# 音频分析及处理作业报告

郑阳 2013013296

王旻晨 2013013302

王沁煜 2013013303

1. **实验目标**

对一段以空白分割的录音，实现音频的自动分段、话者人数估计、话者识别、音频转文本等功能。

1. **运行环境**

Windows 10

Matlab R2015b

Java 1.8

1. **实现方法**
2. **音频分段**

使用基于静音的方法进行分段。将音频中的每一秒分为100帧，计算每一帧的能量，能量小于一定值的帧标记为静音帧，连续的静音帧即静音段。对静音段取补集即语音段。

1. **特征提取**

首先提取各语音段的mfcc特征。对每段音频进行分帧，取帧长为256、步长为128，对每帧的数据提取一个12维的mfcc特征向量。计算各向量的一阶差分并和原有的向量组合得到24维的特征向量。

然后对特征向量进行聚类，通过各个帧的特征得到各段音频的特征。将该段音频的所有提取的帧聚集，在每一维上求均值，并对质心在每一维度上乘1+e、1-e，得到两个质心，然后重新斤算各个点到质心的距离并调整质心，重复上述工作直到质心数量达到64。

1. **语音转文字**

我们调用了百度语音的REST API进行语音识别。调用百度接口可上传一段采样率为8000、单声道的wav文件，并接收到一段json文本。

此外，我们还尝试了多种语音转文字引擎。IBM Watson Speech To Text是我们发现的转换质量相对较高、支持几大主要语言的一个比较好的选择。我们尝试写一个python工程获得转换的文本，但遇到了一定的困难。我们还尝试了Google Cloud Platform，但因没有银行卡卡在了注册阶段。我们成功尝试了著名的Microsoft Speech API（也即SAPI），但由于其转换质量较差放弃。经过比较，我们选择了百度语音的API。

我们制作了一个简易网页用于显示结果。打开该网页，输入上述步骤所获得的音频分段数量、话者顺序字符串，网页将通过jsonp技术读取解析json文本，并显示在网页上。

1. **实验结果**

在测试数据文件夹下保存了测试音频和测试结果网页。

1. **运行方法**

首先，将音频放置在AudioRecognition\code\audio\_data\input.wav，使用Matlab运行AudioRecognition\code\main.m，Matlab将会把切割完毕的音频端放置于AudioRecognition\code\segmentation文件夹下，同时将话者对应编号输出至AudioRecognition\code\recognition\result.txt文件。（备注：由于百度音频识别仅能识别码率8000的音频，故有可能需要将步骤得到的音频手动转码）

然后，运行SpeechRecognition\src\test\Speech2Text.java，将num变量修改为音频分段数量，程序将会把音频通过百度API进行识别，将各音频对应的转换文字json保存在根目录下的txt文件。

最后，打开showResult\showResult.html，将SpeechRecognition\文件夹下的txt文件复制至showResult\文件夹下，并在第一行输入分段数量，第二行输入以空格分割的AudioRecognition\code\recognition\result.txt文件内容。

1. **作业中使用的库**

Matlab中的相关库

百度语音REST API