on Graphics (ToG)*, 34(4):1–9, 2015.
- [66] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 2020.
- [67] Tianhong Li, Lijie Fan, Mingmin Zhao, Yingcheng Liu, and Dina Katabi. Making the invisible visible: Action recognition through walls and occlusions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 872–881, 2019.
- [68] Jie Lian, Xu Yuan, Ming Li, and Nian-Feng Tzeng. Fall detection via inaudible acoustic sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–21, 2021.
- [69] Giuseppe Liistro, Carl Vanwelde, Walter Vincken, Jan Vandevoorde, Geert Verleden, Johan Buffels, et al. Technical and functional assessment of 10 office spirometers. *Chest*, 130(3):657–665, 2006.

- [70] Olof Lippold. The tremor in fatigue. In *Ciba Found Symp*, volume 82, pages 234–248. Wiley Online Library, 1981.
- [71] Jian Liu, Yan Wang, Yingying Chen, Jie Yang, Xu Chen, and Jerry Cheng. Tracking vital signs during sleep leveraging off-the-shelf wifi. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 267–276, 2015.
- [72] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 94–101. IEEE, 2010.
- [73] Zhihan Lv, Liang Qiao, and Qingjun Wang. Cognitive robotics on 5g networks. *ACM Transactions on Internet of Things*, 21(4):1–18, 2021.
- [74] Aicha Maalej and Ilhem Kallel. Does keystroke dynamics tell us about emotions? a systematic literature review and dataset construction. In *2020 16th International Conference on Intelligent Environments (IE)*, pages 60–67. IEEE, 2020.
- [75] Wenguang Mao, Jian He, and Lili Qiu. CAT: high-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 69–81, 2016.
- [76] Sameer K Mathur and William W Busse. Asthma: diagnosis and management. *Medical Clinics*, 90(1):39–60, 2006.
- [77] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. Earfieldsensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1911–1922, 2017.
- [78] Michael J McGeachie, Katherine P Yates, Xiaobo Zhou, Feng Guo, Alice L Sternberg, Mark L Van Natta, Robert A Wise, Stanley J Szeffler, Sunita Sharma, Alvin T Kho, et al. Patterns of growth and decline in lung function in persistent childhood asthma. *New England Journal of Medicine*, 374(19):1842–1852, 2016.
- [79] Roberto Merletti and Philip J Parker. *Electromyography: physiology, engineering, and non-invasive applications*, volume 11. John Wiley & Sons, 2004.

- [80] MicroLife. Pf 100 asthma monitor.
- [81] Martin R Miller, JATS Hankinson, V Brusasco, F Burgos, R Casaburi, A Coates, R Crapo, P vd Enright, CPM Van der Grinten, P Gustafsson, et al. Standardisation of spirometry. *European respiratory journal*, 26(2):319–338, 2005.
- [82] S Morrison, J Kavanagh, SJ Obst, J Irwin, and LJ Haseler. The effects of unilateral muscle fatigue on bilateral physiological tremor. *Experimental brain research*, 167(4):609–621, 2005.
- [83] Abu Saleh Mohammad Mosa, Illhoi Yoo, and Lincoln Sheets. A systematic review of healthcare applications for smartphones. *BMC medical informatics and decision making*, 12(1):1–31, 2012.
- [84] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingero: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1515–1525, 2016.
- [85] Phuc Nguyen, Xinyu Zhang, Ann Halbower, and Tam Vu. Continuous and fine-grained breathing volume monitoring from afar using wireless signals. In *IEEE INFOCOM*, 2016.
- [86] Hayrettin Okut. Bayesian regularized neural networks for small n big p data. *Artificial neural networks-models and applications*, 2016.
- [87] sean olive, omid khonsaripour, and todd welti. a survey and analysis of consumer and professional headphones based on their objective and subjective performances. *journal of the audio engineering society*, october 2018.
- [88] Zhenchao Ouyang, Jingfeng Hu, Jianwei Niu, and Zhiping Qi. An asymmetrical acoustic field detection system for daily tooth brushing monitoring. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.
- [89] Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 1–14, 2007.
- [90] Stefan Poslad. *Ubiquitous computing: smart devices, environments and interactions*. John Wiley & Sons, 2011.

- [91] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. Earsense: earphones as a teeth activity sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–13, 2020.
- [92] Markku Pukkila. Channel estimation modeling. *Nokia Research Center*, 17:66, 2000.
- [93] Philip H Quanjer, Sanja Stanojevic, Tim J Cole, Xaver Baur, Graham L Hall, Bruce H Culver, Paul L Enright, John L Hankinson, Mary SM Ip, and Jinping Zheng. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *European Respiratory Journal*, 40(6):1324–1343, 2012.
- [94] Mohammad Reza Raoufy, Sohrab Hajizadeh, Shahriar Gharibzadeh, Ali R Mani, Parivash Eftekhari, and Mohammad Reza Masjedi. Nonlinear model for estimating respiratory volume based on thoracoabdominal breathing movements. *Respirology*, 18(1):108–116, 2013.
- [95] Helen Reddel, Sandra Ware, Guy Marks, Cheryl Salome, Christine Jenkins, and Ann Woolcock. Differences between asthma exacerbations and poor asthma control. *The Lancet*, 353(9150):364–369, 1999.
- [96] Ravi S Reddy, Khalid A Alahmari, Paul S Silvian, Irshad A Ahmad, Venkata Nagaraj Kakarparthi, and Kanagaraj Rengaramanujam. Reliability of chest wall mobility and its correlation with lung functions in healthy nonsmokers, healthy smokers, and patients with copd. *Canadian respiratory journal*, 2019, 2019.
- [97] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [98] Medical International Research. Mir smart one.
- [99] Paul A Romitti, Yong Zhu, Soman Puzhankara, Katherine A James, Sarah K Nabukera, Gideon KD Zamba, Emma Ciafaloni, Christopher Cunniff, Charlotte M Druschel, Katherine D Mathews, et al. Prevalence of duchenne and becker muscular dystrophies in the united states. *Pediatrics*, 135(3):513–521, 2015.
- [100] Alain Rudiger, Jens P Hellermann, Raphael Mukherjee, Ferenc Follath, and Juraj Turina. Electrocardiographic artifacts due to electrode misplacement and their frequency in different clinical settings. *The American journal of emergency medicine*, 25(2):174–178, 2007.

- [101] Teerayut Sa-Ngiamsak. Assessment of muscle fatigue in work related musculoskeletal disorders by high-density surface electromyography. 2016.
- [102] Kent Sahlin. Muscle fatigue and lactic acid accumulation. *Acta physiologica Scandinavica. Supplementum*, 556:83–91, 1986.
- [103] Chandan K Sahu, Crystal Young, and Rahul Rai. Artificial intelligence (ai) in augmented reality (ar)-assisted manufacturing applications: a review. *International Journal of Production Research*, 59(16):4903–4959, 2021.
- [104] Khairul Muzzammil Saipullah, Ammar Anuar, Nurul Atiqah Ismail, and Yewguan Soo. Measuring power consumption for image processing on android smartphone. *American Journal of Applied Sciences*, 9(12):2052, 2012.
- [105] Mobile Fact Sheet. Pew research center. *Washington DC: Pew Research Center*, 2018.
- [106] Bruno MC Silva, Joel JPC Rodrigues, Isabel de la Torre Díez, Miguel López-Coronado, and Kashif Saleem. Mobile-health: A review of current state in 2015. *Journal of biomedical informatics*, 56:265–272, 2015.
- [107] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. Spirosonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, 2020.
- [108] Larry Squire, Darwin Berg, Floyd E Bloom, Sascha Du Lac, Anirvan Ghosh, and Nicholas C Spitzer. *Fundamental neuroscience*. Academic press, 2012.
- [109] Steven R Steinhubl, Evan D Muse, and Eric J Topol. The emerging field of mobile health. *Science translational medicine*, 7(283):283rv3–283rv3, 2015.
- [110] MJ Stokes and PA Dalton. Acoustic myography for investigating human skeletal muscle fatigue. *Journal of applied physiology*, 71(4):1422–1426, 1991.
- [111] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 591–605, 2018.

- [112] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, pages 3319–3328. PMLR, 2017.
- [113] M Tarata, A Spaepen, and R Puers. The accelerometer mmg measurement approach, in monitoring the muscular fatigue. *Measurement Science Review*, 1(1):47–50, 2001.
- [114] Yonglong Tian, Guang-He Lee, Hao He, Chen-Yu Hsu, and Dina Katabi. Rf-based fall monitoring using convolutional neural networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–24, 2018.
- [115] Jonathan L Ticknor. A bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, 40(14):5501–5506, 2013.
- [116] Mary C Townsend, Occupational, Environmental Lung Disorders Committee, et al. Spirometry in the occupational health setting—2011 update. *Journal of occupational and environmental medicine*, 53(5):569, 2011.
- [117] Hitomi Tsujita and Jun Rekimoto. Smiling makes us happier: enhancing positive mood and communication with smile-encouraging digital appliances. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 1–10, 2011.
- [118] Yu-Chih Tung, Duc Bui, and Kang G Shin. Cross-platform support for rapid development of mobile acoustic sensing applications. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 455–467, 2018.
- [119] Derya Unsal and Kerim Demirbas. Estimation of deterministic and stochastic imu error parameters. In *Proceedings of IEEE/ION PLANS 2012*, pages 862–868, 2012.
- [120] Anran Wang and Shyamnath Gollakota. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.
- [121] Anran Wang, Dan Nguyen, Arun R Sridhar, and Shyamnath Gollakota. Using smart speakers to contactlessly monitor heart rhythms. *Communications biology*, 4(1):1–12, 2021.

- [122] Kai Wang, Xiaojiang Peng, Jianfei Yang, Debin Meng, and Yu Qiao. Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Transactions on Image Processing*, 29:4057–4069, 2020.
- [123] Wei Wang, Alex X Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 82–94, 2016.
- [124] DE Warburton, N Gledhill, VK Jamnik, B Krip, and N Card. Induced hypervolemia, cardiac function, vo2max, and performance of elite cyclists. *Medicine and science in sports and exercise*, 31(6):800–808, 1999.
- [125] Jing Wu, He Li, Zhangxi Lin, and Haichao Zheng. Competition in wearable device market: the effect of network externality and product compatibility. *Electronic Commerce Research*, 17(3):335–359, 2017.
- [126] Meng-Chieh Yu, Jia-Ling Liou, Shuenn-Wen Kuo, Ming-Sui Lee, and Yi-Ping Hung. Noncontact respiratory measurement of volume change using depth camera. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2371–2374, 2012.
- [127] Shichao Yue, Hao He, Hao Wang, Hariharan Rahul, and Dina Katabi. Extracting multi-person respiration from entangled rf signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–22, 2018.
- [128] Sangki Yun, Yi-Chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 15–29, 2015.
- [129] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the 15th annual international conference on mobile systems, applications, and services*, pages 15–28, 2017.
- [130] A Zapletal and J Chalupova. Forced expiratory parameters in healthy preschool children (3–6 years of age). *Pediatric pulmonology*, 35(3):200–207, 2003.
- [131] Hanbin Zhang, Chen Song, Aosen Wang, Chenhan Xu, Dongmei Li, and Wen Yao Xu. Pd vocal: Towards privacy-preserving parkinson’s disease detection using non-speech

- body sounds. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.
- [132] Jie Zhang, Zhanyong Tang, Meng Li, Dingyi Fang, Petteri Nurmi, and Zheng Wang. Crosssense: Towards cross-site and large-scale wifi sensing. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 305–320, 2018.
- [133] Ligang Zhang and Dian Tjondronegoro. Facial expression recognition using facial movement features. *IEEE transactions on affective computing*, 2(4):219–229, 2011.
- [134] Xiao Zhang, Wenzhong Li, Xu Chen, and Sanglu Lu. Moodexplorer: Towards compound emotion detection via smartphone sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–30, 2018.
- [135] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108, 2016.