

《智能网络与计算》期末大报告

2025-2026 学年第一学期

论文 ID: _____ 1 _____

英文题目: __EasySpiro- Assessing Lung Function via Arbitrary Exhalations on
Commodity Earphones__

中文翻译: EasySpiro——通过普通耳机上的任意呼气来评估肺功能__

姓 名: _____ 杨皓翔 _____

学 号: _____ 2023150139 _____

撰写时间: _____ 2026 年 1 月 5 日 _____

其他组员: _____ 张恩实、黄御凯 _____

撰写须知:

1. 请严格按照模板提供的一级大标题撰写报告,可根据需要可自行添加二级、三级小标题。不要更改报告格式、字体等,确保报告格式的统一性。
2. 报告格式统一规定如下:(1)正文部分中文字采用**楷体、小四**,西文字采用 **Times New Roman**;(2)标题采用宋体,各级标题的字号从小二依次递减;(3)正文行间距采用 **1.2 倍行距**,段前 0.25 倍行距,段后 0.25 倍行距。
3. 大报告的评分主要依据是:格式排版、报告内容、独立思考等。其中,报告内容主要从完整性、准确性和丰富性三个方面考察,独立思考则考察是否有对所读内容的批判性思考。**报告中必须包含对于论文中所涉及到的技术点进行展开论述与讲解**。报告尽可能做到图文并茂、内容充实、格式美观。
4. 报告提交时请转化成 **PDF 格式**,并按“paperID-姓名.pdf”方式命名。同时,请将论文 **PPT 一并提交**,PPT 命名方式与大报告命名方式保持一致。
5. 模板中灰色斜体字为报告撰写说明,在最终提交版中需要删除。
6. 除封面和参考文献以外,报告正文部分的页数不少于 **20 页**。
7. 报告提交的截止时间为:**2026 年 1 月 5 日晚 23:59**。

摘要

肺功能检测（PFT）是呼吸系统疾病诊断与长期随访的核心手段，其中最常用的是基于肺量计的用力肺活量测试（spirometry），需要受试者在专业指导下完成“最大吸气后快速、用力、持续 ≥ 6 秒”的呼气动作，以获得 FVC、FEV1、PEF 与 FEV1/FVC 等指标。

然而，最大用力动作对老年人、慢阻肺/哮喘患者等群体负担较大，且居家测量缺少质量控制与反馈，导致数据可用性与依从性受限。

针对这一痛点，EasySpiro 提出一种基于普通耳机（双麦克风+IMU）采集“任意/非最大用力呼气”来估计肺功能指标的新范式：系统首先从耳机音频提取时频特征，并利用 IMU 表征呼气用力程度；随后通过最大用力呼气谱图重建（U-Net+跨模态融合）将“随意呼气”映射为“理想最大用力模式”；最后以 ViT 为骨干进行多指标回归预测。

为解决“用力程度无显式标注”的难题，论文引入自监督 IMU 预训练（LIMU-BERT）+基于 PEF 排序的对比/二分类微调来学习努力程度嵌入。作者与医院合作采集 50 名受试者（含 36 名患者）的多努力等级呼气数据集，并报告总体平均百分比误差约 7%，其中 FVC 8.08%、FEV1 7.12%、PEF 4.50%、FEV1/FVC 6.35%。

总体而言，EasySpiro 把 PFT 从“必须最大用力”推进到“可用随意呼气”，兼顾可用性与一定精度，为可穿戴/居家肺功能评估提供了新的技术路线与研究空间。

1. 研究背景及意义

1.1 研究背景与动机

(1) spirometry 的临床重要性及门槛

spirometry 是最常见的 PFT，输出 FVC、FEV1、PEF、FEV1/FVC 等用于诊断与监测慢阻肺、哮喘等疾病进展。传统流程要求最大努力呼气并持续至少 6 秒，且通常需要临床人员指导以保证质量；标准化与质量控制亦是国际指南长期强调的重点

(2) 居家/移动肺功能评估的需求与不足

居家肺功能监测可以更早发现急性加重、提升随访频率，但现实中“设备成本、使用体验、缺少教练式反馈与质量控制”导致普及受限。SpiroSmart 等工作已指出家庭肺活量测量缺少 coaching/feedback/quality control 机制，影响可接受测量的获得

(3) “最大用力”是移动/可穿戴 PFT 的主要瓶颈

既有移动端肺功能估计多数仍要求“最大用力吹气”或类似受控动作：

- ① SpiroSmart：手机麦克风记录最大用力吹气并结合信号处理估计流量/指标
- ② EarSpiro：耳机记录气流声，在测试动作中估计 F-V 曲线（含吸气/呼气），采用 CNN+RNN 并做分段跟踪
- ③ ExhaleSense：面向手机“吹气式”spirometry，重点在检测高质量用力呼气与质量评估，再回归 FEV1/FVC 等

上述方法对用户动作要求高，且在病患群体中依从性与可重复性仍是挑战。

(4) EasySpiro 动机

EasySpiro 希望把 PFT 从“最大努力呼气”解耦出来，改为利用更自然的、非最大努力的随意呼气来推断肺功能，从根本上降低用户负担与失败率。

论文认为：虽然标准流程要求最大努力，但医学研究表明呼气声学/气流模式与气道阻力、肺功能存在相关性；并提出通过“重建理想最大努力模式”来对齐临床标准，从而提升可解释性与可用性。

1.2 研究意义

- (1) 实用性与可及性提升：以普通耳机为载体（双麦克风+IMU），有望降低额外硬件成本并改善佩戴便利性；“随意呼气”比“最大吹气”更易执行，尤其对老年与呼吸受限人群更友好。
- (2) 方法学创新：提出“最大努力谱图重建→指标预测”两阶段范式，将无标准动作输入映射到可对齐临床的中间表征；并用自监督 IMU 学习解决“努力程度无标签”的关键难点。
- (3) 研究前瞻性：为可穿戴健康从“受控强约束动作”走向“自然交互/被动或低负担采集”提供了可借鉴路线，与多模态表征学习、生成式重建、基础模型等方向天然契合。
- (4) 数据与社区贡献：论文声称开源首个基于呼气声的 PFT 数据集链接，便于后续复现与对比。

1.3 核心思想

- (1) **总体目标**：仅凭多次“非最大努力”呼气声音 + IMU 姿态动态，估计 FVC、FEV1、PEF、FEV1/FVC 四项肺功能指标
- (2) **关键难点**：随意呼气的异质性极强（力度、时长、节律、噪声等），与标准最大努力动作存在“生成差距”；而“努力程度”又缺乏显式标注，难以直接监督学习
- (3) **解决思路**：
 - ① 用 IMU 自监督学习得到“努力程度嵌入”；
 - ② 用 U-Net 重建“理想最大努力”谱图以对齐临床过程；
 - ③ 用 ViT 做全局融合并回归多指标（整体流程如下 Figure 2 所示）

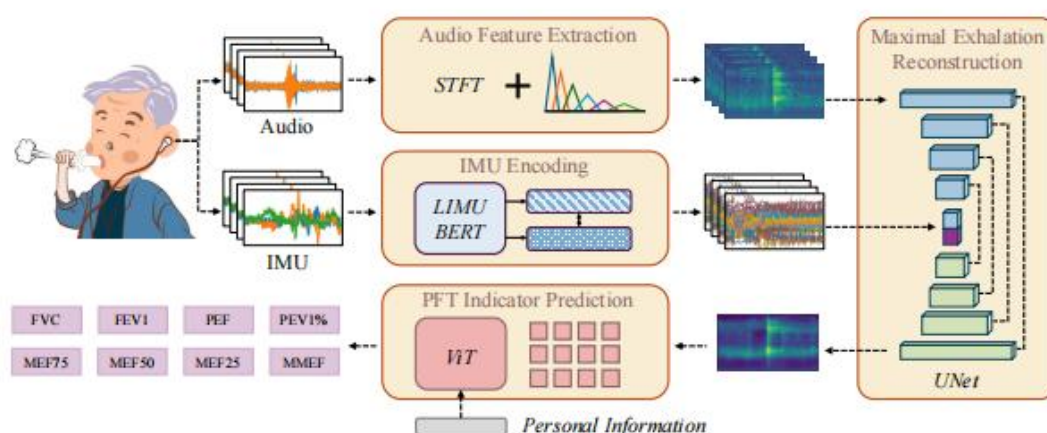


Figure 2: Overview of EasySpiro.

1.4 本文贡献

- (1) 提出 EasySpiro：首个基于耳机、面向非最大努力随意呼气的 PFT 估计系统
- (2) 提出两项核心技术：最大努力谱图重建模型与 IMU 努力编码方法（自监督预训练+对比学习）

(3) 构建并评估医院协作数据集（50 名受试者，含 36 名患者），实现约 7% 平均误差，并给出消融/噪声/人群分层等分析

(4) 提供开源数据集资源，促进后续研究

2. 国内外研究现状

(1) 标准 spirometry 与质量要求

国际指南长期强调：spirometry 是最常用 PFT，需标准化流程、校准与质量控制，最大努力与可重复性是核心关注点。这也决定了很多移动端方法仍试图“复刻最大吹气动作”，否则难以与临床指标对齐。

(2) 手机麦克风“吹气式”spirometry：从信号处理到机器学习

① SpiroSmart (UbiComp'12)：用户对手机麦克风最大用力呼气，通过建模声学传播/声道与环境来估计气流相关指标；该工作明确指出居家测量缺乏教练与质量控制问题

② ExhaleSense (PerCom'20)：把“高保真用力呼气检测”和“质量评估”作为关键步骤，再用正则回归预测 FEV1/FVC 等，强调与 ATS/ERS 质量建议（如 TTPF 等）的一致性。

③ 不足与启示：上述方法对用户动作（最大用力、对准麦克风、时长等）依赖强；一旦吹气不达标，后续预测会显著退化，因此需要复杂的质量筛选与交互引导。EasySpiro 则试图绕开“必须最大用力”，用“重建最大努力中间表征”的方式对齐临床标准，是方向性的差异

(3) 耳机/可穿戴声学：更稳定的采集位置与多模态趋势

① EarSpiro (IMWUT'22)：耳机记录气流声，采用 CNN+RNN 拟合声音到流速/流量-容积（F-V）曲线，并解决吸气信号弱的问题；在 60 名受试者上报告了流速、F-V 曲线误差等。

② 主动声学/超声类（如 SpiroSonic, MobiCom'20）：利用手机声学感知胸壁运动并转换为肺功能指标，强调与临床相关性。

③ 不足与启示：EarSpiro 与 SpiroSonic 等多仍基于“受控测试动作”或相对规范的测量流程；此外，耳机端的优势在于佩戴位置固定、抗手持姿态变化，但也引入硬件差异与噪声鲁棒性挑战。EasySpiro 进一步加入 IMU 以刻画努力程度，体

现出“多模态融合”趋势

(4) 自然语音/咳嗽等更自然的呼吸健康评估

一类研究不再要求吹气，而用语音或咳嗽声推断肺功能/阻塞程度。例如移动端自然语音、单音节语音与自然讲话的肺功能估计，以及基于自愿咳嗽声学特征的阻塞/肺功能相关估计。MMLung 进一步比较多种手机音频任务/模态在同一框架下的效果，并报告其指标误差讨论

不足与启示：语音/咳嗽更自然，但与标准 spirometry 指标之间存在更强的生理间接性、跨人群/跨设备泛化难；“中间可解释表征”与“质量控制”仍是难题。EasySpiro 的“重建最大努力谱图”可视为一种对齐临床过程的解释性尝试

(5) 小结：EasySpiro 相对现有工作的定位差异

现有多数：强约束动作 → 直接预测指标/曲线（或先做质量筛选）。

EasySpiro：弱约束随意呼气 → 重建最大努力“标准形态” → 再预测指标，并用 IMU 自监督学习补齐“努力程度”这一关键隐变量。

这使其在用户体验上更有优势，但也带来“重建误差传播、数据规模与泛化、临床一致性验证”等新的挑战。

3. 研究内容

3.1 主要挑战及创新点

(1) 技术难点总结：

① 技术难点 1：随意呼气的强异质性与“生成差距”

随意呼气在力度、时长、开始瞬间、呼气过程起伏等方面差异显著，导致直接从原始谱图回归肺功能指标容易被“动作差异”淹没。论文提出“生成差距

（submaximal vs maximal）”并用重建模块显式建模该差距，借由重建监督损失提供更强的学习信号

② 技术难点 2：努力程度无显式标签

努力程度是决定呼气声与肺功能映射稳定性的关键变量，但通常没有直接标注。EasySpiro 观察到：用户用力时身体摆动更明显，IMU 信号变化更大，可作为努力程度代理。因此引入 IMU 并设计自监督编码器学习努力表征

③ 技术难点 3：多次呼气数量不稳定、输入缺失与鲁棒性

用户不一定完成 5 次有效呼气。论文通过通道 mask、随机移位等数据增强，模拟“少次呼气/任意顺序/时间偏移”等情况，提高重建模型对输入缺失的鲁棒性。

(2) 创新点总结：

- ① 两阶段学习框架：最大努力谱图重建 + 指标预测，提高可解释性并对齐医学流程。
- ② IMU 努力编码：LIMU-BERT 自监督预训练 + 基于 PEF 排序的对比/二分类微调，将隐变量“努力程度”显式注入模型。
- ③ ViT 用于谱图全局融合：相对 ResNet 更强调全局特征融合能力，并在消融中验证收益。
- ④ 面向真实患者人群的评估与开源数据：医院协作数据集、疾病/年龄/BMI 分层与噪声影响分析。

3.2 技术路线及实现

(1) 硬件与数据采集

- ① 载体：一副改装耳机，配置双麦克风采集呼气声，双 IMU 采集姿态/摆动
- ② 采样率：音频 48 kHz，IMU 50 Hz；用网络时间戳对齐起始点，误差<1ms
- ③ 环境噪声：房间约 60–70 dB，为典型室内环境；呼气声范围约 42–105 dB

(2) 音频预处理与特征提取

- ① 呼气起始点检测：仅靠音频能量峰值在嘈杂环境可能失败，论文利用 IMU 在开始吹气时的显著变化来辅助定位起始点（思想是“吹气时身体/头部会产生瞬时摆动”）。
- ② 时频表示：对齐到起始点后截取固定长度音频段（论文示例为提取 12 秒），STFT 窗口 64ms、相邻窗口重叠 32ms，并用 64 个 Mel 滤波器降维，缓解高维谱图问题。
- ③ 硬件噪声模板相减：由于低成本硬件存在内部噪声，论文在完全静默环境记录“环境/硬件噪声模板”，并从后续录音中相减以降低系统性噪声。

(3) IMU 努力编码

核心假设：更用力的呼气会带来更显著的 IMU 信号变化；IMU 变化程度可帮助判断“当前呼气与最大努力之间的距离/差距”，从而指导谱图重建。

① 自监督预训练：LIMU-BERT

论文采用 LIMU-BERT 作为 IMU 编码骨干，并在大规模 IMU 数据集上以随机 mask 方式预训练以学习通用时序表征。LIMU-BERT 本身是将 BERT 式掩码建模思想适配到 IMU 数据的轻量自监督表征学习框架

② 努力编码微调：基于 PEF 排序的对比/二分类学习

努力程度无标签，但我们常能判断“同一人两次呼气哪次更用力”。论文利用对应的 PEF 值来比较用力大小，并以此构造二分类标签：若 $PEF(IMU_1) > PEF(IMU_2)$ 则 $label=0$ ，反之 $label=1$ 。（对应论文公式 1，如下图所示）

$$label = \begin{cases} 0 & : PEF_{IMU_1} > PEF_{IMU_2} \\ 1 & : PEF_{IMU_1} < PEF_{IMU_2} \end{cases} \quad (1)$$

这里借助医学认识：PEF 出现在呼气早期，且与用力程度相关，常被认为是 effort dependent（努力相关）。通过 BCE 损失微调编码器与 MLP 头，最终得到努力嵌入向量，用于后续跨模态融合。

(4) 最大努力呼气谱图重建（Maximal Exhalation Reconstruction）

① **目标：**输入为同一用户多次随意呼气的谱图（多通道张量）以及对应努力嵌入，输出“理想最大努力呼气谱图”。论文默认使用 5 次非最大努力呼气来重建最大努力模式，理由是临床也常要求多次尝试以保证可用性。

② 模型结构：U-Net + Fusion Block

U-Net 用于像素级（时频格点）重建，具有编码器-解码器与跳连结构，适合保留局部细节并做整体映射（U-Net 经典论文见 ARXIV）。

Fusion Block 将 IMU 努力嵌入映射并与谱图特征在瓶颈/多尺度处融合（如下 Figure 5 展示该结构）

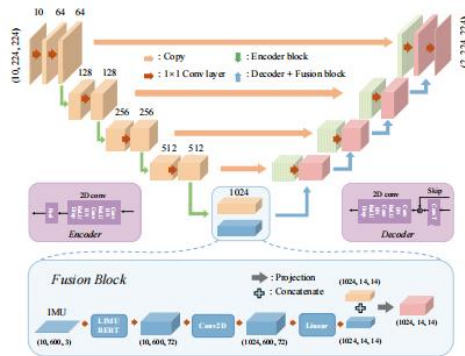


Figure 5: UNet-based reconstruction.

③ 数据增强与鲁棒性设计

为模拟“呼气次数不足/顺序任意/时间偏移/噪声”等实际情况，重建阶段采用：通道顺序 shuffle、随机 mask 若干通道、沿时间轴随机平移谱图、加高斯噪声等增强策略。这也解释了为什么仅 3 次呼气仍能维持可接受误差：因为训练时已通过 mask 增强学到“缺失通道”的重建能力。

(5) 肺功能指标预测 (PFT Indicator Prediction)

重建得到的“最大努力谱图”被当作图像样本输入预测网络。论文采用 ViT 作为骨干，以获得更强的全局融合能力 (ViT 基本思想见 ARXIV)，并输出 FVC、FEV1、PEF、FEV1/FVC 四个连续值。

在消融中，作者对比 ResNet 与 ViT，发现 ViT 在该任务上更能提取谱图全局信息，且不易过拟合于有限数据分布。

(6) 训练与优化要点 (主要设计与优化)

① 两阶段学习带来额外的监督信号：重建阶段的“生成谱图 vs 真实最大谱图”差异损失，为最终预测提供更可控的中间目标，从而提升效果与可解释性。

② 优化器/正则：论文参考 AdamW (解耦权重衰减) 等现代训练策略；AdamW 思想源于“将权重衰减从自适应梯度更新中解耦”

③ 自监督表征：LIMU-BERT 的引入体现“用无标签大数据预训练→小数据微调”的范式迁移

3.3 相关技术介绍

(1) FVC/FEV1/PEF/FEV1-FVC 是什么？

FVC：用力呼气可呼出的总气量（肺活量）。

FEV1：用力呼气第一秒呼出的气量，反映气道阻塞程度。

PEF：峰值呼气流量，通常出现在呼气早期，且与努力相关

FEV1/FVC：阻塞性疾病常用比值指标。

这些指标在临床 spirometry 中是最常用的输出，标准化指南对此有明确规范

(2) 谱图 (spectrogram) 是什么？为什么能用于肺功能？

谱图把声音信号变成“时间-频率-能量”的二维图像。呼气时的湍流、气道狭窄导致的流速变化，会改变呼气声的频谱分布；因此谱图中可能包含与气流、阻塞程度相关的特征。EarSpiro 等工作已用声学到流速/曲线的学习验证该思路

(3) U-Net 为什么适合重建？

U-Net 是经典的编码器-解码器结构，适合把输入图像映射为输出图像（分割/重建等），并通过跳连保留细节信息。EasySpiro 把“随意呼气谱图→最大努力谱图”的映射视为图像到图像问题，因此用 U-Net 自然。

(4) 自监督学习与 LIMU-BERT 在 IMU 上的意义

IMU 数据很难人工标注，但自监督可以通过“遮挡一部分信号并预测它”等任务学习通用表示。LIMU-BERT 把 BERT 的 mask 思想引入 IMU 时序，在多数据集上证明可提升下游任务效果

(5) Vision Transformer (ViT) 为何用于谱图回归？

ViT 把图像切成 patch 序列，用 Transformer 做全局注意力融合，擅长建模远距离依赖。谱图中“不同时间段/不同频带”的联合模式可能与肺功能相关，因此全局建模有潜在优势。

4. 性能评估

4.1 实验设置

(1) 数据集与受试者：与医院合作采集包含 50 名受试者（其中 36 名为多种疾病患者）的多努力等级呼气声音数据集

(2) 采集设备与采样率：音频 48 kHz，IMU 50 Hz；网络时间戳对齐误差<1ms

- (3) 验证方式：整体性能评估使用 LOSO (leave-one-subject-out) 验证；消融实验采用 10-fold 交叉验证
- (4) 对比基线：论文将既有方法在其原始数据集与本数据集上对比（如下图 Table 2），并指出本数据集以非最大努力为主，导致一些“依赖最大努力”的方法性能下降。

Table 2: Comparison with baselines.

| Baselines | SpiroSmart [28] | | ExhaleSense [42] | | EarSpiro [66] | | EasySpiro |
|-----------|-----------------|-------|------------------|-------|---------------|-------|-----------|
| Dataset* | Original | Ours | Original | Ours | Original | Ours | Ours |
| FVC | 5.2% | 22.3% | - [†] | - | 9.9% | 19.3% | 8.1% |
| FEV1 | 4.8% | 24.6% | - | - | 7.8% | 19.8% | 7.1% |
| PEF | 6.3% | 25.8% | - | - | 6.5% | 10.3% | 4.5% |
| FEV1/FVC | 4.0% | 15.4% | 7.57% | 17.3% | 5.1% | 16.3% | 6.3% |

* The original dataset refers to the dataset used in the respective papers. As a comparison, we test the performance of their techniques on our dataset, which contains mostly submaximal exhalation sounds.

[†] Not provided in the paper.

4.2 评价指标

- (1) 百分比误差 (Percentage Error)：论文用该指标衡量预测值与真值（临床设备输出）差异，并用 CDF 与箱线图呈现误差分布（Figure 10）。

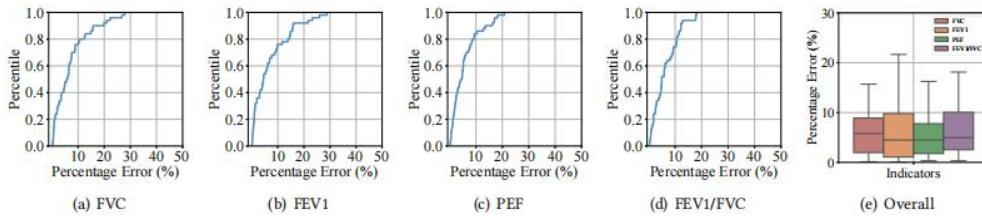


Figure 10: Overall performance. (a)-(d) CDF plots. (e) Box plot.

- (2) 系统延迟 (Latency)：分别在 CPU 与 GPU 上统计预处理、重建、预测以及总耗时 (Table 3)。

Table 3: System delay (in ms)

| Device | Proc | | Recon | Pred | Total |
|--------|---------|--------|--------|--------|---------|
| CPU | 1101.78 | | 990.51 | 625.30 | 2717.59 |
| | AUD | IMU | | | |
| | 474.44 | 627.34 | | | |
| GPU | 471.00 | | 4.74 | 8.54 | 484.28 |
| | AUD | IMU | | | |
| | 465.07 | 5.93 | | | |

Proc: Audio and IMU pre-processing. Recon: maximal effort reconstruction. Pred: PFT indicator prediction.

(3) 鲁棒性/分层分析：按疾病、年龄、BMI、性别分组对比误差 (Figure 13–15)，并分析噪声 SNR 对误差的相关性 (Figure 16)。

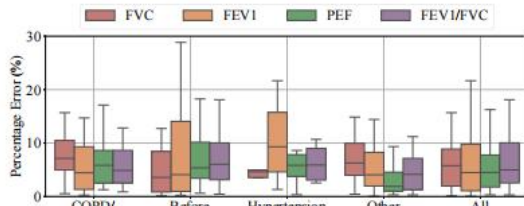


Figure 13: Demographic study - disease.

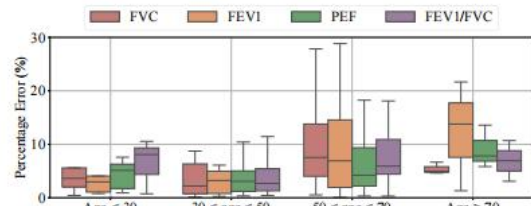


Figure 14: Demographic study - age.

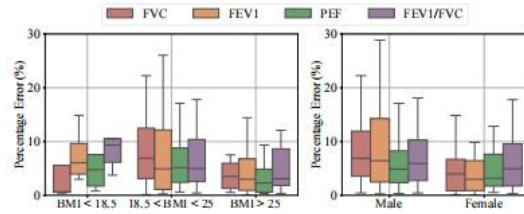


Figure 15: Demographic study - BMI and gender.

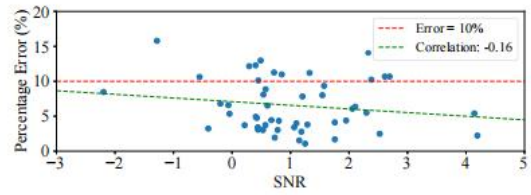


Figure 16: Impact of noise.

4.3 性能评价

(1) 总体预测精度

LOSO 下四项指标平均百分比误差：FVC 8.08%、FEV1 7.12%、PEF 4.50%、FEV1/FVC 6.35% (Figure 10)。

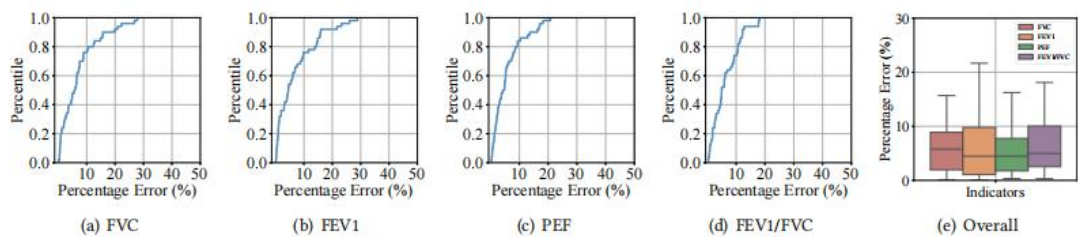


Figure 10: Overall performance. (a)-(d) CDF plots. (e) Box plot.

论文指出相较需要最大努力的相似工作，误差略高，但考虑到本工作使用的是更轻松的非最大努力呼气，仍具有实用价值。此外论文提到医疗级设备常以 5% 误差为设计容忍目标，但约 7% 误差在研究/标准中也被观察到（文中引用其参考文献[3,10,18]）。

(2) 呼气次数减少的影响

在 3/4/5 次呼气输入下，系统误差仍保持可接受，论文解释这是因为重建阶段通过 mask 增强“用少量通道也能重建”的能力（Figure 11）。

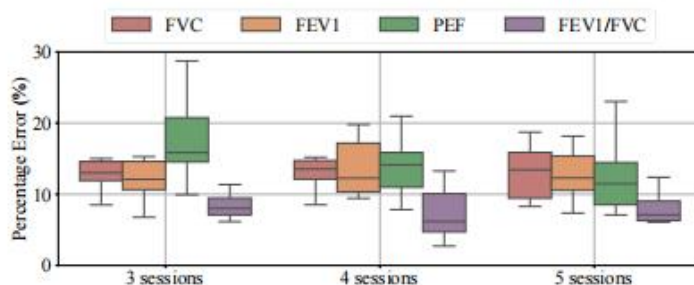


Figure 11: Impact of exhalation sessions.

意义：这直接对应实际落地场景：用户可能只完成 3 次有效呼气，系统仍可工作，提升可用性。

(3) 系统延迟与可部署性

Table 3 显示：CPU 总耗时约 3 秒，GPU 约 0.5 秒；论文认为 PFT 非实时任务，因此该延迟可接受。

(4) 消融实验：关键模块贡献

① 去掉重建模块：性能下降，原因是重建提供“额外监督损失”并与医学流程一致，提升可解释性与最终预测效果。

② 去掉 IMU 或去掉 IMU 编码器：若完全不使用 IMU，平均误差超过 20%；加入 IMU 可降低误差，而使用论文设计的自监督 IMU 编码器可进一步显著降低误差，证明努力编码有效。

③ ResNet vs ViT：ViT 优于 ResNet，论文解释为其更擅长全局融合且不易过拟合，具备更强分布适应潜力

(5) 人群分层与噪声影响

论文按疾病、年龄、BMI、性别分组给出误差箱线图（Figure 13–15），并分析噪声 SNR 与误差的相关性为 -0.16（Figure 16），说明噪声对误差有一定影响但相关性不强。

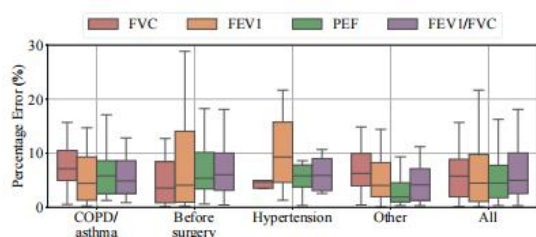


Figure 13: Demographic study - disease.

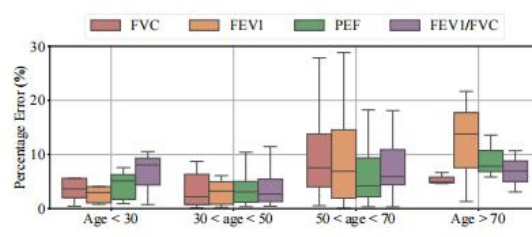


Figure 14: Demographic study - age.

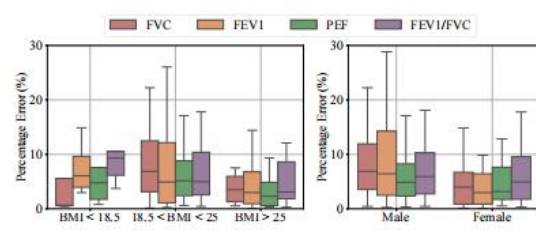


Figure 15: Demographic study - BMI and gender.

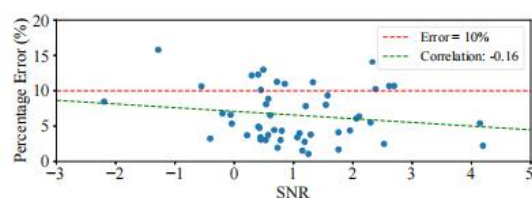


Figure 16: Impact of noise.

5. 论文评述

5.1 论文优点

- (1) 问题定义具有实际价值：从“最大努力”转向“随意呼气”，直击居家/可穿戴 PFT 的体验瓶颈
- (2) 技术路线清晰且可解释：用“最大努力谱图重建”作为中间目标，既增强监督信号，也与临床过程更一致，论文明确强调这提升了解释性
- (3) 努力编码设计巧妙：利用 IMU 变化作为努力代理，并用自监督预训练+PEF 排序弱监督学习努力表征，解决“无标签隐变量”的关键难题
- (4) 评估较全面：LOSO 总体精度、次数鲁棒性、延迟、消融、噪声与人群分层都有覆盖
- (5) 与既有工作的清晰定位：相对 Spir oSmart/EarSpiro/ExhaleSense 等仍依赖受控用力动作的路线，EasySpiro 提出了更“低负担”的新范式

5.2 论文不足

(1) “用 PEF 作为努力排序标签”的循环依赖风险

论文用 PEF 比较两次呼气谁更用力，而 PEF 本身又是最终要预测的指标之一。这可能引入一种“标签泄漏式的偏置”：努力编码器在训练时直接利用了与目标强相关的变量，导致评估时难以判断努力编码是否真正学到了可泛化的“努力”概念，还是学到了与 PEF 相关的捷径。

改进建议：未来可考虑用与肺功能无直接重合的外部努力标注（如口部压力传感、呼气阻力装置的机械反馈）或以“同一人同一时段的相对用力排序”作为更纯粹的监督信号。

(2) 重建阶段的“误差传播”与不确定性缺失

两阶段框架中，重建误差会传递到预测阶段；论文展示“有重建更好”，但缺少对“何时重建失败、失败后预测是否可靠”的不确定性建模。

改进建议：引入不确定性估计（如 MC Dropout、深度集成、分位数回归）并在输出端给出置信区间；或用端到端联合训练+一致性约束减少误差累积。

(3) 数据规模与外部泛化仍不足

数据集为 50 人，其中 36 名患者，规模对深度模型（U-Net+ViT）而言偏小。论文虽用自监督预训练缓解，但跨医院、跨地区、跨语言/口型、跨耳机硬件的泛化仍未充分验证。

改进建议：

- ① 做跨设备/跨场景域适配（domain adaptation）或设备无关特征学习；
- ② 做外部验证集（另一家医院/家庭环境）与前瞻性临床试验；
- ③ 采用更轻量的模型并进行联邦学习/隐私保护学习（也与其引用的移动数据匿名化研究方向相关）。

(4) 与 ATS/ERS 标准的对齐仍主要停留在“形式对齐”

论文强调重建最大努力模式“与医学过程一致”，但对 ATS/ERS 2019 标准中的可接受性/可重复性判据（如起始判据、呼气平台、重复性阈值等）如何在随意呼气场景下重新定义，讨论不够深入

改进建议：提出新的“随意呼气 PFT 质量标准”，或借鉴 ExhaleSense 式质量指标体系，把质量控制模块显式加入系统闭环

(5) 隐私与伦理讨论偏少

呼气声与 IMU 可能携带身份/健康敏感信息，尤其在可穿戴长期采集中更突出；论文未系统讨论数据安全、可撤回机制、端侧处理与最小化采集原则。

改进建议：补充隐私威胁建模与缓解策略，如端侧特征提取、差分隐私、可解释告知与用户控制。

潜在改进方向 / 研究方向：

(1) 端到端联合生成-预测与多任务学习

将重建与预测联合训练：一方面保持“重建最大努力谱图”的中间监督，另一方面加入“直接预测”支路做知识蒸馏/一致性约束，提升鲁棒性并减少误差传播。

(2) 更强的跨域泛化：设备无关、环境无关、人群无关

- ① 设备：不同麦克风频响、IMU 噪声、佩戴位置差异；
- ② 环境：家庭噪声、回声、风噪；
- ③ 人群：年龄、疾病类型、口型/语言习惯。

可引入域对抗学习、自监督对比学习、合成噪声增强等策略，并做跨域基准评测。

(3) 个性化校准（few-shot personalization）

借鉴 EarSpiro 对“日常漏斗/口器替代物”的迁移学习与少量真值校准思路：允许用户在首次使用时做 1-2 次标准测试（或在医院测一次），之后长期用随意呼气跟踪“相对变化”，把“绝对精度”压力转移为“趋势监测”场景，更贴合居家随访。

(4) 不确定性与临床决策支持

在输出指标外给出置信度与“是否建议复测/去医院”的提示；把系统定位为“筛查与趋势监测”而非替代临床确诊，提高落地可行性。

(5) 更自然的交互闭环：从“随意呼气”到“无感呼吸”

结合语音/咳嗽/日常说话模态（相关研究已证明可行但间接性强），与随意呼气形成多模态互补：当随意呼气质量差时，用语音/咳嗽支路提供弱估计或辅助判别，构建“多源一致性”系统。

5.3 课程建议

建议课程在论文阅读之外加入：

- (1) 复现导向的作业（鼓励从开源数据/代码复现一个模块，如 IMU 自监督或谱图重建），提升工程与科研闭环能力；
- (2) 增加医学测量标准与可用性评估专题（如 ATS/ERS 标准解读），帮助同学更好理解“系统指标误差”与“临床可接受性”之间的差异

5.4 其它

个人认为 EasySpiro 最值得学习之处在于：它没有在既有“最大吹气+质量筛选”的路径上做局部改良，而是重新定义输入形态（随意呼气）并用“可解释中间表征(最大努力谱图)+自监督努力编码”把难题拆解成可学习的子问题。这类“从用户体验出发反推算法结构”的系统论文思路，对智能网络与计算领域做可落地健康感知研究非常有启发。

6. 参考文献

[1] Xu and Xie et al. EasySpiro: Assessing Lung Function via Arbitrary Exhalations on Commodity Earphones. ACM MobiCom, 2025.

[2] American Thoracic Society & European Respiratory Society. Standardization of Spirometry 2019 Update: An Official ATS/ERS Technical Statement, 2019.

ATS JOURNALS

[3] E. C. Larson, M. Goel, et al. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. UbiComp, 2012.

UBICOMP LAB

[4] W. Xie et al. EarSpiro: Earphone-based Spirometry for Lung Function Assessment. IMWUT, 2022.

WENTAO XIE

[5] M. M. Rahman et al. ExhaleSense: Detecting High Fidelity Forced Exhalations to Estimate Lung Obstruction on Smartphones. PerCom, 2020.

EECS

[6] R. Tan et al. LIMU-BERT: Unleashing the Potential of Unlabeled Data for IMU Sensing Applications.

TANRUI

[7] A. Dosovitskiy et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (ViT). arXiv:2010.11929.

ARXIV

[8] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597.

ARXIV

[9] I. Loshchilov, F. Hutter. Decoupled Weight Decay Regularization (AdamW). arXiv:1711.05101, ICLR 2019.

ARXIV

[10] ScienceDirect Topics. Peak Expiratory Flow (PEF) — effort dependent portion of expiratory flow.

SCIENCEDIRECT

[11] X. Song, B. Yang, et al. SpiroSonic: Monitoring Human Lung Function via Acoustic Sensing on Commodity Smartphones. MobiCom, 2020.

PITT SITES

[12] M. Mosuily et al. MMLung: Moving Closer to Practical Lung Health Estimation using Smartphones. Interspeech, 2023.

ISCA ARCHIVE

[13] E. Nemati et al. Estimation of the Lung Function Using Acoustic Features of the Voluntary Cough. EMBC, 2020.

EUROPE PMC

7. 附录

A. 缩略语表

PFT: Pulmonary Function Test (肺功能检测)

FVC: Forced Vital Capacity (用力肺活量)

FEV1: Forced Expiratory Volume in 1 second (第一秒用力呼气量)

PEF: Peak Expiratory Flow (峰值呼气流量)

IMU: Inertial Measurement Unit (惯性测量单元)

STFT: Short-Time Fourier Transform (短时傅里叶变换)

ViT: Vision Transformer

LOSO: Leave-One-Subject-Out

B. 插图清单

Figure1: PFT 传统方法与 EasySpiro 方法对比

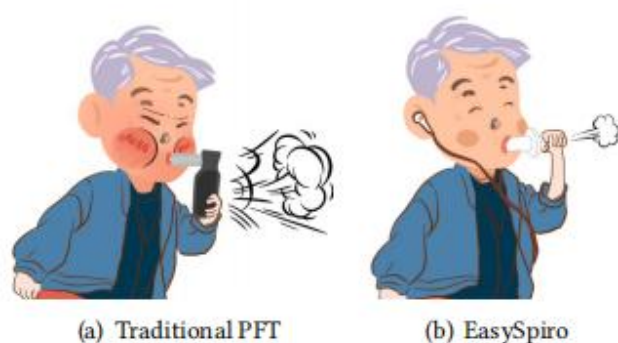


Figure 1: Conducting PFT. (a) Traditional method (b) EasySpiro.

Figure 2: 系统总览

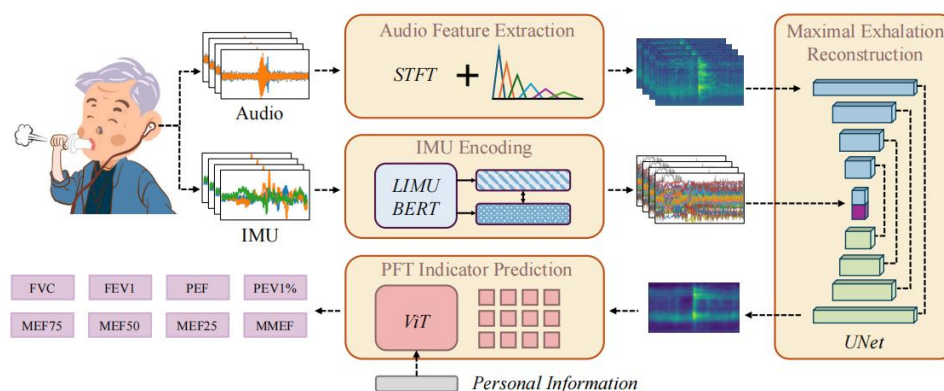


Figure 2: Overview of EasySpiro.

Figure 5: U-Net 重建结构（用于解释 Fusion Block）

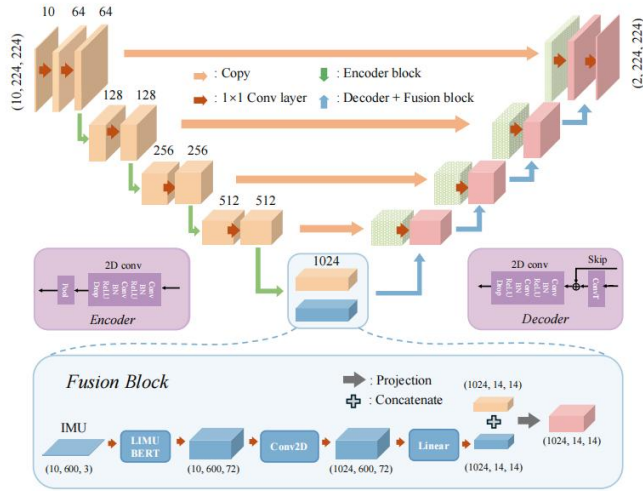


Figure 5: UNet-based reconstruction.

Figure 10: 总体 CDF 与箱线图（用于“性能评估”）

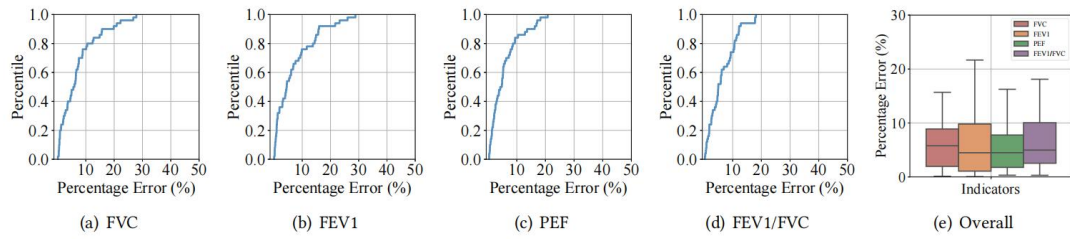


Figure 10: Overall performance. (a)-(d) CDF plots. (e) Box plot.

Figure 11/12: 次数影响与消融（用于“鲁棒性/贡献验证”）

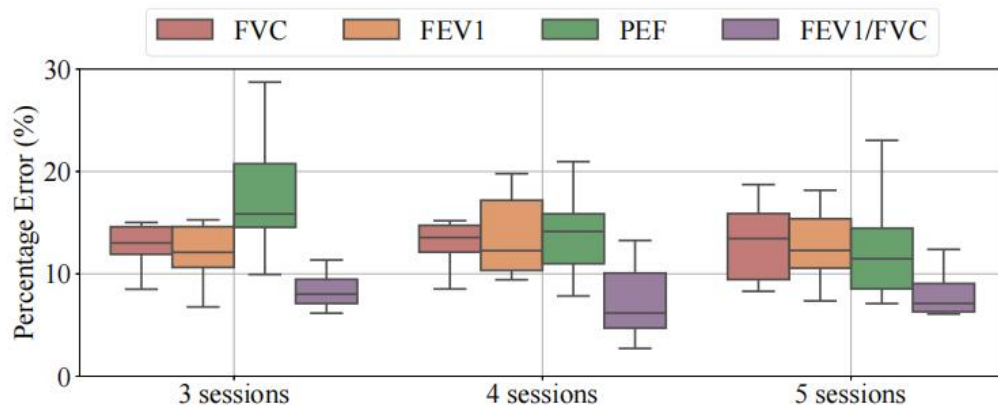


Figure 11: Impact of exhalation sessions.

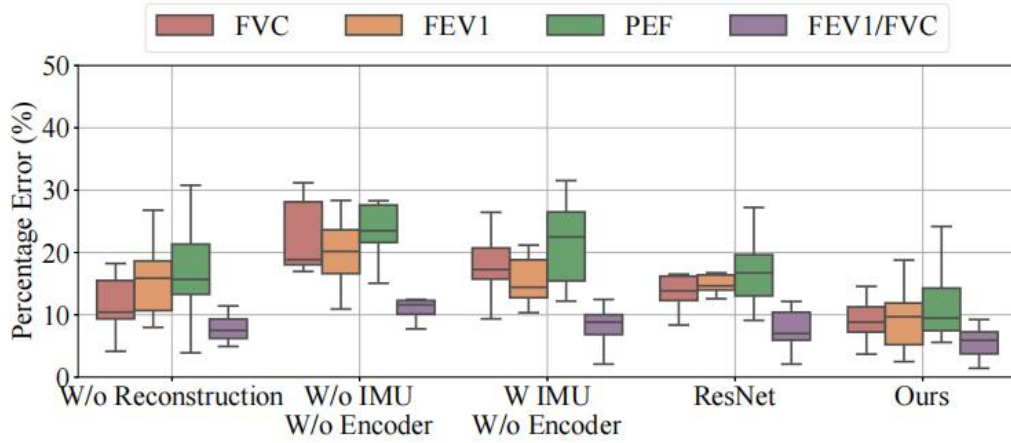


Figure 12: Ablation study.

Figure 13–16: 分层与噪声影响（用于“讨论与不足”）

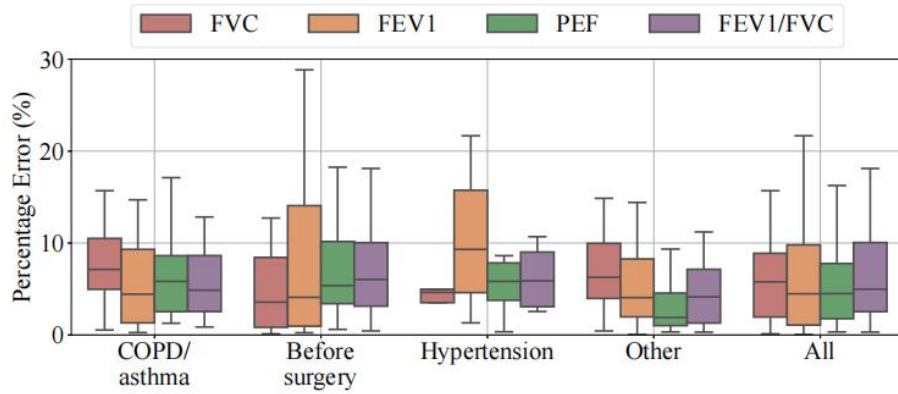


Figure 13: Demographic study - disease.

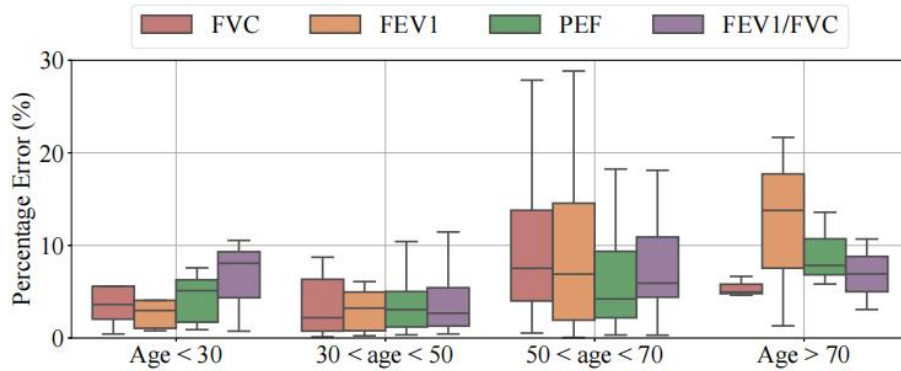


Figure 14: Demographic study - age.

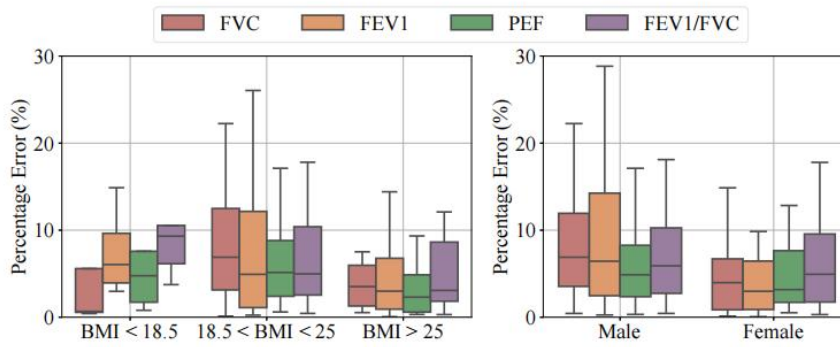


Figure 15: Demographic study - BMI and gender.

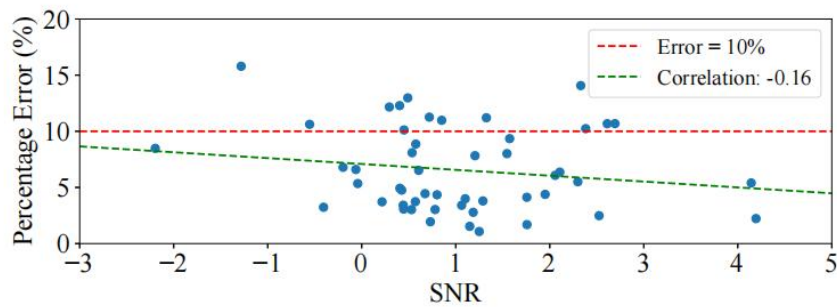


Figure 16: Impact of noise.

Table1: 受试者人群分布及关键肺功能指标统计（人口学与临床特征）

Table 1: Demographics of the subjects.

| Stats | COPD/ asthma | Hyper- tension | Before surgery | Others | All |
|--------------------------|-----------------|-------------------|-------------------|-------------|-------------|
| Population | 18 | 6 | 22 | 14 | 50 |
| Age (years) | 52.8 (18.8) | 62.7 (11.5) | 50.8 (13.1) | 51.4 (14.3) | 51.8 (15.0) |
| Height (cm) | 168.4 (7.0) | 164.9 (9.5) | 163.7 (10.0) | 165.4 (7.5) | 165.2 (8.5) |
| Weight (kg) | 64.6 (9.8) | 73.8 (17.0) | 60.1 (13.1) | 66.0 (8.7) | 63.6 (12.4) |
| BMI (kg/m ²) | 22.7 (2.7) | 26.8 (3.8) | 22.3 (3.7) | 24.1 (2.9) | 23.2 (3.6) |
| FVC (L) | 3.8 (0.7) | 3.0 (0.8) | 3.3 (0.9) | 3.1 (0.8) | 3.3 (0.8) |
| FEV1 (L) | 2.6 (0.7) | 2.2 (0.4) | 2.7 (0.7) | 2.5 (0.9) | 2.6 (0.7) |
| PEF (L/s) | 6.3 (1.4) | 5.6 (1.4) | 6.7 (1.6) | 6.4 (2.3) | 6.4 (1.7) |
| FEV1/FVC | 0.7 (0.1) | 0.8 (0.1) | 0.8 (0.1) | 0.8 (0.1) | 0.8 (0.1) |

Format: mean (standard deviation)

Table2: 与现有基线方法的肺功能指标预测误差对比

Table 2: Comparision with baselines.

| Baselines | SpiroSmart [28] | | ExhaleSense [42] | | EarSpiro [66] | | EasySpiro |
|----------------------|-----------------|-------|------------------|-------|---------------|-------|-----------|
| Dataset [*] | Original | Ours | Original | Ours | Original | Ours | Ours |
| FVC | 5.2% | 22.3% | - [†] | - | 9.9% | 19.3% | 8.1% |
| FEV1 | 4.8% | 24.6% | - | - | 7.8% | 19.8% | 7.1% |
| PEF | 6.3% | 25.8% | - | - | 6.5% | 10.3% | 4.5% |
| FEV1/FVC | 4.0% | 15.4% | 7.57% | 17.3% | 5.1% | 16.3% | 6.3% |

^{*} The original dataset refers to the dataset used in the respective papers. As a comparison, we test the performance of their techniques on our dataset, which contains mostly submaximal exhalation sounds.

[†] Not provided in the paper.

Table3: EasySpiro 系统各模块延迟与总体运行时间统计

Table 3: System delay (in ms)

| Device | Proc | | Recon | Pred | Total |
|--------|---------|--------|--------|--------|---------|
| CPU | 1101.78 | | 990.51 | 625.30 | 2717.59 |
| | AUD | IMU | | | |
| | 474.44 | 627.34 | | | |
| GPU | 471.00 | | 4.74 | 8.54 | 484.28 |
| | AUD | IMU | | | |
| | 465.07 | 5.93 | | | |

Proc: Audio and IMU pre-processing. Recon: maximal effort reconstruction. Pred: PFT indicator prediction.