



中国研究生创新实践系列大赛
“华为杯”第二十三届中国研究生
数学建模竞赛

学 校

长春师范大学

参赛队号

XXXXXX

队员姓名

1.

谢洪博

2.

崔曦月

3.

赵昱森

中国研究生创新实践系列大赛
“华为杯”第二十三届中国研究生
数学建模竞赛

题 目： 全国研究生数学建模竞赛论文标题

摘 要：

XXXX

关键词： XXX

目录

1	问题重述	4
1.1	引言	4
1.2	问题的提出	4
2	模型的假设	5
2.1	建模假设及适用范围	5
2.2	三层掩码与问题一的逐维观测规则	6
2.3	训练集限定的尺度处理与泄漏防护	6
3	符号说明	8
4	问题一：多视图时间锚定与可追溯时序对齐	9
4.1	输入输出与样本处理流程	9
4.2	子模型一：文本、语音和视觉的分层情感表征	9
4.2.1	上下文文本表征与词元归并	9
4.2.2	短时谱、倒谱与韵律的联合声学表征	10
4.2.3	面部动作、头部姿态和几何变化表征	11
4.3	子模型二：受转写约束的单调词级时间锚定	11
4.3.1	词边界后验与占据概率	12
4.4	子模型三：逐维质量加权的时间区间投影	13
4.4.1	计算素材范围与时间归属区间的分离	13
4.4.2	加权投影、局部方差与连续覆盖率	13
4.4.3	多层聚合的一致性与可重构条件	14
4.5	子模型四：固定秒级、词级与多尺度互补表示	15
4.5.1	固定秒级主时间视图	15
4.5.2	词级核验与多尺度上下文	15
4.5.3	保留 50 段相对进程辅助视图	15
4.6	子模型五：边界后验驱动的概率时间投影与低阶动态签名	16
4.6.1	概率时间投影：传播边界不确定性而不重排时间	16
4.6.2	低阶路径签名：补充均值与方差丢失的变化顺序	16
4.6.3	主结果保护与分支状态	17

4.7	特征存储、全量汇总与读取规范	17
4.7.1	按样本存储与视图组织	17
4.7.2	变长序列、填充与来源关系	18
4.7.3	100 条样本与 300 条模态明细	18
4.8	核验指标、典型样本与模型对照	19
4.8.1	样本、时间和对齐准确性	19
4.8.2	典型样本核验卡	19
4.8.3	单因素对照而非复杂模块累加	20
4.9	参数锁定、运行流程与复现说明	20
4.9.1	全量可追溯特征生成流程	20
5	模型评价	22
附录 A	MATLAB 源程序	23
A.1	第 1 问程序	23
附录 B	Python 源程序	24
B.1	第 2 问程序	24

1 问题重述

1.1 引言

1.2 问题的提出

围绕 xxxx，本文依次提出如下问题：

问题一：

问题二：

问题三：

2 模型的假设

2.1 建模假设及适用范围

假设一：样本身份和标签可审计。原始主键由 `video_id` 与 `clip_id` 共同确定。若发现重复主键、文件缺失或标签冲突，须保留异常记录并按来源材料核对，不能按模型预测改写标签。极性与强度的逻辑关系为

$$c_i = \begin{cases} -1, & y_i < 0, \\ 0, & y_i = 0, \\ +1, & y_i > 0. \end{cases} \quad (2.1)$$

数值相等以来源文件的存储精度为准，不自行设置“近零即中性”的标签重编码带。文本、语音和视觉特征提取以及时间定位均不使用情感标签。

假设二：观测状态先于数值变换确定。填充位置、真实缺失、分量自然未定义和正常观测具有不同含义。只有来源说明明确约定“有效时段内全维零向量代表缺失”时，才能对相应预计算文件采用零向量判缺失规则。问题一自生成特征依据解码、检测、估计器状态和逐维定义域判断有效性，不因特征值为零而直接认定缺失。

假设三：不同模态具有部分共享但非完全相同的潜在情感变化。共享动态与私有动态的区分可供后续缺失补全和融合使用；问题一不要求估计共享情感潜变量，也不以强制提高模态间相似性作为对齐目标。讽刺、转折、表情掩饰等跨模态差异不被强制压平。

假设四：局部时间依赖有用，但不保证平稳。短区间缺失可在后续任务中借助邻近位置推断；长区间或多模态同步缺失的不确定性应扩大。线性高斯动态仅作为后续概率补全的可控先验，不能将推断值重新标记为问题一的真实观测。

假设五：解释是对指定模型与指定基线的决策归因。删除、遮蔽和特征置换不是自然世界中的随机干预，相关性归因不自动等于心理或生理因果解释。原始视频映射也不由矩阵形状保证，而由明确的时间区间与来源索引保证。

假设六：物理时间对应与边界不确定性传播严格区分。媒体呈现时间戳、原音轨偏移及可靠词边界构成物理时间依据。边界后验仅用于描述“词位于何处”的时间不确定性，不修改原始媒体时间戳，不跨越已确认的解码缺口或可靠锚点屏障。模态间情感表达不同不被直接解释为音画同步错误；概率时间视图是物理对齐的辅助表示，而不是用语义相似度重新排列时间。

2.2 三层掩码与问题一的逐维观测规则

先依据长度字段、填充规则和结构元数据得到 P_{it}^m 。仅对来源已明确采用全维零向量标记缺失、且一行各维度共同观测的预计算专项文件，使用

$$\begin{aligned} O_{it}^m &= P_{it}^m \mathbf{1}\{\mathbf{x}_{it,\text{raw}}^m \text{ 各分量有限且不全为零}\}, \\ M_{it}^m &= O_{it}^m B_{it}^m. \end{aligned} \quad (2.2)$$

若文件提供更精确的观测元数据，应优先使用元数据。训练完整视图取 $B \equiv 1$ ，人工缺失只作用于原本观测位置；不通过原始文本等辅助字段绕过专项集的既有缺失设置。对问题一自生成特征，逐维定义

$$O_{ird}^m = P_{ir}^m S_{ir}^m E_{ird}^m \mathbf{1}\{x_{ird,\text{raw}}^m \in \mathbb{R}\}. \quad (2.3)$$

这里 S_{ir}^m 表示原始来源可用， E_{ird}^m 表示相应估计成功且该分量有定义，集合 $\mathbb{R}$ 不含非有限数值。正常零值不改变 O_{ird}^m 。原生文本向量可有效而词时间尚不可定位，此时另设时间可定位标记，不删除原生文本，也不将其伪装成已有时间支撑。逐维位置区分 $P = 0$ 的填充、 $P = 1, O = 0$ 的未观测和 $P = 1, O = 1$ 的观测。未观测原因另行区分解码失败、检测失败、自然未定义和时间不可定位。概率边界派生视图有独立状态标记，不覆盖原始 O 。局部覆盖率由式 (4.24) 给出，不能由二值掩码代替。对未说明填充规则的零尾段，不用“最后一个非零位置”推断真实长度。若填充和末尾缺失两种机制可产生相同数组，则单凭数组不能唯一识别二者。此时设置 $U = 1$ ，保守保留为潜在有效但不可观测位置，待合法元数据核验后确定 P ，并单列歧义比例。对 `text_bert` 的使用仅限经确认不会泄露缺失内容的长度或填充元数据；`raw_text` 和词元编号不得进入问题二学生模型，也不得绕过专项集的文本置零直接重编码恢复。预计算特征与原始词元的关系不明时，不擅自建立一一对应。

2.3 训练集限定的尺度处理与泄漏防护

记 Ω_{md}^{tr} 为训练集中模态 m 的第 d 维实际观测位置集合；对于整行共同观测的文件，由式 (2.2) 构造，对于逐维观测的文件，由式 (2.3) 构造。令

$$\begin{aligned} \mu_{md} &= \text{median}_{(i,t) \in \Omega_{md}^{\text{tr}}} x_{itd}^m, \\ r_{md} &= 1.4826 \text{ median}_{(i,t) \in \Omega_{md}^{\text{tr}}} |x_{itd}^m - \mu_{md}|, \\ s_{md} &= \begin{cases} r_{md}, & r_{md} > \varepsilon_s, \\ s_{md}^{\text{fb}}, & r_{md} \leq \varepsilon_s \text{ 且该维并非常量}, \\ 1, & \text{该维为常量}. \end{cases} \end{aligned} \quad (2.4)$$

其中 1.4826 为固定稳健尺度系数， s_{md}^{fb} 为预先声明的回退尺度：具有稳定物理范围的维度采用固定物理尺度，其余连续维度采用训练折标准差或预先固定的参考尺度。这样可避免

“非恒定但 MAD 为零”的稀疏维度被 ε_s 人为放大后全部截断。训练中无有效观测的维度不估计任何尺度，单列状态。对有效维度应用

$$\tilde{x}_{itd}^m = \begin{cases} \text{clip}((x_{itd}^m - \mu_{md})/s_{md}, -b_m, b_m), & O_{itd}^m = 1, \\ 0, & O_{itd}^m = 0. \end{cases} \quad (2.5)$$

这里 $\text{clip}(z, -b, b) = \min\{b, \max\{-b, z\}\}$ 。若合法数据共同逐行观测，令 $O_{itd}^m = O_{it}^m$ 。回退尺度来源、常量维状态及使用次数均写入日志。标准化后的零可能是正常观测，不能再由零值重建掩码。真实填充在变换后置零，且不参与参数估计。原生未标准化特征必须保留。问题一若仅作描述性展示，可以另外保存明确标识为“全量描述性统计”的尺度，但该尺度不得用于声称无泄漏的预测评估。凡进行预测评估，尺度、降维和其他数据驱动参数均只在训练折拟合，验证与测试应用同一变换。按 `video_id` 核验同源片段相关性。已有官方划分时不擅自重划；官方划分若存在同源视频跨集，报告该限制。自行设置问题一的核验或后续实验划分时，以原始视频为分组单位；时间步数增加不等于独立样本数增加。

3 符号说明

表 3.1 符号约定

符号	含义
$i, j; m \in \mathcal{M} = \{T, A, V\}$	样本索引；文本、语音、视觉模态。
$D_i; r, t, k; \tau$	片段实际时长；原生行、目标区间、词或特征维索引；秒级时间。
$\mathbf{X}_i^m \in \mathbb{R}^{L_i^m \times d_m}$	模态特征矩阵；问题一主维度为 $d_T = 768, d_A = 74, d_V = 35$ 。
$P_{it}^m, O_{it}^m, U_{it}^m$	结构有效、整行观测和填充/缺失歧义标记；整行规则仅用于满足其条件的预计算文件。
O_{ird}^m, q_{ird}^m	问题一原生特征第 d 维的有效观测标记和质量权重。
$B_{it}^m; M_{it}^m$	后续训练的人工保留掩码；参与推断的观测掩码。
$\mathcal{R}_{ird}^m; J_{ir}^m$	特征计算使用的素材范围；原生行在时间轴上的归属区间。
$I_{it}; \Delta; h_s$	目标区间；固定秒级步长；多尺度上下文窗口长度。
$\bar{x}_{itd}^m, v_{itd}^m, c_{itd}^m$	物理对齐均值、局部方差和逐维真实时间覆盖率。
$S_{ik}, E_{ik}; \pi_{ik}(\tau)$	第 k 个词的随机起止边界；时刻 τ 属于该词的后验占据概率。
$\alpha_{itk}, \beta_{ikrd}^m$	目标区间-词的期望重叠质量；词-原生来源的后验重叠质量。
$\text{Sig}^{(1)}, \text{LogSig}^{(2)}$	低阶路径签名的一阶增量与二阶有向交互，用于辅助描述窗口内部变化顺序。
$\mathcal{D}_{tr}, \mathcal{D}_{va}, \mathcal{D}_{te}$	附件二的训练、验证和测试划分；不凭问题一重新定义官方划分。
$\mathbf{w}_{it}, A, C_m, Q, R_m$	后续动态模型的增广状态、转移矩阵、观测矩阵及噪声协方差。
$\boldsymbol{\mu}_{it}, \Sigma_{i,tt}$	后续条件平滑后验均值和状态协方差块。
$\rho_{it}^m, \alpha_{it}^m$	可靠度和融合模态权重，不直接等于解释贡献。
p_{ic}, a_{ic}, b_{ic}	极性概率；非中性幅值的 Beta 分布形状参数。
$\hat{c}_i, \hat{y}_i$	最终报告的极性与情感强度。
$F_c, F_y, v(S); \phi_{im}, \psi_{imt}$	待解释函数、联盟价值；模态及模态内位置贡献。
$R_i; \mathcal{T}_i, \mathcal{L}_i$	关键证据位置集合；原始素材映射和处理日志。
ε	数值稳定常数；按变量量纲分别设定并记录。

4 问题一：多视图时间锚定与可追溯时序对齐

4.1 输入输出与样本处理流程

输入为附件一的 100 条原始视频 V_i 、英文转写 W_i 和样本主键。处理算子定义为

$$\mathcal{F}_1 : (V_i, W_i, \text{id}_i) \mapsto \{\mathbf{X}_{i,\text{native}}^m, \mathbf{X}_{i,\text{word}}^m, \mathbf{X}_{i,\text{sec}}^m, \mathcal{X}_{i,\text{multi}}^m, \mathbf{X}_{i,50}^m, \mathbf{X}_{i,\text{post}}^m, \mathbf{X}_{i,\text{dyn}}^m, P_i^m, O_i^m, \mathcal{T}_i, \mathcal{L}_i\}_{m \in \mathcal{M}}. \quad (4.6)$$

其中，固定秒级结果为主时间视图，词级结果用于语义核验，多尺度及 50 段相对进程结果作为辅助；词边界后验视图只在对齐器能够提供可靠路径后验时生成，低阶动态签名作为独立辅助分支保存。空的可选视图以未启用状态表示，不视为样本遗漏。首先建立“标签表一行—原视频文件—输出记录”的双向清单，主键由 `video_id` 和 `clip_id` 共同组成，保留原始字符串及文件哈希。检查标签主键、素材主键与输出主键集合的一致性，分别登记重复、缺少、额外及冲突记录。标签只用于一致性核验，不参与特征定位、异常修正、边界后验或动态分支参数选择。采用 FFmpeg/ffprobe 类媒体工具获取音视频流和时间信息。设共享片段原点为 τ_{i0} ，呈现时间戳和时间基分别为 PTS_{ir}^m 与 TB_i^m ，则

$$\tau_{ir}^m = \text{PTS}_{ir}^m \text{TB}_i^m - \tau_{i0}. \quad (4.7)$$

若音频在共享时间轴上连续，则可写为 $\tau_i^A(n) = \delta_i^A + n/f_s$ 。若存在局部解码缺口，则按第 j 个连续有效音频段 $\mathcal{N}_{ij}$ 分段映射：

$$\tau_i^A(n) = \delta_{ij}^A + \frac{n - n_{ij}}{f_s}, \quad n \in \mathcal{N}_{ij}, \quad (4.8)$$

其中 n_{ij} 为该连续段的起始采样索引， δ_{ij}^A 为该段在共享时间轴上的真实起点。若实现采用保留缺口占位的全长数组，也可继续使用全局索引公式，但缺口位置必须由来源可用标记排除，不能参与声学估计。音频统一解码为单声道 $f_s = 16\,000$ Hz 浮点波形；起始偏移、各连续段边界、重采样映射与原始采样率同时存档。视频按真实时间戳解码，保留可变帧率信息，不以固定帧率公式替换未经核验的时间戳。不进行静音拼接、变速或改变强弱关系的强降噪。无音轨、局部解码缺口和无人脸分别标记，不删除样本，也不把缺口两侧拼接成连续内容。英文转写保留原文副本；工具输入的规范化、字符偏移、子词、原词与标点之间的映射独立保存。

4.2 子模型一：文本、语音和视觉的分层情感表征

4.2.1 上下文文本表征与词元归并

采用冻结的通用预训练 BERT 编码器，模型标识为 `bert-base-uncased`，不使用本题情感标签微调。设原词 w_{ik} 对应子词集合 $\mathcal{B}_{ik}$ ，第 ℓ 层子词表示为 $\mathbf{h}_{ib}^{(\ell)}$ ，定义

$$\mathbf{x}_{ik}^T = \frac{1}{4|\mathcal{B}_{ik}|} \sum_{\ell=L-3}^L \sum_{b \in \mathcal{B}_{ik}} \mathbf{h}_{ib}^{(\ell)} \in \mathbb{R}^{768}. \quad (4.9)$$

对于重复覆盖的子词，先按其距离窗口边界的权重归并多窗口表示，再按式 (4.9) 归并到原词。初始配置采用最多 510 个非特殊子词的窗口和 384 个子词的步长，特殊词元占用模型其余位置；实际分词器版本、最大长度和窗口起止索引均存档。保留否定词、程度词、转折词和重复表达，不机械删除停用词。标点保留在编码输入中，但不单独虚构发音时间；特殊词元不构成独立物理时间步。词时间只表示向量的定位锚，不表示上下文向量仅使用该词。句级汇总可另存，不能替代词级可定位序列。

4.2.2 短时谱、倒谱与韵律的联合声学表征

保留原方案的 74 维可解释主特征。短时频谱配置为窗长 $W = 400$ 点、步长 $H = 160$ 点和 Hann 窗，分别对应 25 ms 与 10 ms。将 FFT 长度单独设置为 $N_{\text{FFT}} = 512$ ，定义

$$S_{ir}[u] = \sum_{n=0}^{W-1} s_i[rH + n]h[n] \exp\left(-\frac{2\pi i u n}{N_{\text{FFT}}}\right),$$

$$E_{irb} = \log\left(\varepsilon_E + \sum_{u=0}^{N_{\text{FFT}}/2} H_b[u]|S_{ir}[u]|^2\right). \quad (4.10)$$

这里 H_b 为第 b 个 Mel 滤波器。主配置不对分析窗作居中平移；末尾不足窗长时仅为计算作右侧零延拓，并保存真实采样范围和填充比例，不把延拓部分记作有效观测时长。Mel 频率映射、滤波器归一化、频率上下限和能量底数均在配置中固定。采用 40 维对数 Mel、13 维 MFCC、13 维 MFCC 一阶差分 and 8 维韵律/谱统计量：

$$\mathbf{x}_{ir}^A = [\text{LogMel}_{40}; \text{MFCC}_{13}; \Delta\text{MFCC}_{13}; \text{Prosody}_8] \in \mathbb{R}^{74}. \quad (4.11)$$

MFCC 采用对数 Mel 的正交归一化第二型离散余弦变换，保留 0 至 12 号系数。差分初始设置为左右各两步的局部线性斜率，内部点使用对称邻域，边界使用实际存在的单侧邻域；不足以估计时逐维标记无效。差分计算的邻域也属于其素材计算范围。8 项统计量依次为对数能量、对数基频、有声概率、谱质心、谱带宽、谱通量、过零率和谐噪比。基频与有声概率采用明确输出这两类量的概率估计器；若工具实际提供的是周期性强度或未校准分数，须按实际含义另命名，不冒称概率。谐噪比采用单独锁定的谐性估计器，并记录单位、基频范围和计算窗口。**不同声学分量允许具有不同计算窗口。**谱特征的 25 ms 窗不自动适用于基频、有声概率或谐噪比。各分量先保留其原生时间锚与计算范围，再依据真实归属区间投影到 10 ms 声学网格，形成式 (4.11)。无声区间中的基频等分量可自然未定义，但不意味着整个音频模态缺失；局部损坏只影响相应局部。全程保留逐维掩码。频谱与倒谱存在冗余，但提供不同形式的谱描述；是否裁减可在合法训练数据上消融。**这里的自定义 74 维向量不是对附件二既有 74 维预计算特征的逐维解释。**冻结的深层语音表示可作为独立辅助分支另存，但不改变主特征维度，也不默认学习公共空间后重新安排时间对应。

4.2.3 面部动作、头部姿态和几何变化表征

使用 OpenFace 类通用面部分析工具提取动作单元、头部姿态、视线和关键点，并保留检测成功标记、置信度和轨迹记录。定义

$$\mathbf{x}_{ir}^V = [\text{AU}_{17}; \text{Pose}_6; \text{Gaze}_6; \text{Geometry}_6] \in \mathbb{R}^{35}. \quad (4.12)$$

17 个 AU 强度按 AU01、02、04、05、06、07、09、10、12、14、15、17、20、23、25、26、45 的顺序保存。具体构建不支持的维度标为不可用，不临时换成其他物理量。Pose 为三维旋转与三维平移，保留工具原始单位再按式 (2.4) 处理；不以二维像素眼间距除以毫米平移。Gaze 为双眼各三维方向，视线估计失败时仅将相关维度标为无效。Geometry 保留六个无量纲量：左右眼开合度、嘴部开合度、嘴角间距、左右眉眼距离。为固定几何定义，设二维关键点按零基编号 $0, \dots, 67$ 记为 $\mathbf{p}_j$ ，距离为 $d_{ab} = \|\mathbf{p}_a - \mathbf{p}_b\|_2$ ，眼中心为

$$\begin{aligned} \mathbf{e}_1 &= \frac{1}{6} \sum_{j=36}^{41} \mathbf{p}_j, & \mathbf{e}_2 &= \frac{1}{6} \sum_{j=42}^{47} \mathbf{p}_j, & \ell &= \|\mathbf{e}_1 - \mathbf{e}_2\|_2, \\ g_1 &= \frac{d_{37,41} + d_{38,40}}{2\ell}, & g_2 &= \frac{d_{43,47} + d_{44,46}}{2\ell}, & g_3 &= \frac{d_{61,67} + d_{62,66} + d_{63,65}}{3\ell}, \\ g_4 &= \frac{d_{48,54}}{\ell}, & g_5 &= \frac{\|\frac{1}{5} \sum_{j=17}^{21} \mathbf{p}_j - \mathbf{e}_1\|_2}{\ell}, & g_6 &= \frac{\|\frac{1}{5} \sum_{j=22}^{26} \mathbf{p}_j - \mathbf{e}_2\|_2}{\ell}. \end{aligned} \quad (4.13)$$

只有关键点完整且 $\ell > \varepsilon_\ell$ 时才计算相应几何量。编号组的左右含义按工具映射锁定，不随画面镜像自动交换。单脸置信度初始门槛可取 $\theta_V = 0.8$ ，作为待核验设计值而非实测最优值。多人画面综合轨迹连续性、位置和口部运动与音频活动的一致性选择主说话者轨迹；选择不唯一时降低权重并保留候选，不逐帧任意切换。归一化差分可作为辅助动态统计，但不跨轨迹切换或缺失段计算。视觉量在后续时间聚合时按几何类型处理：AU、平移量和无量纲几何量可采用质量加权标量均值；头部旋转不直接对欧拉角做全局算术平均，而优先转换为旋转矩阵后在 $SO(3)$ 上求局部加权旋转均值，或在确认角度变化远离 $\pm\pi$ 分支点时使用局部近似；视线方向采用加权合向量并重新单位化，合向量范数过小时标记方向不稳定。这样避免 179° 与 -179° 被错误平均为 0° 。

4.3 子模型二：受转写约束的单调词级时间锚定

已有英文转写时采用强制对齐，不重新识别后覆盖原文。可使用 Montreal Forced Aligner 所代表的声学-音素对齐流程。将原转写转换为按顺序展开的状态图，允许自循环、向后续状态移动和合法非词状态，不允许逆序跳回。给定对齐器声学观测 $\mathbf{a}_r$ ，求解

$$z_{1:R} = \arg \max_{z_{1:R} \in \mathcal{G}(W_i)} \left\{ \log p(z_1) + \sum_{r=2}^R \log p(z_r | z_{r-1}) + \sum_{r=1}^R \log p(\mathbf{a}_r | z_r) \right\}. \quad (4.14)$$

表 4.2 问题一主特征与原始素材追溯字段

模态	维度	原生时间组织	质量依据	追溯字段
文本	768	原词／子词，另存词时间	词元映射、对齐分数及边界不确定性	字符、子词、词区间及编码窗口
语音	74	10 ms 网格；各分量计算范围独立	解码与逐维估计状态	原始采样范围、音轨偏移、分析窗口
视觉	35	原始帧时间戳与归属区间	检测、轨迹和逐维有效性	原始帧号、PTS、轨迹及关键点映射

这里 $\mathcal{G}(W_i)$ 为原转写允许的状态路径集合，对齐器内部声学特征不必等于式 (4.11)。Viterbi 递推为

$$V_r(k) = \log p(\mathbf{a}_r | k) + \max_{j \in \text{Pred}(k)} \{V_{r-1}(j) + \log p(k | j)\}. \quad (4.15)$$

回溯并合并同词音素，得到 $I_{ik}^{\text{word}} = [s_{ik}, e_{ik})$ 。可靠词区间须满足

$$0 \leq s_{i1} < e_{i1} \leq s_{i2} < e_{i2} \leq \cdots \leq s_{iK_i} < e_{iK_i} \leq D_i. \quad (4.16)$$

零时长、逆序或越界结果标记为异常，不直接作为可靠区间使用。词间及首尾未覆盖部分单独保存，区分已确认静音、非词声音和未可靠对齐区域。没有对齐到单词不等于静音，也不能靠拉长相邻词吸收全部间隙。可靠词区间与补充区间共同形成覆盖 $[0, D_i)$ 的时间分区；区间类型不等于情感标签。工具若只提供路径分数，称为对齐质量分数，不称为校准概率。若可在固定状态图上获得前向-后向权重，可另外计算边界均值与分位区间，并与原始最优路径边界分别保存。低质量词保留原文及向量，时间映射标为不确定。完全失败时保留原生文本，时间视图标记不可定位，绝不按词数均分视频后宣称精确对齐。

4.3.1 词边界后验与占据概率

当对齐器能够在固定转写状态图上提供前向-后向后验时，不把单一 Viterbi 边界视为唯一确定事实，而将第 k 个词的起止边界记为随机变量 (S_{ik}, E_{ik}) ，其联合分布由合法单调路径诱导。定义词占据概率

$$\pi_{ik}(\tau) = \Pr(S_{ik} \leq \tau < E_{ik} | A_i, W_i), \quad k = 1, \dots, K_i. \quad (4.17)$$

另设 $\pi_{i0}(\tau)$ 表示状态图中合法静音或非词状态的后验质量；在已建模且媒体来源有效的时间位置满足

$$\sum_{k=0}^{K_i} \pi_{ik}(\tau) = 1. \quad (4.18)$$

解码缺口不属于“被模型观察到的静音”，因此不纳入式 (4.18) 的有效时间域。对固定目标区间 I_{it} ，定义第 k 个词在其中的期望占据时长

$$\alpha_{itk} = \mathbb{E}[|I_{it} \cap [S_{ik}, E_{ik})| \mid A_i, W_i] = \int_{I_{it}} \pi_{ik}(\tau) d\tau. \quad (4.19)$$

由此构造文本后验时间视图

$$\mathbf{x}_{it, \text{post}}^T = \frac{\sum_{k=1}^{K_i} \alpha_{itk} \mathbf{x}_{ik}^T}{\sum_{k=1}^{K_i} \alpha_{itk}}, \quad (4.20)$$

仅在分母大于零时定义。若边界后验退化为确定区间，则 $\pi_{ik}(\tau) = \mathbf{1}\{\tau \in [s_{ik}, e_{ik})\}$ ，式 (4.19) 自动退化为普通区间交长，因此概率视图与物理投影具有一致的极限情形。需要强调，式 (4.20) 是“以期望占据质量归一化”的表示，通常不等于逐条随机边界路径聚合结果的严格后验均值。若需要后者，应从联合合法路径后验采样，在每条路径上先执行完整聚合，再统计样本均值、方差与不可估计比例。边界后验的窄宽也不自动等于现实世界中的校准准确率，仍须通过人工审计边界验证其覆盖率。

4.4 子模型三：逐维质量加权的时间区间投影

4.4.1 计算素材范围与时间归属区间的分离

对每个原生特征同时保存 $\mathcal{R}_{ird}^m$ 与 J_{ir}^m 。前者是第 d 维计算实际使用的采样、帧或上下文范围，允许重叠且可以随维度变化；后者是该原生行在物理时间轴上的归属区间，用于时间质量分配。文本使用可靠词区间作为 J ；音频与视频在每个连续有效段内，按相邻时间锚的中点构造不重叠归属区间，再与该段真实边界相交。已确认的缺口保留为空，不通过中点扩展跨越。区间外推到首尾边界的规则固定并记录。由此，重叠分析窗不会被重复计成多份有效时长。语音分量异步输出时，其原生分支到联合声学网格的投影也使用同一规则，并保留两级来源关系。

4.4.2 加权投影、局部方差与连续覆盖率

给定目标区间 I_{it} ，令

$$\begin{aligned} w_{itrd}^m &= |I_{it} \cap J_{ir}^m| q_{ird}^m O_{ird}^m, \\ W_{itd}^m &= \sum_r w_{itrd}^m, \quad \bar{x}_{itd}^m = \frac{\sum_r w_{itrd}^m x_{ird}^m}{W_{itd}^m} \quad \text{当 } W_{itd}^m > 0. \end{aligned} \quad (4.21)$$

质量权重满足 $0 \leq q_{ird}^m \leq 1$ ，来源和计算公式须记录。无法可靠评估额外质量时，正常有效维可取单位权重，而不是构造未经支持的概率。 $W_{itd}^m = 0$ 时不计算分式，只输出占位值及无观测标记。式 (4.21) 是一维加权最小二乘问题的闭式解：

$$\bar{x}_{itd}^m = \arg \min_{u \in \mathbb{R}} \sum_r w_{itrd}^m (u - x_{ird}^m)^2, \quad W_{itd}^m > 0. \quad (4.22)$$

因此，三模态使用一致的时间投影算子，而不是分别采用无时间依据的帧号平均。原生不可定位的文本不参与该投影。对标量特征，式 (4.22) 直接给出聚合值；对视觉旋转和单位方向量，则沿用相同的时间质量权重，但在各自几何空间内完成聚合，因此“一致的投影算子”指一致的时间质量分配，而不是强迫所有物理量使用同一种欧氏算术平均。另外保存加权局部方差及观测数量：

$$v_{itd}^m = \frac{\sum_r w_{itr}^m (x_{ir}^m - \bar{x}_{itd}^m)^2}{W_{itd}^m}, \quad n_{itd}^m = \#\{r : w_{itr}^m > 0\}. \quad (4.23)$$

上述方差是局部分散程度，不是均值估计的不确定性；仅一个来源时数值可为零，但须标记动态证据不足。均值、方差、掩码和计数分别保存，不改变 768/74/35 的主维度。定义参与估计的真实时间集合及连续覆盖率为

$$\mathcal{U}_{id}^m = \bigcup_{r: O_{ir}^m=1, q_{ir}^m>0} J_{ir}^m, \quad c_{itd}^m = \frac{|I_{it} \cap \mathcal{U}_{id}^m|}{|I_{it}|}. \quad (4.24)$$

$c_{itd}^m \in [0, 1]$ 衡量真实时间支撑，不等于检测置信度。一个 100 ms 区间只有 10 ms 有效来源时，即使均值可计算，其覆盖率也仅为 0.1。目标逐维观测标记取 $O_{itd, \text{sec}}^m = \mathbf{1}\{W_{itd}^m > 0\}$ ，与覆盖率分开保存。区间并集保证重叠来源不重复计时。另存解码可用率与分量定义率，避免将自然未定义误记为模态完全缺失。

4.4.3 多层聚合的一致性与可重构条件

若目标区间 I 被互不重叠的 $I_1, \dots, I_J$ 完整划分，并且各子区间统计量均由同一套底层来源权重产生，则均值满足

$$\bar{x}_d^m(I) = \frac{\sum_{j=1}^J W_d^m(I_j) \bar{x}_d^m(I_j)}{\sum_{j=1}^J W_d^m(I_j)}. \quad (4.25)$$

零权重子区间不参与求和。为使均值与方差能够无损合并，底层同时保存加权充分统计量

$$S_0(I) = \sum_r w_r(I), \quad S_1(I) = \sum_r w_r(I) x_r, \quad S_2(I) = \sum_r w_r(I) x_r^2, \quad (4.26)$$

于是 $\bar{x} = S_1/S_0$ ， $v = S_2/S_0 - \bar{x}^2$ 。等价地，若只保存子区间均值与方差，则大区间方差按

$$v(I) = \frac{\sum_j W_j \left[v(I_j) + (\bar{x}(I_j) - \bar{x}(I))^2 \right]}{\sum_j W_j} \quad (4.27)$$

合并，不能简单平均各子区间方差。上述可合并性只对“完整、互不重叠且与目标边界嵌套”的子区间成立。若新目标边界切穿某个已聚合子区间，则必须回查其原生来源记录重新计算交长，不能把中间格均值当成区间内部均匀分布后再次投影。观测数量 n 也按原生来源索引去重，不能在重复视图之间直接相加。直接原生聚合与充分统计量合并所得结果应在预设数值容差内一致，并作为实现审计条件。

4.5 子模型四：固定秒级、词级与多尺度互补表示

4.5.1 固定秒级主时间视图

采用固定物理步长 $\Delta = 0.1$ s，定义

$$I_{it}^{\text{sec}} = [(t-1)\Delta, \min\{t\Delta, D_i\}), \quad t = 1, \dots, L_i, \quad L_i = \left\lceil \frac{D_i}{\Delta} \right\rceil. \quad (4.28)$$

由题面时长 2.648–34.567 s 可得 $27 \leq L_i \leq 346$ 。最后一个区间保留真实长度，不补出额外物理时长。主特征形状为

$$\mathbf{X}_{i,\text{sec}}^T \in \mathbb{R}^{L_i \times 768}, \quad \mathbf{X}_{i,\text{sec}}^A \in \mathbb{R}^{L_i \times 74}, \quad \mathbf{X}_{i,\text{sec}}^V \in \mathbb{R}^{L_i \times 35}. \quad (4.29)$$

同一样本三模态第 t 行共享同一个物理区间；跨样本的相同步号共享相同名义步长。某模态不具有有效来源时保留该时间位置及掩码，而不是删除一行破坏共同时间轴。

4.5.2 词级核验与多尺度上下文

词级视图直接以 I_{ik}^{word} 为目标，调用式 (4.21)，回答“某原词对应哪些声音采样、视频帧和特征行”。非词区间单独保存类型及来源；未可靠定位的原词仍保留在原生文本记录中。以固定秒级区间的中心 κ_{it} 为锚，构造上下文窗口

$$R_{it}^{(s)} = [\kappa_{it} - h_s/2, \kappa_{it} + h_s/2) \cap [0, D_i), \quad (4.30)$$

$$(h_1, h_2, h_3) = (0.1, 0.3, 0.7) \text{ s}, \quad \mathcal{X}_{i,\text{multi}}^m \in \mathbb{R}^{L_i \times 3 \times d_m}.$$

各尺度直接从原生特征或保存权重的底层统计聚合，并分别保存实际起止时间。边界窗口截断后可不再对称，且 0.1 s 上下文窗口不必与末尾不足步长的主区间相同。较大窗口表示邻域上下文，不宣称是当前 100 ms 内直接观察到的全部内容。各尺度分开存储，不默认加权压成一个新向量。

4.5.3 保留 50 段相对进程辅助视图

同时保存均分 50 段的相对进程辅助视图，定义为

$$I_{it}^{50} = \left[\frac{(t-1)D_i}{50}, \frac{tD_i}{50} \right), \quad t = 1, \dots, 50. \quad (4.31)$$

它表示片段相对进程，而非统一的物理分辨率。题面最短和最长视频的每段分别为 0.05296 s 和 0.69134 s，因此不将其作为唯一主时间轴。所有区间采用左闭右开记法；终止时间戳单独记录，不为包含末端单点而增加虚假时长。50 段视图覆盖全片，不截取前 50 个词；重复使用同一原生来源时保留来源编号，不声称新增了独立观测。式 (4.31) 仅定义本稿问题一的相对进程视图，不解释附件二既有 **aligned** 特征的时间含义。

4.6 子模型五：边界后验驱动的概率时间投影与低阶动态签名

4.6.1 概率时间投影：传播边界不确定性而不重排时间

物理时间投影仍是默认主结果。仅当固定转写状态图能够提供可信的前向-后向后验时，额外生成概率时间视图。对模态 $m \in \{A, V\}$ 的第 r 个原生来源区间 J_{ir}^m ，定义词 k 对该来源的后验重叠质量

$$\beta_{ikrd}^m = q_{ird}^m O_{ird}^m \int_{J_{ir}^m} \pi_{ik}(\tau) d\tau. \quad (4.32)$$

于是词级后验特征为

$$\bar{x}_{ikd,\text{post}}^m = \frac{\sum_r \beta_{ikrd}^m x_{ird}^m}{\sum_r \beta_{ikrd}^m}, \quad \sum_r \beta_{ikrd}^m > 0. \quad (4.33)$$

对固定秒级网格仍使用式 (4.19) 将文本词向量映射至共同时间轴；语音和视觉的媒体时间戳本身不因文本边界不确定而改变。若研究需要把“某词附近的音频/视觉证据”作为词级核验视图，则采用式 (4.32)–(4.33)，而不是依据情感向量相似度移动来源。为审计边界不确定性，保存每个词的后验均值边界、预设分位区间、区间宽度以及人工参考边界是否落入后验区间。若对齐器只能输出单一路径和路径分数，则该概率分支标记为未启用，不以人为高斯扰动伪造后验。

4.6.2 低阶路径签名：补充均值与方差丢失的变化顺序

固定窗口内的均值与方差不能区分“先升后降”和“先降后升”等顺序结构，因此在不改变 768/74/35 主特征维度的前提下，增加一个独立的低维动态辅助分支。对连续有效区间 $I = [a, b]$ ，从少量预先指定且物理含义清楚的动态通道构造时间增广路径

$$\mathbf{z}(\tau) = [\tilde{\tau}, z_1(\tau), \dots, z_p(\tau)], \quad \tilde{\tau} = \frac{\tau - a}{b - a}. \quad (4.34)$$

主配置仅从可解释声学动态量、AU、嘴部开合等通道中选择，选择表和顺序在配置中固定；不在全体 BERT 维度上直接展开高阶交互。一阶签名为各通道总增量

$$S_j(I) = \int_a^b dz_j(\tau), \quad (4.35)$$

二阶迭代积分为

$$S_{jk}(I) = \int_{a < u < v < b} dz_j(u) dz_k(v), \quad (4.36)$$

其反对称部分

$$A_{jk}(I) = \frac{1}{2}(S_{jk}(I) - S_{kj}(I)) \quad (4.37)$$

描述通道间有向变化关系。若含时间坐标共有 $d = p + 1$ 个通道，则二阶对数签名维度为

$$d_{\text{dyn}} = d + \frac{d(d-1)}{2}, \quad (4.38)$$

规模可控，并作为 $\mathbf{X}_{i,\text{dyn}}$ 独立保存。路径签名只在连续有效段内计算：解码缺口、轨迹切换或所选通道未定义时立即分段，不把缺失值填零后连接成虚假轨迹；若窗口内有效连续长度不足预设阈值，则动态分支标记为不可估计。插值仅用于在已观测连续段内定义分段线性路径，插值方法、原生节点和窗口边界全部存档。二阶有向项只作为顺序交互描述，不解释为因果效应。

4.6.3 主结果保护与分支状态

最终交付中， $\mathbf{X}_{i,\text{sec}}^m$ 及其来源链始终作为问题一主结果； $\mathbf{X}_{i,\text{post}}$ 和 $\mathbf{X}_{i,\text{dyn}}$ 均为可选辅助分支。任何概率或动态分支均不得覆盖原始时间戳、真实覆盖率和原生观测掩码。对齐器后验不可得、后验校准审计失败或动态窗口有效长度不足时，只记录“未启用/不可估计”，不使用人为填充值冒充观测。

4.7 特征存储、全量汇总与读取规范

4.7.1 按样本存储与视图组织

采用每个样本一个 HDF5 文件的组织方式，文件路径由无歧义主键清单定位；等价的 NPZ 加变长偏移表也可使用，但不得丢弃变长内容。元数据另存为 `manifest_q1.jsonl`，每个文件包含以下层次。读取顺序为：通过清单主键定位样本文件，选择视图，读取特征、

表 4.3 特征文件的逻辑分组与读取内容

分组	主要内容	读取与核验作用
meta	主键、素材路径、标签、哈希、时长、流偏移、配置标识	判断样本和文件是否一一对应
native/T,A,V	原生特征、 $\mathcal{R}$ 、 J 、逐维掩码、质量及来源索引	保留未聚合数据，允许重建任意视图
views/word	词文本、时间、三模态聚合及非词区间类型	核验原词、声音和视频帧对应
views/sec	固定秒级均值、方差、覆盖率、计数及真实长度	主时序输入；读特征时同时读掩码
views/multi	三尺度特征及每个窗口真实区间	区分局部观测和邻域上下文
views/phase50	相对进程特征及对应区间	固定长度辅助展示，不替代秒级主轴
views/dynamics	选定通道、连续有效段、低阶签名及状态	描述窗口内变化顺序；不覆盖主特征
alignment	硬边界、词边界后验、占据概率、分位区间及后验投影权重	回查确定边界与概率边界，不混淆观察与估计
audit	工具版本、模型哈希、状态、错误码及处理日志	支持复现和异常追踪

时间区间、有效长度及掩码，再按来源索引回查原生行和原素材。时间戳使用足以保持精度的浮点表示；原始采样与帧索引使用整数；质量权重、稀疏矩阵的行列含义、坐标顺序和端点规则均写入模式说明。

4.7.2 变长序列、填充与来源关系

磁盘上的原生和主时序采用真实长度存储；组成批次时才填充到批内最大长度 $L_{\max}$ ，定义

$$P_{it}^{\text{sec}} = \mathbf{1}\{1 \leq t \leq L_i\}, \quad 1 \leq t \leq L_{\max}. \tag{4.39}$$

填充值在标准化后设为零，填充区间时间戳使用无效哨兵并配合 P 读取；不能以 $[0, 0)$ 冒充真实片段。真实时间内的模态缺失仍满足 $P = 1$ ，以 O 及原因字段区分。零占位不表示中性情感。对每个目标区间保存参与聚合的原生行索引及逐维权重；语音分支到联合声学行再到目标区间的两级投影须保留完整链路。多尺度重复使用同一观测只增加视图，不增加独立观测数量。

4.7.3 100 条样本与 300 条模态明细

实际实施后输出 100 行样本级清单和 300 行模态级明细。即使某模态提取失败，也保留相应记录与原因；样本覆盖完整不等于所有模态成功。不得用虚构编号或设计维度表替代真实提取结果。

表 4.4 全量模态明细表的字段规范（非实测结果表）

字段	定义
sample_id, video_id, clip_id	原样保留的唯一主键与来源索引
modality, source_duration	模态类型与原片段时长 D_i
observed_duration	可靠来源归属区间并集长度，另存逐维有效时长
native_length, native_dim	原生行数与实际主维度
aligned_length, aligned_dim	主视图真实长度与维度；不能填批次填充长度
alignment_type, granularity	固定秒级主对齐及辅助视图类型、实际时间粒度
coverage, missing_reason	连续覆盖率、自然未定义及失败原因分列
feature_path, source_hash	特征文件定位与素材校验值
status, config_hash	成功、部分成功或失败；完整配置的校验值

4.8 核验指标、典型样本与模型对照

4.8.1 样本、时间和对齐准确性

先要求标签、素材和输出主键集合相等且分别唯一，再计算覆盖指标。定义

$$C_{\text{id}} = \frac{\#\{\text{有且仅有一条输出记录的原始样本}\}}{100}, \quad (4.40)$$

$$C_{\text{valid},d}^m = \frac{\sum_{i,t} |I_{it}^{\text{sec}}| c_{itd}^m}{\sum_{i,t} |I_{it}^{\text{sec}}|}, \quad C_{\text{valid}}^m = \frac{1}{d_m} \sum_{d=1}^{d_m} C_{\text{valid},d}^m, \quad (4.41)$$

$$E_{\text{bnd}} = \frac{1}{2K_{\text{audit}}} \sum_{k \in \mathcal{K}_{\text{audit}}} (|\hat{s}_k - s_k^{\text{audit}}| + |\hat{e}_k - e_k^{\text{audit}}|). \quad (4.42)$$

覆盖率只在真实主网格求和，不含填充；逐维有效率的均值不替代各维状态明细。 $C_{\text{id}} = 1$ 仍需同时排查额外输出记录；不能仅凭这一比例断言主键集合完全一致。 E_{bnd} 仅在人工建立参考边界且 $K_{\text{audit}} > 0$ 时报告，否则注明未测，不输出虚构误差。所有有效区间须满足 $0 \leq s < e \leq D_i$ ，原生索引存在，主网格覆盖完整，相应归属区间在连续段内不重叠。强制对齐词序单调、概率边界只在合法状态图路径上取值、多尺度聚合满足式 (4.25)，均作为硬性审计条件。概率边界分支不以“更平滑”作为准确性证据，而在人工审计词集合 $\mathcal{K}_{\text{audit}}$ 上同时报告边界点误差与区间校准。设第 k 个词的 $(1 - \alpha)$ 后验边界区间分别为 $[L_k^s, U_k^s]$ 、 $[L_k^e, U_k^e]$ ，定义经验覆盖率

$$C_{\text{post}} = \frac{1}{2K_{\text{audit}}} \sum_{k \in \mathcal{K}_{\text{audit}}} [\mathbf{1}\{s_k^{\text{audit}} \in [L_k^s, U_k^s]\} + \mathbf{1}\{e_k^{\text{audit}} \in [L_k^e, U_k^e]\}]. \quad (4.43)$$

同时报告平均区间宽度

$$W_{\text{post}} = \frac{1}{2K_{\text{audit}}} \sum_{k \in \mathcal{K}_{\text{audit}}} [(U_k^s - L_k^s) + (U_k^e - L_k^e)]. \quad (4.44)$$

覆盖率和宽度必须联合呈现，避免用极宽区间换取表面上的高覆盖。若没有人工参考边界，则只报告模型内部后验宽度，不将其称为真实校准率。动态签名分支则通过与相同主特征基线的独立消融检验其增量价值，不以维度增加本身作为优势。

4.8.2 典型样本核验卡

按时长、对齐质量和缺失情况预先分层选样，至少展示一条真实样本；条件允许时再增加一条困难样本。每张核验卡应给出主键、英文原词、词边界、对应音频采样范围、原始帧时间戳与帧号、特征行编号、权重、覆盖率和异常原因，不只展示最易对齐的情况。本节给出的是模型和交付规范。原始素材未实际处理时，核验卡、100 行真实清单、实测边界误差及特征文件均不能宣称已经生成。

表 4.5 典型样本联合核验卡的填写规范（不含虚构样本）

核验项	必须填写的真实来源内容
样本与原文	双字段主键、原视频哈希、原转写片段与字符位置
文本时间	原词编号、 $[s_{ik}, e_{ik})$ 、边界质量和定位状态
音频对应	音轨偏移、原生采样率、采样点范围及分析窗口记录
视频对应	实际帧号和 PTS、目标轨迹编号，不以假定帧率换算
三模态特征	词级与秒级行索引、768/74/35 维定义及有效维度
对齐权重	物理重叠权重；启用后验分支时另列词占据概率、期望重叠质量与后验区间
核验材料	词边界叠加音频波形、代表帧及异常说明；人工边界另存

4.8.3 单因素对照而非复杂模块累加

采用相同原生特征、主时间网格和观测掩码，设置相互独立的对照：B0 为质量加权物理区间投影；B1 在 B0 上仅加入修正后的尺度处理与视觉几何一致聚合；B2 在 B1 上仅加入词边界后验概率投影；B3 在 B1 上仅加入低阶路径签名动态分支；B4 在 B1 上加入冻结深层语音辅助分支，用于区分“表征增强”与“时间模型增强”的贡献。各版本都比较主键完整性、时间边界误差、连续覆盖、来源可追溯性与人工核验；若后续用轻量预测器检验情感信息保留，则所有尺度、降维和调参只在训练折拟合，并按 `video_id` 分组。不预先宣称某一复杂版本预测性能最高。Soft-DTW、语义相似度驱动的全局重排和公共空间学习不进入默认问题一链路，以免将“表达差异”误当成“时间错误”。若另作研究，只能作为独立对照，不得覆盖问题一保存的物理特征及来源记录。

4.9 参数锁定、运行流程与复现说明

表 4.6 中未给出数值的后验分位水平、动态窗口等参数，应在实施前写入完整配置，说明候选范围和无标签的时间核验选择规则；不能将“待锁定”当作已经复现。工具确切版本与模型哈希应从真实运行环境写入，而不是在未运行时虚构。软件环境不同可能引入数值差异，须预先声明逐位一致或容差一致的复现目标。

4.9.1 全量可追溯特征生成流程

输入：原始视频、转写、标签表和已锁定的配置。

输出：原生层、固定秒级主视图、词级及辅助视图、掩码、来源映射、真实全量清单和日志。

1. 核验双字段主键、文件集合、原标签逻辑及哈希；保留所有失败记录。
2. 对每个样本解析统一时间原点，解码音轨、原始帧和真实时间戳，登记偏移及缺口。
3. 按式 (4.9)、(4.11)、(4.12) 提取主特征，保存逐维有效性、质量和计算素材范围。

表 4.6 关键配置与实际复现时必须锁定的信息

模块	本稿固定或建议配置	实际执行必须记录
媒体与时间	单声道 16 kHz; 原始 PTS 和流偏移	FFmpeg/ffprobe 构建、命令、解码及重采样映射
文本	冻结 BERT; 末四层; 510 子词窗口、384 步长	权重及词典修订号、哈希、分词器与推理依赖版本
谱与倒谱	400 点窗、160 点步、512 点 FFT; 40+13+13 维	窗函数、Mel 映射、DCT、差分、边界及幅度规则
韵律	8 项; 逐分量原生窗口与掩码	估计器名称版本、基频范围、阈值、单位和时间锚
视觉	17+6+6+6 维; 门槛初值 0.8	OpenFace 构建、列序、几何编号、置信度及轨迹规则
词级锚定	固定转写状态图、单调路径	MFA 版本、声学模型和发音词典哈希、异常词处理
时序视图	主步长 0.1 s; 尺度 0.1/0.3/0.7 s; 辅助 50 段	实际窗口端点、归属区间、完整来源权重及长度
边界后验	仅在固定转写状态图可得可靠前向-后向权重时启用	对齐器后验导出方法、分位水平、状态映射、采样数及人工校准结果
动态签名	时间增广、最高二阶; 仅连续有效段计算	通道表、插值规则、最小有效长度、窗口尺度、签名实现与维度
环境与输出	固定随机源; 原生层和派生层分离	操作系统、依赖锁、模型哈希、随机种子、硬件及输出校验值

- 在固定转写状态图上求解式 (4.14), 记录可靠词边界、非词区间与不可定位词; 若可获得前向-后向后验, 则进一步计算式 (4.17)–(4.19), 不以标签修正。
- 建立连续段内不重叠的时间归属区间, 按式 (4.21) 生成物理聚合, 同时计算方差、连续覆盖、观测数量以及式 (4.26) 的充分统计量。
- 根据式 (4.28)、(4.30) 和(4.31) 构造主视图与辅助视图; 对非嵌套目标边界回查原生来源, 不对中间均值二次切分。
- 若词边界后验可用, 生成概率时间视图并保存后验区间与校准字段; 在连续有效段上按式 (4.34)–(4.37) 生成低阶动态签名辅助分支。
- 检查维度、数值、区间、索引、掩码、覆盖与来源链路; 按合法训练规则生成可选标准化层, 批处理时另行填充。
- 写入特征文件、模式说明和日志; 生成真实 100 行样本汇总与 300 行模态明细, 并按质量分层完成典型样本联合回看。

5 模型评价

这里是模型评价

附录 A MATLAB 源程序

A.1 第 1 问程序

code.m

附录 B Python 源程序

B.1 第 2 问程序

mip1.py
