

第三周实验周报

一、项目名称

基于时域分析技术的语音识别系统 项目成员（2305班）：周湛昊、张振鑫、孙鑫磊、王毅

二、当前项目阶段

- 语音信号处理系统功能开发已完成
- 经过消融实验和对比实验，SVM被认为具有明显优势
- 实现从音频录制和数字识别的完整流程
- 对比了三种窗函数的效果
- 集成了多种机器学习算法用于分类，支持多种分类器算法，便于性能对比
- 提供了直观的图形界面和详细文档，模块化设计便于后续功能扩展

三、本周工作完成情况

周湛昊

- 负责整体项目架构设计和代码的编写
- 使用Ableton Live完成音频数据预处理
- 完成实际人声部分的切分
- 实现实时音频录制功能
- 完成Qt图形界面开发
- 围绕实时音频录制稳定性修复
- Qt图形界面交互简化与可视化功能（如音频波形预览）添加
- 各模块衔接调试以减少数据延迟展开优化，为后续功能拓展打基础

张振鑫

- 实现分帧加窗处理
- 完成三种窗函数结果对比功能开发
- 集成时域特征分析算法（短时能量、平均幅度、过零率）
- 协助调试端点检测可视化模块中的时间轴对齐问题
- 检查分帧、窗函数处理与端点检测模块之间的时序一致性
- 优化特征提取模块的稳定性和输出格式，确保提取的时域特征能够准确传递给可视化模块

孙鑫磊

- 优化代码结构，提升系统稳定性和用户体验
- 完善项目文档和用户手册
- 实现自动语音段分割和提取
- 完成相关测试和参数调优
- 围绕端点检测可视化功能包的完善、验证与落地
- 完善功能包鲁棒性，处理异常输入（如空音频、非法时间戳等）
- 补全基础功能
- 优化可视化效果（如颜色区分、语音段时长标注）

- 扩展功能（支持频谱叠加、多格式输出）
- 对长音频绘图性能进行优化

王毅

- 实现端点检测算法
- 集成多种分类算法（SVM、KNN、朴素贝叶斯、Fisher线性判别）
- 利用Python工具库对纯净及带噪语音样本进行实际的频谱分析
- 通过对比不同参数下的频谱图、梅尔频谱等特征提取效果
- 直观评估并确定最适合后续识别模型的最佳变换配置方案

四、技术成果总结

核心功能实现

- **音频处理**：完成WAV文件读取、分帧加窗处理
- **特征提取**：实现时域特征计算（短时能量、平均幅度、过零率）
- **端点检测**：基于双门限法的语音段检测
- **分类识别**：集成多种机器学习算法进行数字识别

系统架构

- **模块化设计**：清晰的代码组织和接口设计
- **用户界面**：Qt图形界面和命令行界面
- **实时处理**：支持实时音频录制和分析

实验验证

- **算法对比**：完成多种分类器性能对比，SVM表现最佳
- **特征分析**：通过消融实验验证特征重要性
- **系统测试**：验证整体性能和稳定性

五、项目中遇到的问题

1. **音频录制与界面衔接问题**
 - 音频录制与界面衔接时出现数据传输延迟
 - 影响实时音频处理的流畅性
2. **窗函数对比问题**
 - 窗函数对比中存在特征差异不明显
 - 不同窗函数的效果区分度不够
3. **算法效率问题**
 - 时域算法集成后计算效率偏低
 - 影响系统整体性能
4. **语音分割准确率问题**

- 语音自动分割提取准确率不足
- 影响后续识别效果

5. 端点检测问题

- 端点检测算法对弱声信号识别不准
- 影响语音段边界的准确检测

六、代码部分

- 代码仓库地址：<https://github.com/ZeroHour-Z/SpeechRecognition>
- 建议：代码详情可见 readme 文件，代码结构可见 PROJECT_STRUCTURE.md 文件