arXiv:2602.03549v1 [cs.SD] 3 Feb 2026

EarResp-ANS：基于音频的耳机端呼吸频率估计与自适应噪声抑制

MICHAEL KÜTTNER，卡尔斯鲁厄理工学院，德国VALERIA ZITZ，卡尔斯鲁厄理工学院，德国SUPRAJA RAMESH，卡尔斯鲁厄理工学院，德国MICHAEL BEIGL，卡尔斯鲁厄理工学院，德国TOBIAS RÖDDIGER，卡尔斯鲁厄理工学院，德国

呼吸频率（RR）是临床评估和心理健康的重要生命体征，但由于缺乏不显眼的传感技术，人们在日常生活中很少对其进行监测。入耳式音频传感具有良好的应用前景，这得益于其较高的社会接受度，以及闭塞效应对生理声音的放大作用；然而，现有方法在现实环境噪声下往往表现不佳，或依赖计算成本高昂的模型。我们提出了 EarResp-ANS，这是首个能够在商用耳机上实现完全设备端实时呼吸频率估计的系统。该系统采用基于LMS的自适应噪声抑制（ANS），在保留呼吸相关声学成分的同时衰减环境噪声，无需神经网络或音频流传输，从而明确应对可穿戴设备在能耗和隐私方面的限制。我们在一项包含18名参与者的研究中，在真实的声学条件下对 EarResp-ANS 进行了评估，测试环境包括音乐、食堂噪声以及最高达80 dB声压级的白噪声。EarResp-ANS 实现了稳健的性能，全局平均绝对误差为0.84每分钟周期数（CPM），通过自动异常值剔除后降至0.47 CPM，同时直接在耳机上运行时的处理器负载低于2%。

CCS 概念：• **Hardware** → **新兴技术**；• **以人为中心的计算** → **泛在与移动计算系统** **与工具**；**泛在与移动设备**；

其他关键词和短语：呼吸频率估计、入耳式麦克风、可穿戴传感、抗噪信号处理、实时系统

1 引言

呼吸频率（RR）是一项重要的生命体征，可为了解一个人的生理状况 [40] 和心理状态 [14]提供关键信息。它在睡眠监测 [15], 和压力估计 [21], 以及身体素质评估 [47, 64]中发挥着重要作用。此外，RR 往往是生理状况恶化的最早期指标之一 [40],，因此对于检测慢性阻塞性肺疾病 [60] 或脓毒症 [70],等新发健康问题尤为重要，这些疾病是导致死亡率的重要因素 [17, 68]。尽管 RR 在临床上十分重要，但其监测仍然不足 [11, 31, 40] ，或在实践中测量不准确 [12, 29]。

在临床环境中，人工计数容易出错 [29]。专业医疗设备通常依赖肺量测定 [61] 或呼吸感应体积描记法（RIP） [7], 等更可靠的技术，但这类设备往往体积庞大、不便于在社交场合使用，也不适合日常连续佩戴。基于惯性测量单元（IMU）等技术的便携式可穿戴方案更为直接，但极易受到运动伪影的影响 [20, 35, 55]。基于光电容积描记法（PPG）的方案对运动的敏感度较低，但通常只能提供间接估计指标 [45]。

基于麦克风的感知提供了一种很有前景的替代方案，因为呼吸声直接承载着呼吸活动的信息 [24, 38, 56]。耳机尤其具有应用前景，因为它们能够在靠近气道的位置持续进行感知，并且日益被视为用于健康监测的多功能感知平台 [10]。耳机的一大优势是，它们属于社会普遍接受、日常佩戴的设备，人们会在工作、通勤或

作者地址：Michael Küttner，michael.kuettner@kit.edu，卡尔斯鲁厄理工学院，卡尔斯鲁厄，德国；Valeria Zitz，valeria.zitz@kit.edu，卡尔斯鲁厄理工学院，卡尔斯鲁厄，德国；SUPRAJA RAMESH，supraja.ramesh@kit.edu，卡尔斯鲁厄理工学院，卡尔斯鲁厄，德国；Michael Beigl，michael.beigl@kit.edu，卡尔斯鲁厄理工学院，卡尔斯鲁厄，德国；TobiasRöddiger，tobias.roeddiger@kit.edu，卡尔斯鲁厄理工学院，卡尔斯鲁厄，德国。

运动时佩戴 [18]。除了主动播放音频外，人们也经常佩戴耳机来营造私密的听觉空间。实证研究表明，57.2% 的用户佩戴个人聆听设备主要是为了隔绝外部噪声，且通常不会播放音频 [16], ，从而减轻感官压力（“随身听效应”） [22, 53, 66]。因此，耳机特别适合作为一种不易察觉的生理感测平台。将呼吸追踪功能整合到耳机中，无需额外使用可能带有污名化意味的医疗设备，即可实现多用途健康监测（例如基于呼吸频率的压力管理） [10, 14]。

与配有醒目绑带的专用穿戴式传感器不同， [18],降噪耳机中使用的入耳式麦克风（IEM）受益于密闭效应 [41]。密闭效应会放大呼吸等内部产生的声音 [25, 42], ，从而与外置麦克风相比提高信噪比（SNR）。然而，要在普通耳机上实现基于麦克风的感知仍面临挑战：

- **声学干扰：** 微弱的呼吸声很容易被语音和环境噪声掩盖 [32, 56, 69]。
- **资源限制：** 对于计算能力有限且能耗预算严格的平台而言，持续进行音频处理的计算开销很大 [54]。
- **隐私与延迟：** 传输原始音频会引发隐私问题，并增加能耗和延迟 [25, 42]。

为应对这些挑战，我们提出了**EarResp-ANS**，一种完全在设备端运行的呼吸频率估计处理流程。我们利用双麦克风配置（IEM/OEM），在推断前抑制环境噪声。该方法采用基于最小均方（LMS）自适应滤波器的自适应噪声抑制（ANS），在消除外部噪声的同时保留由闭塞效应放大的呼吸特征。EarResp-ANS 通过轻量级频谱分析估计呼吸频率，避免神经网络推断和原始音频流传输。此外，它还融合两只耳机采集的双耳信号，以检测不可靠的估计结果。

我们在一项包含18名参与者的研究中，在真实条件下（80分贝音乐、自助餐厅噪声）评估了 EarResp-ANS，并将我们的系统部署于 OpenEarable 2.0 平台 [54]。我们证明：

(1) **EarResp-ANS** 通过资源高效的频谱分析，提供了完全在设备端运行的实时呼吸频率估计处理流程，仅占用耳机处理器不到2%的计算负载。(2) 我们的**双麦克风 ANS 策略**和双耳一致性检查通过保留呼吸相关信号成分，显著提升了系统在高达80多分贝声压级（SPL）的环境噪声下的鲁棒性。(3) 该系统在耳机上实现了噪声鲁棒的设备端呼吸频率感测**全新技术水平** [54], ：自动剔除异常值后，平均绝对误差（MAE）达到0.47每分钟周期数（CPM），超越了此前的基于音频的方法。

2 相关工作

本节回顾了可穿戴呼吸频率监测领域的既有工作，重点介绍支持使用适合日常使用的可穿戴设备及其形态进行连续跟踪的方法（见第2.1节）。此外，我们还回顾了可穿戴音频感测中的噪声抑制与信号增强相关工作，重点关注适用于实时设备端处理的方法（见 第2.2节）。

2.1 可穿戴呼吸监测系统

可穿戴呼吸监测能够在日常条件下持续评估呼吸情况，而环境噪声和声学干扰会对信号可靠性构成重大挑战 [23, 26]。现有方法按传感模态大致可分为基于惯性的测量方法、利用心血管调制的基于心率的方法，以及直接采集呼吸声的基于音频的方法。这些

表1. 代表性呼吸频率监测方法概览。运动和调制类别反映了相应研究报告的实验条件。由于各研究在参考测量、评估协议和运动条件方面存在差异，因此报告的误差指标无法直接比较。

Work	传感器放置位置		运动	调制	方法平均绝对误差		SD	已部署
Angelucci等人 [2]	IMU	多种运动状态	静息、运动	无、运动	信号处理 + 机器学习	–	–	–
Röddiger等人 [55]	IMU	入耳式	静息	None	SP	2.62	2.74	–
OptiBreathe [58]	PPG	入耳式	静息	无，指导	SP	≈ 2.7* / 1.96†	–	–
Romano 等人 [56]	Mic.	Face	静息，运动	无，运动	SP	1.7	–	–
RespEar [38]	Mic.	入耳式	静息、运动	无、运动	信号处理 + 机器学习	1.48‡ / 2.28§	–	手机
Kumar 等人 [32]	Mic.	入耳式	静息、运动	无、运动	ML	–	–	–
Martin 等人 [42]	Mic.	入耳式	静息	None	SP	2.7	1.6	–
BreathPro [25]	Mic.	入耳式	运动	运动	信号处理 + 机器学习	–	–	手机
EarResp-ANS（我们的方法）	**Mic.**	**入耳式**	**静息**	**无、运动**	**SP**	**0.84**	**2.17**	**耳机**

注： * 无调制； † 按指导以每分钟 15 次的频率呼吸； ‡ 静息条件； § 运动条件。

这些模态在信号直接性、抗干扰鲁棒性以及持续设备端运行的适用性方面各不相同。表1 总结了这些类别中的代表性系统。

基于惯性的呼吸频率估计。 基于惯性的方法通过加速度计和陀螺仪捕获呼吸引起的身体运动来估计呼吸频率 [44]。由于惯性测量单元功耗低，且已集成于许多可穿戴设备中，因此广泛用于呼吸监测。为提升佩戴舒适性，研究者探索了不同的佩戴位置，包括头戴式传感 [35], 和耳戴式惯性测量单元 [55]。虽然在受控条件下准确度较高，但姿势变化和大幅度运动会引入与呼吸频段重叠的伪影，从而限制其在日常场景中的鲁棒性 [44, 55]。多传感器配置和活动感知处理可以提升鲁棒性，但会增加系统复杂度和能耗，这与轻量级连续监测的要求相悖 [2]。

基于心率的呼吸频率估计。 基于心脏信号的方法通过分析心电图（ECG）或光电容积描记（PPG）测得的心血管信号中由呼吸引起的调制，间接推断呼吸频率 [8]。从生理角度看，呼吸会引起心搏间期变化（呼吸性窦性心律不齐） [71] ，并导致光电容积描记波形的振幅／基线变化 [8, 30]。基于这些效应，在低运动条件下可通过经典的频谱及时域方法估计呼吸；近期研究则探索多模态融合（例如 ECG+PPG），以提升鲁棒性 [27]。尽管这些方法在低运动条件下效果良好，但在运动、接触压力变化、外周灌注变化及环境光干扰下性能会下降，限制了其在可穿戴设备连续监测中的适用性 [8, 51]。

基于音频的呼吸频率估计。基于音频的方法直接利用麦克风捕获的呼吸声来估计呼吸频率；与惯性或心血管代理信号相比，这提供了一种更直接的传感模态 [32, 56]。早期研究证明了在受控声学条件下使用经典信号处理技术的可行性，但未系统地解决对非平稳环境噪声的鲁棒性问题，也未处理与语音和环境声音的重叠 [42, 56]。这一局限性凸显了明确进行背景噪声抑制的必要性，相关内容将在第 2.2 节中综述。

已有研究探索了不同的麦克风放置方式，包括靠近气道的可穿戴麦克风和头戴式麦克风，以在静息和中等活动期间采集呼吸声 [32, 56]。然而，呼吸音的幅度通常较低，容易被环境噪声、环境声学

干扰以及运动引起的伪影所掩盖，从而导致在真实日常条件下性能大幅下降 [56, 69]。

因此，近期研究日益关注将耳戴式平台作为一种适合长期音频生理感测、社会接受度高且不显突兀的设备形态。入耳式麦克风受益于闭塞效应：密闭的耳道会放大呼吸等体内产生的生理声音；此外，耳机已广泛应用于日常生活，因此成为连续呼吸监测的理想平台 [41]。尽管闭塞效应会放大呼吸等体内产生的生理声音，入耳式麦克风录音仍容易受到环境噪声影响，因为声音会经由耳机外壳、骨传导路径以及不完全密封处发生声学泄漏 [10, 25, 41, 54]。

为应对这些挑战，近期有若干系统采用基于学习的模型，将呼吸声与背景噪声及其他声学事件分离，以测量呼吸量 [39]，或在嘈杂的驾驶环境中重建精细的呼吸波形（据此可推导呼吸频率） [69]。然而，这类方法通常需要大量标注数据集和可观的计算资源，因此难以适用于资源受限的可穿戴设备上的连续监测，并且会引发对能耗、延迟和用户隐私的担忧，尤其是在设备外处理原始音频时 [10, 73]。相比之下，诸如 BreathPro 的系统 [24] 采用基于信号处理的噪声抑制，而非基于学习的降噪，仅在后续呼吸频率估计环节使用轻量级机器学习 [25]。这些局限促使我们采用一种轻量级、基于信号处理的音频方法来估计呼吸频率；该方法能够抵御环境噪声，并直接在耳机上针对连续设备端运行进行评估，考察能耗和延迟；我们的方法正是沿着这一方向展开。

2.2 实时音频背景噪声抑制

在日常环境中，语音、交通噪声或环境声音等背景噪声会显著降低麦克风录音的质量。由于呼吸声振幅较低、频带宽且变化显著，在现实条件下通常具有非常低的信噪比 [49],，如果没有有效的噪声抑制，基于音频的呼吸频率估计往往难以可靠进行。对于可穿戴设备部署，此类方法必须实时运行，在严格的计算和能耗限制下完全在设备端运行，并避免传输原始音频以保护用户隐私。这些要求促使我们进一步研究适用于耳戴式设备呼吸感测的背景噪声抑制技术。

基于音频的降噪：通用单通道方法。在单通道降噪中，目标信号和干扰信号完全混合在一段麦克风录音中；由于缺乏空间、因果或参考信息，源分离本质上成为一个病态问题 [67]。

经典的基于模型的方法，包括谱减法 [4], 维纳滤波 [46, 52], 和最小均方误差（MMSE）估计 [13],，利用平稳性、谱稀疏性或能量占优等假设来衰减不需要的成分。尽管这些方法具有较低的计算复杂度和可预测的延迟，但其有效性在很大程度上取决于基础假设是否成立，而这些假设在动态的现实环境中经常会被违反 [63]。

近期基于深度学习的噪声降低方法，以通过大规模数据集学习的数据驱动映射取代显式信号模型，从而在许多场景中更有效地抑制非平稳干扰 [72]。然而，这些方法仍受到相同单通道观测的根本限制：它们通过学习统计相关性来实现分离，而非利用物理声源分离线索。因此，模型行为与训练数据和评估目标的特征紧密相关，导致其在领域偏移和未见过的声学条件下鲁棒性降低 [34, 36]。在生物声学和音频感知领域，这一局限已得到反复验证：在受控基准测试中表现出色，往往无法转化为真实场景部署中的良好表现 [49]。

对于生理声音等微弱且高度多变的目标信号，这些局限尤为明显。由于缺乏显式参考通道，单通道降噪方法往往会抑制相关信号成分，或引入时间模糊和不稳定，尤其是在信噪比较低且实时约束严格的情况下 [9]。重要的是，这些影响并非源于特定的算法选择，而是单通道表述本身存在根本歧义所致。

基于音频的降噪：基于参考通道的方法。 基于参考通道的噪声抑制通过引入一个额外麦克风来扩展单通道表述，该麦克风捕获与目标信号相关的干扰，从而借助跨通道相关性实现具有物理依据的噪声抑制 [67]。诸如基于LMS的方法等经典自适应滤波技术在实时约束下学习从参考通道到干扰分量的时变映射，从而避免单通道分离的根本歧义，并在低信噪比下提升鲁棒性 [19, 67]。

在耳机与可穿戴耳戴设备系统中，外耳麦克风（OEM）与入耳式麦克风（IEM）之间的物理间隔天然提供了参考—目标配置：OEM主要捕获环境声学干扰，而IEM记录由封闭耳道调制的外部耦合噪声与内部产生的生理声音的混合信号 [5]。已有工作在该配置下探索了用于生理感测的自适应滤波；Martin 等人 [42] 表明，在人为加入的工业噪声下可实现鲁棒的心率估计，但呼吸频率性能显著下降，凸显了保留呼吸特异性信息的必要性。

最近，BreathPro [25] 利用耳外麦克风作为噪声参考，在跑步过程中进行入耳式呼吸感知，但依赖预先计算的特定用户模板，而非持续的在线自适应；当使用通用模板时，性能会下降 [25]。尽管主动噪声消除（ANC）系统传统上旨在实现感知噪声抑制，但其前馈参考通道和自适应滤波原理可直接应用于音频感知中的数值噪声抑制 [33]。然而，现有方法并未明确解决现实且不断变化的声学条件下，耳机端持续、完全设备端呼吸频率估计所需的自适应且保持信号的噪声抑制问题；第 3.2 节对主动噪声控制原理和参考通道自适应滤波进行了技术讨论。

在此基础上，我们重新采用经典基于LMS的自适应滤波，以保留入耳式呼吸信号。我们提出了 EarResp-ANS，这是一种面向普通耳机的集成式设备端处理流程，结合了自适应噪声抑制、呼吸频率估计和基于通道差异的异常值处理。

3 EarResp-ANS：系统概述

本节介绍 EarResp-ANS，这是一种在环境噪声条件下基于音频估计呼吸频率的系统。该系统包含三个阶段：(1) 自适应噪声抑制（见 第3.4节），(2) 呼吸频率估计（见 第3.5节），以及 (3) 基于通道差异的异常值剔除（见 第3.6节）。

3.1 设计目标

本研究旨在设计一种呼吸频率估计系统，使其能够完全在普通的支持主动降噪的耳机设备端运行。与依赖将麦克风信号流式传输至外部设备的方法不同，我们的方法应仅依靠耳机本身运行，且不需要将任何原始音频数据传出设备。为实现始终开启、完全离线运行，所提出的算法经过专门设计，以满足商用支持主动降噪的耳机的计算和内存约束，同时仅需极少的处理和内存开销。除了这些系统级约束外，该方法还应在真实环境噪声条件下（例如背景语音或背景音乐播放）提供可靠的呼吸率估计值。通过避免经由无线接口传输麦克风信号，所提出的

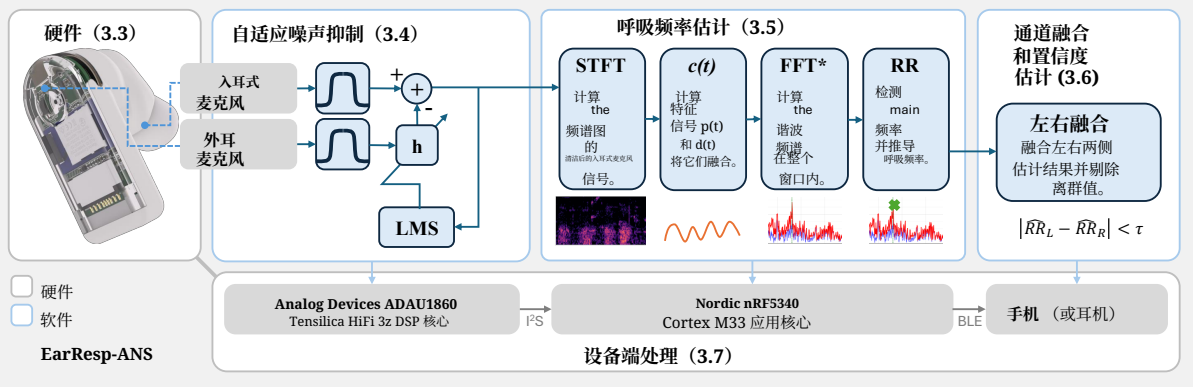


图 1. EarResp-ANS 在 OpenEarable 2.0 上的系统处理流程。自适应噪声抑制在数字信号处理器上使用延迟LMS滤波器对入耳式麦克风信号进行降噪。呼吸频率估计（基于短时傅里叶变换的特征和峰值选择）在微控制器上运行，各耳的估计结果通过蓝牙低功耗传输以进行融合。

系统应扩　　明确解决了与连续音频感知相关的隐私问题，同时
降低了因无线

3.2 自适应滤波背景

本节为适用于耳机系统的基于参考信号的方法建立技术基础。抑制入耳式麦克风（IEM）信号中的外部噪声，与主动噪声消除（ANC）问题密切相关；主动噪声消除旨在通过生成反噪声，降低用户耳膜处感知到的噪声水平。因此，最初为主动噪声消除开发的多种自适应滤波技术，已成功应用于以数值方式抑制入耳式麦克风信号中的外部噪声，例如用于耳内语音增强。 [5, 28, 59]。

在耳机采用的传统主动噪声消除系统中，通过扬声器生成反噪声来抵消外部噪声。要有效抑制外部噪声，必须准确建模从耳外麦克风（OEM）到入耳式麦克风的声学路径（称为主路径），以及从耳机扬声器到入耳式麦克风的声学路径（称为次级路径）。然而，耳机与耳道之间的声学耦合会因用户而有很大差异，并可能因耳塞贴合度或用户运动的不同而随时间变化。因此，固定滤波器或离线调优的滤波器往往不足以准确估计声学路径，这促使人们采用能够持续补偿此类变化的自适应方法 [33]。

滤波-x最小均方（FxLMS）等自适应滤波技术 [33] 广泛应用于主动噪声控制系统中。在其基本形式下，LMS自适应滤波器根据外耳麦克风信号 x（n）估计到入耳式麦克风信号 d（n）的主声学路径。其目标是最小化误差信号 e（n），其定义为

$$e(n) \coloneqq d(n) - \mathbf{h}^T \mathbf{x}(n) \tag{1}$$

通过调整长度为 M 的滤波器$\mathbf{h}$，其中$\mathbf{x}(n) = [x(n)\ x(n-1) \ldots, x(n-M+1)]^\top$ 表示参考信号向量。这相当于最小化瞬时平方误差

$$J \coloneqq \left\| d(n) - \mathbf{h}^T \mathbf{x}(n) \right\|_2^2 . \tag{2}$$

应用随机梯度下降可得 LMS权重更新规则

$$\mathbf{h}(n+1) \coloneqq \mathbf{h}(n) + \mu\, e(n)\, \mathbf{x}(n), \tag{3}$$

其中 μ 表示自适应步长。为了在高幅度噪声条件下提高鲁棒性，可用外耳麦克风信号的瞬时信号功率对梯度进行归一化，从而得到归一化最小均方（NLMS）更新规则

$$\mathbf{h}(n+1) \coloneqq \mathbf{h}(n) + \frac{\mu\, e(n)\, \mathbf{x}(n)}{\epsilon + \mathbf{x}^T(n)\mathbf{x}(n)}, \tag{4}$$

其中 ϵ 是为数值稳定性而加入的一个小常数。此外，可引入泄漏项 γ，以防止权重漂移并缓解系数增长，从而得到泄漏最小均方形式 [62]:

$$\mathbf{h}(n+1) \coloneqq (1-\nu)\, \mathbf{h}(n) + \mu\, e(n)\, \mathbf{x}(n), \qquad \nu = \gamma\mu. \tag{5}$$

传统的主动噪声控制系统施加严格的延迟约束，因为反噪声信号必须以最小延迟生成，才能与入射噪声实现相消干涉。自适应滤波也可不通过扬声器发射反噪声，而是直接在入耳式麦克风信号中以数值方式抑制背景噪声。在此设定下，自适应滤波器估计由OEM捕获的噪声分量，并将其从入耳式麦克风信号中减去，从而无需进行次级通道建模。在外部背景噪声与入耳式麦克风数据中的感兴趣信号不相关这一假设下，该方法能够有效分离期望信号。此类基于参考信号的噪声抑制技术已成功应用，例如用于基于内耳麦克风的语音增强 [5]。

当自适应滤波用于对内置麦克风信号进行后处理以估计呼吸频率，而非用于实时声音消除时，严格的延迟约束得以放宽。在此情境下，数毫秒量级的短延迟是可接受的。这使得延迟LMS自适应滤波器可用于主通道估计。通过向入耳式麦克风信号引入延迟 K，主路径的严格因果性要求得以放宽，从而能够以增加群延迟为代价，获得更稳定且线性相位的滤波器响应。所得误差信号定义为

$$e(n) \coloneqq d(n-K) - \mathbf{h}^T\mathbf{x}(n). \tag{6}$$

总之，这些自适应滤波公式提供了一种高效方法，用于估计并抑制入耳式麦克风信号中与OEM相关的噪声分量。接下来，我们将在此基础上说明如何将延迟LMS实现为面向耳机硬件的设备端自适应噪声抑制（ANS）阶段。

3.3 硬件

所提出的系统在OpenEarable 2.0平台 [54]上实现。音频信号使用*Knowles SPH0641LU4H-1*脉冲密度调制（PDM）麦克风，以48 kHz的采样率从IEM和OEM采集。两个麦克风通过共享PDM总线连接到*Analog Devices ADAU1860* 音频数字信号处理器。*ADAU1860*配备专用于双二阶滤波器运算的*FastDSP*内核，以及基于*Xtensa* 指令集架构 [6]的可编程*Tensilica HiFi 3z* 内核。*HiFi 3z* 内核每个周期最多支持四24×24 次乘累加（MAC）运算[6] ，并以98.304 MHz的时钟频率运行，从而实现高效的实时音频处理。该数字信号处理器连接到*Nordic nRF5340* 片上系统，后者配备128 MHz应用内核和64 MHz网络内核。

3.4 自适应噪声抑制（自适应噪声抑制）

如前所述，尽管耳机的硅胶耳塞能在一定程度上隔绝外界噪声，录音中仍以环境噪声为主，即使噪声水平较低也是如此。为实现准确的呼吸频率估计，必须衰减入耳式麦克风信号中的外部噪声分量。如上所述，

由于运动和耳塞贴合度的变化，从 OEM 到入耳式麦克风的声学传递路径会因用户、每次佩戴以及时间而异。因此，EarResp-ANS 会自适应地估计这条传递路径。

我们采用延迟LMS自适应滤波器来实现这一点，如第 3.2 节所述，用于估计从外侧麦克风到内置麦克风的声学传递路径。该滤波器采用 256阶FIR滤波器实现，延迟为 64 个采样点。将估计出的噪声分量从内置麦克风信号中减去，即可得到降噪呼吸信号。

这种方法基于以下假设：呼吸声音在外侧麦克风中的存在很弱，并且大多被环境噪声掩盖。因此，可以假设呼吸相关成分与参考信号不相关，因而不会在自适应滤波过程中被消除。我们将在第 6 节讨论这一假设的有效性和局限性。

为提高高信号幅度下的稳定性，我们还对滤波器权重施加了较小的泄漏 γ，如第 3.2 节所述。我们发现，使用归一化LMS会降低性能，因为它会减小较强噪声下的梯度。相反，我们仅在信号超过某个阈值时才进行归一化 τ，以防止在幅度极大时发散。由此得到自适应滤波器系数的以下更新规则：

$$\begin{aligned} \mathbf{h}(n+1) &\coloneqq (1-\nu)\,\mathbf{h}(n) + \mu\, s_n\, e(n)\,\mathbf{x}(n), \qquad \nu = \gamma\mu, \\ e(n) &\coloneqq d(n-K) - \mathbf{h}^{\top}(n)\,\mathbf{x}(n), \\ s_n &\coloneqq \min\!\left(1, \frac{\tau}{\epsilon + e(n)\,\mathbf{x}^{T}(n)\mathbf{x}(n)}\right). \end{aligned} \tag{7}$$

其中，$\mathbf{x}(n)$ 表示参考信号向量，$d(n)$ 表示内置麦克风信号，μ 表示自适应步长，γ 表示泄漏因子。误差信号 $e(n)$ 表示噪声抑制后的内部麦克风信号。因此，我们将降噪输出信号定义为

$$y(n) \coloneqq e(n). \tag{8}$$

为聚焦于呼吸相关声学成分，在应用该算法前，先将两个麦克风信号进行带通滤波，滤波范围为 200 Hz 至 1000 Hz。该频率范围既能捕捉具有特征性的呼吸音，又能抑制低频侧的身体噪声和心跳声，以及较高频率的环境噪声。我们进一步发现，在此范围内，麦克风数据与呼吸信号的相关性最高。此外，该范围也允许采用较低的采样率，从而降低计算开销。将该算法应用于麦克风数据的效果可在图 2中直观观察到。

3.5 呼吸频率估计

我们使用固定的窗口大小估计呼吸频率，并采用两种主要特征：对数频谱信号能量，以及分析窗口内与代表性呼吸模式的频谱差异度。

3.5.1 对数谱能量。我们假设，经过带通滤波和自适应噪声抑制后，内置麦克风信号主要由与呼吸相关的声音构成。因此，每次吸气和呼气时，信号能量都会升高，因为两者都会产生明显的呼吸音，如图 2所示。这种特征性的周期性会使能量信号的频谱中出现显著的最大值。我们将对数谱能量信号定义为

$$p(t) \coloneqq \frac{1}{M} \log\left(\sum_{f=0}^{F} |Y(t,f)|^2\right), \tag{9}$$

其中 $Y(t,\ f)$ 表示清理后的 IEM 信号$y(t)$的短时傅里叶变换时频谱图，其中包含 F 个频率箱。短时傅里叶变换采用长度为 512、步长为 64 的汉明窗。

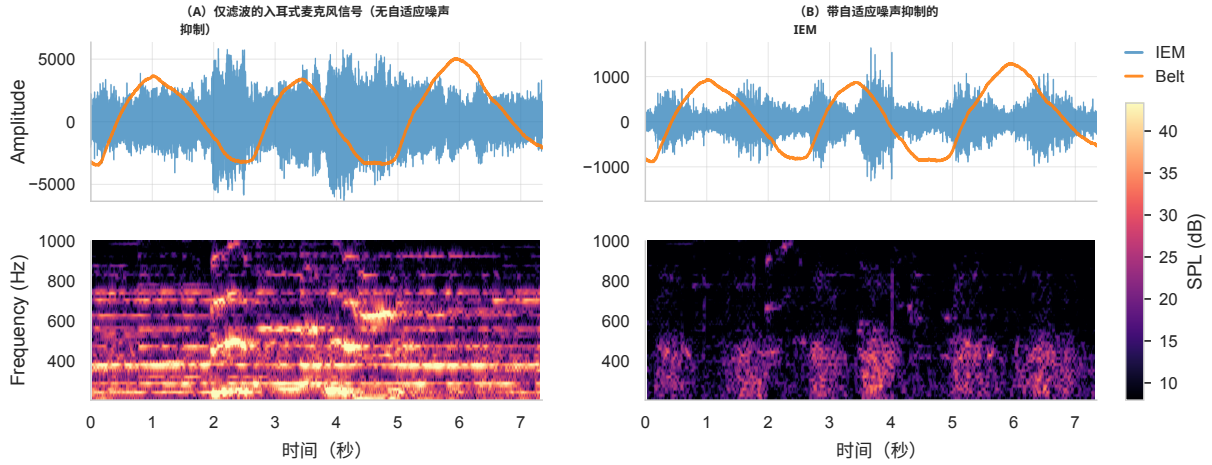


图 2. 音乐声条件下录制的自适应噪声抑制前（A）和后（B）的入耳式麦克风信号及呼吸带真实值数据。图中展示了时域原始信号及其对应的频谱图。图（A）中的吸气和呼气片段被背景噪声淹没，而图（B）中通过自适应噪声抑制将其恢复。原始音频信号和频谱图中均可观察到与呼吸带信号的相关性。

3.5.2 频谱差异度。呼吸频率估计在固定长度的分析窗口内进行。假设经过噪声抑制和带通滤波的信号以呼吸相关声学成分为主，我们仅保留谱能量超过 85%分位数的时频单元。这种选择性掩蔽通过突出最有可能与呼吸事件相关的高能量区域，得以稳健地估计呼吸声成分。选择分位数阈值时需权衡两个相互竞争的目标：既要确保保留的样本很可能对应呼吸相关活动，又要保留足够数量的样本，以稳定地估计平均呼吸频谱。通过实验，我们发现该阈值在所有评估条件下，都能很好地平衡噪声鲁棒性与频谱稳定性。利用保留的时频单元，我们通过对各短时傅里叶变换帧进行归一化并计算其均值，估计每个分析窗口的平均呼吸频谱。此处归一化采用定义如下的 p范数

$$\|\mathbf{x}\|_p \coloneqq \left(\sum_{n=0}^{N} x_n^p\right)^{\frac{1}{p}}, \tag{10}$$

对于某个向量 $\mathbf{x}$。我们不使用最大范数（切比雪夫范数）进行归一化，而是通过选择 $p = \infty$ 来定义，其满足

$$\|Y(t,\cdot)\|_\infty \coloneqq \lim_{p\to\infty} \|\mathbf{Y}(t,\cdot)\|_p = \max_{f\in\{0,\ldots,F-1\}} |Y(t,f)|, \tag{11}$$

我们发现，使用 $p = 8$ 能够在提升对极端频谱峰值的鲁棒性与保持足够的数值稳定性之间取得最佳平衡。因此，我们得到平均呼吸声估计为

$$\overline{\mathbf{Y}} \coloneqq \frac{1}{|Q_{0.85}|} \sum_{t\in Q_{0.85}} \frac{Y(t,f)}{\|\mathbf{Y}(t,\cdot)\|_p}, \tag{12}$$

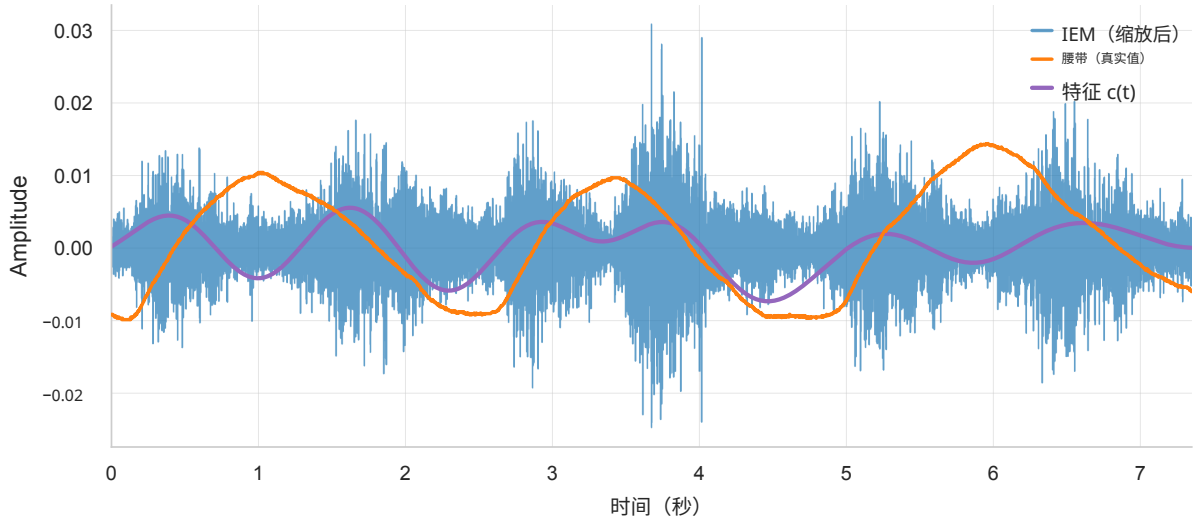


图3。该图 展示了特征信号 $c(t)$ 与去噪后的IEM信号的对比。IEM信号已缩放至 or 合适的显示 范围。特征信号在每次吸气和呼气时都包含局部最大值。

其中，Q_q 表示当前分析窗口内谱能量超过经验 q分位数的时间帧索引集合，即

$$Q_q \coloneqq \{t \in \{0,\dots,N-1\} \mid p(t) \geq \tilde{p}_q\}, \qquad \tilde{p}_q \coloneqq \text{quantile}_q\left(\{p(t)\}_{t=0}^{N-1}\right). \tag{13}$$

对于此窗口内的每个时间帧，都会计算当前帧的频谱与 公式 12 中定义的平均呼吸频谱之间的差异。为减小较大频谱偏差的影响并提高数值稳定性，我们对所得距离度量进行对数压缩。对数会削弱由瞬态频谱分量造成的极端峰值，同时保留较小偏差之间的相对差异。

基于此，我们将频谱差异度量定义为

$$d(t) \coloneqq \frac{1}{F}\log\left(\sum_{f=0}^{F-1}\left(\frac{Y(t,f)}{\|\mathbf{Y}(t,\cdot)\|_p} - \overline{\mathrm{Y}}(f)\right)^2\right) \tag{14}$$

3.5.3 呼吸频率检测。 为检测呼吸频率，将能量信号 $p(t)$ 与频谱差异信号 $d(t)$ 组合为组合特征信号 $c(t)$，由下式给出

$$c(t) \coloneqq a_p \cdot \frac{p(t)}{\|\mathbf{p}(\cdot)\|_2} - a_d \cdot \frac{d(t)}{\|\mathbf{d}(\cdot)\|_2}, \qquad \|\mathbf{x}\|_2 = \sqrt{\sum_{k=0}^{N-1} x^2(k)}, \tag{15}$$

其中我们选择 $a_p = a_d = 0.5$，使两个特征具有相同权重。请注意，由于 $d(t)$ 会在每次吸气和呼气时产生最小值，因此需要反转 a_d 的符号，才能与 $p(t)$ 正确融合。为了从最终的特征信号中估算呼吸频率，我们在应用汉明窗后，使用快速傅里叶变换（FFT）计算 $c(t)$ 的傅里叶变换。为获得更高的频率分辨率，我们在执行 FFT 之前先在时域进行零填充。

呼吸会产生两个显著的频谱分量，因为吸气和呼气都会对应特征信号中的最大值。因此，在呼吸基频及其一次谐波（即呼吸频率的两倍）处均会出现频谱峰值。所得信号的特性可见于图 3。因此，为了提高峰值的可检测性，我们计算定义如下的谐波频谱

$$C^*(f) \coloneqq |C(f)| + |C(2f)|, \quad f \in \{0, \ldots, \tfrac{F}{2} - 1\} \tag{16}$$

其中，$C(f)$ 表示特征信号的傅里叶频谱。选取 $C^*(f)$ 中的主峰作为估计的呼吸频率。随后，估计的呼吸频率可表示为

$$\widehat{\mathrm{RR}} \coloneqq \underset{f \in \{0,\ldots,\frac{F}{2}-1\}}{\arg\max} \; C^*(f). \tag{17}$$

3.6 通道融合与异常值剔除

由于我们的系统由两只耳机组成，每只耳机都会输出一个独立的呼吸频率估计值，因此我们可以将两者融合为一个值，以获得更准确的估计。我们对两个通道进行等权融合，即融合值为

$$\widehat{\mathrm{RR}}_{\mathrm{F}} \coloneqq \frac{\widehat{\mathrm{RR}}_{\mathrm{L}}}{2} + \frac{\widehat{\mathrm{RR}}_{\mathrm{R}}}{2}. \tag{18}$$

此外，我们发现，左右两侧估计值之间绝对差异较大的时间窗口（定义为

$$\Delta\widehat{\mathrm{RR}} \coloneqq |\widehat{\mathrm{RR}}_{\mathrm{L}} - \widehat{\mathrm{RR}}_{\mathrm{R}}|, \tag{19}$$

$\widehat{\ }$往往会导致融合后的呼吸频率 RR_{F}出现较大的估计误差。基于这一观察，我们剔除$\widehat{\ }$满足以下条件的估计值：$\Delta\mathrm{RR}$ 超过预设阈值。该阈值的选取将在第 5.1.6 节中讨论，我们将在该节分析估计准确度与被丢弃时间窗口比例之间的权衡。

3.7 设备端处理

EarResp-ANS 在 OpenEarable 2.0 平台上实现，并分配到板载的*TensilicaHiFi 3z* 数字信号处理器和 *Nordic nRF5340* 应用处理器上。自适应噪声抑制在数字信号处理器上执行，该处理器针对低延迟音频预处理进行了优化；而呼吸频率估计则由微控制器执行，后者专为通用设备端计算而设计。这种划分使计算工作负载能够在两个处理单元之间高效分配。完整的处理流程如图 1所示。

麦克风信号以 48kHz 的采样率采集，并通过数字信号处理器的抽取器模块降采样至 8kHz。后续所有噪声抑制阶段均在*TensilicaHiFi 3z* 数字信号处理器核心上执行。信号处理以每块 24 个样本进行，采用 24 位定点表示。作为初始预处理步骤，IEM 和外耳麦克风信号均使用四阶巴特沃斯滤波器进行带通滤波，该滤波器由双二阶节级联实现。滤波器的通带范围为 200 Hz 至 1000 Hz，如第 3.3 节所述。

噪声抑制采用在 Tensilica HiFi3zDSP 上实现的延迟LMS自适应滤波器完成，具体如第 3.4 节所述。该滤波器使用 256 个 FIR抽头，并对期望信号（内置麦克风信号）施加 64 个采样点的延迟。我们再次采用 24 位定点表示，从而利用数字信号处理器的 24×24 MAC单元。

处理后的音频流通过 $\mathrm{I}^2\mathrm{S}$ 接口传输至 nRF5340，以进行后续的特征提取和呼吸频率估计。为降低计算负载，同时保留与呼吸相关的信息，我们使用抽取器链进一步将信号降采样至 2000Hz 的采样率。呼吸频率估计通过基于短时傅里叶变换的处理流程计算，该流程使用 CMSIS-DSP 库[3] 和 ARM 数字信号处理指令集实现。

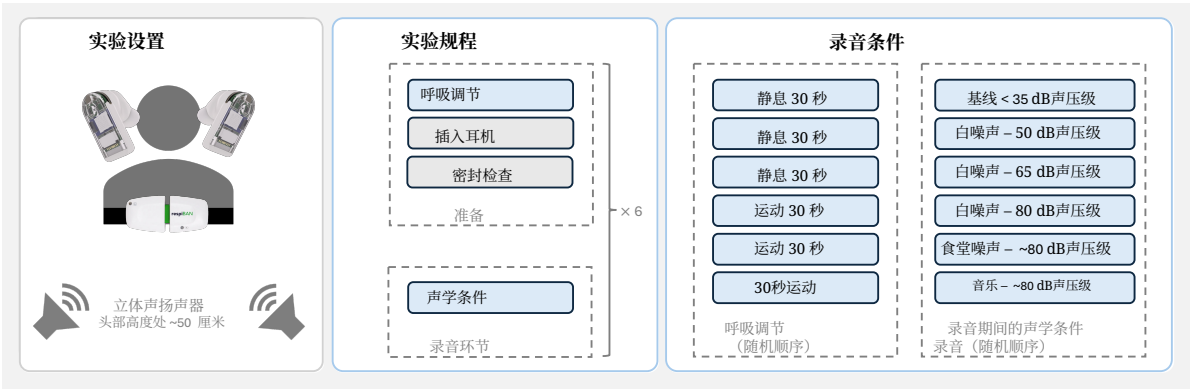


图 4. 实验设置、录音流程和录音条件。参与者坐着佩戴双耳 OpenEarable 耳机和 respiBAN 呼吸带。音频通过立体声扬声器播放（约 50 cm，头部高度）。每位参与者按随机顺序完成六个录音环节。每个环节均设定呼吸调制（无调制，或通过 30 秒开合跳诱发）和一种声学条件。声学条件包括低噪声基线（<35 dB声压级）、白噪声（50/65/80 dB声压级）、食堂噪声（约 80 dB声压级）和音乐（约 80 dB声压级）；每位参与者在每种条件下各录制一次。

计算出特征信号并进行带通滤波后，我们将其按 32 倍降采样至约3。9Hz，以降低谐波频谱计算所需的复杂度和内存。这样仍能捕捉高达 58。6每分钟周期数的呼吸频率，足以覆盖体育活动（例如以 12km/h 的速度跑步）中报告的呼吸频率第 99 百分位数，从而涵盖绝大多数人群 [57]。

随后，每个耳戴设备的呼吸率估计值会通过低功耗蓝牙（BLE）传输到智能手机。在智能手机上，系统会接收来自各耳戴设备的呼吸率估计值，并按照 公式 18 所述的方法将其融合，以获得单一且稳健的输出。此外，还可以根据两个通道估计值之间的差异，按照 第 5.1.6 节 所述的方法对潜在离群值进行滤波。

4 数据采集

本节介绍用于评估 EarResp-ANS 的数据采集流程。我们概述硬件平台和录音设置、实验规程，以及呼吸率真值的推导方法。此外，我们还介绍生成的数据集，以及分析前所执行的数据清理和分段步骤。本研究的实验规程旨在真实声学条件下采集呼吸信号，同时确保结果可复现且参考测量可靠。

4.1 实验规程与设置

图 4 概述了实验设置和录制方案，包括传感器配置、声学条件和会话结构。所有参与者均佩戴配备入耳式麦克风、用于音频采集的 OpenEarable 智能耳机。此外，使用了 PLUX respiBAN BLE 呼吸带 [50] 作为参考系统，通过测量与呼吸相关的胸部扩张来获取呼吸频率真实值。该设备产生的信号示例如图 3 所示。选择该系统而非持续气道正压（CPAP）面罩等其他系统，是因为后者会影响呼吸音 [65]。音频数据采集使用了 OpenEarable 2。0 [54] 平台，相关介绍见

第 3.3 节。录制期间，麦克风信号通过级联抽取级降采样至 8000Hz，每一级均应用二阶巴特沃斯抗混叠滤波器。随后将降采样后的数据录制到 microSD 卡。

在所有录音过程中，参与者均保持坐姿，并被要求避免大幅度动作。不过，为了在录音期间保持自然舒适的姿势，参与者可以在不做大幅度动作的前提下阅读手机或笔记本电脑上的内容。

每次录音前，都会进行密封检查，以确保耳道得到适当密闭。具体做法是通过耳机扬声器播放多音信号，并分析入耳式麦克风测得的频率响应。利用明显的低频成分来确认密闭情况，从而确保不同录音过程中设备与耳道之间的声学耦合一致。录音在以下声学条件下进行：

- 环境噪声低于 35 dB声压级（SPL）的基线条件
- 以 50 dB、65 dB 和 80 dB声压级播放白噪声
- 约为 80.0 dB声压级的食堂噪声（±5.2dB 的时间标准差），通过人工声音播放生成
- 约为 80.1 dB声压级的音乐条件（±5.8dB 的时间标准差），播放内容为流行音乐

所有声学条件均通过市售立体声扬声器播放。扬声器放置在头部高度，距离参与者约 50cm，以确保不同录音中的声音暴露一致。

每位参与者总共完成六个录音环节，分别对应六种声学条件。各条件的顺序在参与者之间随机安排。为增加呼吸频率的变异性，每位参与者随机选取三个录音环节，在运动诱发的呼吸调制条件下录制；具体做法是在录音前立即进行 30 秒开合跳。每次录音前，参与者均按照指导取下并重新戴上耳机，以增加耳机在耳道内贴合度的变异性。

研究开始前，对所有声音播放音量进行校准，以确保不同条件下的声压级准确且可复现。校准时，将声级计放置在参与者所在位置的耳朵高度处。调整扬声器的播放增益，使声源距离 50cm 处达到目标声压级。对于非平稳声学条件（食堂噪声和音乐），校准采用等效连续声压级（等效连续 A 计权声级）进行。

4.2 真实值推导

评估在 20 秒分析窗口上进行。对于每个窗口，通过对呼吸信号应用汉明窗并进行快速傅里叶变换，提取主导谱峰，从而估计呼吸率真实值。为提高频率分辨率，在快速傅里叶变换之前将信号零填充至其原始长度的 32 倍。必须执行此步骤，因为 20 秒的窗口时长本身只能产生 0.05Hz 的频率分辨率（相当于 3 每分钟周期数），不足以进行准确的呼吸频率估计。零填充有效地对频谱进行插值，从而能够更精确地定位主导呼吸频率，而不改变底层信号内容。

在数据预处理过程中，将连续记录切分为长度为 20 s、重叠率为 50% 的时间窗口。根据真实信号，剔除了呼吸频率低于 7.5 CPM 或高于 30 CPM、不合乎常理的时间窗口。此外，对于真实值呼吸信号的快速傅里叶变换（FFT）结果中无法识别出与呼吸频率对应的明确主峰的时间窗口，也予以剔除。这一步确保仅保留具有可靠参考呼吸频率的片段用于评估。由于真实信号不可靠，共排除了 19.3% 的帧。

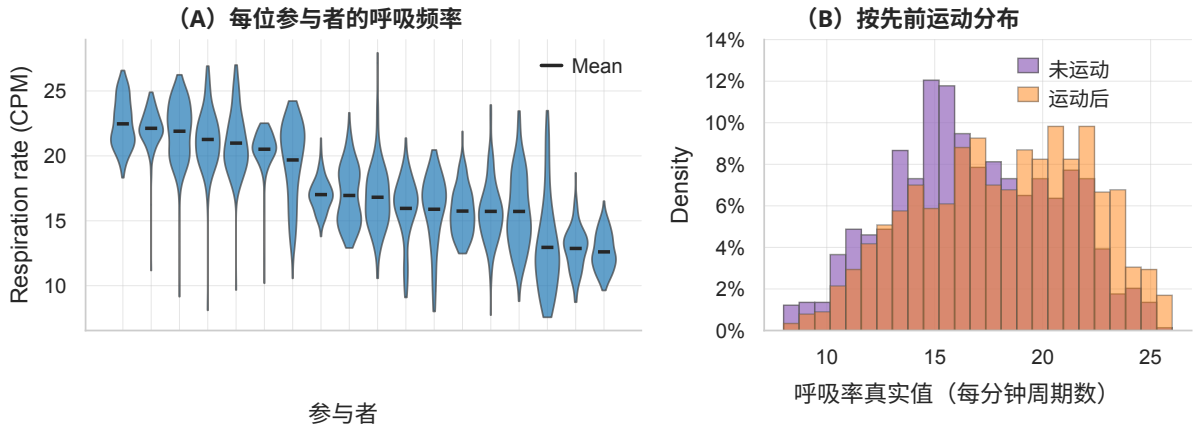


图 5. (A) 小提琴图展示了每位参与者的呼吸频率真值分布。根据所有录音中的平均呼吸频率，事后对参与者进行排序。(B) 直方图展示了有无先前体力活动情况下的呼吸频率真值分布。

4.3 数据集

数据采集自 20 名参与者，其中两个完整的采集时段因真实信号设备故障而损坏，最终从 18 名参与者（男性：14 名，女性：6 名）中获得 2812 个有效的 20s 时间窗口。参与者年龄介于 22 至 46 岁之间（$M = 26.55$，$SD = 5.29$），呼吸频率介于 7.57CPM 至 27.94CPM。图 5 展示了各参与者以及不同活动条件下的呼吸频率分布。所得数据集的呼吸频率中位数为 17.97CPM，标准差为 3.84CPM。更多详细信息见附录 A 中的表 6。由于体能会影响活动后呼吸频率的变化，我们记录了参与者平时的运动频率：9 人报告没有运动，5 人每周运动 1–2 次，5 人每周运动 3–4 次，1 人每周运动超过 4 次。

5 评估

我们重点从两个方面评估 EarResp-ANS：(i) 呼吸频率（RR）估计算法的总体性能，以及 (ii) 噪声抑制阶段的有效性及其对不同声学条件下呼吸相关信号质量的影响。

5.1 呼吸频率估计性能

为评估呼吸频率估计算法的性能，我们定义了以下指标。所有指标均基于估计的呼吸频率（以每分钟周期数表示）计算，并与真实参考值进行比较。一个呼吸周期定义为一个完整的吸气-呼气序列。周期时长从一次吸气开始时刻测量至下一次吸气开始时刻。此定义符合呼吸频率的标准生理学解释 [37]。

作为评估指标，我们考察平均绝对误差（MAE）和均方根误差（RMSE）。平均绝对误差定义为估计的呼吸频率与真实值之间的绝对差，单位为每分钟周期数（CPM）；均方根误差定义为估计的呼吸频率与真实值之间均方误差的平方根，单位为 CPM。

表 2. 各条件下的呼吸频率估计误差。报告所有有效窗口的平均绝对误差（MAE）和均方根误差（RMSE），以及仅使用分类为置信的窗口时的估计误差（剔除了 14.4%）。对于离群值检测，$\tau = 0.52$选用 CPM，作为所有内群值的 3σ 区间限值。对于单通道和融合估计，括号中显示的是按照 3σ 式 21 中定义的方法进行 MAD 异常值剔除后的内群值表现。

	平均绝对误差（CPM）			均方根误差（CPM）		
条件	**单通道（3σ）**	**融合（3σ）**	**置信单一（3σ）**	**融合（3σ）**		**置信**
基线（<35分贝）	0.94 (0.15)	0.85 (0.14)	0.70	2.52 (0.20)	2.27 (0.19)	2.21
50 分贝白噪声	0.52 (0.11)	0.47 (0.11)	0.29	1.79 (0.14)	1.56 (0.13)	1.22
65 分贝白噪声	0.56 (0.14)	0.55 (0.14)	0.36	1.85 (0.19)	1.67 (0.18)	1.35
80 分贝白噪声	1.54 (0.16)	1.43 (0.15)	0.75	3.25 (0.22)	2.84 (0.20)	2.25
65 分贝自助餐厅	1.03 (0.18)	0.95 (0.17)	0.39	2.50 (0.24)	2.07 (0.22)	1.18
80 分贝音乐	0.85 (0.16)	0.81 (0.15)	0.41	2.52 (0.20)	2.30 (0.19)	1.46
总体	0.90 (0.15)	0.84 (0.14)	0.47	2.45 (0.19)	2.16 (0.18)	1.65

5.1.1 总体性能。我们首先使用经噪声抑制的入耳式麦克风信号，评估 EarResp-ANS 在呼吸频率（RR）估计方面的总体性能。所有声学条件下的误差指标详见表 2，误差分布及其与真实值的一致性则通过图 6和图9进行可视化。

在所有测试条件下，该系统的单通道设备端估计 MAE 为 0.90，经过双耳融合后为 0.84。如表 2所示，应用基于通道差异的异常值检测（见第3.6节）可进一步提高估计准确度，将 MAE 降至 0.47。

图图6 （**A**）中的 Bland–Altman 图证实了这些发现：大多数估计值聚集在零误差线附近，表明在整个观测呼吸率范围（7.57 至 27.94 每分钟周期数）内系统性偏差较小（$\mu = 0.05$）。如后滤波 Bland–Altman 图所示，基于通道差异识别不可靠窗口可有效剔除大量离群值，见图6 （**B**）。这使一致性界限从一致性界限 = [−4.18, 4.27]每分钟周期数降至一致性界限 = [−3.11, 3.34]每分钟周期数。第 5.1.6节对异常值剔除策略进行了更详细的分析。

5.1.2 不同参与者间的性能。为评估跨参与者的泛化能力，我们采用方差分解方法分析有符号预测误差，该方法将受试者内变异性与受试者间变异性区分开来。具体而言，这种分解可量化残差误差在多大程度上由参与者特异性特征所驱动。在来自 18 位参与者的所有评估窗口中，MAE 的受试者间方差为 $\sigma^2_{\text{between}} = 0.39\text{CPM}^2$,，而受试者内方差则显著更大，为 $\sigma^2_{\text{within}} = 3.96\,\text{CPM}^2$。由此得到的可推广性比率定义为

$$G \coloneqq \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{between}} + \sigma^2_{\text{within}}}, \tag{20}$$

of $G = 0.089$，这表明总误差方差中可归因于受试者特定效应的部分不足 10%。这些发现表明，尽管不同参与者的平均误差存在差异，但变异性的主要来源在于参与者自身。因此，所提出的方法在不同用户之间展现出很强的泛化能力，残差误差主要由瞬态信号特征

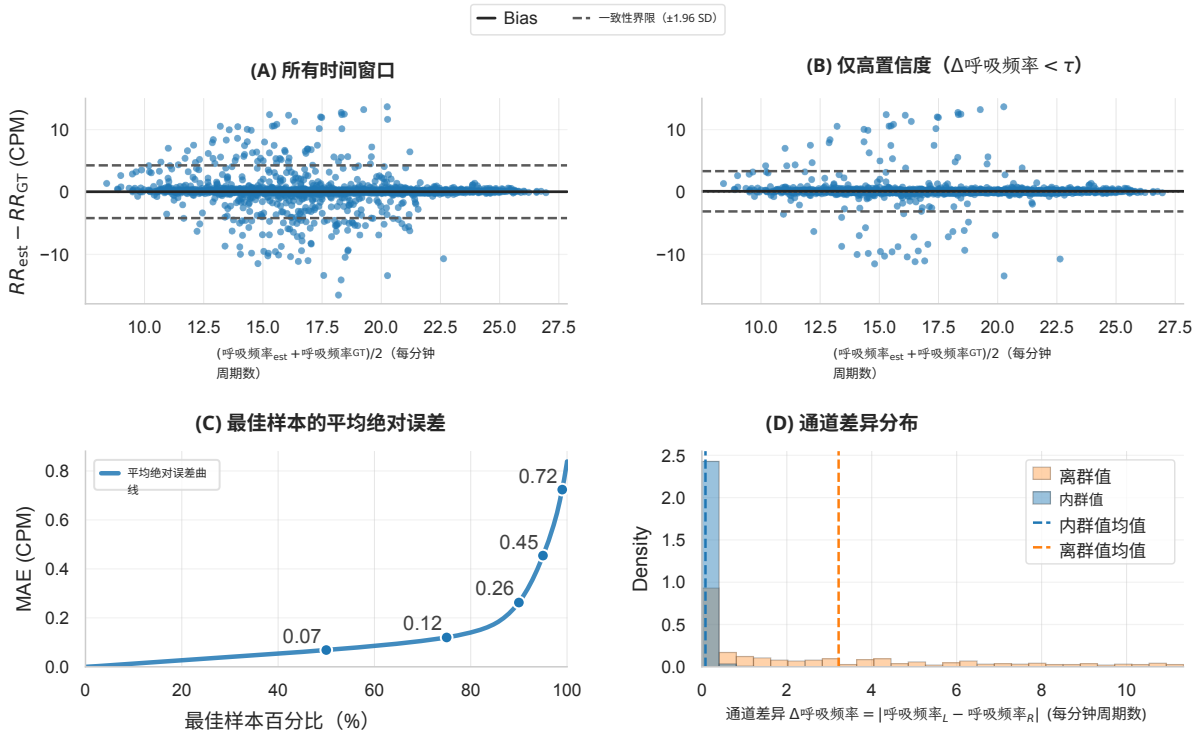


图 6. 融合后的呼吸率估计值与基于置信度的筛选分析。**(A)** 所有有效分析窗口的 Bland–Altman 图（偏差 $\mu = 0.05$ 每分钟周期数，一致性界限 $= [-4.18, 4.27]$ 每分钟周期数），展示融合估计值与呼吸率真实值之间的一致性。**(B)** 置信度筛选后的 Bland–Altman 图（偏差 $\mu = 0.12$ 每分钟周期数，一致性界限 $= [-3.11, 3.34]$ 每分钟周期数），仅保留通道间差异较小（$\Delta RR < \tau = 0.52$ 每分钟周期数）的窗口。**(C)** 平均绝对误差（MAE）随最佳样本百分比变化的情况，展示随着不可靠窗口逐步剔除，估计准确度如何提升。(D) </style)离群点窗口与非离群点窗口的绝对通道间差异 $\Delta RR = |$呼吸频率$_L$ − 呼吸频率$_R|$ 分布，突出显示置信度标准所产生的区分效果。

或录音条件，而非系统性的受试者相关偏差。每位参与者的误差分布如图7所示。

5.1.3 不同条件下的呼吸频率估计性能。 图 8 总结了不同声学条件下的呼吸频率估计误差。更详细的结果见 表 8 ，位于 附录 A。我们比较了简单带通滤波信号的性能与采用自适应噪声抑制（ANS）所取得的结果。此外，我们还将本方法与 Hu 等人在 BreathPro 中采用的噪声抑制方法进行基准比较。 [24]，该方法依赖于一个在所有参与者的数据上离线训练的固定滤波器。在 BreathPro 中，噪声抑制采用基于短时傅里叶变换的滤波方法在频域中进行，随后将去噪信号变换回时域，以估计呼吸频率。此外，我们还将延迟 LMS 方法与 Martin 使用的归一化 LMS 方法进行比较 [43]。为单独评估各种噪声降低方法的有效性，我们在所有基准比较中（包括 BreathPro 方法）均采用 EarResp-ANS 中相同的呼吸频率提取流程。这样可以确保性能

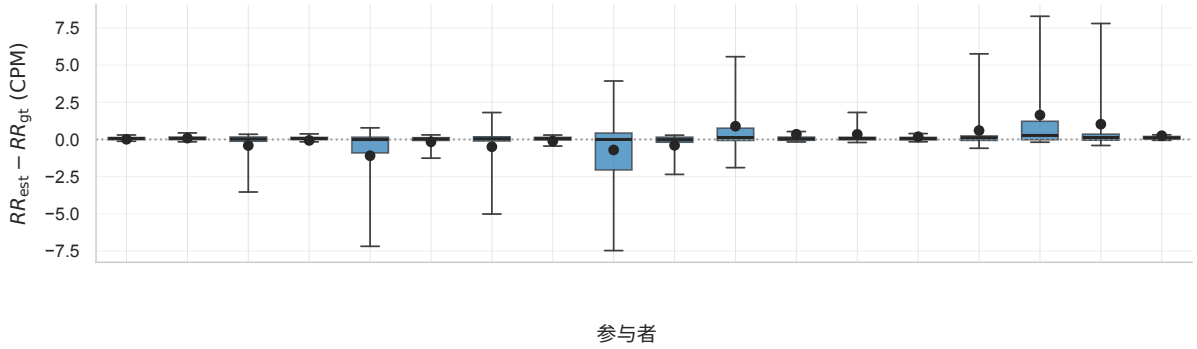


图 7. EarResp-ANS 在各参与者中的呼吸频率估计误差分布，以箱线图展示 .

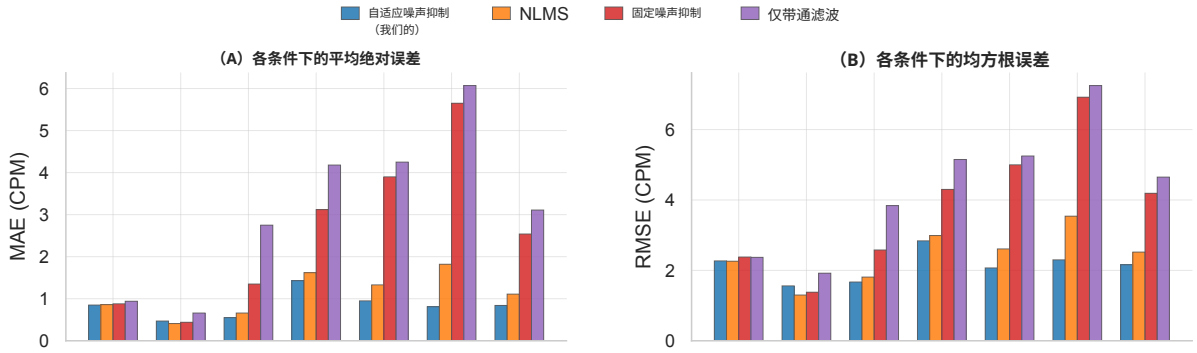


图 8. 使用 EarResp-ANS 的估计阶段进行呼吸频率估计时的误差，对比不同噪声抑制方法：ANS（本文方法）、归一化最小均方（NLMS）、BreathPro 中使用的离线训练滤波器 [25]（固定噪声抑制，FixedNS）以及仅带通滤波（BPF-only）基线。

图中所示的差异 图10 完全源于信号清理质量，而非估计算法的差异。

总体而言，应用 ANS 可显著提升性能，将平均绝对误差（MAE）从 3.11 降至 0.84。在基线条件下，噪声抑制方法和仅滤波方法的性能相当（ANS 的 MAE：0.85；滤波后的 MAE：0.94）。相比之下，随着噪声水平升高，仅对滤波后数据进行处理的性能显著下降，在音乐条件下 MAE 达到 6.07。应用 ANS 后，所有噪声条件下的性能均显著提升，估计准确度接近基线测量（音乐条件下 MAE 为 0.81）。这表明所提出的方法能够适应不同噪声水平，具有良好的鲁棒性。白噪声 50 dB（MAE 为 0.47）和白噪声 65 dB（MAE 为 0.55）条件下的性能最高；音乐（MAE 为 0.81）和食堂噪声（MAE 为 0.95）条件下的性能略低。白噪声 80 dB 条件下的性能最低（MAE 为 1.43），这与信号质量下降相符（见第 5.2.2 节）。

与BreathPro的噪声抑制方法相比， [25], 我们的自适应方法始终取得更好的总体性能，平均绝对误差为0.84，而对比结果为2.54。对于基线条件，

BreathPro的表现与之相当，但略差于自适应噪声抑制（自适应噪声抑制的平均绝对误差：0.85，BreathPro的平均绝对误差：0.88）。在更响的声学条件下，BreathPro的性能显著下降。这一影响在音乐条件下最为明显，BreathPro的平均绝对误差达到5.65，而所提出的自适应方法为0.81。

归一化最小均方（NLMS）方法取得了与所提出的自适应噪声抑制方法相当的性能，但在所有条件下始终略差。尤其是在音乐条件下，平均绝对误差显著更高（自适应噪声抑制平均绝对误差：0.81，nLMS平均绝对误差：1.82）。这种表现可归因于归一化，它提高了数值稳定性，但同时降低了有效自适应速度。因此，nLMS滤波器对快速变化的非平稳信号自适应更慢，导致在动态声学条件下性能下降。

5.1.4 体力活动的影响。 为评估EarResp-ANS对呼吸动态变化的鲁棒性，我们分析其在不同生理状态下的性能：静息时以及体力活动后立即。使用融合估计，该算法在两种条件下均取得相当的准确度，表明其在更广范围的呼吸频率值上具有鲁棒的泛化能力。

具体而言，体力活动后进行的测量中，融合平均绝对误差为0.61每分钟周期数，静息条件下为1.10每分钟周期数。尽管运动后状态的准确度略高，但0.49每分钟周期数的差异仍远低于各组内部的变异性，相应的标准差也反映了这一点（运动后标准差：1.87每分钟周期数，静息标准差：2.13每分钟周期数）。这表明所观察到的差异并不意味着静息时存在系统性的性能下降。体力消耗后准确度略有提高，可能归因于更明显且在声学上更为显著的呼吸模式，这有助于更可靠地进行呼吸相关信号提取。总体而言，这些结果证实所提出的方法在不同呼吸频率下保持稳定。更多细节见附录 A 中的表7。

5.1.5 离群值分布。分析估计误差的分布后，我们发现，基于中位数绝对偏差（MAD）的稳健标准差估计$\hat{\sigma} = 0.185$明显小于传统标准差 2.17。这表明误差分布具有重尾特征，且主要由少数离群值造成；这一点可从图6(C)中观察到。

总体而言，所有估计值中有 80.3% 落在以中位数为中心的稳健三西格玛等效区间内，对应的绝对误差小于0.612。将评估范围限制在此置信区间内，其定义为

$$\mathcal{I}_{\mathrm{MAD}} \coloneqq [\mathrm{median}(x) - 3\hat{\sigma},\ \mathrm{median}(x) + 3\hat{\sigma}], \quad \hat{\sigma} \coloneqq 1.4826 \cdot \mathrm{median}(|x_i - \mathrm{median}(x)|), \tag{21}$$

使平均绝对误差从 0.84 降至 0.14。

进一步检查离群值分布后发现，离群值往往聚集在真实呼吸频率的两倍和一半附近。这一现象反映在图9所示的百分比误差分布中，表现为靠近 −50% 和 100%的明显簇群。这些簇群表明存在谐波估计误差，即算法错误地选择了次谐波或谐波分量，而非呼吸基频。

5.1.6 基于通道差异的异常值剔除。最后，我们分析通道差异$\widehat{\ }$呼吸频率 Δ与公式 19中定义的离群值出现情况之间的关系。具体而言，我们计算左右耳机所得呼吸率估计值之间的绝对差值。我们发现，通道间差异较大的分析窗口往往具有更高的估计误差，如图6(D) 所示。这一发现支持将通道一致性用作置信度指标，并促使我们将其应用于异常值抑制。然而，部分离群值的通道差异仍然较小，这表明仅凭这一标准无法消除所有错误估计值。在图6所示的异常值剔除前后 Bland–Altman 图中也可以看到这一现象：大多数离群值已被抑制，但仍有少数残留。尽管存在这一局限，所提出的异常值剔除策略仍显著提升了

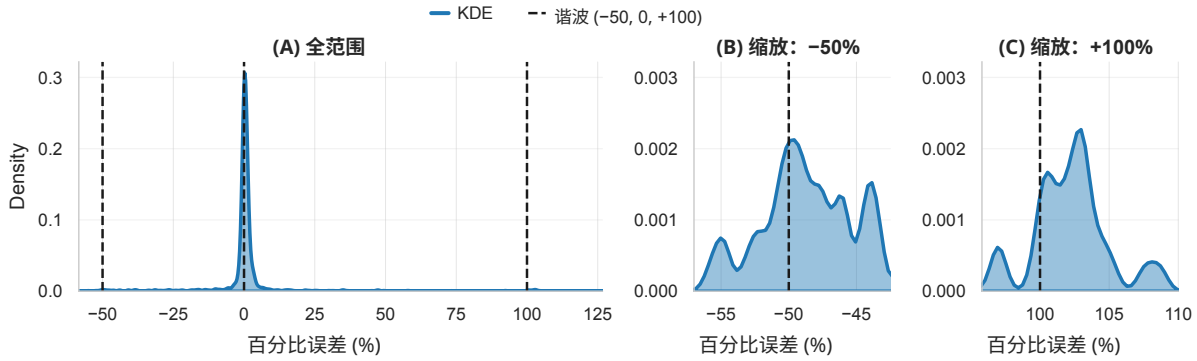


图9. 百分比呼吸率估计误差的分布。（**A**）展示了在观测值全范围内百分比误差的核密度估计（KDE）。右侧两个面板分别提供了以 −50%（**B**）和 +100%（**C**）为中心的显著谐波结构的放大视图。

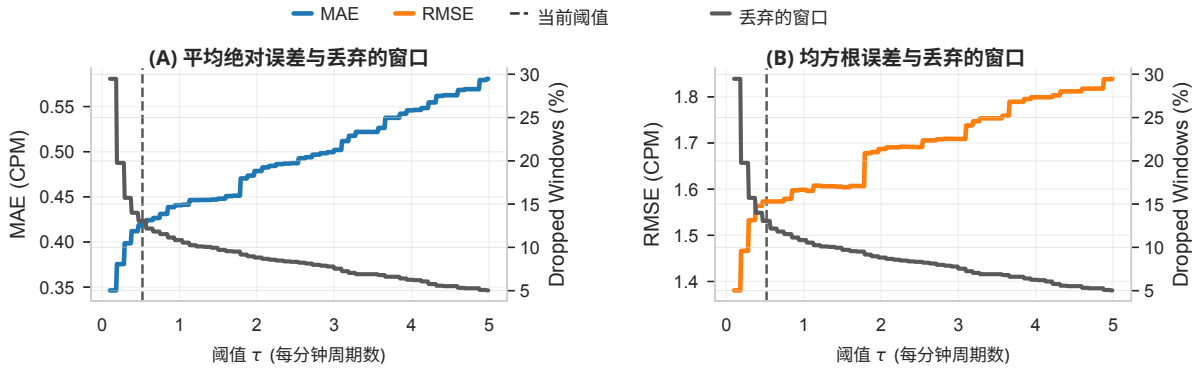


图10. 该图展示了阈值 τ（用于基于通道差异的异常值剔除）$\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}$（定义为 Δ呼吸频率 = |呼吸频率$_L$ −呼吸频率$_R$|（第3.6节））与所得平均绝对误差（**A**）和均方根误差（**B**）的影响之间的关系。

估计准确度，将平均绝对误差从 0.84 降至 0.47。差异阈值对估计误差和被拒绝窗口比例的影响进一步见图10。

5.2 噪声抑制性能

我们首先研究对内置麦克风信号应用基于延迟LMS的降噪所带来的益处。从信号能量降低和呼吸相关信息保留情况两方面评估噪声抑制性能。为测量呼吸信息的保留情况，我们监测其与呼吸信号的相关性，并评估该方法在降噪后数据上的表现。

表 3. 不同声学条件下基于中位数的相关性与噪声降低分析。分别报告滤波信号、ANS 信号及其比值（NS / Filtered）的相关性指标。噪声抑制以能量降低1 − （E_{ns}/E_{filter}）及相应的衰减（分贝）表示，后者按 $10\log_{10}$（E_{ns}/E_{filter}）计算。

	相关性（中位数）			噪声抑制	
条件	**滤波**	**ANS**	**比率**	**[%]**	**[dB]**
基线（<35分贝）	0.31	0.31	1.0	0.6	−0.03
50 分贝白噪声	0.33	0.41	1.2	23.4	−1.2
65 分贝自助餐厅	0.01	0.38	30	98.9	−19.6
65 dB白噪声	0.06	0.44	7.4	86.7	−8.8
80 dB白噪声	0.01	0.36	33	99.1	−20.4
80 分贝音乐	0.01	0.38	44	98.6	−18.5

5.2.1 噪声抑制。为量化噪声抑制效果，我们比较应用延迟LMS滤波器前后内置麦克风录音的信号能量。对于每个分析窗口，均在与呼吸相关的频带内计算信号能量。因此，噪声抑制效果定义为：

$$\mathrm{NR} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1} e^2(n)}{\sum_{n=0}^{N-1} d^2(n)}\right)\ \mathrm{dB}. \tag{22}$$

5.2.2 呼吸相关信号相关性。 除了噪声抑制之外，保留呼吸相关信息也至关重要。从生理学角度看，胸部运动的速度与气流速度相关，而气流速度又会调制内置麦克风捕获的呼吸声音的振幅。因此，呼吸相关信号的时间包络预计与真实值呼吸信号的斜率相关。为捕捉这种关系，我们计算麦克风信号均值能量与真实值呼吸信号绝对斜率之间的相关性，其中使用 $K = 100$ 计算割线斜率，相当于在 400Hz 采样率下产生 0.25s 的时间延迟。基于此，我们正式定义呼吸信息（RI）指数作为评估指标：

$$\begin{aligned} \mathrm{RI} &\coloneqq \mathrm{corr}\big(E_y, s_g\big) \\ E_y(t) &\coloneqq \frac{1}{T}\sum_{k=0}^{K-1} y^2(t-k) \\ s_g(t) &\coloneqq \frac{|g(t) - g(t-K)|}{N}, \end{aligned} \tag{23}$$

其中， $corr(x, y)$ 表示 x与 y之间的相关性。

5.2.3 结果。 定量结果汇总于表 3。在所有声学条件下，应用噪声抑制后，信号能量均持续降低，表明外部噪声成分得到了有效衰减。测得的能量降低范围为 −0.03分贝，对应基线条件（基线 <35 分贝）至 −20.4分贝，对应 80 dB 白噪声。

在应用噪声抑制之前，只有在基线条件（0.31）和 50 dB白噪声测量（0.33）中，才观察到内置麦克风信号与呼吸信号之间存在明显相关性。在其余声压级更高的条件下，与呼吸信号的相关性较弱。音乐条件下这一现象尤为明显，相关性指标（RI）降至 0.01，

表4. 消融实验：分析 ANS、特征组合 $c(t)$、通道融合和置信度筛选对呼吸频率估计性能的影响。

ANS	特征 $c(t)$		融合	置信度	平均绝对误差 (CPM)	均方根误差 (CPM)
	$\mathbf{d(t)}$	$\mathbf{p(t)}$				
×	✓	✓	×	×	3.26	5.01
✓	✓	×	×	×	1.10	2.79
✓	×	✓	×	×	0.90	2.55
✓	✓	✓	×	×	0.90	2.35
✓	✓	✓	✓	×	0.84	2.26
✓	✓	✓	✓	✓	**0.47**	**1.65**

这很可能是由于噪声具有强烈的非平稳特性。应用 ANS 后，我们观察到所有声学条件下的相关性数值均稳定高于 0.31。这表明自适应滤波器能够有效保留呼吸相关信号成分。值得注意的是，所得相关性数值仍与基线测量（0.31）相当，这说明噪声去除后保留的信号质量不逊于近乎理想的录音条件。除基线外，在所有声学条件下，噪声抑制都显著增强了呼吸相关信息，增强幅度从 1.2 倍（50 dB白噪声）到 44 倍（音乐）不等。这些结果凸显了所提方法在不利声学条件下增强呼吸相关信息的有效性。在平稳声学条件下，信号质量最高，尤其是在 50 dB 和 65 dB 白噪声条件下。相比之下，80 dB白噪声条件下的信号质量最低。在这一条件下，降噪后的信号保留的原始信号能量不足 1%，表明呼吸信号受到外部噪声的强烈干扰。这可能导致该条件下的呼吸频率估计性能略低于其他白噪声条件。

综合来看，这些结果表明，固定的离线训练滤波器无法充分适应不同耳塞贴合情况下的声学条件。此类滤波器在安静或轻度噪声环境下可能表现尚可，但在较强且非平稳的噪声下，无法稳健地恢复呼吸相关信号。相比之下，所提出的自适应噪声抑制（ANS）在所有评估条件下都能持续保留呼吸相关信息，凸显了在线自适应对于实现稳健且与参与者无关的入耳式呼吸感知的重要性。

5.3 消融实验

各系统组件对总体性能的贡献汇总于表4。结果证实了前文各节报告的观察结果。在本研究考察的录音条件下，ANS 对性能的影响最大，使平均绝对误差从3.26 降至 0.90。此外，将对数谱能量特征 $p(t)$ 与谱不相似度特征 $d(t)$ 结合使用，始终优于仅依赖任一特征的配置，这表明这两种表征捕获了互补的呼吸相关信息。融合两只耳机的呼吸频率估计值还可进一步提升性能。然而，除 ANS 外，最显著的改进来自剔除低置信度窗口，这使平均绝对误差进一步从0.84 降至 0.47。

表 5. 各个系统组件的内存占用、实时指数（实时指数）以及延迟（由处理引起的延迟和缓冲区延迟），这些组件包括在数字信号处理器上执行的自适应噪声抑制、短时傅里叶变换的计算，以及以 1 赫兹速率执行的特征信号提取与呼吸频率估计。

组件	内存	实时指数	计算延迟	缓冲区延迟
ANS	3488 字节	1.20 %	–	11 毫秒
STFT	768 字节	1.56 %	122 μ秒	8 ms
特征与呼吸频率	12000 字节	0.10 %	1007 μ秒	1000 毫秒

5.4 运行性能与延迟

我们评估了所提出系统的运行时性能和端到端延迟，分别针对在数字信号处理器上执行的 ANS 阶段，以及在微控制器上执行的呼吸频率估计流水线。结果汇总于表 5。

5.4.1 数字信号处理器运行时间与音频延迟。ANS 算法在 *Tensilica HiFi 3z* 数字信号处理器上以 8000 Hz 的采样率运行，相对于呼吸相关频带实现了四倍过采样。在此配置下，延迟LMS滤波器引入的数字信号处理器负载较低，仅占可用处理能力的约 1.2%。音频处理以每块 24 个样本进行，因此在 8000 Hz 下会产生 24 个样本的固有处理延迟，即 3 毫秒。此外，延迟LMS滤波器还会引入 64 个样本（8 毫秒）的显式延迟，以确保估计的主路径满足因果性。与基于分块的缓冲延迟相比，额外的处理时间可以忽略不计。因此，降噪后的麦克风信号总音频延迟约为 11 毫秒。包括滤波器系数以及所需的输入与梯度缓冲区在内，ANS 阶段需要约 3.5 kB 的内存。

5.4.2 呼吸频率估计的微控制器运行时间。呼吸频率估计在*NordicnRF5340* 微控制器上执行，使用采样率为 2000 Hz 的降采样后的麦克风信号。短时傅里叶变换（STFT）采用窗口长度为 128 个样本、帧移为 16 个样本，并使用 16 位定点精度进行计算。在此配置下，STFT 计算的平均实时因子为 1.56%，对应于每个短时傅里叶变换帧的平均执行时间约为 122 μs。包括快速傅里叶变换计算和加窗所需的所有输入与输出缓冲区在内，该阶段的内存占用为 768 字节。使用 32 位浮点精度逐窗口计算谐波频谱（包括峰值检测和呼吸频率估计），运行频率为 1 Hz。该阶段额外产生 0.10% 的计算负载，对应于每次更新的平均执行时间约为 1007 μs。存储 20 s 窗口内的降采样短时傅里叶变换结果及相应的特征信号，并提取特征信号，总共需要 12 kB 内存。

5.4.3 实时适用性。 总体而言，测得的运行时间和延迟表明，所提出的系统在商用嵌入式硬件上运行时完全满足实时约束。该算法在微控制器和数字信号处理器上的计算开销都微乎其微，分别仅占可用处理器时间预算的平均 1.56% 和 1.20%。此外，该方法仅需使用微控制器随机存取存储器（512KB）的 2.50% [48] 以及数字信号处理器随机存取存储器（224 KB）的 1.56% [1]。这些结果证实，在不影响系统响应能力或其他感知任务的资源可用性的前提下，完全在设备端进行实时呼吸监测是可行的。

6 讨论

这项工作为在普通耳机上实现无感、完全设备端的呼吸感测奠定了基础，既推动了感知流程的技术扩展，也为围绕放松和健康的应用场景探索提供了可能。EarResp-ANS 专为静息、恢复或放松等低运动场景下的呼吸监测而设计。尽管 EarResp-ANS 在各种声学条件下均表现出稳健性能，但仍需考虑若干局限性。下文将讨论这些局限，并概述潜在的缓解策略以及有前景的未来研究方向。

运动状态下的性能。当前评估的一个主要局限在于运动。数据采集是在坐姿下进行的，仅允许极少量活动。因此，所呈现的评估未能充分反映较强运动伪影的影响，例如步行或跑步造成的伪影。尽管所提出的系统适用于在休息或恢复阶段（例如运动后冷却）监测呼吸，但其在活动状态下的性能仍有待探究，需要进一步研究。

密封状态对闭塞效应的影响。除运动外，信号质量还会受到耳机与耳道之间物理耦合的显著影响。数据采集期间，参与者插入耳机后会确认耳道已正确密封。然而，在实际使用中，密封不严可能会同时减弱闭塞效应以及 IEM 与外部噪声源之间的物理隔离。尽管日常场景中的平均外部噪声水平通常低于超过 80 dB声压级的极端条件，但相对于残余背景噪声，呼吸相关声学成分是否会变得过弱、甚至接近麦克风的灵敏度极限，从而降低噪声抑制算法恢复信号的可靠性，目前仍不明确。不过，值得强调的是，本研究通过 IEM 自动测量耳机扬声器的入耳频率响应，完成了密封检查。这表明系统可能能够检测到佩戴不当，并在实际部署中据此进行调整。

呼吸相关信号占主导的假设。 所提出的方法假设，在应用 ANS 后，呼吸相关的声学成分在降噪后的入耳式麦克风信号中占主导地位。当这一假设不成立时——例如由于耳道密封不当或高振幅干扰——残余噪声可能会在信号中占主导地位，从而导致不可靠的呼吸率估计值。可以通过监测每个分析窗口内所提取呼吸相关片段的频谱一致性来应对这一局限。通过将这些片段与平均呼吸频谱模板进行比较，可以抑制由残余噪声主导的窗口。或者，也可以将所提出的方法与基于其他传感模态的互补呼吸频率估计技术进行融合。该假设定义了所提出系统的运行范围：只有当呼吸相关成分在滤波信号中占据足够主导地位时，才能获得可靠的呼吸率估计值。当这一条件不满足时，估计误差和离群值自然会增加，这一点已在我们的评估中观察到，并将在下文讨论。

异常值剔除。 无论上述讨论的具体失效模式如何， 第 5.1.5 节 中观察到的误差分布表明，少量离群点窗口对总体性能产生了不成比例的影响。这些离群值可能由残余噪声、拟合相关效应以及参考信号不准确等因素共同导致。尽管基于通道差异的置信度估计已经能够过滤掉一部分不可靠窗口，但该机制无法消除所有错误估计值。因此，未来有必要研究扩展的置信度感知和多模态融合策略，例如将所提出的基于音频的方法与基于惯性测量单元（IMU）或光电容积描记（PPG）等互补模态的呼吸感测相结合，从而进一步提升系统在瞬态干扰和复杂声学条件下的鲁棒性。

真实值的有效性。 在解读报告的结果时，除了系统和算法层面的局限性，还必须考虑参考信号的有效性。由于真实值呼吸估计不可靠，共有 19.3% 的分析窗口被舍弃；不可靠的判定依据是缺少清晰的呼吸

模式。可能的原因包括胸部腰带张力变化，或参与者意外接触传感器等瞬态干扰，这些因素可能会干扰测量，并降低参考信号的可靠性。我们的分析表明，依据呼吸频率的显著性进一步筛选真实值窗口，可大幅减少离群值，并使总体估计性能明显提升。这一观察结果表明，剩余离群值中可能有一部分源于真实值提取不准确，而非所提出的呼吸频率估计算法本身失效。

基线条件下的性能。 我们观察到，在部分基线录音中应用噪声抑制后，IEM 信号幅度与真实值呼吸信号之间的相关性略有下降。一种可能的解释是，在非常安静的条件下，外部麦克风信号中也可能包含呼吸相关成分，因此被自适应滤波器部分去除。另一种可能是，外部噪声较少时，自适应滤波器无法充分适应，导致其对短暂瞬态干扰的抑制效果降低。这些影响表明，自适应噪声抑制在存在外部噪声时最为有益，而在近乎静音条件下可能不那么有利。未来的工作可以评估为自适应噪声抑制模块采用电平相关控制或旁路机制，以便在近乎静音的环境中主动避免过度滤除外部麦克风中存在的呼吸成分，从而优化信号完整性。

通过耳机播放音乐。 除了已评估的感知场景外，还可以考虑进一步扩展系统。本研究中的所有音乐条件均通过外部扬声器播放。尽管我们证明，音乐等非平稳声学内容并不会从根本上限制所提出的方法，但直接通过耳机本身播放音乐会带来额外挑战，包括声反馈路径和自发声。虽然音乐的频率特性本身不会对该方法造成问题，但本研究并未明确评估这些方面。不过，考虑到人们经常佩戴耳机但不播放音频，解决这些问题将是新增的一项能力，而非呼吸监测的先决条件。因此，纳入感知播放的信号分离，主要是为了拓宽所支持用例的范围，而不是解决所提出方法的根本局限。

应用驱动的方向。 从应用角度来看，本文提出的系统尤其适用于这样的场景：用户已经佩戴耳机进行感官调节或放松，且不播放音乐。在此类场景中，持续的呼吸率监测可作为反映放松程度和心理状态的轻量级生理指标，为实时反馈、引导式呼吸干预以及压力相关趋势的长期追踪创造机会。此外，观察到的误差特征表明，与互补模态（例如基于惯性测量单元或光电容积描记的呼吸监测）进行置信度感知融合可能带来益处，尤其有助于处理仅靠通道差异无法完全解决的瞬态离群值。

进一步扩展。 除呼吸频率估计外，提取的呼吸相关信号还可能用于估计其他呼吸参数，例如吸气与呼气的时长或潮气量。尽管本文未对这些参数进行评估，但它们是所提出处理流程的自然延伸，并可进一步提升其生理相关性。

综上，这些方向凸显了所提出的处理流程如何：(1) 在普通耳机上实现完全设备端的呼吸频率估计；(2) 支持在低运动量且以放松为导向的场景中进行可靠监测；(3) 为日常耳戴式设备上的更丰富呼吸分析及应用驱动的扩展奠定基础。

7 结论

本文介绍了 EarResp-ANS，这是一种轻量且稳健的系统，可在商用耳机上进行设备端呼吸频率估计。通过利用主动降噪耳机的入耳式与外耳双麦克风配置

，该系统采用基于LMS的自适应噪声抑制，降低外侧麦克风采集到的环境噪声，同时保留入耳信号中的呼吸相关声学成分。在一项于真实声学条件下开展、涉及18名参与者的综合用户研究中，我们证明，即使在超过80 dB声压级的高噪声暴露环境下，自适应噪声抑制仍能有效恢复呼吸音，而离线训练的固定滤波器方法在这种情况下已达到极限。在所有评估条件下，EarResp-ANS 均能保持较高的呼吸频率估计准确度，采用双耳融合时的平均绝对误差（MAE）为0.84 CPM。通过基于通道差异进行自动异常值剔除，MAE 可进一步降至0.47 CPM。此外，在稳健的 3σ-区间内，MAE 最低可达0.14 CPM，展现了结合互补感知模态进行置信度感知融合的潜力。我们还证明，整个处理流程可在支持主动噪声控制的耳机上实时运行，系统计算负载低于2%，且无需将任何音频数据传出设备。我们通过在 OpenEarable 2.0 平台上完整部署 EarResp-ANS 验证了这一点。总体而言，EarResp-ANS 既满足了可穿戴设备严格的资源限制，也解决了连续音频感知所涉及的隐私问题。这些结果凸显了自适应信号处理作为日常耳机上稳健、无感生命体征监测的可行基础的潜力，并为未来开展先进呼吸分析及健康应用研究开辟了道路。除呼吸感测外，所提出的噪声抑制方法还可扩展至其他基于音频的生理信号，例如心音，从而支持未来的心血管健康监测应用。

致谢

本研究部分由德国研究联合会（DFG，德国研究联合会）通过研究生院项目 GRK2739/1（项目编号：447089431）——KD2School：面向经济决策的自适应系统设计，以及卡尔·蔡司基金会通过“JuBot——与机器人一起保持年轻”项目资助。

参考文献

[1] Analog Devices, Inc. 2024. ADAU1860：SigmaDSP 低功耗音频编解码器（含 DSP）。https://www.analog.com/en/products/adau1860.html. 访问日期：2026-01-31。[2] Alessandra Angelucci 和 Andrea Aliverti. 2023. 一种用于静态和动态条件下呼吸频率估计的基于惯性测量单元的可穿戴系统。《心血管工程与技术》 14, 3 2023 351 363 10 1007 13239 023 00657 3 （6月），–。[3] [ ] https://doi.org/。/s - - - Arm Ltd. n. d.。CMSIS-DSP 软件库。https://arm-software.github.io/CMSIS /DSP/html/index.html. 访问日期：-01-20。[4] S. Boll. 1979. 使用谱减法抑制语音中的声学噪声。IEEE声学、语音与信号处理汇刊 27, 2 1979 113 120 10 1109 1979 1163209 （），–。https://doi.org/。/TASSP.。Rachel E Bouserhal、Tiago H Falk 和 Jérémie Voix.。通过自适应滤波和人工带宽扩展增强入耳式麦克风语音质量。《美国声学学会杂志》141, 3 (2017), 1321–1331. [6] Cadence Design Systems, Inc. [n. d.]。Tensilica HiFi 3z DSP。https://www.cadence.com/en_US/home/tools/silicon-solutions/compute-ip/hifi-dsps/hifi-3z.html. 访问日期：2026-01-20。[7] Pierre-Yves Carry、Pierre Baconnier、Andre Eberhard、Pierre Cotte 和 Gila Benchetrit. 1997. 呼吸感应体积描记法评估：呼吸波形分析的准确性。《胸部》111, 4 (1997), 910–915. [8] Peter H Charlton、Timothy Bonnici、Lionel Tarassenko、Jordi Alastruey、David A Clifton、Richard Beale 和 Peter J Watkinson. 2017. 从心电图和光电容积描记图中提取呼吸信号：技术与生理决定因素。《生理测量》38, 5 (2017), 669. [9] Jingdong Chen、J. Benesty、Yiteng Huang 和 S. Doclo. 2006. 关于维纳噪声抑制滤波器的新见解。IEEE/ACM音频、语音与语言处理汇刊 14, 4 2006 1218 1234 10 1109 2005 860851 （），–。https://doi.org/。/TSA.。Logan Clarke、Christopher Liau、Vivian Wei Wei 和 Caizhi Liao.。利用耳戴式设备进行医疗健康监测：迈向数字健康的范式转变。《生物医学分析》 2, 4 2025 207 222 10 1016 2025 12 002 （），–。https://doi.org/。/j.bioana.。Michelle A Cretikos、Rinaldo Bellomo、Ken Hillman、Jack Chen、Simon Finfer 和 Arthas Flabouris. 2008. 呼吸频率：被忽视的生命体征。《澳大利亚医学杂志》188, 11 (2008), 657–659。

[12] Gordon B Drummond、Darius Fischer 和 DK Arvind. 2020. 当前临床呼吸频率测量方法所得数值不精确。ERJ Open Research 6, 3 (2020)。[13] Y. Ephraim 和 D. Malah. 1984. 使用最小均方误差短时谱幅度估计器进行语音增强。IEEE声学、语音与信号处理汇刊 32, 6 1984 1109 1121 10 1109 1984 1164453 （），–。[14] 1983 https://doi.org/。/TASSP.。Paul Grossman.。呼吸、压力与心血管功能。《心理生理学》 20, 3 1983 284 300 （），–。[15] Guillermo Gutierrez、Jeffrey Williams、Ghadah A Alrehaili、Anna McLean、Ramin Pirouz、Richard Amdur、Vivek Jain、Jalil Ahari、Amandeep Bawa 和 Shawn Kimbro. 2016. 无阻塞性睡眠呼吸暂停的睡眠成人呼吸频率变异性。《生理学报告》4, 17 (2016), e12949。[16] Gabriel Haas、Evgeny Stemasov 和 Enrico Rukzio. 2018. 你听不见我吗？个人声景策划研究。载于《第17届移动与泛在多媒体国际会议论文集》。59–69。[17] David MG Halpin 和 Marc Miravitlles. 2006. 慢性阻塞性肺疾病：疾病及其对社会的负担。《美国胸科学会学报》3, 7 (2006), 619–623。[18] S Harshitha 和 Azeem Ayesha Siddiqua. 2017. 青少年耳机使用患病率及影响调查。Int J Appl Res 3, 12 (2017), 438–42。[19] S Haykin. 2001. 最小均方误差自适应滤波器。《自适应滤波器理论》，第4版。Prentice Hall, 上萨德尔河 (2001), 183–228。[20] Javier Hernandez、Daniel McDuff 和 Rosalind W. Picard. 2015. BioWatch：利用腕部运动估计心率和呼吸频率。载于《第9届医疗保健泛在计算技术国际会议论文集》（土耳其伊斯坦布尔）（PervasiveHealth '15）。ICST（计算机科学、社会信息学与电信工程研究所），布鲁塞尔，比利时，169–176。[21] Alberto Hernando、Jesús Lázaro、Adriana Arza、Jorge Mario Garzón、Eduardo Gil、Pablo Laguna、Jordi Aguiló 和 Raquel Bailón. 2015. 情绪压力期间的呼吸变化。载于《2015年心脏病学计算会议》（CinC）。IEEE, 1005–1008。[22] Shuhei Hosokawa. 1984. 随身听效应。《流行音乐》4 (1984), 165–180。[23] Wei-Ting Hsueh、Cheng-Xue Yu、Hung-Chi Cheng、Mu-Yen Chen、Hsing-Meng Wang 和 Lung-Ming Fu. 2025. 无创生物传感可穿戴设备综合综述。《TrAC分析化学趋势》193 (2025), 118425. https://doi.org/10. 1016 2025 118425 /j.trac.。[24] Changshuo Hu、Thivya Kandappu、Yang Liu、Cecilia Mascolo 和 Dong Ma. 2024. Breathpro：通过耳戴式设备监测跑步时的呼吸模式。《ACM交互式、移动、可穿戴与泛在技术汇刊》8, 2 (2024), 1–25。[25] Changshuo Hu、Thivya Kandappu、Yang Liu、Cecilia Mascolo 和 Dong Ma. 2024. BreathPro：通过耳戴式设备监测跑步时的呼吸模式。《ACM交互式、移动、可穿戴与泛在技术汇刊》 8, 2, 71 2024 25 10 1145 3659607 论文（5月），页码。https://doi.org/。[26] 2021 / Malte Jacobsen、Till A. Dembek、Guido Kobbe、Peter W. Gaidzik 和 Lutz Heinemann.。使用可穿戴设备无创连续监测生命体征：是否适用于医疗用途？《糖尿病科学与技术杂志》15, 1 (2021年1月), 34–43. https: //doi.org/ 10 1177 1932296820904947 . / Arlene John、He Wang、Barry Cardiff、Keshab K. Parhi 和 Deepu John. 2025. 可穿戴传感中的稳健呼吸频率估计多模态融合。《信息融合》 123 2025 103253 10 1016 2025 103253 （11月），。[27] 2020 https://doi.org/。/j.inffus.。Priyanka Kajla 和 Nithin V. George.。采用双通道稀疏自适应滤波方法进行语音质量增强。《应用声学》 158 2020 107035 10 1016 2019 107035 （），。https://doi.org/。/j.apacoust.。Noa Kallioinen、Andrew Hill、Melany J Christofidis、Mark S Horswill 和 Marcus O Watson. 2021. 定量系统综述：人工测量成人呼吸频率数据不准确的来源。《高级护理杂志》77, 1 (2021), 98–124。[30] Walter Karlen、Christopher J Brouse、Erin Cooke、J Mark Ansermino 和 Guy A Dumont. 2011. 利用光电容积描记法中的呼吸性窦性心律不齐估计呼吸频率。载于《 2011 年 IEEE 医学与生物学工程学会年度国际会议》。IEEE, 1201–1204。[31] Susan A Kayser、Rachel Williamson、Gabriela Siefert、Dan Roberts 和 Angela Murray. 2023. 呼吸频率监测及病情恶化早期检测实践。《英国护理杂志》32, 13 (2023), 620–627。[32] Agni Kumar、Vikramjit Mitra、Carolyn Oliver、Adeeti Ullal、Matt Biddulph 和 Irida Mance. 2021. 从可穿戴麦克风采集的呼吸音频中估计呼吸频率。载于《2021年第43届IEEE医学与生物学工程学会年度国际会议》（EMBC）。IEEE, 7310–7315。[33] Sen M Kuo 和 Dennis R Morgan. 2002. 主动噪声控制：教程式综述。《IEEE学报》87, 6 (2002), 943–973。[34] Jinkyu Lee 和 Hong-Goo Kang. 2019. 深度学习单通道语音增强系统中复数时频掩码的联合学习算法。IEEE/ACM音频、语音与语言处理汇刊27, 6 (2019), 1098–1108. https://doi.org/ 10 1109 2019 2910638 . /TASLP.。Yin Li、Javier Hernandez、James Rehg 和 Rosalind Picard. 2015. 使用头戴式运动传感器估计心脏和呼吸参数。《EAI泛在健康与技术汇刊》（ 05 2015 10 4108 1 1 2）。[35] 2023 [36] https://doi.org/。/phat.。。e Shaoxiong Lin、Wangyou Zhang 和 Yanmin Qian.。采用多帧滤波的两阶段单通道语音增强。《应用科学》13, 8 (2023), 4926。

[37] Haipeng Liu、John Allen、Dingchang Zheng 和 Fei Chen。2019。呼吸频率测量技术的最新进展。《生理测量》40，7（2019年8月），07TR01。https://doi.org/10.1088/1361-6579/ab299e [38] Yang Liu、Kayla-Jade Butkow、Jake Stuchbury-Wass、Adam Pullin、Dong Ma 和 Cecilia Mascolo。2025。Respear：基于耳戴式设备的稳健呼吸频率监测。载于《2025年IEEE普适计算与通信国际会议》（PerCom）。IEEE，67–77。[39] Yang Liu、Qiang Yang、Kayla-Jade Butkow、Jake Stuchbury-Wass、Dong Ma 和 Cecilia Mascolo。2025。EarMeter：利用耳戴式设备进行连续呼吸量监测。《ACM交互式、移动、可穿戴与泛在技术汇刊》9，4，第198篇文章（2025年12月），29页。https://doi.org/ 10 1145 3770655 。/ Patrick C Loughlin、Frank Sebat 和 John G Kellett。2018。呼吸频率：被遗忘的生命体征——让它发挥作用！《联合委员会质量与患者安全杂志》44，8（2018），494–499。[41] Dong Ma、Andrea Ferlini 和 Cecilia Mascolo。2021。OESense：利用闭塞效应进行入耳式人体感知。载于《第十九届移动系统、应用与服务年度国际会议论文集》（线上活动，威斯康星州）（MobiSys '21）。计算机协会，纽约，纽约州，美国，175 187 10 1145 3458864 3467680 – 。https://doi.org/ 。/ 。Alexis Martin 和 Jérémie Voix。 。入耳式音频可穿戴设备：用于健康与安全监测的心率和呼吸频率测量。《IEEE生物医学工程汇刊》65，6（2017），1256–1263。[43] Rainer Martin。2002。采用短时最小均方误差谱估计和伽马分布语音先验进行语音增强。载于《2002年IEEE声学、语音与信号处理国际会议》，第1卷。IEEE，I–253。[44] Rodrigo Martins、Fátima Rodrigues、Susana Costa 和 Nelson Costa。2024。基于惯性传感器的人体呼吸模式评估：系统性文献综述。《算法》17, 6 2024 10 3390 17060223 （）。https://doi.org/ 。/a David J Meredith、D Clifton、Peter Charlton、J Brooks、CW Pugh 和 L Tarassenko。 。光电容积描记法推导呼吸频率：相关生理学综述。《医学工程与技术杂志》36，1（2012），1–7。[46] Nayan Modhave、Yepuganti Karuna 和 Sourabh Tonde。2016。用于助听器语音增强的多通道维纳滤波器设计及噪声抑制技术。载于《2016年绿色工程与技术在线国际会议》（IC-GET）。1–4。https://doi.org/ 10 1109 2016 7916626 。/GET。 。Andrea Nicolò、Carlo Massaroni、Emiliano Schena 和 Massimo Sacchetti。2020。呼吸频率监测的重要性：从医疗保健到体育与运动。《传感器》20, 21 2020 6396 10 3390 20216396 （11月）， 。https://doi.org/ 。/s Nordic Semiconductor ASA。 。nRF 5340 DK：双核蓝牙低功耗SoC。https://www.nordicsemi.com/Products/nRF5340。访问日期：2026-01-31。[49] Geofrey Owino 和 Bernard Shibwabo。2025。抗噪生物声学特征提取方法及其对音频分类性能的影响：系统综述。《JMIR生物医学工程》10 16 2025 80089 10 2196 80089 （12月），e 。https://doi.org/ 。/ PLUX Biosignals。 。RespiBAN BLE — 呼吸感应体积描记传感器。https://www.pluxbiosignals.com/products/respiban-ble。访问日期：2026-01-15。[51] Aarthy Prabakaran 和 Elizabeth Rufus。2022。综述：采用稳健运动伪影抑制技术的可穿戴医疗保健监测系统。《传感器评论》42，1（2022），19–38。[52] U Purushotham 和 K Suresh。2018。利用自适应谱减法提升移动通信中的语音质量。《国际通信网络与分布式系统杂志》21，3（2018），297–314。[53] Jenni Radun、Iida-Kaisa Tervahartiala、Ville Kontinen、Jukka Keränen 和 Valtteri Hongisto。2024。主动降噪耳机是否会影响办公环境中的表现、压力或体验？《建筑与环境》266（2024），112102。[54] Tobias Röddiger、Michael Küttner、Philipp Lepold、Tobias King、Dennis Moschina、Oliver Bagge、Joseph A Paradiso、Christopher Clarke 和 Michael Beigl。2025。Openearable 2.0：用于耳部生理感知的开源耳机平台。《ACM交互式、移动、可穿戴与泛在技术汇刊》9，1（2025），1–33。[55] Tobias Röddiger、Daniel Wolffram、David Laubenstein、Matthias Budde 和 Michael Beigl。2019。探索利用入耳式耳机中的惯性测量单元监测呼吸频率。载于《第一届耳戴计算国际研讨会论文集》。48–53。[56] Chiara Romano、Andrea Nicolò、Lorenzo Innocenti、Marco Bravi、Sandra Miccinilli、Silvia Sterzi、Massimo Sacchetti、Emiliano Schena 和 Carlo Massaroni。2023。利用麦克风录制的呼吸声估算步行与跑步期间的呼吸频率。《Biosensors》13，6（2023），637。[57] Chiara Romano、Andrea NicolÃ2、Lorenzo Innocenti、Marco Bravi、Sandra Miccinilli、Silvia Sterzi、Massimo Sacchetti、Emiliano Schena 和 Carlo Massaroni。2023。利用麦克风录制的呼吸声估算步行与跑步期间的呼吸频率。《Biosensors》13, 6 2023 10 3390 13060637 （）。https://doi.org/ 。/bios Julia Romero、Andrea Ferlini、Dimitris Spathis、Ting Dang、Katayoun Farrahi、Fahim Kawsar 和 Alessandro Montanari。2024。OptiBreathe：一种基于耳戴式设备的光电容积描记系统，用于连续监测呼吸频率、呼吸相位和潮气量。载于《第二十五届移动计算系统与应用国际研讨会论文集》。99–106。[59] Akila Sayoud、Mohamed Djendi、Soumia Medahi 和 Abderrezak Guessoum。2018。一种用于盲语音质量增强的双重快速归一化最小均方自适应滤波算法。《Applied Acoustics》135 2018 101 110 10 1016 2018 02 002 （），– 。https://doi.org/ 。/j.apacoust. 。 。Syed Ahmar Shah、Carmelo Velardo、Andrew Farmer 和 Lionel Tarassenko。 。慢性阻塞性肺疾病急性加重：利用数字健康系统进行识别和预测。《医学互联网研究杂志》19，3（2017），e7207。

[61] 美国胸科学会等。2019。肺量测定标准化：2019年更新，美国胸科学会和欧洲呼吸学会官方技术声明。《美国呼吸与重症监护医学杂志》200，8（2019），E70–E88。[62] Orlando José Tobias 和 Rui Seara。2005。泄漏型 FXLMS 算法：高斯数据和次级路径建模误差的随机分析。《IEEE语音与音频处理汇刊》13，6（2005），1217–1230。[63] Navneet Upadhyay 和 Abhijit Karmakar。2015。采用谱减型算法进行语音增强：比较与仿真研究。《Procedia Computer Science》54（2015），574–584。https://doi.org/10.1016/j.procs.2015.06.066 第十一届国际通信网络会议，ICCN 2015，2015年8月21日至23日，印度班加罗尔；第十一届国际数据挖掘与数据仓库会议，ICDMW 2015，2015年8月21日至23日，印度班加罗尔；第十一届国际图像与信号处理会议，ICISP 2015，2015年8月21日至23日，印度班加罗尔。[64] Diana Vitazkova、Erik Foltan、Helena Kosnacova、Michal Micjan、Martin Donoval、Anton Kuzma、Martin Kopani 和 Erik Vavrinsky。2024。呼吸监测的进展：可穿戴与远程技术综合综述。《 Biosensors》14，2（2024年2月 2024 90 10 3390 14020090 ），。https://doi.org/ 。/bios Tung Xuan Vuong 和 Ahmed M. Al-Jumaily。2016。[65] CPAP设备噪声产生的研究：实验与数值研究。《Applied Acoustics》 114 2016 281 293 10 1016 2016 08 005 （），– 。https://doi.org/ 。/ j.apacoust. 。 。David Waldecker。 。公共空间中的耳朵与耳机。载于《封闭社会——德国社会学会大会》，第38卷。[67] B. Widrow、J.R. Glover、J.M. McCool、J. Kaunitz、C.S. Williams、R.H. Hearn、J.R. Zeidler、Jr. Eugene Dong 和 R.C. Goodlin。1975。自适应噪声消除：原理与应用。《Proc. IEEE》 63, 12 1975 1692 1716 10 1109 1975 10036 （），– 。https://doi.org/ 。/PROC. 。 。Bradford D Winters、Michael Eberlein、Janice Leung、Dale M Needham、Peter J Pronovost 和 Jonathan E Sevransky。 。脓毒症的长期死亡率与生活质量：系统综述。《重症监护医学》38，5（2010），1276–1283。[69] Xiangyu Xu、Jiadi Yu、Yingying Chen、Yanmin Zhu、Linghe Kong 和 Minglu Li。2019。Breathlistener：利用声学信号在驾驶环境中进行精细化呼吸监测。载于《第十七届移动系统、应用与服务年度国际会议论文集》。54–66。[70] Albert C Yang、Wei-Ming Ma、Dung-Hung Chiang、Yi-Ze Liao、Hsien-Yung Lai、Shu-Chuan Lin、Mei-Chin Liu、Kai-Ting Wen、Tzong-Huei Lin、Wen-Xiang Tsai 等。2025。利用单时间点无创生命体征和XGBoost模型早期预测脓毒症及其与C反应蛋白和降钙素原的相关性：多中心研究。《Intelligence-Based Medicine》11（2025），100242。[71] Fumihiko Yasuma 和 Jun-ichiro Hayano。2004。呼吸性窦性心律不齐：为什么心跳会与呼吸节律同步？《Chest》125，2（2004），683–690。[72] Sally Taha Yousif 和 Basheera M Mahmmod。2025。语音增强算法：系统性文献综述。《算法》18，5（2025），272。[73] Bonan Zhang、Chao Chen、Ickjai Lee、Kyungmi Lee 和 Kok-Leong Ong。2025。可穿戴健康监测设备的安全与隐私问题综述。《Computers & Security》155（2025年8月），104453。https://doi.org/10.1016/j.cose.2025.104453

A 附录

A.1 数据集

表 6. 按运动类别划分的呼吸率真实值分布（每分钟周期数）。

类别	Mean	中位数	标准差	Min	最大范围	
未运动	16.82	16.53	3.77	7.57	26.24	18.67
运动后	18.20	18.44	3.97	7.77	27.94	20.17
总体	17.57	17.47	3.94	7.57	27.94	20.37

A.2 按活动评估

表7. 按活动划分的呼吸频率估计误差。报告了单通道、融合和置信估计的平均绝对误差和均方根误差。数值以均值 ±标准差表示，单位为每分钟周期数。

	平均绝对误差（每分钟周期数）			均方根误差（每分钟周期数）		
活动	**单一**	**融合**	**置信**	**单一**	**融合**	**置信**
未运动	1.22 ± 2.50	1.10 ± 2.13	0.56 ± 1.69	2.78 ± 2.78	2.39 ± 2.39	1.78 ± 1.78
运动后	0.63 ± 2.03	0.61 ± 1.87	0.41 ± 1.50	2.13 ± 2.12	1.96 ± 1.96	1.55 ± 1.55
总体	0.90 ± 2.27	0.84 ± 2.00	0.47 ± 1.58	2.45 ± 2.45	2.16 ± 2.17	1.65 ± 1.65

A.3 按条件评估

表 8. 按条件统计的融合估计呼吸频率估计误差。报告了自适应噪声抑制、nLMS、Breath Pro 和仅带通滤波（仅 BPF）基线的平均绝对误差和均方根误差。

	平均绝对误差（每分钟周期数）				均方根误差（每分钟周期数）			
条件	**自适应噪声抑制**	**nLMS**	**BreathPro**	**仅 BPF**	**自适应噪声抑制**	**归一化最小均方**	**BreathPro**	**仅 BPF**
基线（<35分贝）	0.85	0.86	0.88	0.94	2.27	2.26	2.38	2.37
50 分贝白噪声	0.47	0.41	0.44	0.66	1.56	1.30	1.38	1.92
65 dB白噪声	0.55	0.66	1.35	2.75	1.67	1.81	2.58	3.84
80 dB白噪声	1.43	1.62	3.12	4.18	2.84	2.99	4.30	5.15
65 分贝自助餐厅	0.95	1.33	3.90	4.25	2.07	2.61	5.00	5.25
80 分贝音乐	0.81	1.82	5.65	6.07	2.30	3.54	6.92	7.25
总体	**0.84**	**1.11**	**2.54**	**3.11**	**2.16**	**2.52**	**4.19**	**4.65**

A.4 逐参与者评估

表 9. 逐参与者呼吸频率估计性能。报告单通道、融合（左右声道均值）和置信估计的平均绝对误差（MAE）与均方根误差（RMSE）。括号中的数值表示依据 3σ MAD 准则剔除异常值后的性能（单通道和融合）。所有数值均以每分钟周期数（CPM）表示。

参与者	平均绝对误差（CPM）			均方根误差（CPM）		
	单一（3σ）	融合（3σ）	置信度	单一（3σ）	融合（3σ）	置信度
P01	0.28 (0.11)	0.22 (0.10)	1.05	1.14 (0.14)	0.66 (0.13)	1.35
P02	1.60 (0.20)	1.55 (0.18)	1.68	3.71 (0.26)	3.16 (0.23)	2.86
P03	0.35 (0.15)	0.34 (0.14)	1.10	1.29 (0.19)	1.24 (0.18)	2.04
P04	0.59 (0.11)	0.58 (0.11)	1.59	2.14 (0.15)	1.66 (0.14)	2.35
P05	2.80 (0.19)	2.51 (0.18)	2.42	4.97 (0.28)	4.36 (0.25)	4.06
P06	0.69 (0.13)	0.68 (0.12)	1.38	2.22 (0.17)	1.90 (0.16)	2.41
P07	2.26 (0.20)	1.93 (0.18)	2.49	4.20 (0.27)	3.55 (0.24)	3.31
P08	0.53 (0.14)	0.51 (0.13)	1.55	1.57 (0.19)	1.21 (0.17)	1.90
P09	0.77 (0.23)	0.76 (0.22)	1.43	1.77 (0.30)	1.74 (0.29)	2.07
P10	2.75 (1.91)	2.44 (1.55)	3.50	4.14 (2.75)	3.73 (2.24)	4.66
P11	0.61 (0.12)	0.58 (0.11)	1.71	1.94 (0.16)	1.60 (0.14)	2.77
P12	2.59 (1.21)	2.23 (1.04)	2.29	3.93 (1.88)	3.26 (1.59)	3.17
P13	0.72 (0.11)	0.71 (0.10)	1.20	2.38 (0.14)	1.88 (0.13)	1.97
P14	0.47 (0.12)	0.47 (0.11)	1.81	1.63 (0.15)	1.47 (0.14)	3.15
P15	1.50 (0.19)	1.29 (0.18)	3.26	3.09 (0.25)	2.68 (0.23)	4.02
P16	2.09 (0.37)	2.04 (0.36)	3.65	3.98 (0.54)	3.61 (0.53)	4.77
P17	1.23 (0.26)	1.21 (0.25)	1.94	3.08 (0.35)	2.84 (0.34)	3.01
P18	0.35 (0.14)	0.34 (0.14)	0.34	1.43 (0.17)	1.43 (0.17)	1.43
总体	1.21 (0.16)	1.10 (0.15)	0.58	2.92 (0.21)	2.51 (0.19)	1.92

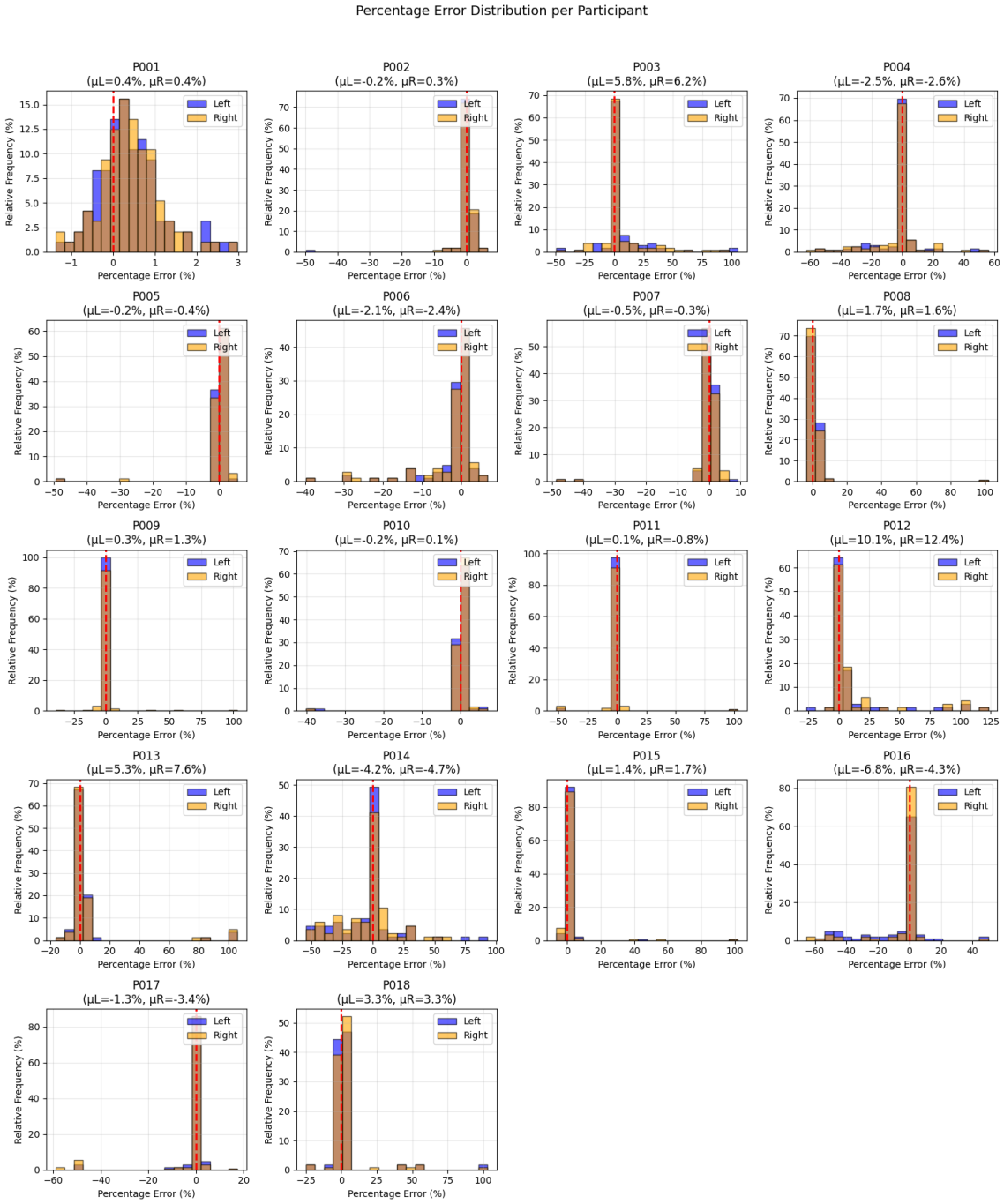


图 11. 每位参与者的呼吸率百分比误差分布