

EarSpiro: Earphone-based Spirometry for Lung Function Assessment

WENTAO XIE, The Hong Kong University of Science and Technology, Hong Kong and Southern University of Science and Technology, China

QINGYONG HU, The Hong Kong University of Science and Technology, Hong Kong

JIN ZHANG*, Southern University of Science and Technology, China

QIAN ZHANG*, The Hong Kong University of Science and Technology, Hong Kong

Spirometry is the gold standard for evaluating lung functions. Recent research has proposed that mobile devices can measure lung function indices cost-efficiently. However, these designs fall short in two aspects. First, they cannot provide the flow-volume (F-V) curve, which is more informative than lung function indices. Secondly, these solutions lack inspiratory measurement, which is sensitive to lung diseases such as variable extrathoracic obstruction. In this paper, we present EarSpiro, an earphone-based solution that interprets the recorded airflow sound during a spirometry test into an F-V curve, including both the expiratory and inspiratory measurements. EarSpiro leverages a convolutional neural network (CNN) and a recurrent neural network (RNN) to capture the complex correlation between airflow sound and airflow speed. Meanwhile, EarSpiro adopts a clustering-based segmentation algorithm to track the weak inspiratory signals from the raw audio recording to enable inspiratory measurement. We also enable EarSpiro with daily mouthpiece-like objects such as a funnel using transfer learning and a decoder network with the help of only a few true lung function indices from the user. Extensive experiments with 60 subjects show that EarSpiro achieves mean errors of 0.20L/s and 0.42L/s for expiratory and inspiratory flow rate estimation, and 0.61L/s and 0.83L/s for expiratory and inspiratory F-V curve estimation. The mean correlation coefficient between the estimated F-V curve and the true one is 0.94. The mean estimation error for four common lung function indices is 7.3%.

CCS Concepts: • **Human-centered computing** → **Mobile devices**; • **Applied computing** → **Health informatics**.

Additional Key Words and Phrases: smart healthcare, wearables, earphone, lung function, flow rate, flow-volume curve

ACM Reference Format:

Wentao Xie, Qingyong Hu, Jin Zhang, and Qian Zhang. 2022. EarSpiro: Earphone-based Spirometry for Lung Function Assessment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 188 (December 2022), 27 pages. <https://doi.org/10.1145/3569480>

*Corresponding authors.

Authors' addresses: Wentao Xie, wxieaj@cse.ust.hk, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering and Southern University of Science and Technology, China; Qingyong Hu, qhuag@cse.ust.hk, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong; Jin Zhang, zhang.j4@sustech.edu.cn, Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, China; Qian Zhang, qianzh@cse.ust.hk, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

2474-9567/2022/12-ART188 \$15.00

<https://doi.org/10.1145/3569480>

Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 6, No. 4, Article 188. Publication date: December 2022.

1 INTRODUCTION

Spirometry is the most common test to assess lung function by measuring how much air and how fast one can breathe in and out of his/her lung. Spirometry is often used to diagnose chronic respiratory diseases (CRDs), such as chronic obstructive pulmonary disease (COPD) and asthma. In addition, health experts and governments call for regular spirometry screening for those in high-risk occupations, such as coal workers, even if they have no symptoms of lung function disorders [1, 15]. Also, doctors suggest smokers take spirometry tests regularly to track lung function impairments before developing any symptoms.

A standard spirometry test requires a subject to expire and inspire forcefully [30] while a spirometer, an airflow measurement device, measures his/her lung function. Although there are various types of spirometers on the market, few of them are suitable for daily use. Traditionally, spirometry is performed in clinics and hospitals using conventional spirometers. Although accurate, these spirometers are often too large and heavy and are only available in clinics. Alternatively, portable spirometers are much smaller than those in clinics but have similar accuracy, which makes them more suitable for home use than clinical spirometers. However, their costs are too high (normally more than 2,000 USD [12, 14]), and this is a huge obstacle for users, especially for low-income families.

Recently, researchers have developed smartphone-based systems to conduct spirometry [27, 33, 37, 50]. These designs significantly reduce the cost of spirometry while preserving an acceptable level of accuracy. However, the main target of these designs is to measure four expiratory lung function indices, and there is still a gap between these systems with a real spirometer. First, in addition to the lung function indices, a real spirometer can also measure the F-V curve (introduced in Sec. 2). Second, these designs can only support expiratory measurement, while a real spirometer also contains inspiratory measurement. The following explains why these two aspects are important for spirometry. (i) The F-V curve is important information in spirometry for two reasons. First, its shape is a direct indicator of certain lung function impairments [58] (Fig. 2). For example, a concave shape of the expiratory limb may imply obstructive lung function (Fig. 2(b)). Secondly, the F-V curve is required in clinics for quality control [30]. That is to say, after a subject performs a spirometry test, doctors should assess the quality (*i.e.*, usability) of the spirometry result by observing the shape of the F-V curve because different deficiencies when conducting the spirometry will produce F-V curves of different shapes (Fig. 3). For example, the ATS/ERS standard requires the subject to make his/her maximal effort in the expiration phase so that a clear peak should be observed in the F-V curve (Fig. 2(a)). If the user fails to do so, the F-V curve will display a rounded shape instead of a clear peak (Fig. 3(a)). This is important, especially for home spirometry, where assistance from doctors and technicians is absent. Imagine a scenario when a patient conducts spirometry at home and uploads the result to his/her doctor remotely. If the uploaded data only contains the lung function indices, the doctor can only assume the patient performs the spirometry correctly and use these data to assess the user's lung condition. Nevertheless, if the F-V curve is also included, the doctor can evaluate whether the spirometry is correctly conducted and further decide whether the spirometry data is reliable. (ii) The inspiratory measurement is important in spirometry because it is sensitive to some diseases such as variable extrathoracic obstruction [22, 36, 56]. For the patients with this disease, the inspiratory limb of the F-V curve shows a flattened pattern while the expiratory limb is almost the same with a healthy subject (*e.g.*, Fig. 2(d)). Additionally, inspiratory measurement can also be used to evaluate whether the air in the lung is expired completely [30]. That is to say, if the user does not expire the air out of his/her lung completely, but the inspiration is conducted correctly, the expiratory limb and the inspiratory limb will be mismatched (*e.g.*, Fig. 3(c)).

To solve the above limitations, we ask: *Can we design a mobile spirometry system that can measure the full F-V curve, including both the expiratory and inspiratory measurements, just like a real spirometer?* Inspired by previous works [27, 37], our observation to design such a system is that airflow will generate sounds when passing through a constrictive area, such as an individual's airway [26, 35], and this airflow sound correlates with

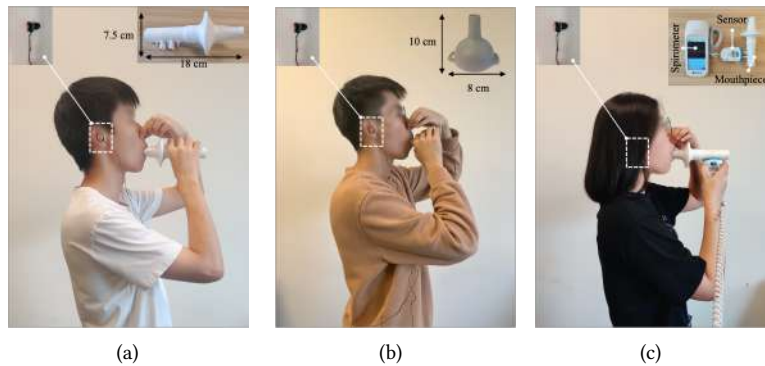


Fig. 1. Users using EarSpiro. (a) Use EarSpiro with a standard mouthpiece. (b) Use EarSpiro with a funnel. (c) Ground truth data collection.

the flow rate [24]. When a mouthpiece controls the user’s mouth posture, the produced airflow sound mostly depends on the flow rate. Therefore, we can interpret the airflow sound signal into the corresponding real-time flow rate by capturing such a correlation and further derive the F-V curve. We propose to use microphones embedded in earphones to collect the airflow sound. The benefits of this design are two-fold. First, the collected airflow sound is of high quality. This is because the airflow sound generated from the respiratory system can easily propagate through the tissues and bones of the head and the auditory tube through bone conduction [46]. Therefore, an earphone can collect airflow sound even if the airflow sound is weak, as in the case of inspiration. Intuitive evidence of this claim is that one can easily hear his/her breathing sound even if the breathing rate is slow. Second, if an earphone is used to collect the airflow sound, the sound propagation distance is fixed, and the user does not need to manually maintain the distance between the mouth and the microphone since the sound intensity changes with distance otherwise.

With the above observations, we propose EarSpiro, an earphone-based spirometry solution that can measure the whole F-V curve, including both the expiratory and the inspiratory measurements. To use EarSpiro, a user needs to wear a pair of microphone-equipped earphones and perform a standard spirometry maneuver (as introduced in Sec. 2) through a mouthpiece (Fig. 1(a)). The mouthpiece is used to maintain the user’s mouth posture to ensure the consistency of the airflow sound. Apart from the standard mouthpiece, we envision that EarSpiro can also be used with any customized, mouthpiece-like object such as a funnel (Fig. 15) since these objects can also control the user’s mouth posture. In this way, users of EarSpiro can use daily funnel-like objects to perform the lung function assessment (Fig. 1(b)).

Although promising, designing such a system is challenging. The first challenge is that interpreting airflow sound to flow rate is hard since there is no well-developed solution to derive flow rate directly from airflow sound. We use a deep learning approach to model this complex correlation to resolve this challenge. The second challenge is that the inspiratory measurement is difficult because the inspiratory sound can be too weak to collect due to the low flow rate. In this case, environmental noise can easily mask the inspiratory signal. To resolve this challenge, we design a clustering-based segmentation algorithm to precisely extract the inspiratory signal from the entire signal. The third challenge is extending EarSpiro’s function so that EarSpiro can work with not only the standard mouthpiece but also daily funnel-like objects. This is difficult because the airflow sound generated using a funnel differs from that of a standard mouthpiece (see Sec. 4.3.3). This means a new model needs to be trained for EarSpiro every time a new funnel is used. However, the ground truth flow rate when using a funnel

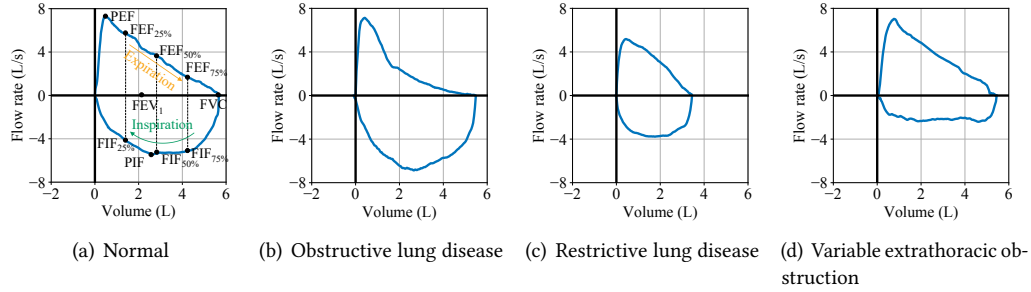


Fig. 2. The F-V curves of different lung conditions. (a) Normal. (b) Obstructive lung disease. The expiratory limb exhibits a steep shape [30]. (c) Restrictive lung disease. The shape of the curve is reduced in size [58] (d) Variable extrathoracic obstruction. The inspiratory limb is flattened [51].

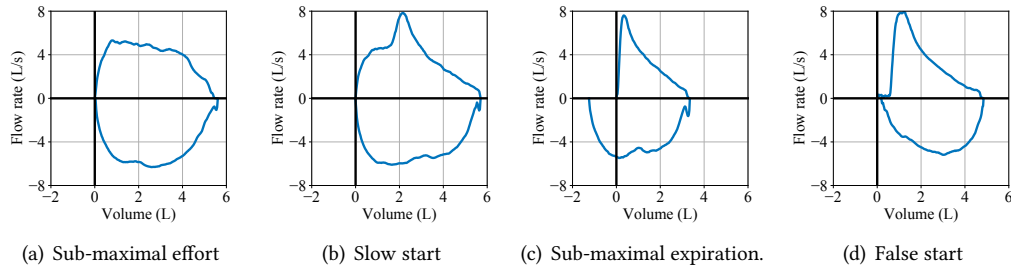


Fig. 3. F-V curves of unacceptable maneuvers. (a) Sub-maximal expiratory effort. The shape of the expiratory limb is rounded [30]. (b) Slow start. The peak occurs late in the expiratory limb [30]. (c) Sub-maximal expiration. The inspiratory volume is larger than the expiratory volume [30]. (d) False start. The hesitation time at the beginning of the F-V curve is overlong [30].

is unavailable because no spirometer can match with an arbitrary funnel. Therefore, training a new model is infeasible. Besides, it is impractical to train a new model for every funnel-like object. To resolve this challenge, we leverage the fact that most patients know their lung function indices (from previous clinic visits), and healthy users can easily obtain them after a clinic visit. Therefore, we use a deep decoder network to generate virtual flow rate samples from the lung function indices. Then we use transfer learning with the virtual samples to adapt the basic deep learning model to the new funnel.

The contributions of this work are summarized as follows. First, to the best of our knowledge, EarSpiro is the first earphone-based solution for spirometry measurement. In addition, EarSpiro is the first mobile spirometry system to estimate the entire F-V curve, including expiratory and inspiratory measurements. Secondly, we design algorithms to solve practical challenges in our system, including a clustering-based segmentation algorithm to extract the weak inspiratory signal, a hybrid CNN-RNN deep learning architecture to model the complex correlation between the flow rate and the airflow sound, a transfer learning framework to enable EarSpiro on any funnel-like objects leveraging the virtual flow rate samples generated from a decoder network. Finally, we build a prototype for EarSpiro and conduct extensive experiments with 60 subjects to evaluate the performance of EarSpiro.

The rest of this paper is organized as follows. Sec. 2 introduces the background information and Sec. 3 provides an overview of the design of EarSpiro. Sec. 4 elaborates on the design details and Sec. 5 introduces our implementation of the EarSpiro prototype. Sec. 6 evaluates the performance of EarSpiro. Sec. 7 discusses

the limitations of this research and provides future research directions. Sec. 8 discusses the related work of this research and Sec. 9 concludes this paper.

2 BACKGROUND

In this section, we first briefly discuss spirometry's principles, process and outcomes. Then, we discuss why we can estimate spirometry by analyzing the airflow sound caused by forced expiration and inspiration.

2.1 Spirometry

Spirometry is a standard method to assess lung function. During a spirometry test, a subject is asked to maintain an upright sitting posture and perform the spirometry maneuver, which includes the following three steps: (1) maximal inspiration, (2) expiration with maximal effort until no more air can be breathed out, (3) inspiration with maximal effort until no more air can be breathed in [30]. A spirometer records the real-time flow rate and air volume during the above maneuver. When the maneuver is over, the spirometer outputs an F-V curve whose shape can reveal a characteristic pattern suggestive of different lung diseases. For example, Fig. 2 shows the F-V curves of different diseases. Also, an F-V curve can be used to evaluate the acceptability and usability of the performed maneuver. Fig. 3 shows several examples of unacceptable F-V curves. Usually, a set of lung function indices are extracted from the F-V curve as indicators of some lung diseases like asthma and COPD [18]. The most common lung function indices are listed below. Their graphical explanations are shown in Fig. 2(a).

- **Forced vital capacity (FVC).** Total air volume expelled.
- **Forced expiratory volume in one second (FEV1).** Air volume expelled in the first second.
- **FEV1/FVC.** The ratio of FEV1 and FVC.
- **Peak expiratory flow (PEF).** The maximum expiratory airflow rate.

Apart from the above four indices, other important lung function indices are also measured by clinical spirometers. These indices are summarized as follows.

- **Forced expiratory flow at x% point of FVC (FEF_{x%}).** The instantaneous expiratory flow rate at x% of FVC where x is usually 25, 50 and 75 [10, 12].
- **Forced mid-expiratory flow (FEF_{25-75%}).** The averaged expiratory flow rate from the 25% FVC to 75% FVC [47].
- **Forced inspiratory vital capacity (FIVC).** Total air volume inhaled in the inspiratory phase [30].
- **Peak inspiratory flow (PIF).** The maximum airflow rate during the inspiratory phase [10].
- **Forced inspiratory flow at x% point of FIVC (FIF_{x%}).** The instantaneous inspiratory flow rate at x% of FIVC where x is usually 25, 50 and 75 [10].

These indices are also crucial for lung function assessment. For example, FEF_{25%}, FEF_{50%} and FEF_{75%} serve as landmarks of the middle phase of expiration, which can indicate the status of small airways [13], FEF_{25-75%} is a helpful predictor for the development of COPD [23]. Apart from expiratory indices, some inspiratory indices also help to assess lung function. For example, PIF is a sensitive indicator to variable extrathoracic obstruction [51].

2.2 Correlation between Airflow Speed and Airflow Sound

When an individual performs the expiratory maneuver, turbulence will form in the constrictive areas of the respiration system, such as the trachea and mouth [35]. The turbulence flow will generate sound, which is often called expiratory sound [35]. It has been shown that the intensity of forced expiratory sound generated by an individual's trachea is proportional to the flow rate that passes through the trachea [24]. However, the tracheal sound intensity is also a function of the subject's tracheal size and shape [24]. Since the shape of people's trachea is complex and various [19, 41], there is no well-developed theory to predict flow rate directly based on expiratory sound. Apart from the trachea and mouth, when a mouthpiece is used, turbulence flow will also form inside

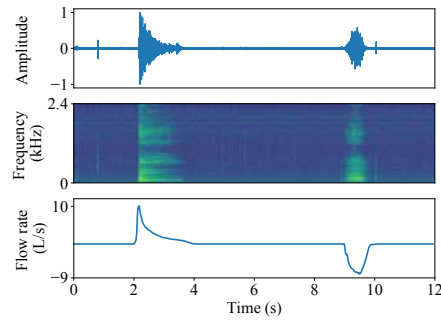


Fig. 4. The correlation between flow rate and airflow sound. From top to bottom are time-domain, frequency-domain, and flow rate signals.

the mouthpiece, generating airflow sound whose intensity is proportional to flow rate [54]. As discussed in Sec. 1, EarSpiro leverages such a correlation between airflow speed and sound to estimate the flow rate. An intuitive example of this correlation is shown in Fig. 4 where the time and frequency domain of airflow sound of a spirometry maneuver, as well as the ground truth flow rate, are present together. It is easy to observe the consistency between the airflow sound and flow rate.

3 SYSTEM OVERVIEW

In this section, we give an overview of EarSpiro. We design three modules for EarSpiro. (i) **Audio feature extraction** module. This module extracts time-frequency features from the audio recordings because the raw expiration and inspiration signals have certain frequency characteristics. (ii) **Expiratory and inspiratory phase segmentation** module. We design this module because it is not always easy to localize the airflow signal from the entire audio recording, especially for the inspiratory signal. Therefore, we use this module to separate the expiratory and inspiratory signals from other signals. (iii) **Deep learning-based F-V curve estimation** module. Since the correlation between airflow sound and flow rate is complex, we use deep learning techniques to model this correlation. In addition, when using EarSpiro with new mouthpiece-like objects (*e.g.*, a funnel), we use deep learning techniques to adapt EarSpiro’s functionality to this new mouthpiece with as little prior knowledge as possible. The system diagram of EarSpiro is shown in Fig. 5. The following briefly introduces the working pipelines of the three modules.

- (1) **Audio feature extraction.** This module takes the raw audio recordings as the input. First, an energy-based method is used to roughly locate the expiratory and inspiratory signals. Then, a short-time Fourier transform (STFT) and a Mel filter bank are applied to extract the time-frequency features from the audio signal.
- (2) **Expiratory and inspiratory phase segmentation.** This module operates on the obtained Mel spectrogram. First, an energy-based algorithm is used to localize the expiratory phase. Then, a K-means clustering-based segmentation algorithm is applied to localize the weak inspiratory signal.
- (3) **Deep Learning-based F-V Curve Estimation.** A CNN-GRU model is trained by the segmented features and the corresponding ground truth flow rate with data augmentation. The flow rate output by the model is then transformed into an F-V curve. To adapt the model to a new mouthpiece, virtual flow rate samples are first generated by a decoder network from a few true lung function indices from the user. Then, the model is fine-tuned by the Mel spectrogram features generated from the new mouthpiece and the virtual flow rate samples.

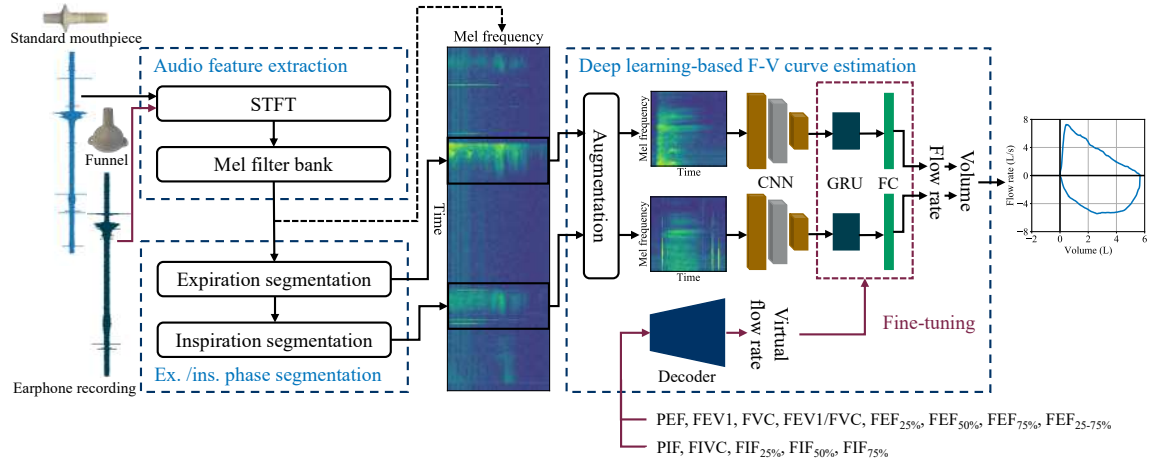


Fig. 5. The workflow of EarSpiro.

4 SYSTEM DESIGN

In this section, we elaborate on the design details of EarSpiro. The following three subsections discuss the three main modules of EarSpiro.

4.1 Audio Feature Extraction

After receiving the raw audio recordings from the earphone, EarSpiro first applies STFT to extract time-frequency features because, as validated by medical research, the forced breathing sound has distinct frequency characteristics [26]. Fig. 6(a) shows the spectrogram after applying STFT to the audio signal. The powerful regions around 3-6 seconds and 8-9 seconds represent the expiratory and inspiratory phases, respectively.

It is worth noting that the output spectrogram is usually high-dimensional because of the large number of frequency samples. Therefore, we apply a Mel filter bank to the spectrogram to reduce the feature size. An advantage of the Mel filter bank is that it is a set of filters whose bandwidth is increased exponentially with frequency. In other words, it puts more weight on the lower frequencies. Since the spectral features of the human voice are centered at the lower frequencies, the Mel filter bank can emphasize the low-frequency features. To compute the Mel filter bank, we first transform the target frequency range (*i.e.*, 0.5-15kHz) from the Hertz scale to the Mel scale and equally divide the frequency range into M central frequencies for the filters where M is the number of filters. Then the central frequencies are transformed back to the Hertz scale. The transformation functions are $M(\cdot)$ and $M^{-1}(\cdot)$ where $M(\cdot)$ is given by $M(f) = 1125 \cdot \log(1 + f/700)$. We denote the central frequencies in the Hertz scale as f_m with ascending order where m takes from 1 to M . We denote the boundary frequencies as f_0 and f_{M+1} . Then, the frequency response of each filter in the filter bank is computed as [32]:

$$H_m(k) = \begin{cases} \frac{2(k-f_{m-1})}{(f_{m+1}-f_{m-1})(f_m-f_{m-1})}, & f_{m-1} \leq k \leq f_m \\ \frac{2(f_{m+1}-k)}{(f_{m+1}-f_{m-1})(f_m-f_{m-1})}, & f_m \leq k \leq f_{m+1} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $m = 0$ to M , M is the number of filters, and f_m is the central frequency of the m th filter. All the f_m 's are equally spaced in Mel scale. Then, the Mel spectrogram is computed as $\mathbf{M} = \mathbf{H}^T \mathbf{S}$, where \mathbf{M} , \mathbf{H} and \mathbf{S} are the Mel spectrogram, the Mel filter bank, and the original spectrogram. The generated Mel spectrogram is shown in Fig.

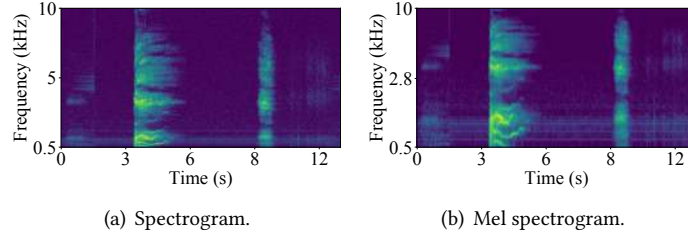


Fig. 6. Spectrogram and Mel spectrogram.

Table 1. Parameters in the audio feature extraction module.

| Para | Explanation | Value |
|------------------|---|--------|
| f_s | Sampling frequency. | 48000 |
| T_{win} | Length of each STFT segment. | 50ms |
| N_{FFT} | FFT point. | 2400 |
| $\rho_{overlap}$ | Overlap ratio between segments. | 75% |
| f_{low} | The lowest frequency. | 0.5kHz |
| f_{high} | The highest frequency. | 15kHz |
| M | Number of filters in the Mel filter bank. | 100 |

6(b). The parameters used in this module are summarized in Tab. 1. The first four parameters are of common values, and we select the value of the last three parameters empirically.

4.2 Expiratory and Inspiratory Phase Segmentation

In this section, we discuss the methods and algorithms to localize the expiratory phase and the inspiratory phase from the features derived from the last module. Although Fig. 6(b) implies that it is straightforward to identify the expiratory and inspiratory phases, it is not always easy to observe such a clear pattern. First, there is environmental noise. Although, as we shall discuss in Sec. 6 that we assume the user is in a quiet environment when using EarSpiro, it is still inevitable to record environmental noise. For example, when the user expires or inspires with maximal efforts, noise can accidentally be generated by friction between the user and the chair or the floor. Also, inertial body sound such as teeth knocking can happen during spirometry. Secondly, inspiratory sound can be too weak to identify [27, 37], especially for users with limited lung function. Therefore, we design special algorithms to extract the expiratory and inspiratory signals.

4.2.1 Expiration Segmentation. To localize the expiration signal, we first identify the start time of expiration. This is not difficult since the spirometry maneuver starts with a blast of expiration where the audio recording shows a sudden energy jump (Fig. 7(a)). Therefore, we first obtain the energy profile of the Mel spectrogram by taking the average along the frequency axis (Fig. 7(b)). Note that, generally, there is noise in the Mel spectrogram, which results in spikes in the energy profile. These spikes can be removed by a Hampel filter (brown line in Fig. 7(b)). Then, we use a gradient search algorithm to find the expiratory starting point based on the energy profile. The algorithm, detailed in Algorithm 1, searches along the time axis starting from the maxima until the gradient decreases to a predefined threshold. The estimated start time of the example above is shown in Fig. 7(b) with a triangle marker.

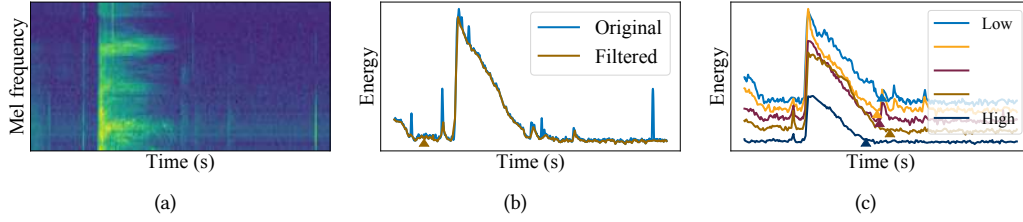


Fig. 7. Expiratory phase segmentation. (a) Mel spectrogram. (b) Energy profile of the Mel spectrogram. (c) Energy profiles of the five sub-bands.

Next, we estimate the end time of the expiratory phase. Different from the start of the expiration, where all the frequencies consistently show a sharp rise, at the end of the expiration phase, different frequency components vanish at different times, as shown in Fig. 7(a). Therefore, to estimate the expiratory end time, we divide the Mel frequency band into five sub-bands and make the end time estimations separately using the gradient search algorithm discussed above. The five energy profiles of these five bands are shown in Fig. 7(c). The end times of these energy profiles estimated by the gradient search algorithm are marked with triangles in Fig. 7(c). We select the largest prediction as the end time estimation of the expiratory phase, *i.e.*, the brown triangle in Fig. 7(c). The algorithm for expiratory phase segmentation is described in Algorithm 1. The parameters in this algorithm are selected empirically.

Algorithm 1 Expiratory Phase Segmentation

Input: $E[n]$: energy profile; $E_k[n]$: energy profiles of the target sub-bands; f_k : central frequencies of the target sub-bands; d : step size; w : window length; c : stop criterion.

Output: n_0 : Estimated start time; n_1 : Estimated end time.

```

1:  $n_0 \leftarrow \text{SEARCH}(E[n], -d, w, c)$  ▷ Execute the gradient search algorithm.
2: for  $k = 1$  to  $K$  do
3:    $n_k \leftarrow \text{SEARCH}(E_k[n], d, w, c)$  ▷ Execute the gradient search algorithm on each sub-band.
4: end for
5:  $n_1 \leftarrow \max(\{n_k\})$ 
6: return  $n_0, n_1$ 
7:
8: procedure  $\text{SEARCH}(E[n], d, w, c)$  ▷ The gradient search algorithm.
9:    $n \leftarrow \text{argmax}(E[n])$ 
10:  while  $E[n] \geq \frac{1}{2}\max(E[n])$  or  $\frac{E[n+w]-E[n]}{w} \geq c$  do
11:     $n \leftarrow n + d$ 
12:  end while
13:  return  $n$ 
14: end procedure

```

4.2.2 Inspiration Segmentation. After the expiratory phase is derived, we extract the inspiratory phase from the remaining signal. Unlike the expiratory phase, where the time-frequency feature is clear and regular, the inspiratory signal is easily masked by noise because the inspiratory sound is usually weak. For example, Figs. 8(a) and 8(f) show two examples of the Mel spectrogram of the inspiratory signal where the red rectangles mark the actual inspiratory interval according to the ground truth flow rates which are shown in Figs. 8(b) and 8(g). In

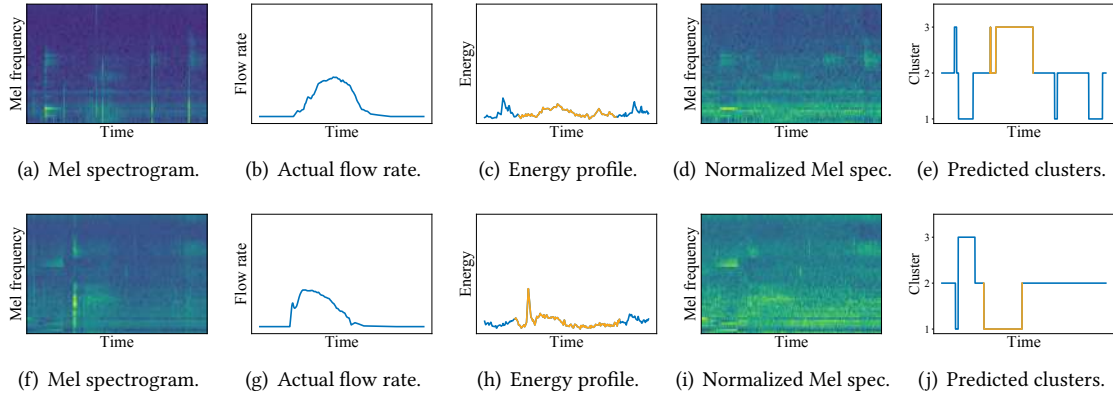


Fig. 8. Two examples of inspiratory phase segmentation. (a-e) The first example. (f-j) The second example.

Algorithm 2 Inspiratory Phase Segmentation

Input: $M[n]$: Mel spectrogram; N_{ext} : extension time.

Output: n_0 : The estimated inspiratory start time; n_1 : The estimated inspiratory end time;

- 1: **for** $n = 1$ to N **do**
 - 2: $M[n] \leftarrow \text{ZSCORE}(M[n])$ ▷ Compute the Z-score normalization.
 - 3: **end for**
 - 4: $\{c_{in}, c_{noise}, c_{null}\} \leftarrow \text{KMEANS}(M[n])$ ▷ Execute the K-means clustering algorithm.
 - 5: $c_{in} \leftarrow$ Select the target cluster using the criteria in Sec. 4.2.2
 - 6: $n_0 \leftarrow \min(\arg(C[n] = c_{in}) - N_{ext})$
 - 7: $n_1 \leftarrow \max(\arg(C[n] = c_{in}) + N_{ext})$
 - 8: **return** n_0, n_1
-

these two examples, the energy of the inspiratory signal is lower than the environmental noise, as shown in Figs. 8(c) and 8(h). Therefore, the energy-based method introduced in the previous subsection is no longer suitable for extracting the inspiratory interval.

To solve this challenge, we observe that although it is difficult to distinguish the inspiratory signal from noise signals solely based on the energy characteristics, the inspiratory signal still has different frequency distribution compared with the noise signal. Let us consider the frequency distribution at every time instances of the Mel spectrogram. There are usually three types of signals: (i) inspiration, where the frequency distribution is consistent, and the power decreases smoothly as the frequency increases, (ii) noise, where the energy distribution is usually random, and (iii) null, where there is neither inspiratory signal nor noise signal and the power across all frequencies are low. To separate the inspiratory sound from the other two, we use the k-means clustering algorithm [31] to cluster all the time samples into three clusters. The feature to compute the per-sample distance in the k-means clustering algorithm is the Z-score normalized frequency distribution, as shown in Figs. 8(d) and 8(i). The clustering results are shown in Figs. 8(e) and 8(j) with blue lines.

Although the clustering algorithm can separate the inspiratory signal from the environmental noise, the algorithm does not output which cluster represents the inspiratory signal. Therefore, because the power of the inspiratory signal is consistent while the power of noise is usually spike-like and the null signal has little power, we design the following screening process to identify the cluster that represents inspiration. First, we compute

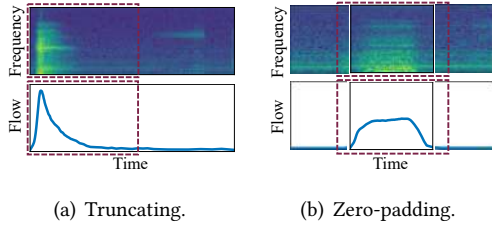


Fig. 9. Data augmentation.



Fig. 10. Flow rate estimator architecture.

the mean energy of the three clusters based on the energy profile. Then, we select the cluster that has the largest mean energy as the inspiration signal. Note that the estimated inspiratory signal may not be continuous in time, as shown in Figs. 8(e). Therefore, we group the nearby inspiratory intervals into one. The predicted inspiratory intervals are shown with orange lines in Figs. 8(e) and 8(j) and denoted with yellow squares in Figs. 8(a) and 8(f). Since the start/end time predicted by the above algorithms always underestimate the actual inspiratory interval according to our experiments, we extend the start/end time estimation by an additional one second (dashed yellow squares in Figs. 8(a) and 8(f)). Although this extension may accidentally include other signals, we will discuss in Sec. 4.3 that by using a certain data augmentation technique, EarSpiro can tolerate this error. The inspiratory phase segmentation algorithm is summarized in Algorithm 2.

4.3 Deep Learning-based F-V Curve Estimation

After extracting the expiratory and inspiratory features from the audio signal, we use a deep learning approach to estimate the flow rate. Specifically, we use a CNN-GRU-based flow rate estimator to interpret the segmented Mel spectrogram into flow rate. The rationale behind this design is two-fold. First, the Mel spectrogram is image-like data. Therefore, we use CNN to process it. Second, random inertial body sounds can coexist with the airflow sound (e.g., Fig. 8(a)). Since recent research has shown that RNN can filter out non-Gaussian noise [49], we connect a GRU after the CNN. Apart from the model architecture, we also need to consider the data quality when training the model. During the expiration phase, the airflow is mostly slow except for the beginning part. Therefore, a large portion of Mel spectrogram shows nearly no energy (e.g., Fig. 9(a)). Hence, training with the original expiratory data can make the model overfit to the low-flow rate part. Meanwhile, since the inspiratory interval tends to be over-estimated (explained in Sec. 4.2.2), the trained model must tolerate this error. To improve the data quality, we adopt two data augmentation strategies, that is, truncating and zero-padding, to enhance the dataset.

Another objective of this module is to adapt the deep learning model to any mouthpiece or even other mouthpiece-like objects, such as a funnel. The benefit of this design is that users of EarSpiro can use daily funnel-like objects to perform lung function assessment instead of being limited to the same mouthpiece used in training data collection. However, as discussed in Sec. 1, the challenge is that the ground truth flow rate, in this case, is unavailable. Therefore, training a new model for each new funnel is infeasible. Our observation to solve this challenge is that a user can easily know his/her lung function indices (after a visit to the clinic), and we can use a deep learning-based decoder to generate virtual flow rate samples from the lung function indices as implied by previous research [16]. These virtual flow rate samples can serve as the ground truth for transfer learning. In this way, we can adapt the model to any funnel. The detailed model adaptation framework is discussed in Sec. 4.3.3.

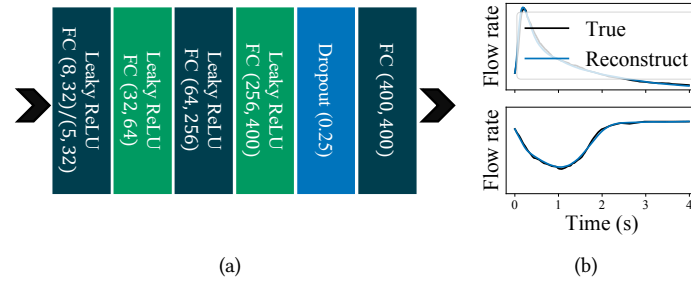


Fig. 11. Flow rate decoder. (a) The architecture. (b) Reconstructed flow rate samples.

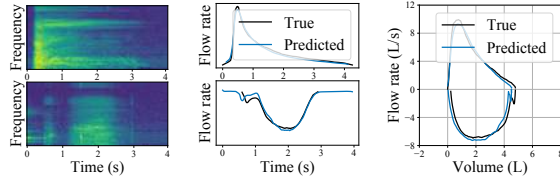
4.3.1 Data Augmentation. As discussed above, we use truncating and zero-padding to enhance the expiratory and inspiratory datasets. To truncate an expiratory sample, we first discard the tail portion where the flow rate is less than $0.2L/s$. Then, we extend the endpoint of this sample by a random 0-1 second. For an inspiratory sample, we randomly pad 0-1 second of zeros at the two sides of the ground truth. We also extend the corresponding Mel spectrogram of this sample to the same length. Examples of the data augmentation process are shown in Fig. 9.

4.3.2 Flow Rate Estimator Design. The flow rate estimator in EarSpiro is a hybrid model which contains a CNN followed by a GRU. The model architecture is shown in Fig. 10. The CNN model consists of three convolutional layers followed by a leaky ReLU activation with a negative slope of 0.01. Since EarSpiro leverages two earphones to collect airflow sound, we use two identical CNNs to process the two audio signals separately. The outputs of the two CNNs are concatenated and fed to the GRU model sequentially along the time axis. Two consecutive fully-connected layers further process the output of each GRU cell for flow rate regression. A dropout layer follows the first fully-connected layer. After the flow rate estimation is obtained, we take the cumulative summation of the predicted flow rate to obtain the airflow volume. Finally, we construct the F-V curve on the flow-volume domain.

We train two separate estimators for expiration and inspiration. The loss function used to train the models is the mean squared error (MSE) loss. The optimizer is Adam [34] with a 0.001 learning rate. We empirically set the dropout rate to be 0.5 and 0.2 for the expiratory and inspiratory estimators, respectively. Examples of flow rate estimation based on the Mel spectrograms in Fig. 12(a) are shown in Fig. 12(b). The corresponding F-V curve estimation is shown in Fig. 12(c).

4.3.3 Adaptation to Any Funnel-like Objects. When a new funnel (e.g., Fig. 15) is used as the mouthpiece, the produced airflow sound has different spectral characteristics compared with the standard mouthpiece (Fig. 13(a)). Therefore, if we directly leverage the deep learning model trained in the previous section to estimate the F-V curve, the result will be erroneous (brown curves in Figs 13(b) and 13(c)). As discussed above, we apply transfer learning to adapt the basic model to new funnel-like objects using virtual flow rate samples generated by a decoder network. We first introduce the decoder network design. Then we discuss the transfer learning framework we use for model adaptation.

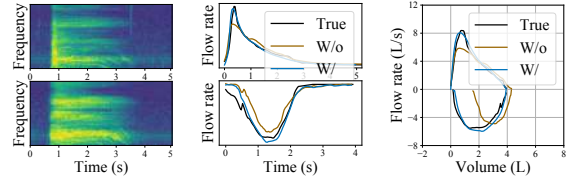
The decoder architecture is shown in Fig. 11(a). It is composed of five stacked fully-connected layers. We use separate decoders for expiration and inspiration. The input for this decoder is the eight expiratory indices or the five inspiratory indices as listed in Sec. 2.1. The output of this decoder is a signal of 400 samples representing the reconstructed flow rate sample of four seconds. We limit the output size to four seconds because we find that most airflow sound generated through spirometry is within four seconds. An example of the decoded flow rate sample is shown in Fig. 11(b).



(a) Mel spectrogram.

(b) Flow rate.

(c) F-V curve.



(a) Mel spectrogram.

(b) Flow rate.

(c) F-V curve.

Fig. 12. F-V curve estimation. (a) Mel spectrogram (Top: expiration. Bottom: inspiration). (b) Flow rate estimation (Top: expiration. Bottom: inspiration). (c) F-V curve estimation.

Fig. 13. Model adaptation. (a) Mel spectrogram (Top: mouthpiece. Bottom: funnel). (b) Flow rate estimation (Top: expiration. Bottom: inspiration). (c) F-V curve estimation.

After the decoder generates the virtual flow rate samples, we use these samples together with the Mel spectrogram generated through the funnel to perform transfer learning. Specifically, we randomly pair the available virtual flow rate samples and the Mel spectrogram to form a dataset. We then freeze the flow rate estimator's CNN layers and fine-tune the GRU and fully-connected layers. After this adaptation process, the user can directly perform the spirometry maneuver through the funnel, and the adapted model will produce the user's F-V curve accordingly. The flow rate estimation after model adaptation is shown in Fig. 13(b) with blue curve and the final F-V curve estimation is shown in Fig. 13(c) with blue curve.

5 IMPLEMENTATION

The prototype of EarSpiro is shown in Fig. 14(a). Although some commercial earphones are equipped with microphones [3], the raw audio recordings from these microphones are not accessible to developers. Therefore, we integrate a MEMS microphone [11] on a pair of commercial earphones [6]. This earphone is chosen because a recent work [38] reports that it has an excellent structural property suitable for integrating an additional microphone. The microphones are connected to the interface circuit connected by a Sseed ReSpeaker voice accessory board [9] for data acquisition. The system is powered by a Raspberry Pi 3 Model B+ [8]. We use the Audacity software [4] to record audio signals, which are saved as a WAV file in the disk drive of the Raspberry Pi for further processing. We implement the software of EarSpiro on a laptop and train the deep learning models on Google Colab [28] with an Nvidia T4 GPU. The software is written in Python 3.7, and the deep learning module is implemented with PyTorch 1.7 [7].

6 EVALUATION

In this section, we evaluate the performance of EarSpiro with extensive experiments under various parameter configurations and several real-world scenarios.

6.1 Data Collection

We recruit 60 subjects to participate in our experiments. The subjects are all students and staff on our campus. Six of the recruited subjects self-reported having asthma. One of these asthma patients is female, and the others are males. The age of these patients is from 21 to 27 years old. The demographics of the recruited subjects are summarized in Tab. 2. Unless otherwise specified, all the experiments are conducted in a quiet conference room with normal furniture and facilities. The subjects are asked to perform the spirometry maneuver three times through a mouthpiece connected to a certified spirometer [12] while wearing the EarSpiro prototype. The spirometer collects the ground truth flow rate while EarSpiro records audio signals.

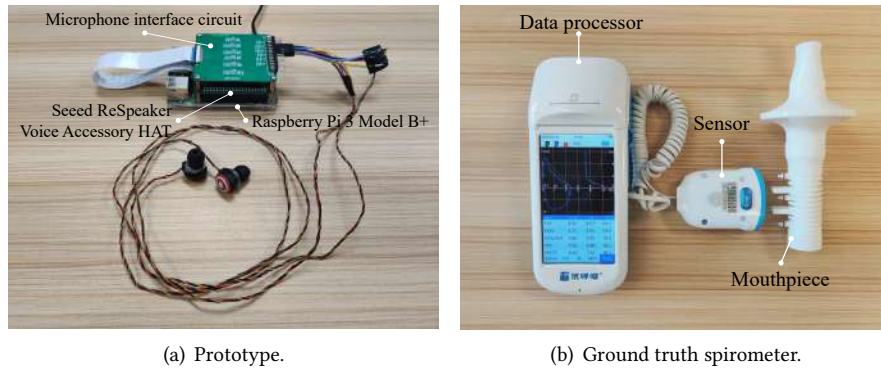


Fig. 14. EarSpiro implementation. (a) The prototype. (b) The spirometer used for ground truth collection.



Fig. 15. A funnel that can serve as mouthpiece.

Table 2. Demographics of the participants.

| Category | Value ^a |
|--------------------------|--------------------|
| Age (years) | 27.4 (20-70) |
| Female | 21 (35%) |
| Height (cm) | 171.6 (156-188) |
| Weight (kg) | 63.7 (43-92) |
| BMI (kg/m ²) | 21.5 (16.2-27.7) |

^aAverage values are given in each category.

Before each data collection session, the researcher conducts informed consent and explains the spirometry protocol to the subject according to the ATS/ERS standard [30]. A video clip recorded by a professional pulmonologist [2] is also presented to the subject for a better understanding of the procedure. Nevertheless, We notice some maneuvers from several subjects that do not strictly comply with the guideline. We keep these data in our dataset because we want the model to learn how to predict the correct F-V curve even if the maneuver contains deficiencies. Sec. 6.9 evaluates EarSpiro’s performance on these data.

We have collected 180 samples in total. Since the spirometer and the earphone are not synchronized, we manually align the audio recordings with the ground truth. The experiments are approved by the IRBs of our institutions¹.

6.2 Measurement Accuracy

In this section, we evaluate the general performance of EarSpiro. We first discuss EarSpiro’s accuracy of flow rate estimation and F-V curve estimation. Then, we discuss the accuracy of predicting lung function indices. In this section, unless otherwise noted, we use leave-one-subject-out (LOSO) validation to evaluate the performance.

6.2.1 Flow Rate Estimation. Since the primary output of EarSpiro is the expiratory/inspiratory flow rate, in this subsection, we evaluate EarSpiro’s performance of flow rate estimation. Because EarSpiro performs segmentation before estimating the flow rate, the predicted and actual flow rates can be misaligned. Therefore, we manually pad zeros before and after the predicted flow rate signal to align the predicted and the actual flow rate. We use

¹HKUST Human Research Ethics Protocol No. HREP-2021-0152 and SUSTech IRB No. 20210082.

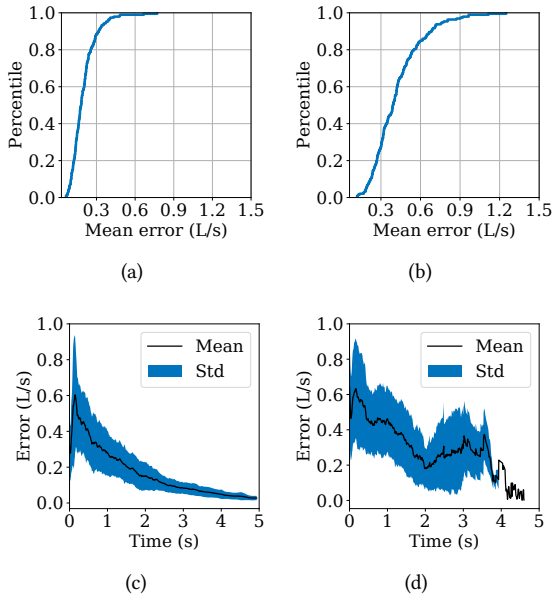


Fig. 16. Flow rate estimation error. (a) CDF plot of mean absolute error of expiration estimation. (b) CDF plot of mean absolute error of inspiration estimation. (c) Expiratory flow rate error distribution over time. (d) Inspiratory flow rate error distribution over time.

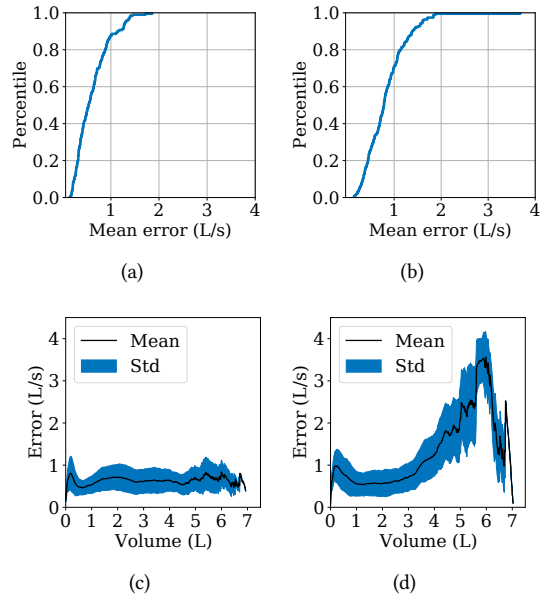


Fig. 17. F-V curve estimation error. (a) CDF plot of mean absolute error of expiratory limb estimation. (b) CDF plot of mean absolute error of inspiratory limb estimation. (c) Expiratory limb's error distribution over the volume-axis. (d) Inspiratory limb's error distribution over the volume-axis.

absolute mean error as the metric to measure the error. Specifically, for each pair of actual flow rates and its prediction, we compute the absolute error of each sampling point and take the average of all sampling points. The CDF plots of the error are shown in Figs. 16(a) and 16(b). The mean errors of flow rate estimation are $0.20L/s$ for expiration and $0.42L/s$ for inspiration. Note that the expiration estimation is more accurate than the inspiration. We believe this is because the inspiratory airflow sound is much lower than that of expiration, thus making it harder for inspiration estimation. We also show the error distributions over time in Figs. 16(c) and 16(d). It is worth noting that in Fig. 16(d), the tail portion of the plot has no standard deviation, this is because there is only one inspiration sample that takes almost 5 seconds while others are all within 4 seconds.

6.2.2 Flow-Volume Curve Estimation. We use mean absolute error to measure EarSpiro's performance of F-V curve estimation. Specifically, when computing the mean absolute error between a true F-V and the predicted one, we first compute the point-wise absolute error over the volume axis with a step size of $0.01L$ and then take the average of all the samples as the error of this pair of F-V curves. The CDF plots of the errors of expiration limb and the inspiration limb are shown in Fig. 17(a) and 17(b). The mean expiratory limb estimation error is $0.61L/s$, and the error for the inspiratory limb is $0.83L/s$. This error is almost twice as large as the flow rate estimation error presented in the last section. This result is reasonable because before generating the F-V curve, EarSpiro computes airflow volume by taking the cumulative summation of the flow rate, and errors in flow rate estimation will accumulate by the cumulative summation operation. Figs. 17(c) and 17(d) also present the error distributions of F-V curve estimation along the volume axis. Note that because of error accumulation, the end of the inspiratory limb can have an error of more than $3L/s$. Nevertheless, this huge error is rare, as indicated in the

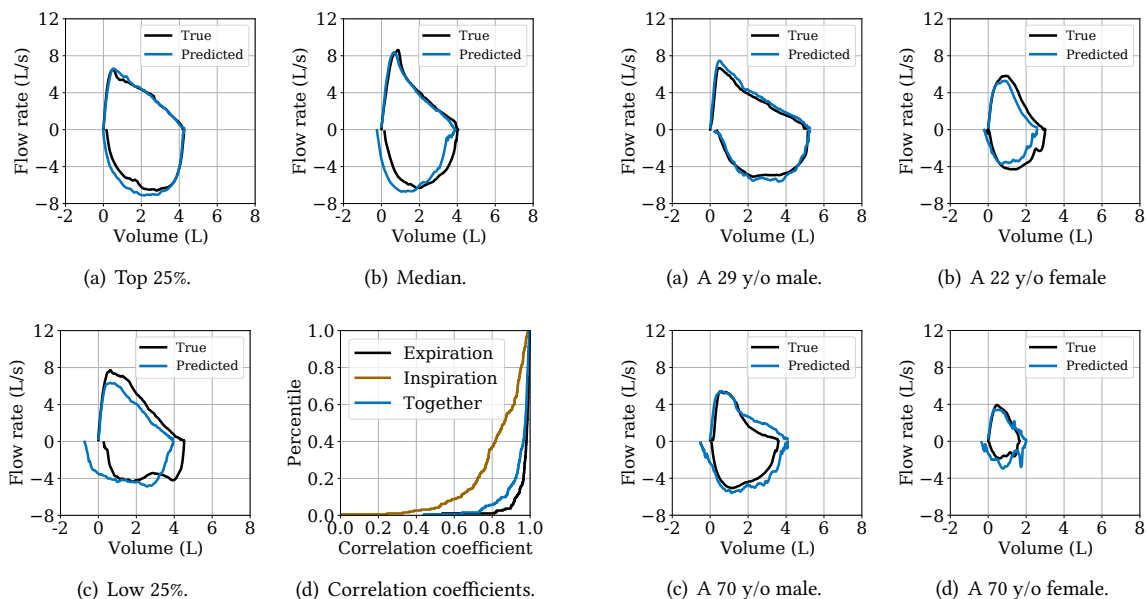


Fig. 18. F-V curve estimation. (a-c) Examples of estimated F-V curve ranked by mean absolute error. (d) Pearson correlation coefficients of F-V curve estimation.

Fig. 19. Examples of F-V curve of different age groups. (a) Sample of a 29 y/o male. (b) Sample of a 22 y/o female. (c) Sample of a 70 y/o male. (d) Sample of a 70 y/o female.

CDF plot. We rank all the estimated F-V curves by their average mean absolute error, and we present the sample ranked at the top 25%, top 50%, and low 25% in Figs. 18(a), 18(b), and 18(c).

We also use the Pearson correlation coefficient to measure the similarity between the true F-V curve and the predicted one. The result is shown in Fig. 18(d). In summary, EarSpiro can estimate the F-V curve with a mean correlation coefficient of 0.94 compared with the ground truth. Specifically, the correlation coefficients of the expiratory limb and the inspiratory limb are 0.97 and 0.82, respectively. This result indicates that EarSpiro can reliably capture the basic shape of a user's F-V curve. Fig. 19 shows additional examples of the F-V curve estimation of subjects of different age groups.

6.2.3 Lung Function Indices Estimation. Fig. 20 shows the CDF plots of the percent estimation error of four primarily used lung function indices, that is, PEF, FVC, FEV1, and FEV1/FVC. The mean errors of these four lung function indices are 6.5%, 9.9%, 7.8%, and 5.1%. The overall mean error of these four indices is 7.3%. The FEV1/FVC has the lowest error among these four indices. We believe this is because of the composite nature of this index, where the fraction of FEV1 and FVC cancels out the correlated errors from each other. Note that this error of 5.1% is very close to in-clinic spirometers whose errors are usually around 5% [27, 50]². Considering the widely accepted standard for COPD diagnosis is to examine the FEV1/FVC value [48], we believe EarSpiro can provide a usable and cost-efficient solution for out-of-clinic lung function tracking. We also present our evaluation result to a pulmonologist. The pulmonologist states that although the accuracy of our system is not as good as that of a conventional spirometer, it can be used in home scenarios for everyday lung function monitoring. Therefore, we

²The authors in [27] compared the perspective measures of two clinical spirometers based on the same group of subjects. They found that the measurement differences of PEF, FVC, FEV1, FEV1/FVC are 9.2%, 5.2%, 5.1% and 3.2% respectively.

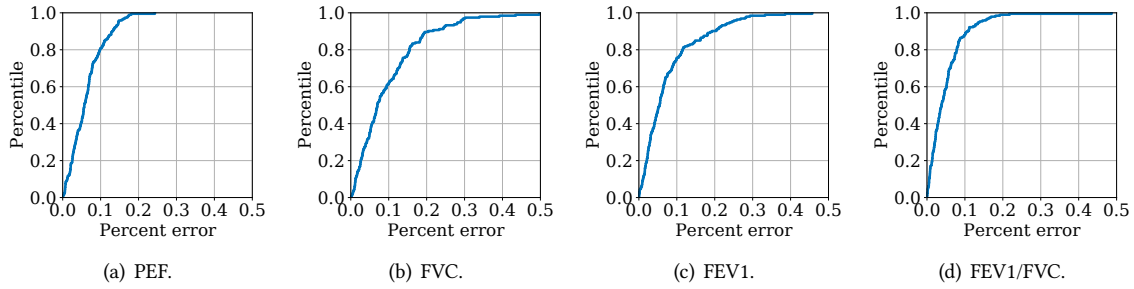


Fig. 20. Estimation error of four common lung function indices.

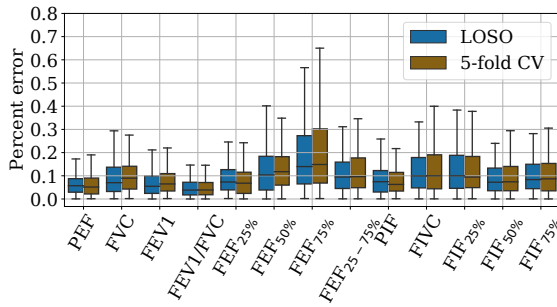


Fig. 21. Lung function index estimation error.

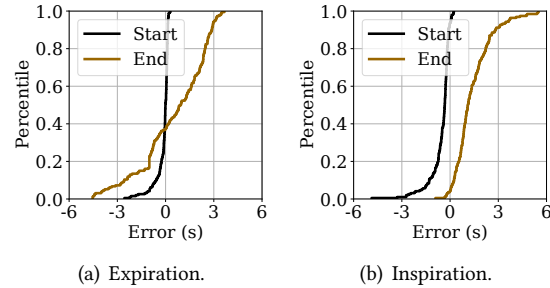


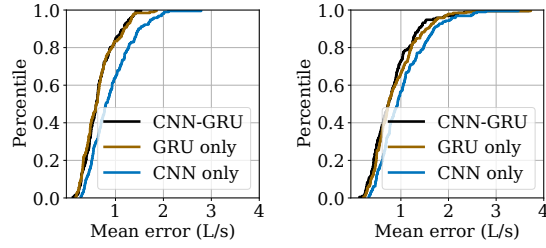
Fig. 22. Segmentation error of the expiration phase and inspiration phase.

believe users can use EarSpiro as a tracking tool for early warning signs of abnormal lung function. The user could use EarSpiro regularly at home and go to a clinic for further assessment if EarSpiro detects abnormalities. We note that the error of FVC is slightly higher than the other three. This is because the output of the deep learning model is the flow rate, and FVC is computed by taking the summation of the flow rate estimation. In this way, errors are accumulated.

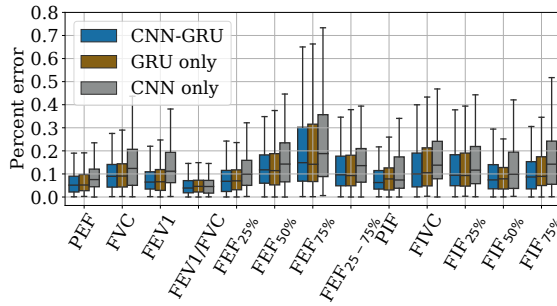
In addition to the above four commonly used lung function indices, in-clinic spirometers usually provide other lung function indices that are also sensitive to lung diseases, as discussed in Sec. 2.1. Therefore, since EarSpiro can estimate the full F-V curve, we also evaluate EarSpiro’s accuracy in estimating these additional lung function indices. The result is shown in Fig. 21. Note that in addition to leave-one-subject-out validation, we also present the result of 5-fold cross-validation (CV) where data from the same subject are either in the training or testing sets. This result indicates that apart from the expiratory lung function indices, EarSpiro can also estimate the inspiratory indicators. Again, the inspiratory error is higher than the expiratory indicators because the inspiratory measurement is harder than expiration. We observe that the errors of FEF and FIF increase proportionally to volume. This is because the error of volume estimation accumulates over time, so the larger the volume is, the larger the error will be.

6.3 Accuracy of Expiratory and Inspiratory Phase Segmentation

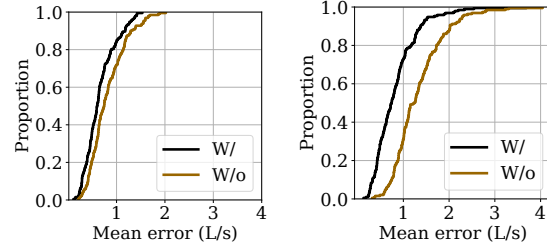
In this section, we evaluate the performance of the expiratory and inspiratory phase segmentation algorithms. We compute the absolute errors of the expiratory and inspiratory phases’ estimated start time and end time. The



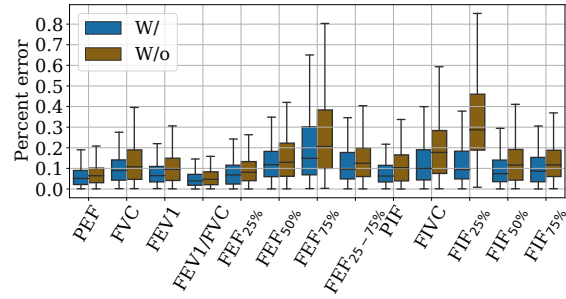
(a) Error of expiratory limb. (b) Error of inspiratory limb.



(c) Error of lung function index estimation.



(a) Error of expiratory limb. (b) Error of inspiratory limb.



(c) Error of lung function index estimation.

Fig. 23. Effectiveness of the deep learning architecture.

Fig. 24. Effectiveness of the data augmentation.

CDF plots of the results are shown in Fig. 22. Since the expiration signal contains a sharp energy rise, it is easy to locate the start of the expiratory phase. Therefore, the estimated expiratory start time is almost perfect. However, since the airflow sound is weak at the end of expiration, the error of end time estimation can be large. As for the inspiratory phase, although the inspiratory signal can be very weak sometimes, by leveraging our segmentation algorithm, EarSpiro can still achieve an accurate segmentation. Note that because we manually extend the start time estimation and the end time estimation of an inspiration sample for one second (discussed in Sec. 4.2), the start time estimation is generally before the actual start time and the end time estimation generally exceeds the actual end time. As we discussed in Sec. 4.3.1, the augmentation techniques can tolerate this error.

6.4 Benefits of the CNN-GRU architecture

As discussed in Sec. 4.3, the deep learning model used in EarSpiro contains CNN and GRU units. To evaluate the effectiveness of this design, we conduct an ablation experiment to test the gain of each unit. Specifically, we test the performance when only CNN is used, only GRU is used, and the hybrid model is used. We use 5-fold cross-validation for evaluation. The result is shown in Fig. 23. The mean F-V curve estimation errors when the above three models are used are $0.905L/s$, $0.654L/s$ and $0.652L/s$ for the expiratory limb, and $1.038L/s$, $0.880L/s$ and $0.824L/s$ for the inspiratory limb. This result indicates that the CNN-GRU architecture achieves much better performance than CNN. Compared with the GRU architecture, the hybrid architecture has little improvement in expiration estimation, but one can observe a larger enhancement in inspiration estimation.

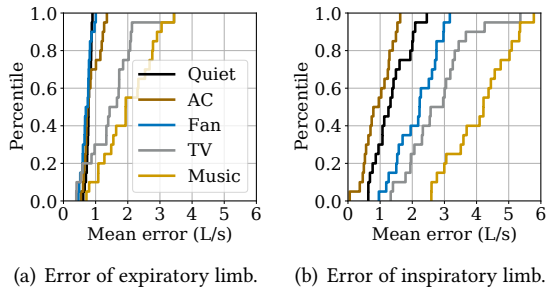


Fig. 25. Impact of background noise.

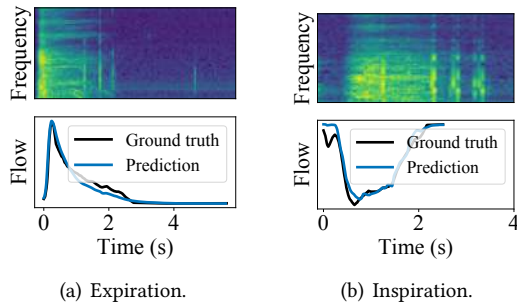
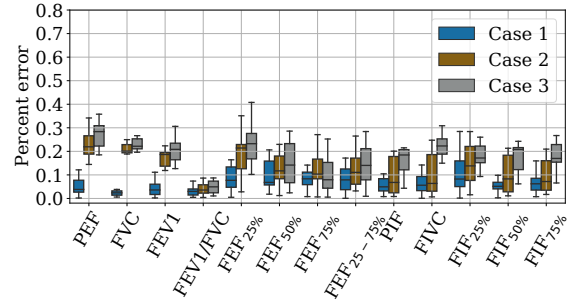
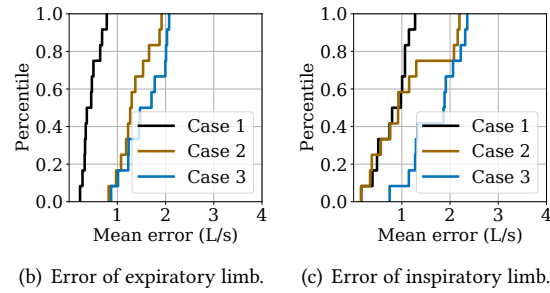


Fig. 26. Interference of teeth knocking.



(a) Error of lung function index estimation.



(b) Error of expiratory limb. (c) Error of inspiratory limb.

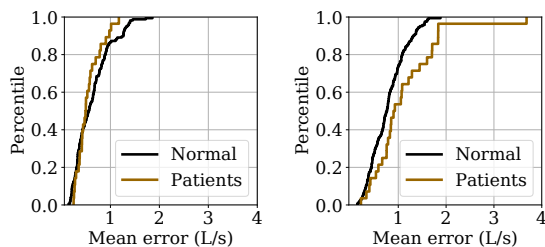
Fig. 27. Impact of Earphone position.

6.5 Benefits of Data Augmentation

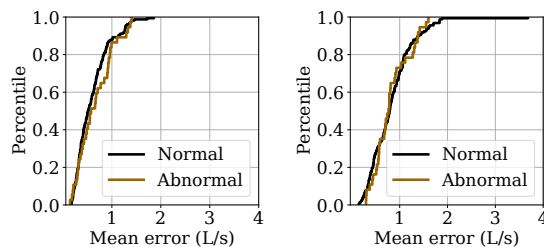
As discussed in Sec. 4.3, we use two data augmentation techniques to train the expiratory and inspiratory models. As a consequence, the dataset size increased from 180 samples to 900 samples. In this subsection, we evaluate the effectiveness of these designs. We use 5-fold cross-validation for this evaluation. We retrain the expiratory and inspiratory models without data augmentation. The performance of the new models is summarized in Fig. 24. Specifically, if trained without data augmentation, the mean error of the expiratory limb increases from $0.51L/s$ to $0.65L/s$, and the mean error of the inspiratory limb increases from $0.82L/s$ to $1.33L/s$. This result indicates that the data augmentation techniques can significantly improve the performance of the deep learning models.

6.6 Impact of Background Noise

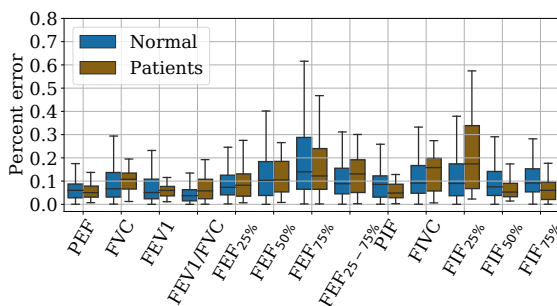
In the above experiments, spirometry is conducted in a quiet environment. In this subsection, we evaluate the performance of EarSpiro when there is additional noise in the environment. We create four noisy environments with the following noise sources: an air conditioner, a fan, a television, and a loudspeaker that plays music. We use a smartphone app [55] to measure the noise level in these noisy environments. Specifically, the noise level of a quiet room is $17dB$, and for the above four noisy environments, the noise levels are $22dB$, $31dB$, $42dB$, and $64dB$, respectively. We recruit four subjects to conduct five spirometry tests in the four noisy and standard quiet environments. The result is shown in Fig. 25. This result indicates that EarSpiro can reliably estimate the four common lung function indices under the first noisy environment. However, when the surrounding is too noisy, say a fan is running, or a TV is open, the F-V curve generated by EarSpiro is less reliable, especially for the inspiratory limb. Furthermore, when the environmental noise is too loud, say, reaches $64dB$ in our experiment, the lung function indices estimated by EarSpiro are unreliable.



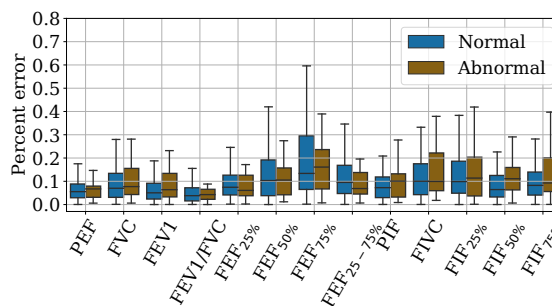
(a) Error of expiratory limb. (b) Error of inspiratory limb.



(a) Error of expiratory limb. (b) Error of inspiratory limb.



(c) Error of lung function index estimation.



(c) Error of lung function index estimation.

Fig. 28. Evaluation on six patients.

Fig. 29. Evaluation on abnormal F-V curves.

As discussed in Sec. 4.3, the deep learning model can filter out the noise caused by inertial body sounds such as teeth-knocking sounds. Here, we use examples to show EarSpiro's ability to mitigate such noises. We recruit one subject to participate in this experiment. We ask the subject to perform the maneuver while intentionally knocking the mouthpiece with her teeth. Fig. 26 shows the corresponding expiration and inspiration measurement. These two examples imply that EarSpiro can tolerate inertial body sounds.

6.7 Impact of Earphone's Position

When using EarSpiro, the user must wear a pair of earphones. In this section, we evaluate the impacts of the earphones' position on EarSpiro's performance. We create the following three scenarios for this evaluation. (i) The user wears the earphones correctly, (ii) The user hangs one earphone out of his/her ear while wearing the other earphone correctly. Without loss of generality, we assume the right earphone is worn correctly. (iii) The user hangs both of the earphones out of his/her ears. We recruit four subjects to participate in this evaluation. The F-V curve estimation errors in the above three scenarios are shown in Figs. 27(b) and 27(c). The lung function index estimation errors are shown in Fig. 27(a). This evaluation indicates that to let EarSpiro work properly, it is essential to instruct the user to wear both of the earphones correctly, *i.e.*, both earbuds must be worn in the ear canals.

6.8 Performance on Subjects with Lung Function Impairments

As explained in Sec. 6.1, our dataset contains six self-reported asthma patients. In this subsection, we show EarSpiro's performance on this group of subjects. Note that these subjects only contribute dozens of spirometry samples which may not be enough to prove EarSpiro's performance on lung disease patients. We present this

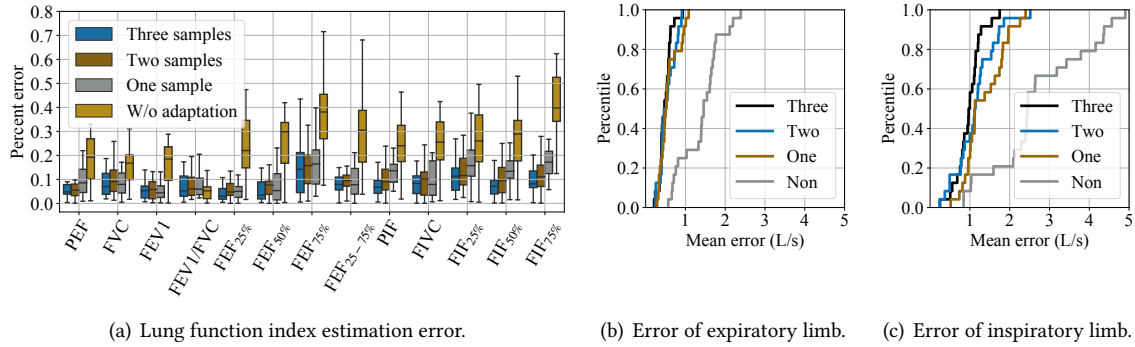


Fig. 30. EarSpiro's performance when used with a funnel.

result to showcase EarSpiro's potential for measuring the F-V curve of lung disease patients. As we discuss in Sec. 7, more data from patients are required to evaluate EarSpiro's clinical value on lung disease patients. We compute the mean absolute error of the estimated F-V curve from these patients, and the errors of the expiratory and inspiratory limbs are shown in Figs. 28(a) and 28(b). We also summarize the prediction error of the lung function indices as shown in Fig. 28(c). The above figures also include the corresponding results from normal subjects for reference.

6.9 Performance on Abnormal F-V Curves

Spirometry requires a subject to expire and inspire with his/her maximal efforts. A subject may produce a less useful result if there is a lack of practice. One of the common spirometry deficiencies is sub-maximal expiration effort, characterized by a rounded shape of the expiratory limb where no distinct peak can be found (e.g., Fig. 3(a)). We evaluate EarSpiro's performance on this specific abnormal F-V pattern in this subsection. We have identified 37 samples that fall into this category, and we evaluate EarSpiro's accuracy on these samples. Figs. 29(a) and 29(b) show the expiratory error and the inspiratory error of the estimated F-V curve. Fig. 29(c) shows the accuracy of lung function index prediction. Again, these figures include the corresponding performance of normal F-V curves.

6.10 Accuracy with Funnel-like Objects

Apart from the standard mouthpiece, we evaluate EarSpiro's performance on a funnel shown in Fig. 15. We recruit six subjects (five males and one female) to participate in this evaluation. We first ask the participants to perform three spirometry tests with the standard spirometer to collect three sets of lung function indices. Then, we ask the participants to perform seven spirometry maneuvers through the funnel. When using the funnel, we let the participants hold the funnel tip with their mouth and exhale and inhale according to the standard spirometry protocol while wearing the earphone. Among the seven maneuvers, the data from the first one, two, and three maneuver(s) is(are) used, together with the reconstructed flow rate data, for model adaptation. Finally, we use the data from the last four maneuvers to test the performance of the adapted model. The result is shown in Fig. 15. The mean F-V curve estimation errors when three samples, two samples, one sample and no sample are used are 0.49L/s, 0.54L/s, 0.57L/s and 1.40L/s for the expiratory limb, and 0.98L/s, 1.14L/s, 1.36L/s and 2.68L/s for the inspiratory limb.

7 LIMITATIONS AND DISCUSSION

Though we have shown that EarSpiro can provide comprehensive and reliable spirometry for common use cases, this work still has its limitations. In this section, we discuss these limitations, and we point out future research directions to improve EarSpiro.

- (1) **Variety of subjects.** In this work, we recruit 60 subjects who are all college students or staff. Most subjects are healthy adults. However, To better evaluate EarSpiro's robustness, more subjects with more age groups and social backgrounds need to be included in the future study. Also, it is better to conduct a clinical study to thoroughly verify the clinical value of EarSpiro for CRD patients. We leave this to our future work.
- (2) **Accuracy of some lung function indices.** As discussed before, all the volume-related indices (such as FVC, FEF_{x%}) are less accurate than the others. This is because the primary objective of the deep learning model is to estimate the flow rate. Since the volume is computed by taking the accumulative summation of the flow rate estimation, the errors of the flow rate estimation will accumulate, which results in more significant errors in the volume estimation. To solve this issue, future research can use multi-task learning techniques [60] to predict the flow rate and volume at the same time. In this way, there is less dependency between flow rate and volume, and thus the errors in volume prediction might be lower. A recent work [16] uses an end-to-end deep learning model to predict all the lung function indices and another deep learning model to regress the F-V curve from the indices. This is also a potential solution to the problem.
- (3) **Accuracy of inspiratory measurement.** As shown in Sec. 6, the performance of the inspiratory measurement of EarSpiro is worse than that of the expiratory measurement. This is because the inspiratory flow rate is lower than expiration; thus, it is hard to predict the flow rate based on weak audio signals. Future research can implement weight sharing between the expiratory and the inspiratory models. This might improve the performance of the inspiratory model since, despite the low flow rate of inspiration, the expiratory sound and inspiratory sound still share some similarities. Sharing weights can make the best use of the available data.
- (4) **Adaptation to funnel-like objects.** In the current design, once the flow rate estimator is adapted to a funnel-like object by a user, the funnel can only be used by this user. However, it would be valuable that once the model is adapted to an object by a user, the user can share the model parameters with other users so that the other users can use EarSpiro with the same funnel-like object (not the exact same one but the same product produced by the same manufacturer). Future research can use a more sophisticated model adaptation framework to realize this potential design.

8 RELATED WORK

In this section, we review the research works related to this paper. They fall into the following two categories: mobile spirometry systems and earphone-based sensing systems.

8.1 Mobile Spirometry Systems

Recent efforts have shown that spirometry can be conducted with mobile devices. A comparison of these works in terms of functionality is shown in Tab. 3. Furthermore, since the target of these works is to predict the four common expiratory lung function indices, we also compare the prediction accuracy of these indices between EarSpiro and these works in Tab. 4. SpiroSmart [37] is a smartphone-based solution that predicts lung function indices by analyzing the recorded expiratory sound. SpiroCall [27] further extends the idea of SpiroSmart to telephony and implements the spirometry system as a phone call service so that the system can also be used in situations where a smartphone is not available. EarSpiro is different from these two works. Instead of treating the audio recording as a whole, EarSpiro captures the real-time correlation between flow rate and airflow sound so that the F-V curve can be estimated. Although SpiroSmart and SpiroCall can also produce an F-V curve, it is

Table 3. Comparison with related works.

| Work | Modality | Measure | Lung function index measurement | | F-V curve measurement | |
|---------------------------|------------------------------------|-------------------|--|---|-----------------------|-------------|
| | | | Expiration | Inspiration | Expiration | Inspiration |
| SpiroSmart [37] | Smartphone & mouthpiece (optional) | Audio | PEF, FEV1, FVC, FEV1/FVC | None | No* | No |
| BKSpiro [53] | Smartphone | Audio | PEF, FEV1, FVC | None | No* | No |
| SpiroCall [27] | Any phone & mouthpiece (optional) | Audio | PEF, FEV1, FVC, FEV1/FVC | None | No* | No |
| Karser <i>et al.</i> [33] | Smartphone & vortex whistle | Audio | PEF, FEV1, FVC, FEV1/FVC, FEF _{25-75%} | None | No* | No |
| Thap <i>et al.</i> [52] | Smartphone | Audio | FEV1/FVC | None | No* | No |
| SpiroSonic [50] | Smartphone | Chest wall motion | FEV1, FVC, FEV1/FVC, PEF [†] | None | No | No |
| mmFlow [16] | Millimeter wave radar | Vibration | PEF, FEV1, FVC, FEV1/FVC, FEF _{25%} , FEF _{50%} , FEF _{75%} , MMEF | None | Yes | No |
| EarSpiro | Earphone & mouthpiece | Audio | PEF, FEV1, FVC, FEV1/FVC, FEF _{25%} , FEF _{50%} , FEF _{75%} , FEF _{25-75%} | PIF, FIVC, FIF _{25%} , FIF _{50%} , FIF _{75%} | Yes | Yes |

* A curve is produced for visualization purpose. No quantitative evaluation is provided.

† The performance of PEF prediction is not evaluated.

mainly for visualization, and no evaluation of the accuracy of the F-V curve is presented. Besides, while these works focus on expiratory measurement, EarSpiro also supports inspiratory measurement. SpiroSonic [50] uses a smartphone to capture chest motions during spirometry using acoustic sensing methods. SpiroSonic can estimate the lung function indices by analyzing the chest motion patterns. Again, the objective of SpiroSonic is to predict the expiratory lung function indices, while in EarSpiro, we are also interested in the F-V curve and inspiratory measurement. Kaiser *et al.* [33] designs a set of vortex whistles that can produce sound whose frequency is proportional to the airflow speed. Therefore, the F-V curve can be estimated by analyzing the frequency of the recorded expiratory sound. The difference between this work and our work is that EarSpiro only requires a mouthpiece available in the market, and EarSpiro supports inspiratory measurement. mmFlow [16] uses a millimeter-wave (mmWave) radar to conduct spirometry. It requires a subject to perform the expiratory maneuver to the radar while the radar measures its vibration during the maneuver and transforms the vibration signal to the F-V curve. Although mmFlow can estimate the F-V curve as well, our work is different from mmFlow in that EarSpiro is a low-cost solution while a mmWave radar is usually expensive, and mmFlow can only measure expiration while EarSpiro can measure both the expiratory phase and the inspiratory phase. Plus, mmFlow is only evaluated with a flow-volume calibrator [5] while we evaluate EarSpiro on 60 subjects. RF-RVM [21] is a respiratory volume monitoring system based on RFID. RF-RVM tracks the respiratory volume by processing the

Table 4. Comparison with related works in expiratory index measurement.

| Metric | Work | Lung function index | | | |
|--------------------------------------|-----------------------------|---------------------|---------------|---------------|--------------|
| | | PEF | FVC | FEV1 | FEV1/FVC |
| Mean percentage absolute error | SpiroSmart [37] | 5.2% | 4.8% | 6.3% | 4.0% |
| | BKSpiro [53] | 7.9% | 6.7% | 8.0% | / |
| | SpiroCall [27] ** | 6.6% or 8.5% | 6.5% or 7.0% | 8.5% or 7.1% | 5.5% or 8.9% |
| | Thap <i>et al.</i> [52] | / | / | / | 7.1% |
| | EarSpiro | 6.5% | 9.9% | 7.8% | 5.1% |
| IQR [†] of percentage error | Karser <i>et al.</i> [33] * | 20% | 17% | 12% | 8.0% |
| | EarSpiro | 11% | 14% | 10% | 7.8% |
| Median percentage absolute error | SpiroSonic [50] **‡ | / | 2.1% or 11.5% | 2.5% or 15.5% | 1.6% or 8.5% |
| | EarSpiro | 5.7% | 7.0% | 5.5% | 3.8% |
| Median absolute error | mmFlow [16] ◊ | 0.40L/s | 0.03L | 0.05L | / |
| | EarSpiro | 0.45L/s | 0.30L | 0.15L | / |

* The result is estimated by reading the error chart in the paper.

** The error of using or not using a whistle is given separately.

† IQR: interquartile range.

‡ The errors are computed separately based on two dataset. One dataset contains five healthy adults and the other one contains 83 pediatric patients.

◊ The error is computed by simulation using a 3-L flow-volume calibrator.

phase derived from several RFID tags attached to the user. EarSpiro is different from RF-RVM in that EarSpiro provides spirometry during forced expiration and inspiration while RF-RVM is designed to track normal breathing.

8.2 Earphone-based Sensing Systems

Recently, plenty of earphone-based sensing systems have been proposed. These systems mainly fall into three categories: human-computer interaction (HCI), human authentication, and health monitoring. We first summarize the earphone-enabled HCI systems. Ear-AR[59] is proposed to support acoustic augmented reality (AAR) by jointly using the inertial sensors in earphones and smartphones, and acoustics in the earphones. EarFieldSensing [40] introduces a gesture-based input system by sensing the facial muscle movements from the electric field changes with the electrodes placed inside the ear canal. OESense [38] is a human motion sensing system leveraging the occlusion effect inside the human ear. EarSense [45] is a teeth gesture sensing system powered by vibration sensors in the earphone. Next, we summarize the earphone-based system for user authentication. Takayuki *et al.* [17], EarEcho [25] and EarDynamic [57] are similar authentication systems that recognize a legitimate user by extracting the ear canal signature by analyzing the acoustic features from the ear canal. For health-monitoring applications, LIBS [42] can monitor whole-night sleep staging by incorporating electronic components from the earphone. WAKE [43] is designed to detect microsleep events by capturing electronic signals behind the ears. Hearables [29] is a multimodal physiological sensing system integrated with multiple sensors including in-ear electroencephalogram (EEG) sensors and microphones. Heartphones [44] monitors heart rate by embedding a

photoplethysmography (PPG) sensor. eBP [20] measures blood pressure from the user's ear by integrating a pulse sensor, a digital air pump, and a PPG sensor. Martin *et al.* [39] uses an in-ear microphone to measure the heart and breathing rates by analyzing the audio features.

9 CONCLUSION

In this paper, we present EarSpiro, an earphone-based spirometry solution that can estimate the whole F-V curve, including both the expiratory and the inspiratory limb. EarSpiro uses an earphone-equipped microphone to collect expiratory and inspiratory sound and leverages deep learning techniques to interpret the sound signal into an F-V curve. We build a prototype for EarSpiro, and Extensive experiments with 60 subjects show that EarSpiro can accurately estimate expiratory and inspiratory flow rates with mean errors of 0.20L/s and 0.42L/s, respectively. EarSpiro can also reliably estimate the F-V curve with mean errors of 0.61L/s for the expiratory limb and 0.83L/s for the inspiratory limb. The mean correlation coefficient between the estimated F-V curve and the true one is 0.94. The mean estimation error for four standard lung function indices is 7.3%.

ACKNOWLEDGMENTS

This work was supported in part by Hong Kong RGC under Contract CERG 16203719, 16204820, 16206122, Contract R8015 and Contract R6021-20, in part by Shenzhen Science, Technology and Innovation Commission Basic Research Project under Grant No. JCYJ20180507181527806. The authors would like to thank Mr. Linfei Ge for his help in the prototyping of this work.

REFERENCES

- [1] 2018. Coal Workers' Health Surveillance Program: Spirometry. <https://www.cdc.gov/niosh/topics/cwhsp/coalminerhealth.html>. Accessed Mar 28, 2022.
- [2] 2018. Taking a Spirometry Test. <https://www.youtube.com/watch?v=Zs8Fs5HajHs>. Accessed Apr 7, 2022.
- [3] 2021. AirPods Pro — Apple (UK). <https://www.apple.com/uk/airpods-pro/>. Accessed Aug 25, 2021.
- [4] 2021. Audacity. <https://www.audacityteam.org>. Accessed Aug 15, 2021.
- [5] 2021. Flow-Volume Calibrator (FVC-3000). <https://www.jonesmedical.com/product/flow-volume-calibrator-fvc-3000/>. Accessed Sep 6, 2021.
- [6] 2021. MINISO Marvel Earphones. <https://www.miniso-au.com/en-au/product/145169/marvel-earphones>. Accessed Aug 14, 2021.
- [7] 2021. PyTorch. <https://pytorch.org>. Accessed Aug 15, 2021.
- [8] 2021. Raspberry Pi 3 Model B+. <https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/>. Accessed Aug 15, 2021.
- [9] 2021. ReSpeaker 6-Mic Circular Array kit for Raspberry Pi — Seed Wiki. https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/. Accessed Aug 15, 2021.
- [10] 2021. Spirolab — Spirometry, Oximetry, Mobile Health. <https://www.spirometry.com/prodotti/spirolab/>. Accessed Aug 23, 2021.
- [11] 2021. SPU0414HR5H-SB-7. <https://www.digikey.com/en/products/detail/knowles/SPU0414HR5H-SB-7/2420969>. Accessed Aug 15, 2021.
- [12] 2021. UBREATH Spirometer System (PF680). <https://www.e-linkcare.com/spirometer-system-pf680-product/>. Accessed Aug 23, 2021.
- [13] 2021. *What is a Pulmonary Function Test?* Retrieved September 1, 2021 from <https://www.morgansci.com/support/what-is-a-pulmonary-function-test/>
- [14] 2022. EasyOne Air. <https://nddmed.com/products/spirometers/easyone-air>. Accessed Aug 4, 2022.
- [15] 2022. Occupational Lung Diseases. <https://www.cedars-sinai.org/health-library/diseases-and-conditions/o/occupational-lung-diseases.html>. Accessed Mar 28, 2022.
- [16] Aakriti Adhikari, Austin Hetherington, and Sanjib Sur. 2021. mmFlow: Facilitating At-Home Spirometry with 5G Smart Devices. In *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 1–9.
- [17] Takayuki Arakawa, Takafumi Koshinaka, Shohei Yano, Hideki Irisawa, Ryoji Miyahara, and Hitoshi Imaoka. 2016. Fast and accurate personal authentication using ear acoustics. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 1–4.
- [18] Timothy Barreiro and Irene Perillo. 2004. An approach to interpreting spirometry. *American family physician* 69, 5 (2004), 1107–1114.
- [19] Eamann Breatnach, Gypsy C Abbott, and Robert G Fraser. 1984. Dimensions of the normal human trachea. *American Journal of Roentgenology* 142, 5 (1984), 903–906.

- [20] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. 2019. 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*. 1–17.
- [21] Xiangmao Chang, Jiahua Dai, Zhiyong Zhang, Kun Zhu, and Guoliang Xing. 2021. RF-RVM: Continuous Respiratory Volume Monitoring With COTS RFID Tags. *IEEE Internet of Things Journal* 8, 16 (2021), 12892–12901.
- [22] Arvind K Das, Lawrence D Davanzo, George J Poiani, Peter G Zazzali, Anthony T Scardella, Martha L Warnock, and Norman H Edelman. 1999. Variable extrathoracic airflow obstruction and chronic laryngotracheitis in Gulf War veterans. *Chest* 115, 1 (1999), 97–101.
- [23] Yong Jun Choi Do Sun Kwon, Tae Hee Kim, Min Kwang Byun, Jae Hwa Cho, Hyung Jung Kim, and Hye Jung Park. 2020. FEF25-75% values in patients with normal lung function can predict the development of chronic obstructive pulmonary disease. *International Journal of Chronic Obstructive Pulmonary Disease* 15 (2020), 2913.
- [24] AI Dyachenko, GA Lyubimov, IM Skobeleva, and MM Strongin. 2011. Generalization of the mathematical model of lungs for describing the intensity of the tracheal sounds during forced expiration. *Fluid Dynamics* 46, 1 (2011), 16–23.
- [25] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
- [26] AY Glazova, VI Korenbaum, AE Kostiv, OI Kabancova, AA Tagiltcev, and SN Shin. 2018. Measurement and estimation of human forced expiratory noise parameters using a microphone with a stethoscope head and a lapel microphone. *Physiological measurement* 39, 6 (2018), 065006.
- [27] Mayank Goel, Elliot Saba, Maia Stiber, Eric Whitmire, Josh Fromm, Eric C Larson, Gaetano Borriello, and Shwetak N Patel. 2016. Spirocall: Measuring lung function over a phone call. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5675–5685.
- [28] Google. 2021. *Google Colab*. <https://colab.research.google.com/>
- [29] Valentin Goverdovsky, Wilhelm Von Rosenberg, Takashi Nakamura, David Looney, David J Sharp, Christos Papavassiliou, Mary J Morrell, and Danilo P Mandic. 2017. Hearables: Multimodal physiological in-ear sensing. *Scientific reports* 7, 1 (2017), 1–10.
- [30] Brian L Graham, Irene Steenbruggen, Martin R Miller, Igor Z Barjaktarevic, Brendan G Cooper, Graham L Hall, Teal S Hallstrand, David A Kaminsky, Kevin McCarthy, Meredith C McCormack, et al. 2019. Standardization of spirometry 2019 update. An official American thoracic society and European respiratory society technical statement. *American journal of respiratory and critical care medicine* 200, 8 (2019), e70–e88.
- [31] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* 28, 1 (1979), 100–108.
- [32] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, and Raj Reddy. 2001. *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR.
- [33] Spencer Kaiser, Ashley Parks, Patrick Leopard, Charlie Albright, Jake Carlson, Mayank Goel, Damoun Nassehi, and Eric C Larson. 2016. Design and learnability of vortex whistles for managing chronic lung function via smartphones. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 569–580.
- [34] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [35] Vladimir I Korenbaum, Irina A Pochekutova, Anatoly E Kostiv, Veronika V Malaeva, Maria A Safronova, Oksana I Kabantsova, and Svetlana N Shin. 2020. Human forced expiratory noise. Origin, apparatus and possible diagnostic applications. *The Journal of the Acoustical Society of America* 148, 6 (2020), 3385–3391.
- [36] Meir Kryger, Frederick Bode, Ral Antic, and Nicholas Anthonisen. 1976. Diagnosis of obstruction of the upper and central airways. *The American journal of medicine* 61, 1 (1976), 85–93.
- [37] Eric C Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N Patel. 2012. SpiroSmart: using a microphone to measure lung function on a mobile phone. In *Proceedings of the 2012 ACM Conference on ubiquitous computing*. 280–289.
- [38] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. OESense: employing occlusion effect for in-ear human sensing. *arXiv preprint arXiv:2106.08607* (2021).
- [39] Alexis Martin and Jérémie Voix. 2017. In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring. *IEEE Transactions on Biomedical Engineering* 65, 6 (2017), 1256–1263.
- [40] Denys JC Matthies, Bernhard A Strecker, and Bodo Urban. 2017. 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*. 1911–1922.
- [41] S Mehta and HM Myat. 1984. The cross-sectional shape and circumference of the human trachea. *Annals of the Royal College of surgeons of England* 66, 5 (1984), 356.
- [42] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-Kashani, Ann C Halbower, and Tam Vu. 2016. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 230–244.
- [43] Nhat Pham, Tuan Dinh, Zohreh Raghebi, Taeho Kim, Nam Bui, Phuc Nguyen, Hoang Truong, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, et al. 2020. WAKE: a behind-the-ear wearable system for microsleep detection. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 404–418.

- [44] Ming-Zher Poh, Kyunghye Kim, Andrew D Goessling, Nicholas C Swenson, and Rosalind W Picard. 2009. Heartphones: Sensor earphones and mobile application for non-obtrusive health monitoring. In *2009 International Symposium on Wearable Computers*. IEEE, 153–154.
- [45] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. 2020. EarSense: earphones as a teeth activity sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.
- [46] GA Pressler, JP Mansfield, H Pasterkamp, and GR Wodicka. 2002. Detection of respiratory sounds within the ear canal. In *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society*[Engineering in Medicine and Biology, Vol. 2. IEEE, 1529–1530.
- [47] Rundong Qin, Jiaying An, Jiaying Xie, Renbin Huang, Yanqing Xie, Li He, Hui Xv, Geng Qian, and Jing Li. 2021. FEF25-75% is a more sensitive measure reflecting airway dysfunction in patients with asthma: a comparison study using FEF25-75% and FEV1%. *The Journal of Allergy and Clinical Immunology: In Practice* 9, 10 (2021), 3649–3659.
- [48] Klaus F Rabe, Suzanne Hurd, Antonio Anzueto, Peter J Barnes, Sonia A Buist, Peter Calverley, Yoshinosuke Fukuchi, Christine Jenkins, Roberto Rodriguez-Roisin, Chris Van Weel, et al. 2007. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American journal of respiratory and critical care medicine* 176, 6 (2007), 532–555.
- [49] Francisco Perdigón Romero, David Castro Piñol, and Carlos Román Vázquez-Seisdedos. 2021. DeepFilter: an ECG baseline wander removal filter using deep learning techniques. *arXiv preprint arXiv:2101.03423* (2021).
- [50] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
- [51] James B Sterner, Michael J Morris, Joshua M Sill, and Jackie A Hayes. 2009. Inspiratory flow-volume curve evaluation for detecting upper airway disease. *Respiratory care* 54, 4 (2009), 461–466.
- [52] Tharoeun Thap, Heewon Chung, Changwon Jeong, Ki-Eun Hwang, Hak-Ryul Kim, Kwon-Ha Yoon, and Jinseok Lee. 2016. High-resolution time-frequency spectrum-based lung function test from a smartphone microphone. *Sensors* 16, 8 (2016), 1305.
- [53] Hai Anh Tran, Quynh Thu Ngo, and Huy Hoang Pham. 2015. An application for diagnosing lung diseases on Android phone. In *Proceedings of the Sixth International Symposium on Information and Communication Technology*. 328–334.
- [54] F Van Herpe and DG Crighton. 1994. Noise generation by turbulent flow in ducts. *Le Journal de Physique IV* 4, C5 (1994), C5–947.
- [55] Chrystian Vieyra and Rebecca Vieyra. 2021. *Physics Toolbox Suite*. Vieyra Software. <https://www.vieyrasoftwre.net/physics-toolbox-sensor-suite>
- [56] Daphna Vilozni, Ori Efrati, Asher Barak, Yakov Yahav, Arie Augarten, and Lea Bentur. 2009. Forced inspiratory flow volume curve in healthy young children. *Pediatric pulmonology* 44, 2 (2009), 105–111.
- [57] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.
- [58] Karen L. Wood. 2020. Airflow, Lung Volumes, and Flow-Volume Loop — Pulmonary Disorders — MSD Manual Professional Edition. <https://www.msmanuals.com/professional/pulmonary-disorders/tests-of-pulmonary-function-pft/airflow,-lung-volumes,-and-flow-volume-loop>. Accessed Aug 23, 2021.
- [59] Zhijian Yang, Yu-Lin Wei, Sheng Shen, and Romit Roy Choudhury. 2020. Ear-ar: indoor acoustic augmented reality on earphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
- [60] Yu Zhang and Qiang Yang. 2021. A Survey on Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering* (2021), 1–1.