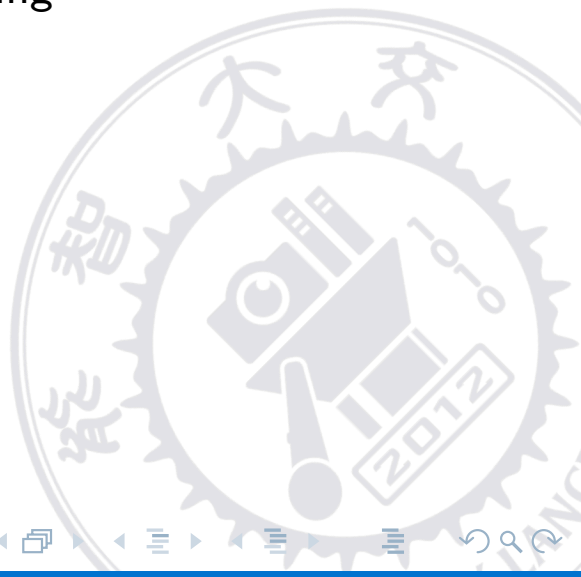
# Project 1: 语音端点检测

## Voice Activity Detection

Kai Yu and Yanmin Qian

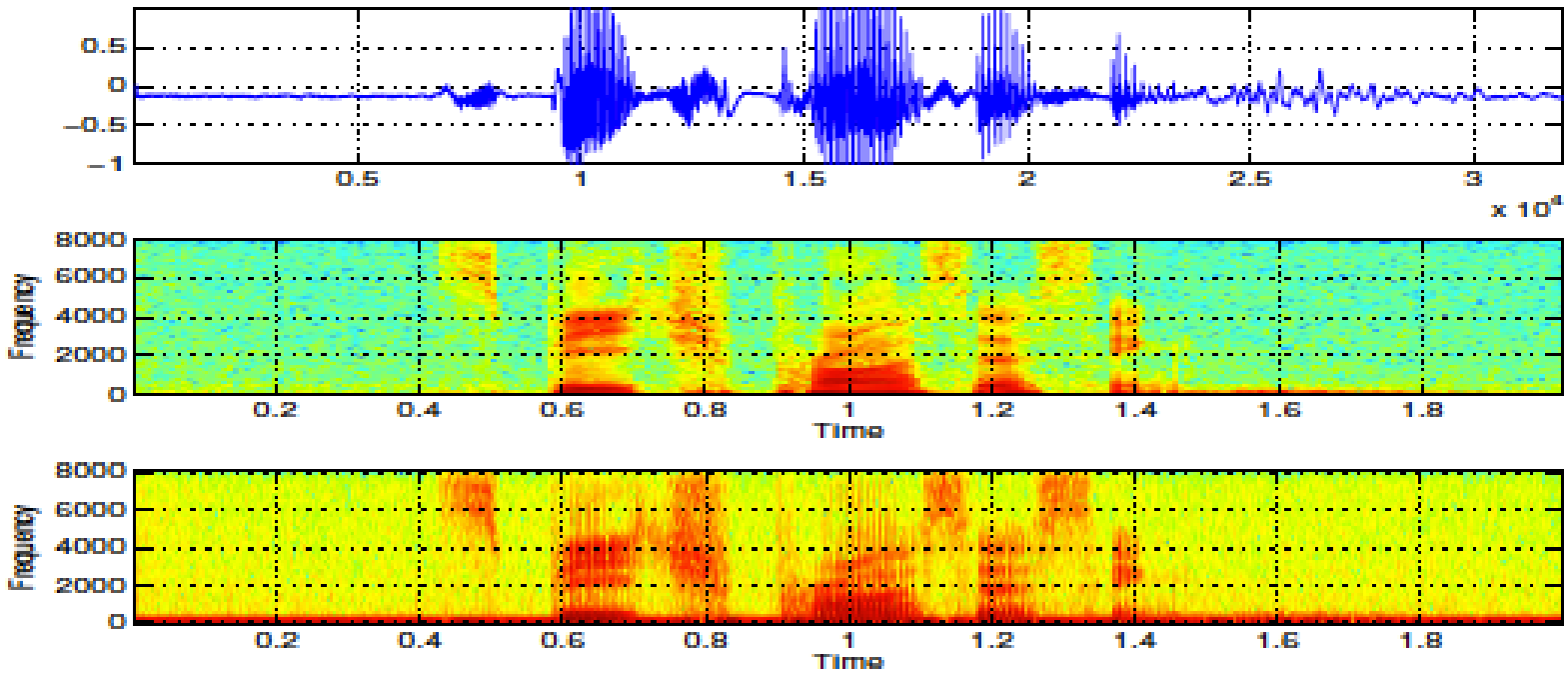
Cross Media Language Intelligence Lab (X-LANCE)
Department of Computer Science & Engineering
Shanghai Jiao Tong University

Spring 2024



# 任务背景

语音端点检测（voice activity detection，VAD），即对语音的每一帧判断其是否属于语音段（speech）或者静音段（silence）。这是语音信号处理中非常常见的预处理操作，旨在帮助理解语音信号基础知识和短时特征以及机器学习、统计学习基本方法。

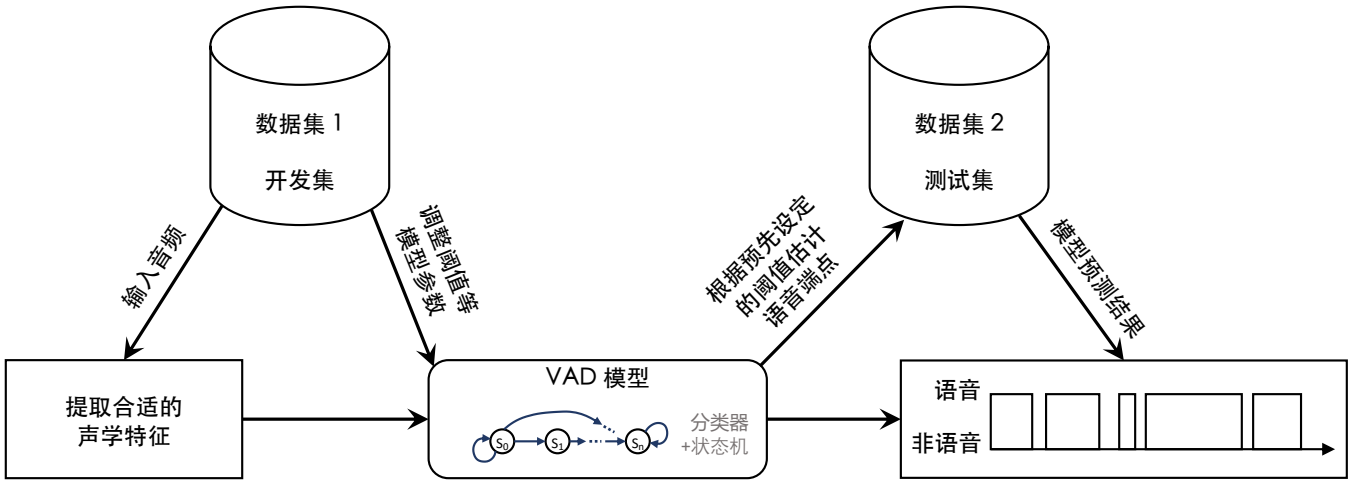


# 任务要求

1. 基于线性分类器和语音短时信号特征的简单语音端点检测算法
   - 利用语音的短时信号特征（短时能量，过零率，短时频谱，以及基频等)
   - 简单线性分类器的使用（如：阈值分类器）
2. 基于统计模型分类器和语音频域特征的语音端点检测算法
   - 利用语音的频域特征（MFCC, PLP, FBank 等）
   - 统计模型分类器的使用（如 GMM, DNN）

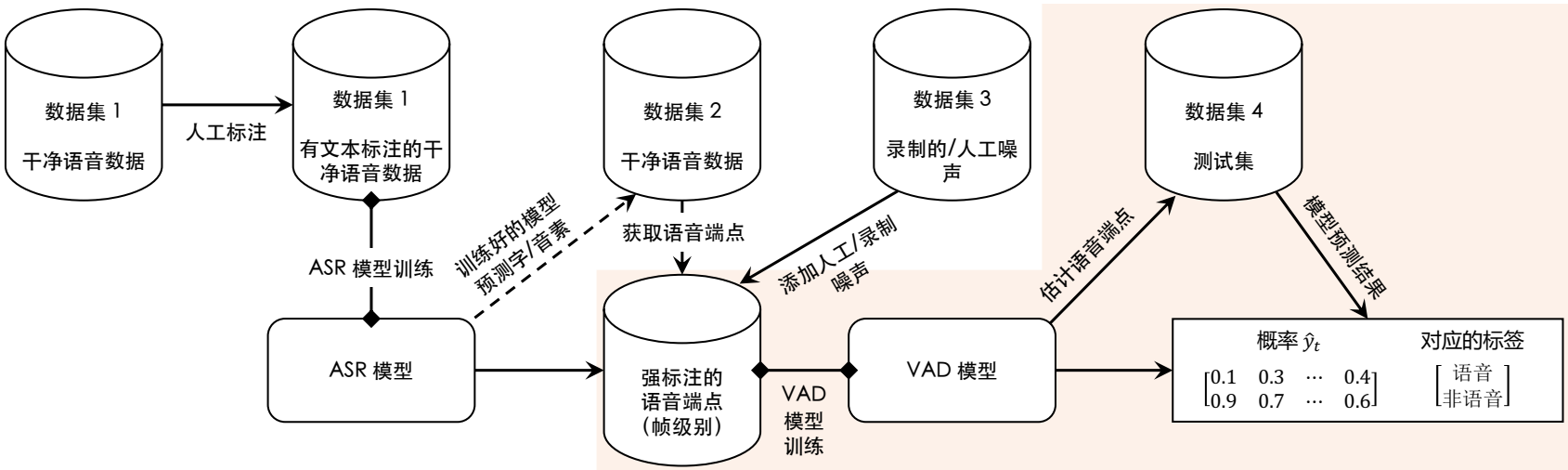
# 任务框图-1

任务 1 的主要内容如图所示：



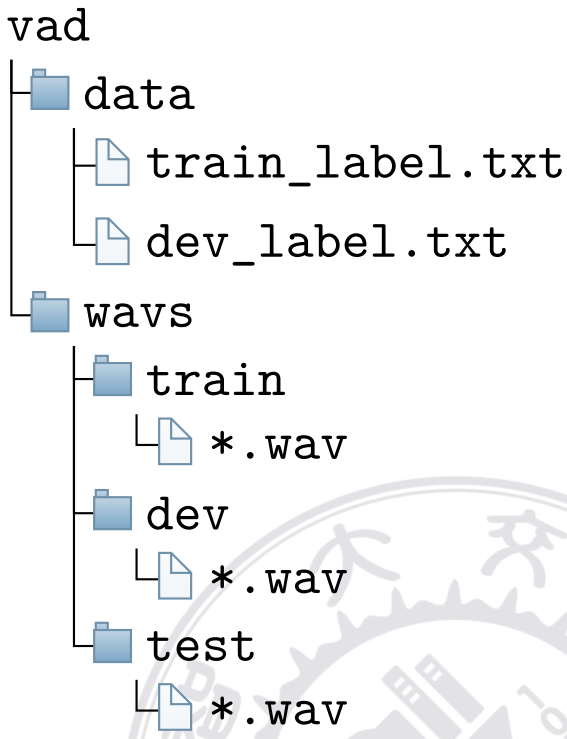
# 任务框图-2

传统有监督 VAD 模型的训练流程如下，任务 2 的主要内容如高亮部分所示：



# 数据介绍

- 数据集（1.7GB）：
  - 训练集 (train)：3600 条音频，长度为 10–20 秒，有语音段时间标注
    - 仅用于任务 2的模型训练
  - 开发集 (dev)：500 条音频，长度为 10–18 秒，有语音段时间标注
  - 测试集 (test)：1000 条音频，长度为 10–18 秒
  - 所有音频数据的采样率均为 16 kHz
  - 音频中可能存在背景噪声



# 数据格式介绍

vad/data/目录下的两个 txt 文件分别为训练集和开发集的标签，标注了每条音频的语音段起止时间戳。

如 train_label.txt 的第一行：

```
100-122655-0035 0.14,1.79 1.82,2.88 ... 11.43,13.72
```

- 标签文件的每一行有多列，用空格分隔
- 第一列为每条音频的唯一 ID，与音频文件名相同
- 后面每一列表示该音频中所有语音段的起止时间（秒），格式为 “X,Y”，其中英文逗号左边的数字 X 表示该段语音的起始时间，右边的数字 Y 代表该段语音的截止时间
- 数据集压缩包中也提供了标签文件格式转换的 Python 脚本，仅供参考

# 模型评估指标

Accuracy: $\mathrm{Acc} = \dfrac{\text{所有样本中预测正确的总帧数}}{\text{所有样本的总帧数}}$

Receiver Operating Characteristic curve: ROC

area-under-ROC-curve: $\mathrm{AUC} = \mathrm{ROC}$曲线下的面积

equal error rate: $\mathrm{EER} = \mathrm{ROC}$曲线上$\mathrm{FNR}$与$\mathrm{FPR}$相同时的取值

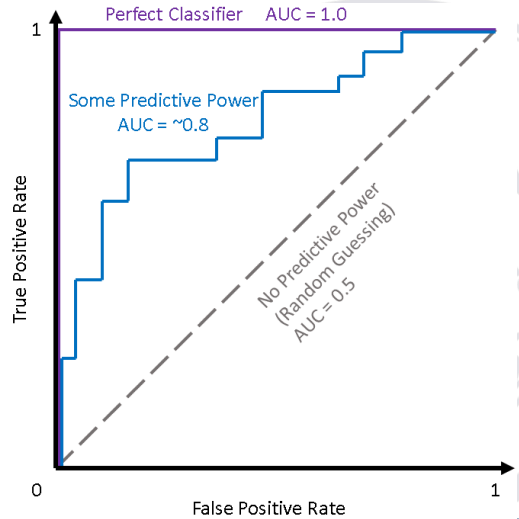
**以上指标的相关计算脚本（Python）会和数据集一同发布。**

此外也建议大家参考这篇关于 VAD 评估指标的论文：EVALUATING VAD FOR AUTOMATIC SPEECH RECOGNITION

召回率 True Positive Rate (TPR) $= \dfrac{\text{所有样本中预测正确的语音段总帧数}}{\text{所有样本中实际标注为语音段的总帧数}}$

误报率 False Positive Rate (FPR) $= \dfrac{\text{所有样本中预测错误的非语音段总帧数}}{\text{所有样本中实际标注为非语音段的总帧数}}$

ROC 曲线：所有可能的阈值情况下的 (FPR, TPR) 构成的曲线



# 提交要求

- 在测试集数据上生成标注文件，每个任务各一个，格式应与前面的介绍一致
- 报告（中文）采用 LaTeX 编写，提交 PDF 格式
  - 模板下载地址：
    `https://latex.sjtu.edu.cn/read/vckssbvjpfvg`
  - 只需要提交一份报告，包含 task1 和 task2（“学号-姓名.pdf”）
  - 需给出开发集上的测试结果，并按照上一页的要求计算相关指标
- 报告、代码和测试集结果文件一起打包提交，压缩包命名格式为“Project1-学号-姓名.zip”，压缩文件里面包含报告和两个文件夹“task1”，“task2”，每个文件夹包含代码和 test_label.txt

# 提交要求

- **提交方式**：`Canvas`
- **预计截止时间**：2024 年 5 月 5 日
- **作业分值**：本次 project 占课程总分值 25%
  - VAD 性能评估预计占 10%，其中两个子任务的性能各 5%。
  - 提交报告（含代码）预计占 15%，其中两个子任务各占 5% 和 10%。