声乐呼吸音的无监督分割

arXiv:2304.03758v1 [eess.AS] 7 Apr 2023

摘要—**呼吸是人类生存不可或缺的一部分，携带着有关个人生理和心理状态的信息。通常，在用于任何任务之前，专家都会标注呼吸边界。本研究提出了一种无监督呼吸边界检测算法，用于检测在口部录制的呼吸音，即本文所称的口腔呼吸音（VBS）。本研究使用在口部录制的呼吸音，因为与气管呼吸音和肺部呼吸音相比，它们更容易录制，且无需接触。利用呼吸信号能量的周期性来分割呼吸边界。采用动态规划，并利用呼吸相位数量（P）和呼吸阶段时长（d）的先验信息来寻找边界。本研究预测了来自60名受试者（31名健康受试者、29名患者）的367个呼吸边界，涵盖307次呼吸。使用所提出的方法，得到M（89%）、I（13%）、D（11%）和S（79%）。所提出的方法的表现优于本研究使用的基线方法。即使使用所提出的方法估计的边界来区分哮喘患者与健康受试者，其分类性能也与真实边界相当。**

索引词—**动态规划、呼吸音、哮喘、分割、发声**

I. 引言

呼吸是人类赖以生存、不可或缺的过程。呼吸频率异常是重要生命体征之一，可提示潜在的不良心理状态，如压力、焦虑，以及心搏骤停、哮喘、慢性阻塞性肺疾病（COPD）等生理状况 [1]。为了评估肺部健康状况，呼吸音分析正成为医生和研究人员关注的一项新兴技术。呼吸音可从胸部、气管、鼻部和口部录制。从胸部录制的呼吸音也称为肺音 [2]。健康者与患有肺炎、哮喘等肺部疾病的患者，其肺音的声学特性有所不同。临床上根据偶发肺音在呼吸周期中的持续时间对其进行描述 [3]。Forgacs等人 [4] 注意到，在口部录制的呼吸音中，健康受试者与患者之间也呈现出类似的声学特性差异趋势。

我们的研究兴趣之一是基于人声的哮喘监测与诊断。该研究方向中的第一项工作是利用在口部记录的持续发声、咳嗽和呼吸音，对哮喘患者与健康受试者进行分类 [5]。Yadav 等人 [5] 已表明，在所有声音中，呼吸音的表现最佳。在 Yadav等人的研究中 [5] ，呼吸音通过目视检查声谱图并进行听辨来手动标记，这是一项非常耗时的任务。Yadav 等人的另一项工作 [6] 已表明，与针对相同任务从连续呼吸周期的呼吸信号中随机选取的片段相比，在预先获知呼吸边界的情况下，哮喘患者与健康受试者之间的分类性能更好。

已有多种方法用于分割从胸部和气管录制的呼吸音，但对从口部录制的呼吸音进行分割的研究较少。Palaniappan等人 [7] 使用归一化平均功率谱密度进行呼吸相位检测（相位指呼吸的吸气和呼气部分），并利用自适应神经模糊推理系统进行分割。Aras等人 [8] 利用动态时间规整算法，根据呼吸信号功率谱包络的相似性确定呼吸周期的边界。研究利用单通道肺音短时能量的准周期性，设计带通滤波器来估计呼吸相位边界 [9]。Cam等人 [10]在贝叶斯框架中使用三元马尔可夫链确定呼吸边界并检测呼吸相位。Feng等人 [11]提出使用沃尔什基函数对气管呼吸音及其相位进行分割。Jin等人 [12] 提出通过多种群遗传算法对单通道气管音进行呼吸相位分割，并以样本熵作为评估指标。Yap等人 [13] 提出一种时域技术，利用气管呼吸音的方差分形维数检测呼吸起始。Yahya等人 [14] 利用鼻气流寻找呼吸边界，并使用语音活动检测器检测静音和呼吸相位。随后使用检测到的有声区域峰值，通过支持向量机检测吸气和呼气阶段。

在这项工作中，我们提出了一种对在口部录制的呼吸音（称为口腔呼吸音（VBS））进行分段的算法。与胸部和气管处的呼吸音不同，只需付出很少努力且无需与患者进行身体接触，麦克风就能轻松录制口部的呼吸音。本工作使用的口腔呼吸音数据是在医院自然嘈杂的环境中录制的，因此所提出的方法适合实时部署。数据库由健康患者和哮喘患者的呼吸样本组成；因此，所提出的方法也考虑了正常和异常呼吸频率的变化。由于大部分呼吸信号信息都存在于最高 2kHz 的频率范围内，因此所有呼吸样本都经过 2kHz 低通滤波。我们利用 VBS 能量近乎周期性的特点，找出 VBS 的边界及其阶段（“吸气”和“呼气”）。VBS 边界采用动态方法进行估计。

基于先验呼吸时长和呼吸次数信息的基于编程的方法。本研究还提出了一种估计呼吸时长和呼吸次数的方法。由于尚无与VBS 分割相关的现有研究，我们使用了无监督音素分割[15] 和肺音分割技术 [8] 作为基线方法。采用评估指标来量化匹配、缺失、插入和删除的边界，如Ghosh 等人 [16]的研究中所述。这些指标是根据预测边界与真实边-界之间的距离（有助于量化匹配边界），以及真实边界数量与预测边界数量之间的差异（用于量化缺失、插入和删除的边界）计算得出的。对于匹配情况，使用了另一项评估指标重叠率[17] 来确定预测边界与真实边界之间的重叠程度。通过所提出的方法，我们发现边界匹配率为89%，其中79%实现了片段匹配，重叠率（均值（标准差））为88(±13)%。所提出的方法优于两种基线方法。我们还报告了健康组和哮喘组之间所提出的方法的性能，以分析该算法对由哮喘导致的可变呼吸频率的鲁-棒性。研究发现，使用所提出的方法通过估计边界对哮喘患者和健康受试者进行分类时，其分类性能与使用真实边界进行分类时具有可比性。

II. 数据集

在这项工作中，我们共从60名（24名女性，36名男性）受试者中记录了数据，其中包括31名（12名女性，19名男性）健康对照组和29名（12名女性，17名男性）患者。受试者的年龄范围为15岁至53岁，患者的平均年龄为37.17岁，对照组的平均年龄为30.38岁。我们在医院的嘈杂环境中记录了数据，其中包括人声、风扇声、空调声、电话铃声等噪声。每位受试者在录音前均签署了知情同意书。所有患者的录音均在医生指导下于圣约翰医学院医院完成。圣约翰医学院医院位于印度卡纳塔克邦班加罗尔，其伦理委员会批准了该研究（方案编号：IECstudyrefno.382/2018），批准日期为 12^{th}，2019年2月。所有录音均使用ZOOMH6便携式录音机麦克风，以44.1kHz采样率和16位深度完成。录音时，麦克风与口部保持3cm至5cm的距离。录音过程中，使用鼻夹闭合受试者的鼻部，使其仅通过口部呼吸。记录了受试者的深呼吸。平均每位受试者记录5个呼吸音。共记录了对照组的151个呼吸音和患者的156个呼吸音。因此，总数为307个。呼吸音的平均持续时间为2.97秒，标准差为1.47秒。呼吸的最短和最长持续时间分别为1.125秒和11.356秒。图1绘制了健康受试者和哮喘患者呼吸频率（以每分钟呼吸次数表示）的直方图。图1显示，对于哮喘患者和健康受试者而言，呼吸频率高度重叠。大多数受试者的呼吸频率介于每分钟15至30次呼吸之间。

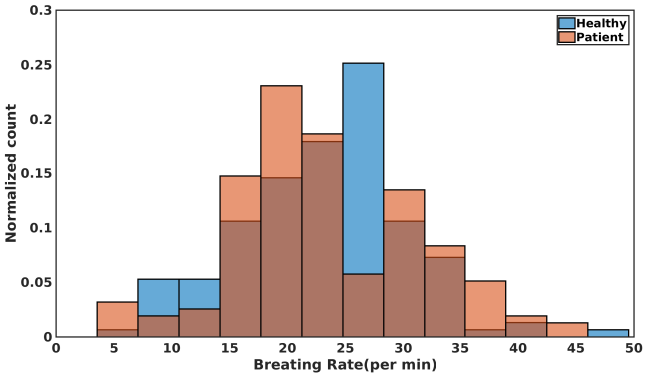


图 1. 患者与健康受试者的呼吸率直方图。

受试者之间的吸气持续时间从 0.50 秒到 4.62 秒不等，平均持续时间为 1.3 秒；而对于呼气，持续时间从 0.65 秒到 8.09 秒不等，平均持续时间为 1.62 秒。两名标注员通过使用 Audacity 检查频谱图和波形，对录音数据进行了标注[18]。图中给出了一个标注波形的示例。2。图中展示了一个包含 3 次呼吸的噪声呼吸样本。2。对于每次呼吸，吸气和呼气边界分别进行了标记（以红色显示），呼吸边界以绿色显示。

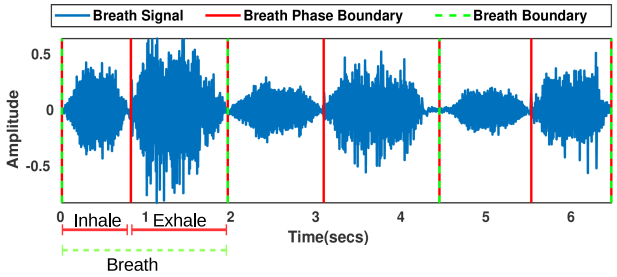


图 2. 人工标注的呼吸声样本文件。吸气和呼气边界以绿色显示，呼吸边界以红色显示。

III. 所提出的呼吸分段

我们所提出的方法的框图见图 3。所提出的方法包含三个主要步骤。第一步是预处理，有助于去除高频噪声并估计信号的能量。第二步包括估计录音中的平均呼吸时长和呼吸次数，并利用这些信息通过动态规划找到边界。下面将详细解释框图的各个部分。

A. 预处理

1) 低通滤波：呼吸信号的能量（记为 $B[u]$，其中 u 表示样本索引）主要存在于2kHz以下 [3]。因此，记录的呼吸音经过截止频率为 2kHz的低通滤波。由于我们的数据是在嘈杂环境下记录的，此步骤对于去除 2kHz 以上的所有噪声、增强算法的鲁棒性是必要的。

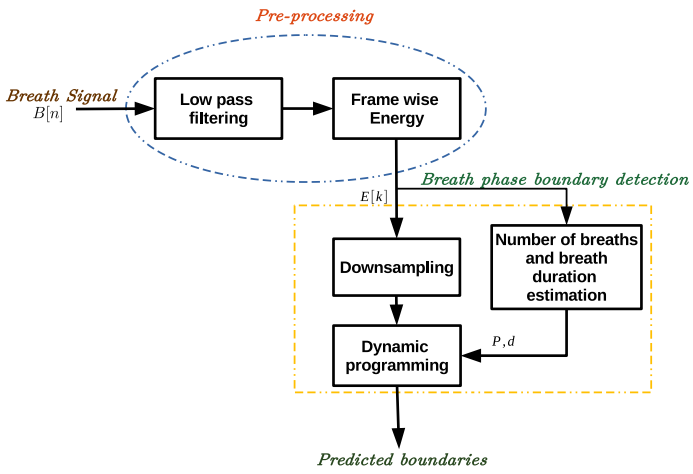


图 3. 我们所提出的方法的框图。 $\boldsymbol{B}[\boldsymbol{u}]$， $\boldsymbol{E}[\boldsymbol{n}]$， $\boldsymbol{P}$ 和$\boldsymbol{d}$，分别表示呼吸信号、能量信号、呼吸相位数量和呼吸阶段时长。 $\boldsymbol{u}$ 和 $\boldsymbol{n}$，分别表示样本索引和帧索引。

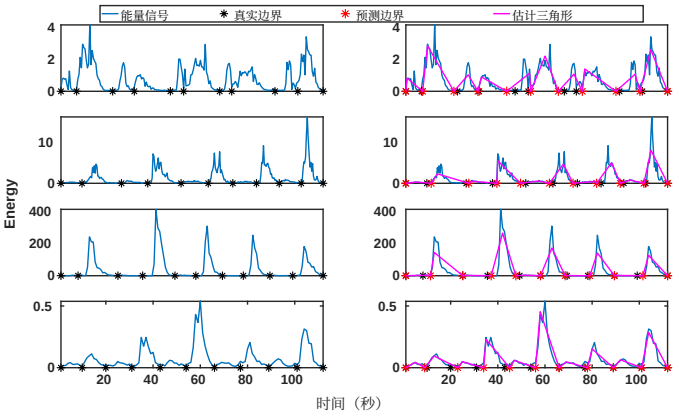


图4。 能量信号样本。左列显示带有真实值边界的短时能量信号，右列显示使用三角形拟合得到的相应信号及其预测边界。

2) 逐帧能量计算：滤波后的信号已使用窗口大小（w_s）和重叠大小（w_o）进行分帧。已计算每个帧的能量，并获得能量信号（$E[n]$），其中 n 表示帧索引。由于与噪声很大的原始信号相比，它具有更稳定的呼吸周期结构，因此本工作使用 $E[n]$ 来查找边界。

B. 呼吸阶段边界预测

1) 呼吸相位建模：一次呼吸包含吸气和呼气两个部分。取自4名受试者的呼吸信号能量如图 4 所示。在图 4 中，左列显示标有呼吸相位真实边界的能量信号，而右列除真实边界外，还显示了使用所提出的方法得到的相应预测边界。从图 4 中我们可以看到，能量信号在本质上具有一定的周期性。然而，由于记录的噪声特性以及不同受试者的呼吸模式，短时能量的振幅变化很大，并且在信号本身中。在本研究中，我们利用信号的这种周期性以及能量包络形状。在本研究中，我们用持续时间可变的三角形对呼吸的每个相位进行建模，其中三角形的端点给出了边界。因此，呼吸能量信号的短时能量轮廓可以近似表示为多个三角形之和，其中三角形的数量等于信号中呼吸相位的数量，而三角形的端点将给出呼吸相位边界。为了理解如何对每个相位进行三角形拟合，考虑一个短时能量轮廓 $x[k]$ ，即 $k = 0$的范围，1、2、...M，其中 k表示样本索引。为了将三角形拟合到 $x[k]$，在任意三个点（k_1，0）、（k_2，$x[k_2]$）和（k_3，0）之间，其中 $k_1 < k_2 < k_3$ 由 $F[k]$ 给出，而 $x[k_2]$ 表示为 α。 $F[k]$ 如下所示。

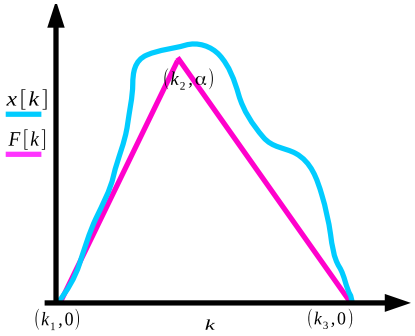


图 5. 三角形拟合示例。

$$F[k] = \begin{cases} \frac{\alpha(k-k_1)}{k_2-k_1}, & k_1 \le k \le k_2, \\ \frac{\alpha(k-k_3)}{k_2-k_3}, & k_2 \le k \le k_3 \end{cases}$$

要找到最优的 (k_2, α)，我们需要最小化以下目标函数。

$$J(k_1,k_3) = \underset{k_2,\alpha}{\text{minimize}} \sum_{k=k_1}^{k_2} (x[k] - \frac{\alpha(k-k_1)}{k_2-k_1})^2 + \sum_{k=k_2+1}^{k_3} (x[k] - \frac{\alpha(k-k_3)}{k_2-k_3})^2 \quad (1)$$

对公式 1 关于 α 求导并令其等于零，可得到以下关于 k_2 和 α的方程，

$$\alpha = \frac{\sum_{k=k_1}^{k_2} \frac{x[k]k}{k_2} + \sum_{k=k_2+1}^{k_3} \frac{x[k](k-k_3)}{k_2-k_3}}{\sum_{k=k_1}^{k_2} \frac{k^2}{k_2^2} + \sum_{k=k_2+1}^{k_3} \frac{(k-k_3)^2}{(k_2-k_3)^2}} \quad (2)$$

根据等式2 可知，(k_2, α) 无法通过解析方法求解，因此 k_2 从1 变化到 $k_3 - 1$ ，并针对 k_2的每个给定值， α 均已计算得出。因此，对于给定的(k_2, α) 值，可以计算等式1 的值。根据等式1中计算出的所有目标函数值， α 和k_2 对应的值达到最小值时，即为最佳拟合。

2) 呼吸信号相位分割的目标函数：在这项工作中，我们有两个假设。第一个假设是，呼吸从第一个音频信号样本开始，到最后一个样本结束，这意味着信号的开头和结尾都没有静音。其次，呼吸

相位是连续的，这意味着呼气的终点就是下一次吸气的起点，而吸气的终点就是下一次呼气的起点。如第 节所述，三角形拟合是寻找边界的一种有效方法；因此，我们可以说 $E[n]$ 是一系列三角形，相邻三角形共用边界。因此，寻找边界可表述为最小化以下目标函数。假设我们已知短时能量信号 P中的相位数量（ $E[n]$，其中 $0 \leq n \leq M$，那么呼吸相位的边界 n_1^*、n_2^*、$...n_{P+1}^*$可通过求解式3来估计。

$$n_1^*, n_2^*, ...n_{P+1}^* = \underset{n_1, n_2 ... n_{p+1}}{\operatorname{argmin}} \sum_{p=1}^{P} J(n_p, n_{p+1}) \qquad (3)$$
$$\text{subject to} \quad n_1 = 0, n_{P+1} = M,$$
$$\{n_2, n_3..n_p\} \in \{2, ...M-1\}, n_{p+1} > n_p$$

公式 3 是一个迭代代价函数，取决于拟合先前呼吸相位的代价；因此，公式 3可以通过动态规划（DP）问题求解。DP[19] 用于求解可分解为子块的问题，并利用每个子块的解来求得整个问题的解。同样，公式 3 可以通过优化各个相位来进行优化，进而可以使用最小距离路径回溯最优边界（公式 3）。使用 DP 求解公式 3 的步骤如算法1所示。

算法 1 通过求解进行呼吸阶段边界检测方程 Eq. 3.

初始化：
P=阶段数量，d=阶段时长，δ=.3
$E_d[n]$=降采样能量信号，$0 \leq n \leq M$
$O(1, n_2) = \underset{n_2}{\text{最小值}}\ J(1, n_2), (Eq.1) \quad \lfloor d(1-\delta) \rfloor \leq n_2 \leq \lfloor d(1+\delta) \rfloor$
$I(1, n_2) = \underset{n_2}{\text{参数最小值}}\ J(1, n_2), \quad \lfloor d(1-\delta) \rfloor \leq n_2 \leq \lfloor d(1+\delta) \rfloor$
递归：
对于 k 从 2 变化到 P 执行
$O(k, n_{k+1}) = \min_{k\lfloor d(1-\delta) \rfloor \leq n_k \leq k\lfloor d(1+\delta) \rfloor} \{O(k-1, n_k) + J(n_k, n_{k+1})\}$
$I(k, n_{k+1}) = \arg\min_{k\lfloor d(1-\delta) \rfloor \leq n_k \leq k\lfloor d(1+\delta) \rfloor} \{O(k-1, n_k) + J(n_k, n_{k+1})\}$
$\forall\ n_{k+1} \in (k+1)\lfloor d(1-\delta) \rfloor \leq n_{k+1} \leq (k+1)\lfloor d(1+\delta) \rfloor$
结束循环
向后追踪：
$n_{P+1}^* = M$
对于从 P 到 1 的每个阶段 (l) cuntegn
$n_l^* = I(l, n_{l+1}^*)$;
结束循环
返回：$n_1^*, n_2^*, ..., n_{P+1}^*$

3) 呼吸阶段数量 *(P)*：正如我们在上一部分中所解释的，可以通过使用动态规划找出 $B[u]$ 的边界，但动态规划需要呼吸阶段平均持续时间（称为 d）和呼吸阶段数量（称为 P）的信息。由于 $E[n]$ 是一个近似周期性的信号，因此在该信号的这一频率处应出现一个信号幅度谱的峰值。$E[n]$ 的这一特性已被用于查找 P。从 $E[n]$的频谱中，在我们数据的最小和最大呼吸频率之间选取了峰值。通过人工标注数据计算得出，最大和最小呼吸时长分别为11.2秒和1.2秒。因此，呼吸频率分别为0.833 Hz和0.089 Hz。所选峰值的频率被用作呼吸频率，通过将其加倍来估计 P。通过将 $E[n]$ 的长度除以 P来估计 d。这些估计的 d 和 P 值被用作动态规划。

C. 降低复杂度

为降低动态规划的时间复杂度，对 $E[n]$ 进行降采样，并通过对 d 添加额外约束，缩小寻找最优 n_k 的搜索范围。这两种方法都大幅缩短了算法寻找边界所需的时间。下面将分别介绍这两种方法。

1) 降采样：动态规划通过寻找最小距离路径，高效估计呼吸阶段边界。由于 $B[u]$ 的采样频率非常高，在 $E[n]$ 中需要搜索的点数也非常多，因此穷举搜索会耗时很长，增加了寻找边界所需的时间。为解决这一问题，将 $E[n]$ 按 10 倍因子降采样。我们测试了从 2 到 14 、步长为 2 的不同降采样因子，结果表明 10 的表现最佳。

2) 呼吸时长估计：使用我们的方法估计P，计算出 d 的值。由于 d 的先验信息能够显著减少动态规划查找边界所需的时间，方法是将穷举搜索范围限制为适合该三角形的范围。还使用了另一个称为 δ 的变量，该变量对照组 d周围的搜索范围。

IV. 实验

A. 基线

据我们所知，我们没有发现任何对在口部记录的呼吸音进行分割的相关工作。因此，在本工作中，我们使用了两种用于呼吸音以外信号的基线分割技术。此外，由于我们所提出的方法是无监督的，为了进行公平比较，所选择的两种基线方法也都是无监督的。第一种基线（B1）是由Dusan等人所提出的方法，用于使用TIMIT语料库 [20] 进行音素分割 [15]。B1使用频谱转移度量作为特征，以找出音素之间的转移时刻 [15]。频谱转移度量在音素边界处达到峰值。随后提取这些转移峰，并进行后处理，以确定音素边界。

第二个基线（称为 B2）由 Aras等人 提出，用于对胸部记录的呼吸声音进行分割。该工作利用给定信号中呼吸声音重复出现的能量模式。通过使用动态时间规整，在信号中找到了相似重复模式的边界。由于该方法只能进行呼吸分段，无法进行呼吸阶段（吸气和呼气）分割，因此，为了进行比较，我们仅使用了呼吸声音结果。

B. 评估指标

为了衡量算法的性能，我们按照 Ghosh等人 [16]所述，通过比较预测边界和真实边界的位置，计算匹配、删除、片段匹配和插入。如果真实边界位于预测边界的 $\pm$ 阈值范围内，则称为一个

表 I

标注者间一致性：匹配(M)、插入(I)、删除(D)、片段匹配(S)的均值，以及片段匹配呼吸音的重叠率(OvR)的均值和标准差。

总计边界	M(%)	D(%)	I(%)	S(%)	OvR 均值（标准差）
367	97	3	3	82	92（11）

匹配（M）；否则，真实边界视为已删除（D）。两个连续的匹配称为片段匹配（S）。所有未与任何真实边界匹配的预测边界称为插入边界（I）。M、D、S 和 I 均以百分比表示。

另一项评估指标——重叠率（简称OvR），用于量化片段匹配的呼吸与其预测对应片段之间的重叠程度。OvR 已被用于音素分割工作的评估指标 [17]。OvR 定义为参考边界与预测边界之间的共同持续时间除以呼吸可能达到的最大持续时间所得的比值。OvR 的取值范围为 0 到 1。当 OvR 为 1 时，共同持续时间与最大可能持续时间相等；如果共同持续时间为 0，则 OvR 为 0。在本研究中，我们分别分析了呼吸、吸气和呼气分割的 OvR 均值和标准差。

C. 标注者间差异

在当前工作中，呼吸边界由两名标注员进行标注。为了确定标注员间一致性达到何种程度，使用了M、I、D和S。为获得上述指标，将一名标注员标记的呼吸边界视为参考，将另一名标注员标记的边界视为预测值。M、I、D、S 的均值以及 OvR 已被报告。从表 I中可以观察到，在总计 367个呼吸边界中，涉及 307次呼吸的边界具有非常高的 M（97%），片段匹配率为82%。这表明标注员间一致性良好。对于这项工作，结果仅使用一名标注员的边界作为真实值进行报告。

D. 实验设置

对于这项工作，呼吸音信号（$B[u]$）使用 6_{th} 阶巴特沃斯滤波器进行了低通滤波，截止频率为 2kHz。为了逐帧计算能量，$B[u]$ 已按 $N_w = .1sec$ 分帧，帧移为 $N_w = 10ms$ 。计算每帧的能量以求得信号 $E[n]$。为降低动态规划的时间复杂度，$E[n]$ 从 100 个采样点/秒降采样至 10 个采样点/秒。设置 $\delta = .3$ 以降低动态规划的复杂度。为估计 P 和 d，对 $E[n]$ 进行了快速傅里叶变换，快速傅里叶变换点数为 $E[n]$ 长度的两倍。从 $E[n]$ 频谱中分别选取最大和最小呼吸频率 0.833 Hz 和 0.089 Hz 之间的峰值。

分类设置与 [5]中给出的类似。对于哮喘患者和健康受试者之间的分类任务，梅尔频率倒谱系数（MFCC）的统计量，即均值、中位数、众数、均方根、方差和标准差已被用作特征，支持向量机已被用作分类器。分类准确率已被用作评估指标。梅尔频率倒谱系数矩阵在 20 毫秒窗口、10 毫秒移位下计算。每个训练集和测试集分别包含 50名和10 名受试者。在六折中的五折里，训练集中有26 名对照组和24名患者，相应的测试集中有5名患者和 5 名对照组。剩余的一折中，测试集有 6名对照组和 4 名患者，训练集有 25 名患者和 25 名健康受试者。

B1 的实现方式与Ghosh等人所描述的类似 [16]。在 B2 中，用于寻找能量信号极小值的搜索范围设定为 1.2 秒至 12 秒，B2 算法的其余步骤保持不变。

V. 结果

A. 基线方法和所提出的方法的比较

表 II

匹配(M)、插入(I)、删除(D)、片段匹配(S)，以及使用基线方法和所提出的方法对片段匹配的呼吸音进行匹配时的重叠率(OvR)均值和标准差（STD）。

方法	预测的边界	M(%)	D(%)	I(%)	S(%)	OvR 均值(std)
B1	1028	88	12	192	74	92(6)
B2	428	49	51	68	48	75(31)
所提出的方法	375	89	11	13	79	88(13)

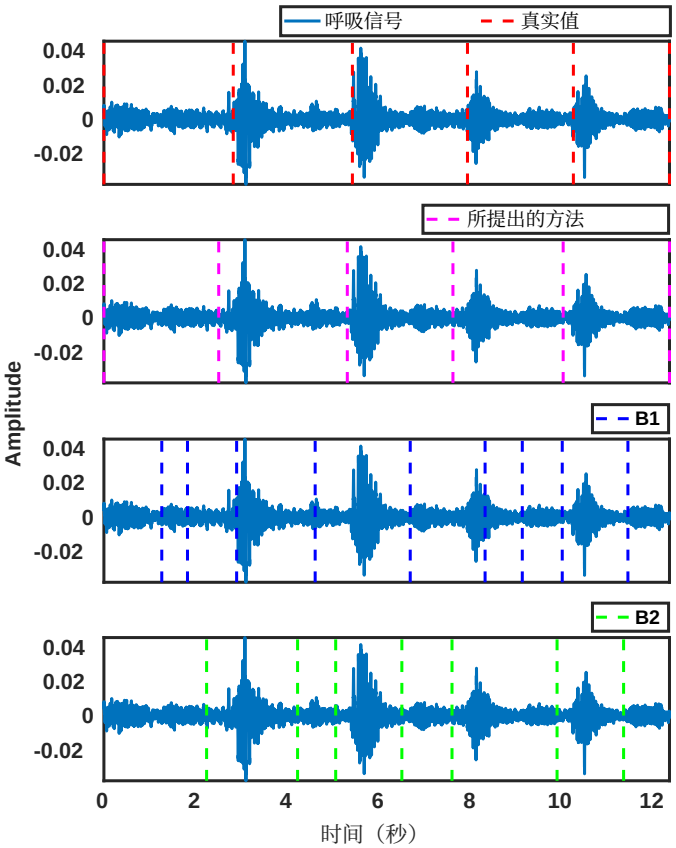


图 6. 用于比较所提出的方法、B1 和 B2 性能的示例。

表 III
B1、B2 和所提出的方法在吸气分割中的表现。

方法 used	预测边界	M(%)	D(%)	I(%)	S(%)	OvR 均值（标准差）
B1	**1936**	**89**	**11**	**226**	**81**	88(7)
B2	**740**	**42**	**58**	**79**	**25**	61(37)
所提出的方法	**630**	**73**	**27**	**30**	**52**	85(15)

表 IVB1、B2 和所提出的方法在呼气分割中的性能。

方法 used	预测值边界	M(%)	D(%)	I(%)	S(%)	OvR 均值（标准差）
B1	**1936**	**89**	**11**	**226**	**81**	88(6)
B2	**740**	**49**	**51**	**72**	**30**	68(35)
所提出的方法	**630**	**78**	**22**	**25**	**61**	84(13)

使用所提出的方法、B1 和 B2 进行呼吸分段的结果如表 II 所示。数据中的呼吸总数为 307，呼吸边界为 367。同样，每个阶段的边界总数为 614。从表 II 中可以看出，在所有方法中，B2 的表现最差，因为与另外两种方法 B1（M (88%)& S (74%)) 和所提出的方法（M (89%)& S (79%)) 相比，其 M (49%) 和 S (48%) 较低。就 M 和 S 而言，与 B2 和所提出的方法相比，B1 的表现相对较好，但其插入边界（I）数量非常高，192% ，是 B2 的两倍多。然而，B2 和所提出的方法的插入边界分别为 68% 和 13%。由于 TIMIT 数据库较为干净，而我们的数据库存在噪声，因此在计算频谱转移度量时会出现许多伪峰；这些伪峰可能被检测为错误的峰边界，从而导致 I 度量较高。所提出的方法、B1 和 B2 的性能示例见图 6。从图 6 中可以看出，所提出的方法预测的边界与真实边界相当，而 B1 和 B2 预测出的边界比真实边界更多且不正确。

1）相位分割性能：吸气和呼气分割的结果见表 III 和 IV。从表 III 可以看出，B2 的表现最差，因为它的 M 和 S 最低，而 D 较高。比较 B1 和所提出的方法可知，尽管 B1 的 I 较高，导致 M 和 S 也较高，所提出的方法仍优于 B1。相比之下，所提出的方法中的 I 较低，为 B1 的 30% ，而 B1 为 226% 。呼气分割中也观察到了类似趋势 IV：由于 I 较高，B1 的 S 和 M 较高。结果表明，所提出的方法优于所有其他方法，但仍有改进空间，因为吸气和呼气分割中的 S 分别为 52% 和 61% 。此外，在呼气分割中，B2 同样是所有方法中表现最差的。呼吸分割和相位分割的结果呈现出非常相似的趋势。在所有三种情况下，B2 都是三种方法中表现最差的，而所提出的方法表现最佳。

B. 与真实值和预测的P 及d进行性能比较

使用预测边界和真实边界进行呼吸分段的结果 P 和 d 见表 V。从表 V可以看出，与使用估计的 P相比，使用真实边界时 M 和 S 更高，尽管均值 OvR 几乎相同，分别为 88% 和 90%。

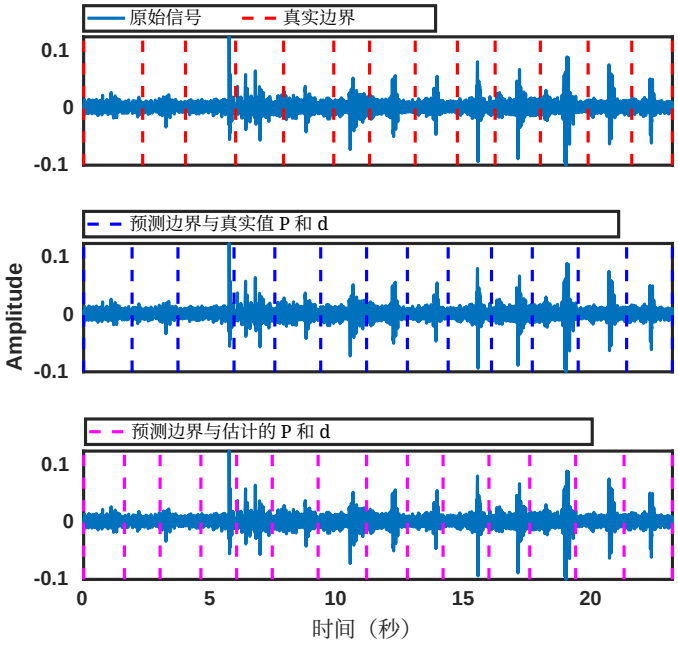


图 7. 使用预测的预测边界 $\boldsymbol{P}$和 $\boldsymbol{d}$ 相比真实值表现较差 $\boldsymbol{P}$和 $\boldsymbol{d}$。

全部 60 名受试者的实际呼吸次数和预测呼吸次数分别为 307 和 315。尽管使用真实值时的性能 P 很高，但我们可以看到，所有边界和片段匹配都未接近 100%。这里给出了一个样本文件的示例，其中真实值 P 由于第一个吸气声比其他吸气声更长而导致预测效果较差，从而产生了错误的呼吸边界预测。因此，所有呼吸边界都发生了偏移。在这种情况下，7个呼吸边界中有 3 个匹配成功、4 个被删除，并有 1 个片段匹配。该示例表明，我们所提出的方法容易受到 d 和 P的影响。这一点也可以从图 7所示的另一个示例中看出，其中对 P 和d 的估计不正确，导致边界预测错误；而使用正确的 P和 d 预测时，结果得到了改善。在图 7中，预测的 P 为14，而其真实值为 13，这导致 M 和 D 分别为 7 和 7；而使用真实的 P时，M 和 D 的值分别为 12 和 2。

C. 所提出的方法在患者与健康受试者之间的比较

由于我们的数据同时包含健康者和患者，因此我们分析了所提出的方法在不同组别之间的性能。表 VI 展示了所提出的方法在两组中的结果。从结果可以看出，性能更好

表 V

匹配(M)、插入(I)、删除(D)、片段匹配(S)以及重叠率(OvR)的均值和标准差（以%表示），针对使用估计值和真实值 P 以及 d 进行片段匹配的呼吸音。

P 和 d	M(%)	D(%)	I(%)	S(%)	OvR 均值(std)
估计值	**89**	**11**	**13**	**79**	88(13)
真实值	**93**	**7**	**7**	**86**	90(7)

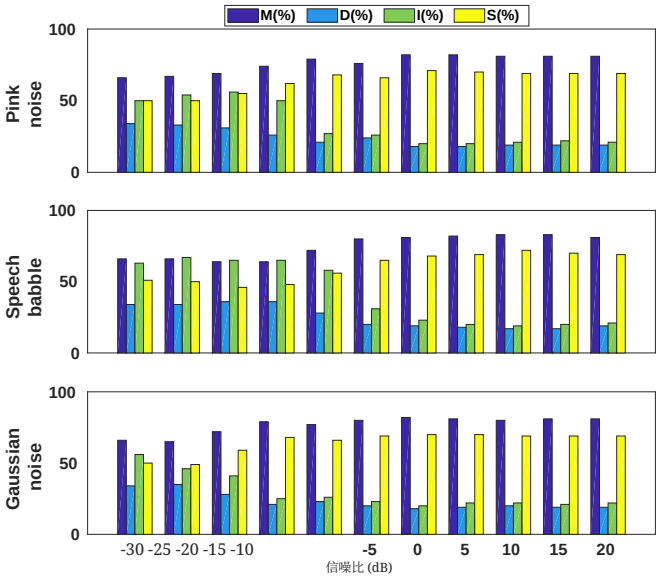


图 8. 不同噪声条件下评估指标匹配(M)、删除(D)、插入(I)和片段匹配(S)的比较，即在信噪比(dB)（横轴所示）下加入粉红噪声、语音 babble 噪声和高斯噪声。信噪比范围为 -30 dB 至 20 dB，步长为 5 dB。

对照组的匹配率高于患者，分别为 98% 和 81%。对于患者，片段匹配率为 63%，与对照组的 95% 相比非常低。我们还观察到，与健康受试者相比，患者的删除率和插入率都非常高。患者表现不佳的原因是对 P 的预测错误，因此导致 d。患者连续呼吸信号中的呼吸时长不规则，可能导致对 P 的预测不佳；与健康受试者相比，这是由呼吸困难造成的。然而，健康受试者的呼吸时长变化不大。

D. 所提出方法的加性噪声鲁棒性

关于实时应用，大多数呼吸音都是在非常嘈杂的环境中录制的；因此，通过添加三种类型来分析所提出的方法的性能

表 VI呼吸音片段匹配中的匹配数（M）、插入数（I）、删除数（D）、片段匹配数（S）以及重叠率的均值和标准差（%）(OvR)，适用于患者和健康者。

方法 used	预测边界	M(%)	D(%)	I(%)	S(%)	OvR 均值（标准差）
患者	**193**	**81**	**19**	**24**	**63**	86(17)
对照组	**182**	**98**	**2**	**2**	**95**	90(10)

表 VII

使用估计边界和真实边界时，每折分别显示哮喘患者与健康受试者之间的总分类准确率（TCA）（%）。最后一列表示所有折的 TCA 均值（标准差）。

呼吸边界	折1	折2	折3	折4	折5	折6	**Mean** (标准差)
真实值	60	80	90	70	81.81	66.67	75 (±11)
估计的	60	70	100	60	71.57	66.66	72 (± 15)

噪声的影响已得到研究。将三种噪声，即语音 babble 噪声、高斯噪声和粉红噪声，以不同信噪比添加到语音信号中。信噪比范围为 -30dB 至 20dB，步长为 5dB。所有噪声信号在添加到信号之前均为 16kHz。添加这三种噪声后信号的性能指标如图 8所示。从图 8 可以观察到，随着信噪比增加，M 增大，S 也随之增加。例如，在粉红噪声下，当信噪比分别为 -30dB 和 20dB 时，M 从 66% 增加到 81%，S 则从 50% 增加到 69%。语音 babble 噪声和高斯噪声的情况也呈现类似趋势。另一个观察结果是，随着信噪比增加，I 减小，这表明在正信噪比下，对 P 的估计有所改善，从而实现更好的边界预测。例如，在语音 babble 噪声下，当信噪比分别为 -30dB 和 20dB 时，I 从 63% 降至 21%。通过这些实验，我们得出结论：所提算法的性能对 P的估计非常敏感。随着信噪比增加，预测呼吸次数与真实值相等，从而获得更好的性能。

E. 哮喘患者和健康受试者分类

使用估计边界和真实边界进行逐折分类的准确率如表中所示 VII。使用估计边界的平均 TCA 为 72%(± 15%)，而使用真实边界时为 75%(± 11%)。使用估计边界进行分类的结果与使用真实边界时的结果接近。因此，所提出的方法可用于分割，分割后的数据可用于分类。

VI. 结论

在本研究中，提出了一种呼吸分割算法。所提算法用于寻找呼吸及其吸气和呼气相位的边界。利用信号能量的周期性，通过动态规划寻找边界。我们观察到，呼吸信号的周期性足以用于寻找呼吸边界。研究发现，与哮喘患者相比，所提出的方法在健康组中的表现良好，因为患者的呼吸阶段时长比健康受试者具有更大的变化。因此，为了获得更好的算法性能，需要对呼吸阶段时长和呼吸相位数量进行稳健估计。使用该方法时，哮喘患者与健康受试者之间的分类性能被发现具有可比性

估计边界和真实边界。所提出的方法具有噪声鲁棒性，因此适用于实时部署。

VII. 利益冲突

未声明任何利益冲突。

参考

[1] ThomasRitz and Walton T Roth。哮喘行为干预：呼吸训练。《行为矫正》，27(5):710–730，2003。[2] BarretZimmerman and Donna Williams。肺音。载于 StatPearls *[互联网]*。StatPearls Publishing，2019。[3] Malay Sarkar、Irappa Madabhavi、NarasimhaluNiranjan 和MeghaDogra。呼吸系统听诊。《胸腔医学年鉴》，10(3):158，2015。[4] Paul Forgacs、AR Nathoo 和 HD Richardson。呼吸音。《胸腔》，26(3):288–295，1971。[5] Shivani Yadav、NK Kausthubha、Dipanjan Gope、Uma Maheswari Krish- naswamy 和 Prasanta Kumar Ghosh。健康受试者与哮喘患者之间咳嗽、喘鸣和持续发声的自动分类比较。载于第40届国际医学与生物工程年会（EMBC），第1400–1403页。电气电子工程师学会，2018。[6] Shivani Yadav、Dipanjan Gope、K UmaMaheswari和PrasantaKumar Ghosh。呼吸阶段与呼吸边界在哮喘患者和健康受试者分类中的作用。载于2021年第43届电气电子工程师学会医学与生物学工程学会国际年会（EMBC），第870–873页。电气电子工程师学会，2021。[7] Rajkumar Palaniappan、Kenneth Sundaraj和SebastianSundaraj。用于呼吸相位检测和呼吸周期分割的自适应神经模糊推理系统。《生物医学中的计算机方法与程序》，145:67–72，2017。[8] Selim Aras、Mehmet ¨OZT¨URK 和 AliGangal。从肺部录制的单通道声音中自动检测呼吸周期。《土耳其电气工程与计算机科学杂志》，26(1)，2018。[9] Isa Yildirim、RashidAnsari 和 ZahraMoussavi。仅使用胸部声音信号的自动呼吸阶段与起始检测。载于2008年第30届IEEE医学与生物学工程学会国际年会，第2578–2581页。电气电子工程师学会，2008。[10] Steven Le Cam、ChCollet 和 FabienSalzenstein。呼吸声学信号分析与阶段检测。载于2008年IEEE声学、语音与信号处理国际会议，第3629– 3632页。电气电子工程师学会，2008。[11] JinFeng、FarookSattar和 Moe Pwint。基于沃尔什变换的方法在气管呼吸音信号分割中的应用。2008。[12] Feng Jin、FarookSattar和 DanielYT Goh。一种使用遗传方法的呼吸声学阶段分割算法。《医学与生物工程及计算》，47(9):941–953，2009。[13] YeeLengYap 和ZahraMoussavi。使用方差分形维数进行呼吸起始检测。载于2001年IEEE医学与生物学工程学会第23届国际年会会议录，第2卷，第1554–1556页。电气电子工程师学会，2001。[14] Omar Yahya和 Miad Faezipour。声学呼吸周期的自动检测与分类。载于美国工程教育学会2014年一区会议论文集，第1–5页。电气电子工程师学会，2014。[15] Sorin Dusan 和Lawrence Rabiner。最大谱过渡位置与音素边界之间的关系。载于第九届国际口语处理会议，2006。[16] PrasantaKumar Ghosh。使用基于极值的信号轨迹长度度量进行语音分割。载于2007年IEEE声学、语音与信号处理国际会议-ICASSP'07，第4卷，第IV–1065页。电气电子工程师学会，2007。[17] S´ergioPaulo 和 Lu´ıs C Oliveira。自动语音对齐及其置信度度量。载于国际自然语言处理会议（西班牙），第36–44页。施普林格，2004。[18] D Mazzoni和 RDannenberg。Audacity [软件]。匹兹堡，2000。[19] Richard Bellman。动态规划理论。《美国数学学会通报》，60(6):503–515，1954。[20] John W Lyons。DARPATIMIT声学语音连续语料库。美国国家标准与技术研究院，1993。

arXiv:2304.03758v1 [eess.AS] 7 Apr 2023

亮点

- 口部录制的呼吸音（人声呼吸音（VBS））在肺部疾病（如哮喘）的诊断和监测中的作用尚未得到充分研究。我们之前的研究表明，人声呼吸音（VBS）能够提供区分哮喘患者和健康受试者的线索。研究发现，了解人声呼吸音（VBS）的确切位置可以带来更好的分类性能。

- 在这项工作中，已使用基于动态规划的方法，通过利用呼吸信号的能量曲线结构来分割 VBS。

- 据我们所知，这项工作是同类研究中的首例，完成了VBS的分割。