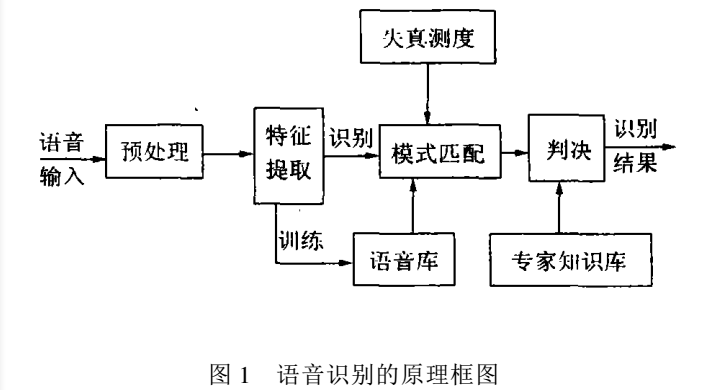
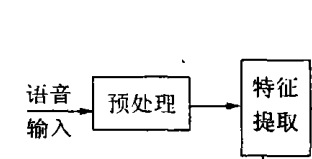
周报

在这一周的时间里，主要完成的任务是观看莫烦的教学视频，安装配置python，阅读了《基于声音信号的特征提取方法的研究》\_李宏松、《基于Matlab的声音信号采集与分析处理》\_陈家焱、《基于特征分析的环境声音事件识别算法》\_刘波霞这三篇与声音信号特征提取相关的论文。

语音识别：[语音识别技术](https://baike.baidu.com/item/%E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB%E6%8A%80%E6%9C%AF" \t "https://baike.baidu.com/item/%E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB/_blank)就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高技术，主要包括特征提取技术、模式匹配准则及模型训练技术三个方面。我们可以借鉴语音识别中的特征提取技术对我们所采集到的声音进行特征提取，然后将其特征作为已经完成训练的模型的输入进行分类。在《基于声音信号的特征提取方法的研究》中提到了语音识别的一个根本问题是合理的选用特征，其特征向量一般分为两类：第一类为时域特征向量，通常将帧语音信号的各个时域采样值直接构成一个向量，其优点在于计算简单，缺点在于不能压缩维数且不适于表征幅度谱特性；第二类为变换域特征向量，即对一帧语音信号进行某种变换后产生的特征向量，其缺点是计算复杂，优点是能从不同的角度反映幅度谱的特征。同样，对于我们所进行的声学场景分类而言，合理的特征提取是非常关键的。



上图是语音识别的原理图，我们可以借鉴其特征提取的部分，即：



声音信号的采集：（所查找的论文都是基于matlab实现的）

1. 由Windows录音机预先录制的声音文件（\*.wav），再由matlab语言调入。其调入语句为：wavread(‘d:\你好.wav)

此语句的功能是调入保存在d盘根目录下的‘你好’语音文件，并将其转化为matlab数据类型，然后对其进行分析。

2、由matlab语言直接通过windows的音频设备实现录音功能，并将采样结果转换为MATLAB数据类型。其录音语句为:

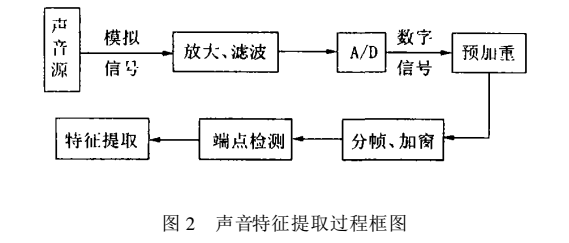
Fs=11025

X=wavread(5\*Fs,Fs,’double’)

此代码是通过windows的录音设备用单声道，11025Hz的采样频率录制5s的数据，采样精度为16位，存储格式为双精度型，通过以上语句调入或录制后，就可以在matlab语言中进行下一步的数据处理工作。

预处理分析：

语音信号特征的提取如图：在特征提取之前为声音信号预处理过程。



1. 放大、滤波、A/D转换

采样Windows自带的音频设备及MATLAB语言即可实现，其采用频率自由设定，精度为16bit。

PC机的声卡可以完成放大和滤波的反混叠过程，并可以完成A/D转换。

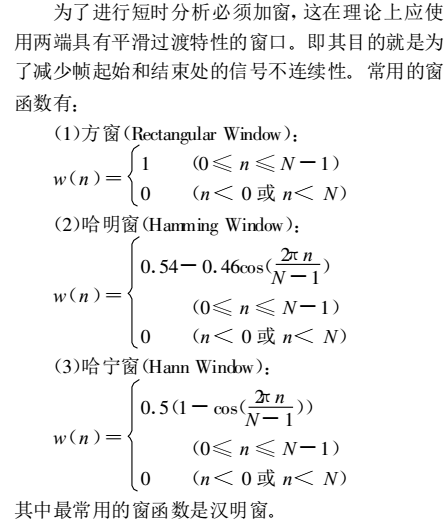
2、预加重

预加重是一阶网络，用来对声音的高频部分进行加重，以增强声音的高频分辨率

3、分帧

虽然声音信号是非线性时变信号，但它同时具有短时平稳的特点，所以对其分帧可以提取其短时特性。在处理语音信号时一般取10~30ms，为了避免帧与帧之间的特性变化太大，所以帧移一般取帧长的1/3或1/2.

4、加窗

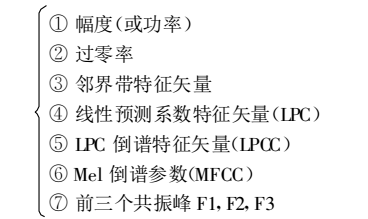


1. 端点检测

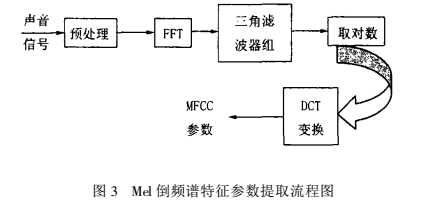
端点检测即在实时输入的声音信号中，区分背景噪声和环境噪声，准确判断出声音信号的开始点和结束点。

特征提取：

代表性特征：



着重于梅尔倒谱系数：



由于语音信号在时域上的变化快速而不稳定，所以通常将它转换到频域观察。此时他的频谱会随着时间做缓慢变化，所以通常将加窗后的帧经过快速傅里叶变换（FFT），求出每帧的频谱参数，再将每组的频谱参数通过一组N个（N一般为20—30个）三角形带通滤波器所组成的梅尔频率滤波器，将每个频带的输出取对数，求出每一个输出的对数能量Ek，再将此N个参数进行余弦变换，求出L阶的Mel-scale cepstrum参数。

余弦转换公式如下：