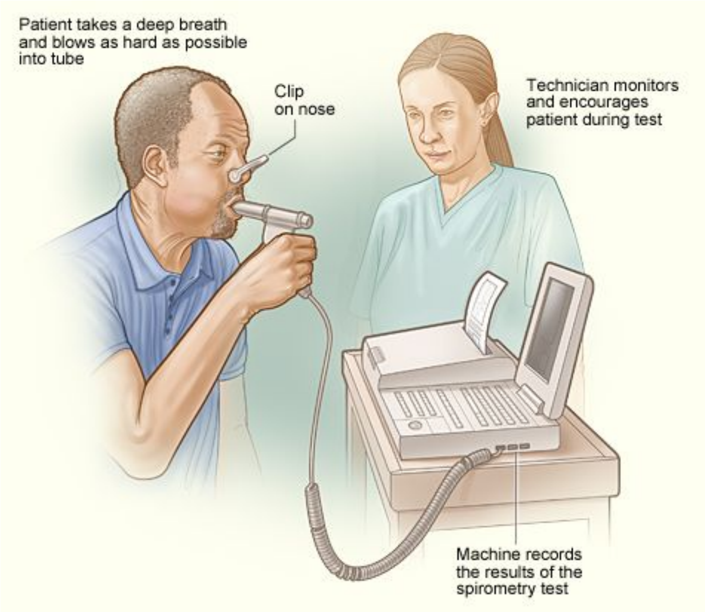
# EarSpiro: Earphone-based Spirometry for Lung Function Assessment

Presentor: Kechen Liu

# 1.1 Motivation: Clinical Motivation of Spirometry

Spirometry is the gold standard for evaluating lung function. It measures how much and how fast a person can breathe in and out.

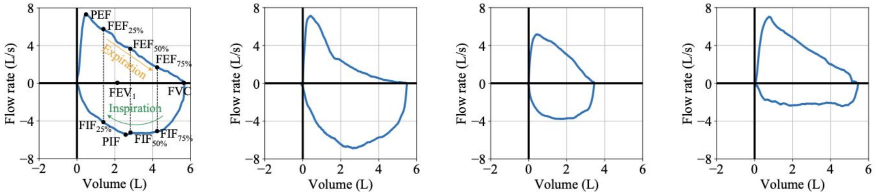
It is critical for diagnosing chronic respiratory diseases (CRDs) like COPD and asthma, and is recommended for high-risk groups (e.g., coal workers, smokers) even before symptoms appear



# 1.2 Spirometry Basics

**Procedure**: A spirometry test requires:

1. Maximal inspiration,
2. Forceful expiration until no more air remains,
3. Maximal inspiration again




(a) Normal (b) Obstructive lung disease (c) Restrictive lung disease (d) Variable extrathoracic obstruction

Fig. 2. The F-V curves of different lung conditions. (a) Normal. (b) Obstructive lung disease. The expiratory limb exhibits a steeple 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].

**Outputs**:

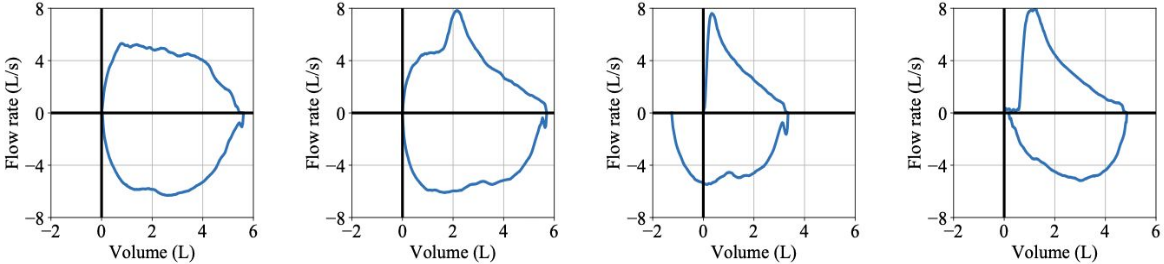
- ✓ **Flow–Volume (F-V) curve**: A central tool. Its *shape* can reveal lung conditions (normal, obstructive, restrictive, extrathoracic obstruction). Figures 2 and 3 illustrate diagnostic patterns and unacceptable maneuvers

- ✓ **Indices extracted**:
  - Core: FVC (forced vital capacity), FEV1, FEV1/FVC, PEF (peak expiratory flow).
  - Additional: FEFx% (flow at 25%, 50%, 75% of FVC), FEF25–75%, FIVC (forced inspiratory vital capacity), PIF (peak inspiratory flow), FIFx% (flow at % of inspiratory volume)

# 1.3 limitations of current system

- **Conventional spirometers**: accurate but bulky, clinic-only, and expensive portable models (≈$2000+) limit accessibility
- **Smartphone-based spirometry** (SpiroSmart, SpiroCall, etc.): reduce cost, but only measure **four expiratory indices** (FVC, FEV1, FEV1/FVC, PEF). They miss two critical aspects:

  1. **F-V (flow–volume) curve** – essential for:
     - Diagnosing impairments (e.g., concave expiratory limb = obstructive disease).
     - **Quality control** of maneuvers (Fig. 3 shows unacceptable curves).
     - Remote/home use where doctors must validate maneuver correctness
  2. **Inspiratory measurement** – vital for detecting conditions like **variable extrathoracic obstruction**, where the *inspiratory limb flattens* while the expiratory limb appears normal




(a) Sub-maximal effort (b) Slow start (c) Sub-maximal expiration. (d) False start

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].

# 1.3 Research question + background

**Can we design a low-cost, mobile system that measures the full F-V curve (expiratory + inspiratory), comparable to a clinical spirometer?**

Airflow creates **sounds** when passing through constricted areas like the airway. These sounds correlate with flow rate

**Earphones with microphones** can capture these sounds effectively because:

- Sounds propagate through bone/tissue to the ear canal (bone conduction).
- Even weak signals (inspiration) are perceivable at the ear.
- Fixed placement in the ear eliminates distance variability (unlike phone mics)

# 1.3.1 Airflow sound as signal

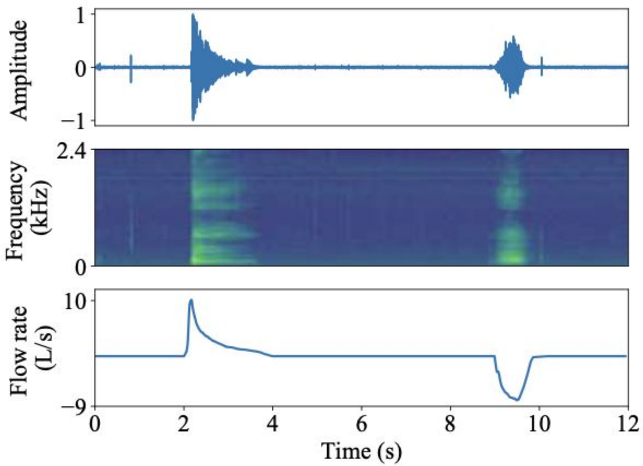



Fig. 4. The correlation between flow rate and airflow sound. From top to bottom are time-domain, frequency-domain, and flow rate signals.

# 2. System overview

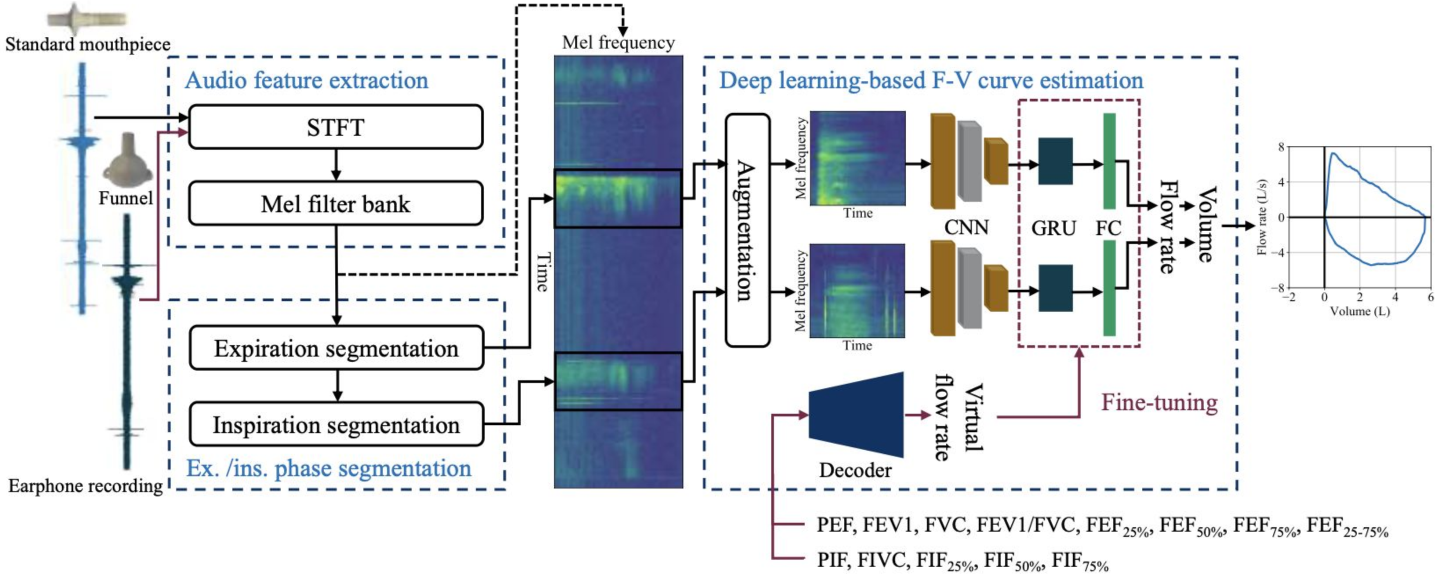



Fig. 5. The workflow of EarSpiro.

# 2.1: Audio Feature Extraction

Step 1: STFT (Short-Time Fourier Transform)

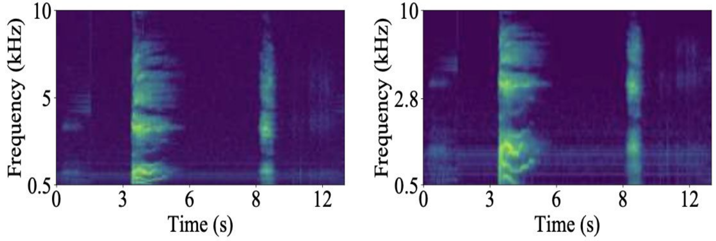
Step 2: Mel Filter Bank

Purpose: Audio feature extraction converts noisy, high-dimensional raw sound into a compact, meaningful Mel spectrogram representation that preserves breathing-related features and discards irrelevant detail.

$$M(f) = 1125 \cdot \log(1 + f/700).$$

$$\mathbf{M} = \mathbf{H}^{\top}\mathbf{S},$$

where M, H and S are the Mel spectrogram, the Mel filter bank, and the original spectrogram




(a) Spectrogram.

(b) Mel spectrogram.

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. | 50$ms$ |
| $N_{FFT}$ | FFT point. | 2400 |
| $\rho_{ovlp}$ | Overlap ratio between segments. | 75% |
| $f_{low}$ | The lowest frequency. | 0.5$kHz$ |
| $f_{high}$ | The highest frequency. | 15$kHz$ |
| $M$ | Number of filters in the Mel filter bank. | 100 |

## 2.2 Expiratory and Inspiratory Phase Segmentation

**Input feature**: the Mel spectrogram from §4.1.

**Goal**: localize the expiratory and inspiratory intervals (time ranges) within the recording.

**Challenges:** (i) environmental/inertial noise (friction, chair/floor contact, teeth knocking); (ii) weak inspiration (often below the noise floor), so the “clear” patterns seen in Fig. 6(b) are not guaranteed in practice. Hence, tailored algorithms are needed for each phase.

## 2.2.1 Expiration Segmentation

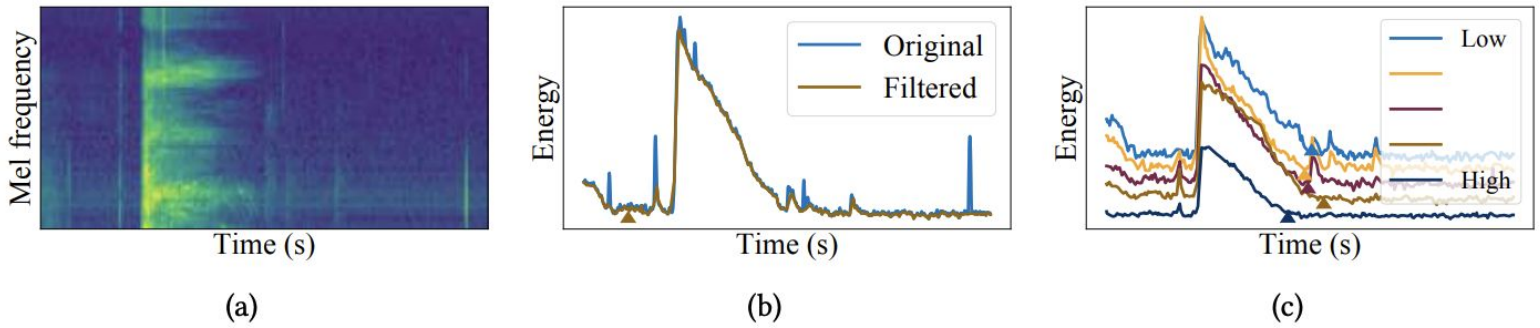



Fig. 7. Expiratory phase segmentation. (a) Mel spectrogram. (b) Energy profile of the Mel spectrogram. (c) Energy profiles of the five sub-bands.

# Expiration Segmentation

**Start detection**:

- Expiration begins with a sharp energy burst.
- Compute energy profile from Mel spectrogram → smooth with **Hampel filter** → apply **gradient search** to find the start

**End detection**:

- Different frequency bands decay at different rates.
- Split Mel spectrogram into **five sub-bands**, apply gradient search to each → take the **latest end time** as expiration end (Fig. 7c).
- Input: energy profile, sub-band energies.
- Output: estimated start & end indices

---

**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.

$n_0 \leftarrow \text{SEARCH}(E[n], -d, w, c)$ ▷ Execute the gradient search algorithm.

**for** $k = 1$ to $K$ **do**

  $n_k \leftarrow \text{SEARCH}(E_k[n], d, w, c)$ ▷ Execute the gradient search algorithm on each sub-band.

**end for**

$n_1 \leftarrow \max(\{n_k\})$

**return** $n_0, n_1$

**procedure** SEARCH$(E[n], d, w, c)$ ▷ The gradient search algorithm.

  $n \leftarrow \text{argmax}(E[n])$

  **while** $E[n] \geq \frac{1}{2}\max(E[n])$ **or** $\frac{E[n+w]-E[n]}{w} \geq c$ **do**

    $n \leftarrow n + d$

  **end while**

  **return** $n$

**end procedure**

---

# 2.2.2 Inspiration Segmentation

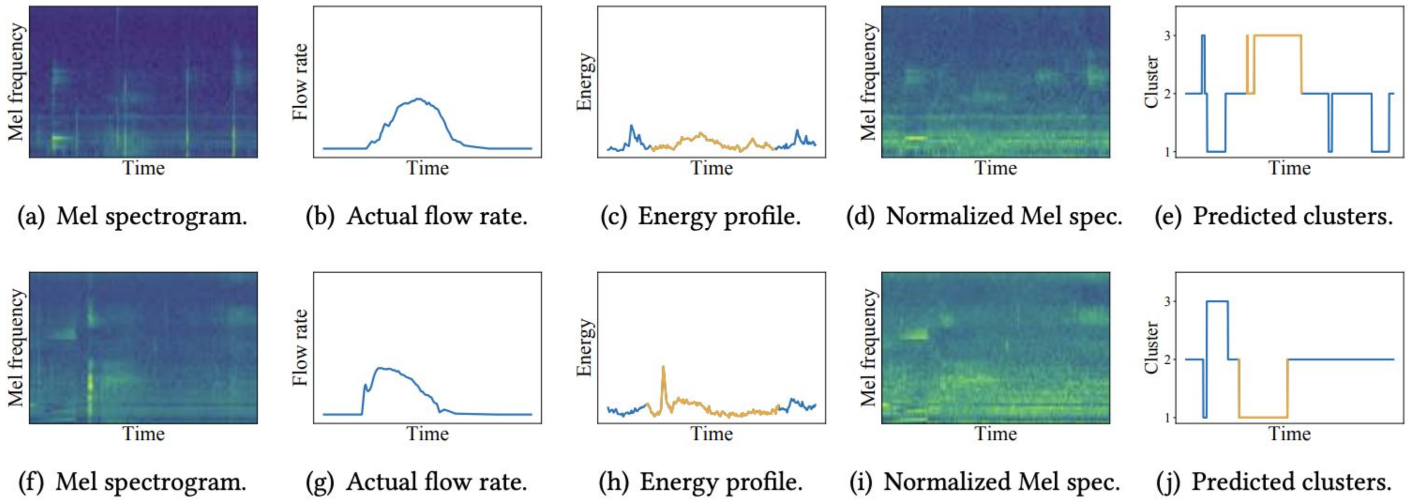



Fig. 8. Two examples of inspiratory phase segmentation. (a-e) The first example. (f-j) The second example.

# Inspiration Segmentation

Inspiratory sounds are weaker and irregular.

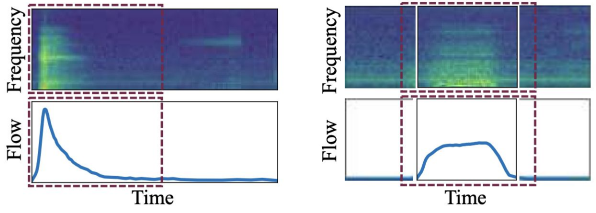
Method:

1. **Z-score normalize** Mel spectrogram frames.
2. Treat each frame as a probability distribution over frequency.
3. Apply **K-means clustering** → clusters correspond to background noise, weak inspiration, etc.
4. Extract the segment belonging to inspiration cluster (Algorithm 2).

# 2.3 : Deep Learning–based F-V Curve Estimation

**CNN → GRU → Fully Connected layers**:

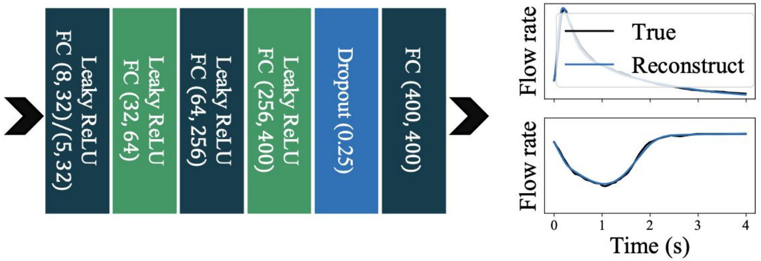
- CNN extracts spatial features from Mel spectrogram (both ears).
- GRU (RNN) captures temporal correlation, robust to non-Gaussian noise.
- FC layers regress **flow rate over time**




(a) Truncating. (b) Zero-padding.

Fig. 9. Data augmentation.


Fig. 10. Flow rate estimator architecture.




(a) (b)

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

# 3. Implementation

**Prototype hardware**

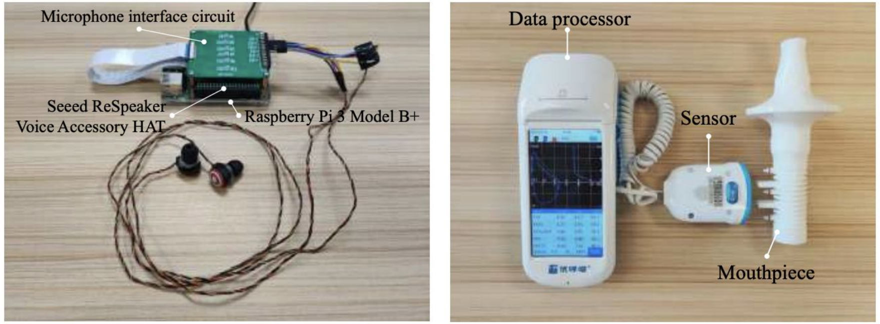
- Commercial earphones + **integrated MEMS microphones**
- Interface: **Seeed ReSpeaker voice board**
- Powered by **Raspberry Pi 3 B+**

**Data acquisition**

- Audio captured with **Audacity**
- Stored as **WAV files** for processing

**Software & training**

- Processing software: **Python 3.7**
- Deep learning: **PyTorch 1.7**
- Training on **Google Colab (Nvidia T4 GPU)**




(a) Prototype. (b) Ground truth spirometer.

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

# 4. Evaluation

## Flow rate estimation

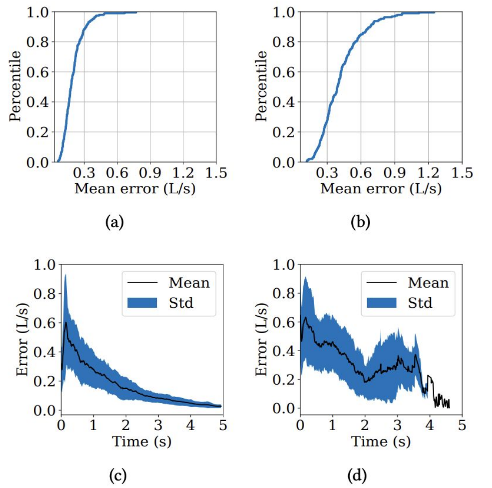



Fig. 16. Flow rate estimation error. (a) CDF plot of mean absolute error of expiration estimation. (b) CDF plot of mean absolute error of inspiration estimation. (c) Expiratory flow rate error distribution over time. (d) Inspiratory flow rate error distribution over time.

## Accuracy of Expiratory and Inspiratory Phase Segmentation

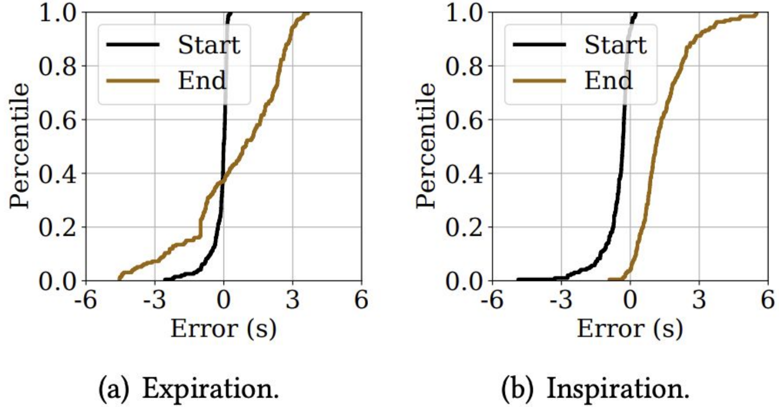
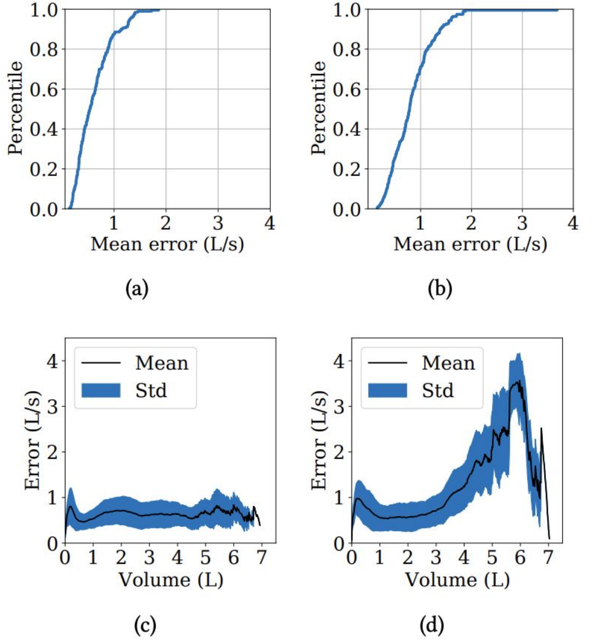



Fig. 22. Segmentation error of the expiration phase and inspiration phase.

# Flow-volume curve estimation



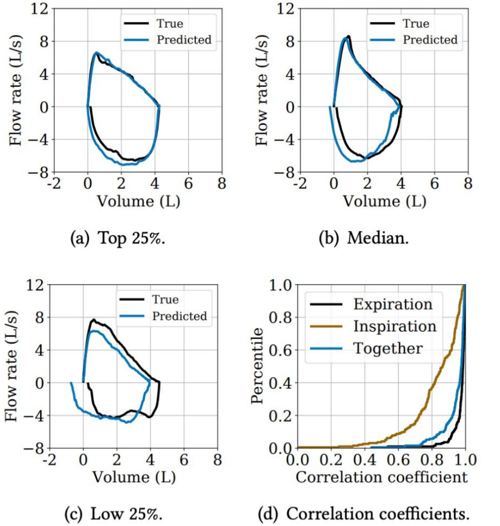




Fig. 18. F-V curve estimation. (a-c) Examples of estimated F-V curve ranked by mean absolute error. (d) Pearson correlation coefficients of F-V curve estimation.

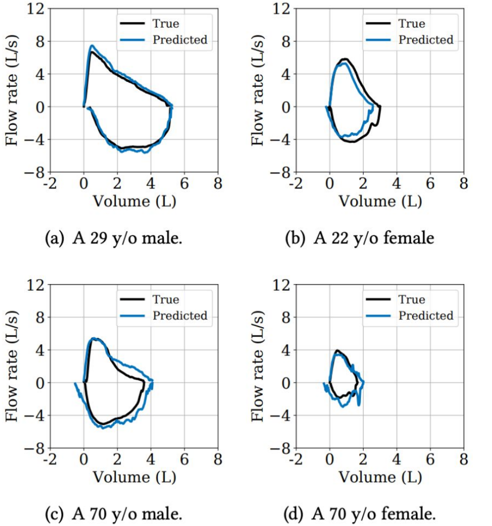
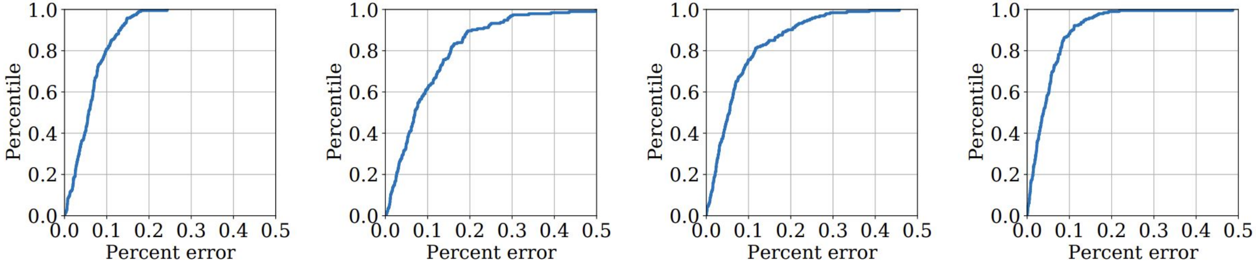



Fig. 19. Examples of F-V curve of different age groups. (a) Sample of a 29 y/o male. (b) Sample of a 22 y/o female. (c) Sample of a 70 y/o male. (d) Sample of a 70 y/o female.

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)

# Lung function indices estimation




(a) PEF. (b) FVC. (c) FEV1. (d) FEV1/FVC.

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

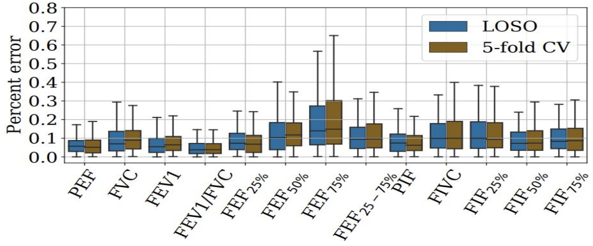
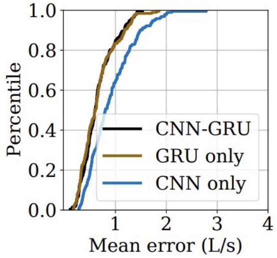
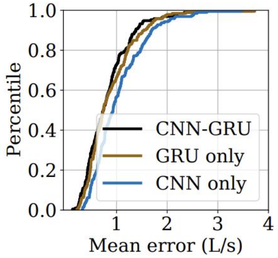



Fig. 21. Lung function index estimation error.

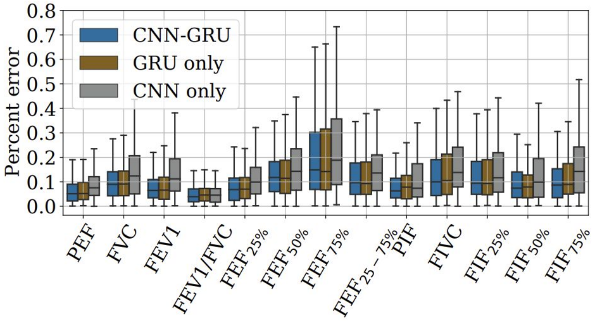
# Benefits of the CNN-GRU architecture




(a) Error of expiratory limb.



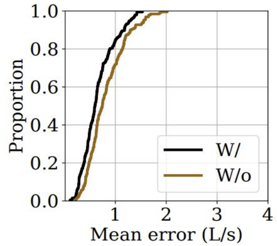

(b) Error of inspiratory limb.



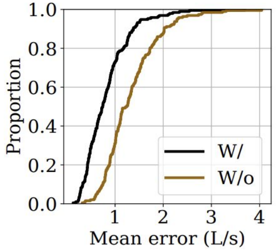

(c) Error of lung function index estimation.

Fig. 23. Effectiveness of the deep learning architecture.

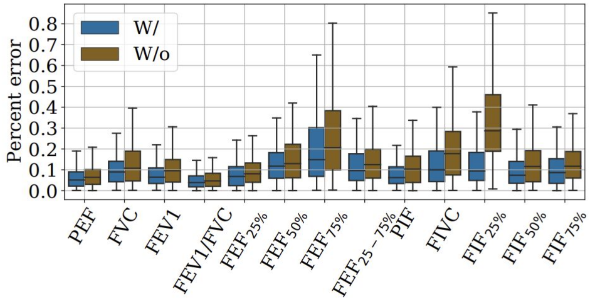
# Benefits of Data Augmentation




(a) Error of expiratory limb.



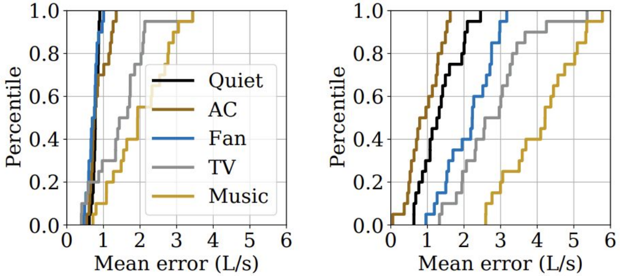

(b) Error of inspiratory limb.




(c) Error of lung function index estimation.

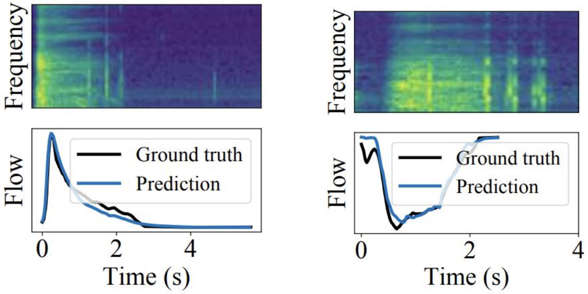
Fig. 24. Effectiveness of the data augmentation.

# Impact of Background Noise




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

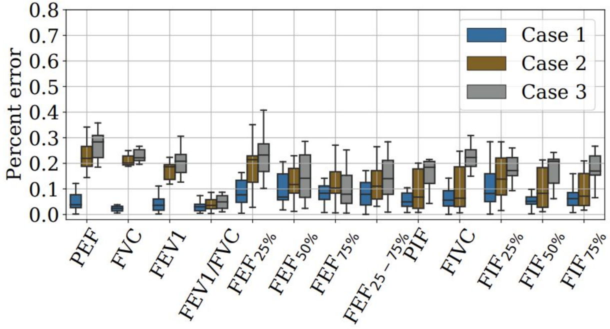
Fig. 25. Impact of background noise.



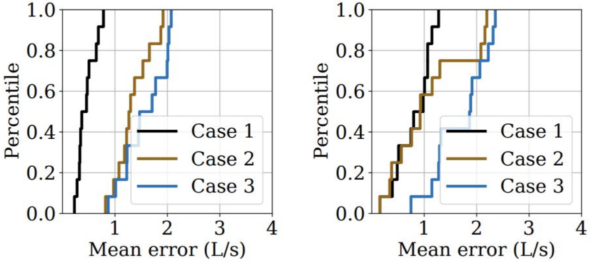

(a) Expiration. (b) Inspiration.

Fig. 26. Interference of teeth knocking.

# Impact of earphone's position



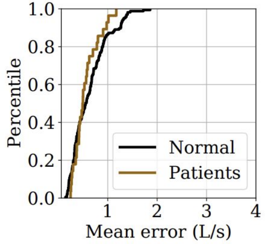

(a) Error of lung function index estimation.



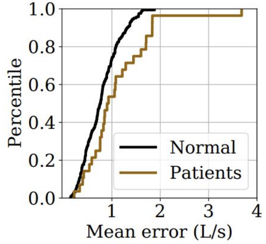

(b) Error of expiratory limb. (c) Error of inspiratory limb.

Fig. 27. Impact of Earphone position.

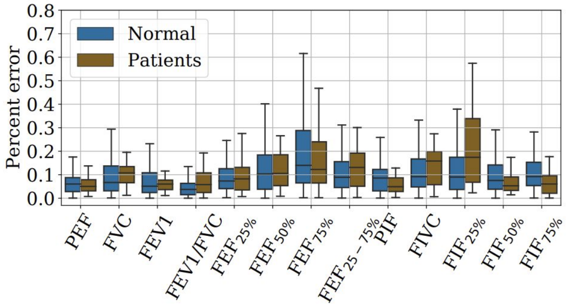
# Performance on Subjects with Lung Function Impairments




(a) Error of expiratory limb.



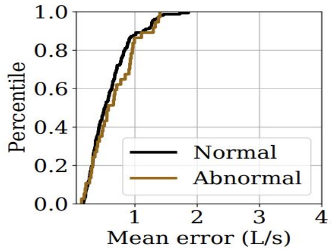

(b) Error of inspiratory limb.



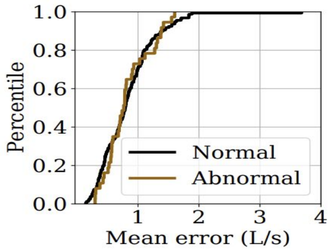

(c) Error of lung function index estimation.

Fig. 28. Evaluation on six patients.

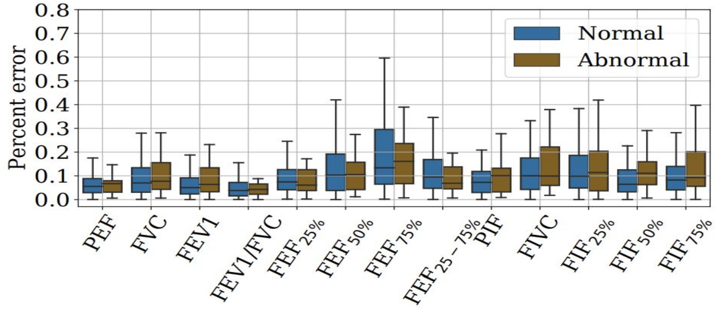
# Performance on Abnormal F-V Curves




(a) Error of expiratory limb.




(b) Error of inspiratory limb.




(c) Error of lung function index estimation.

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

# Limitations & Discussion

Environment: Assumes quiet settings; noise can still affect weak inspiration.

Hardware: Prototype uses custom MEMS mic (not yet standard in commercial earbuds).

Calibration: Current system requires manual alignment for ground truth; automation needed for deployment.

Accuracy gap: Inspiratory errors remain higher than expiratory (due to weaker signals).

Future directions:

- Better noise cancellation & segmentation.
- Explore consumer earbud integration.
- Refined model adaptation across diverse funnels/objects.

Let’s check perusall comments!