

语音信息处理实验一

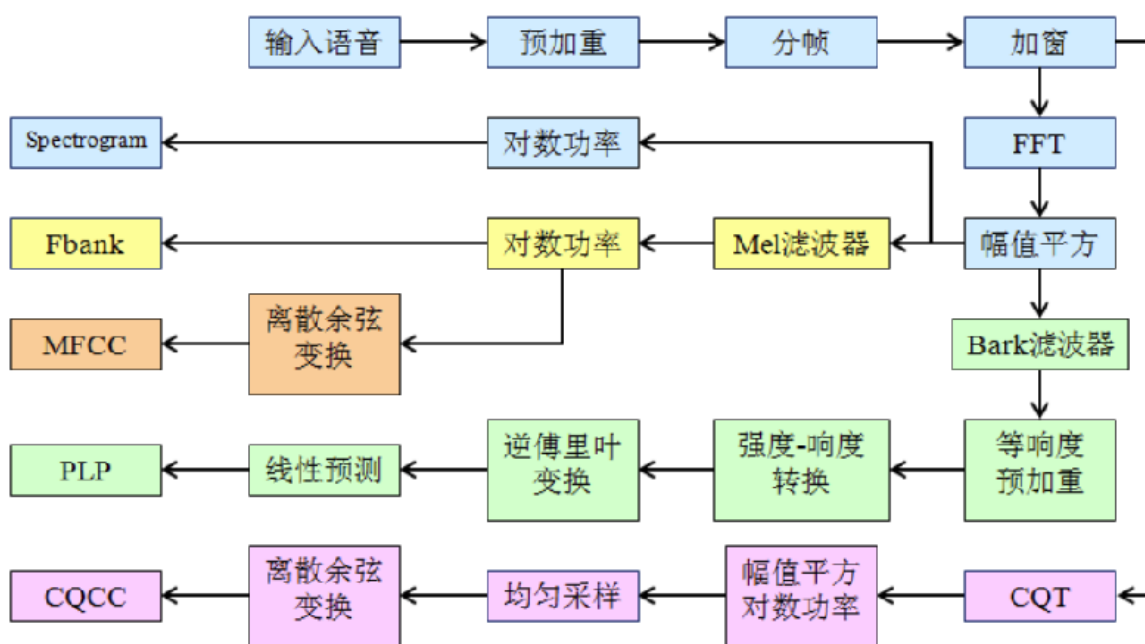
姓名：高祎珂

学号：2011743

专业：计算机科学与技术

实验描述

- 给定一段语音信号（16KHZ Wav PCM），这里是已经给出的我爱南开.wav语音，提取80维Log Mel Spectrogram（Fbank）特征，并画图。
- 根据图示给出的步骤进行特征提取



实验过程

在这里进行处理的核心代码主要是基于给出的basic_operator文件中所提供的函数。这些函数按照下图中的方法处理原始音频信号，从而提取出与语音相关的特征。在使用这些函数时，可以通过调用不同的函数组合来实现对特定语音任务的定制化处理。

其中我利用给出的函数主要实现了Fbank，Spectrogram，MFCC特征的提取，并进行了可视化。

实验代码

```
import basic_operator as bo
import wave
import scipy.io.wavfile
# f = wave.open('./wav/我爱南开.wav', 'rb')
# print(type(f))
sample_rate, signal = scipy.io.wavfile.read('./wav/我爱南开.wav')
# print(type(signal))
bo.plot_freq(signal, sample_rate, '频域图')
```

```

bo.plot_time(signal,sample_rate,'时域图')
signal = bo.pre_emphasis(signal) # 预加重
signal = bo.framing(signal,sample_rate) # 分帧
signal = bo.add_window(signal, sample_rate) # 加窗
# signal = bo.my_fft(signal) # FFT
signal = bo.stft(signal) # FFT+幅值平方
signal = bo.mel_filter(signal,sample_rate) # Mel滤波器
signal = bo.log_pow(signal) # 对数功率
# bo.plot_spectrogram(signal,"spec","spec")
bo.plot_spectrogram(signal,"fbank","fbank")
signal=bo.discrete_cosine_transform(signal)
bo.plot_spectrogram(signal,"mfcc","mfcc")

```

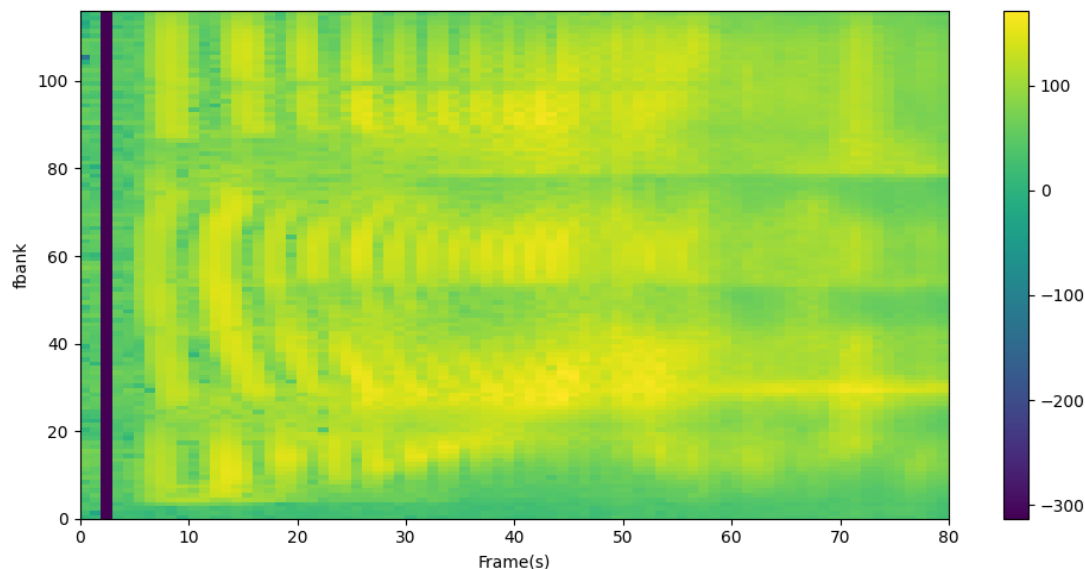
实验结果

FBrank

```

signal = bo.pre_emphasis(signal) # 预加重
signal = bo.framing(signal,sample_rate) # 分帧
signal = bo.add_window(signal, sample_rate) # 加窗
signal = bo.stft(signal) # FFT+幅值平方
signal = bo.mel_filter(signal,sample_rate) # Mel滤波器
signal = bo.log_pow(signal) # 对数功率
bo.plot_spectrogram(signal,"fbank","fbank")

```

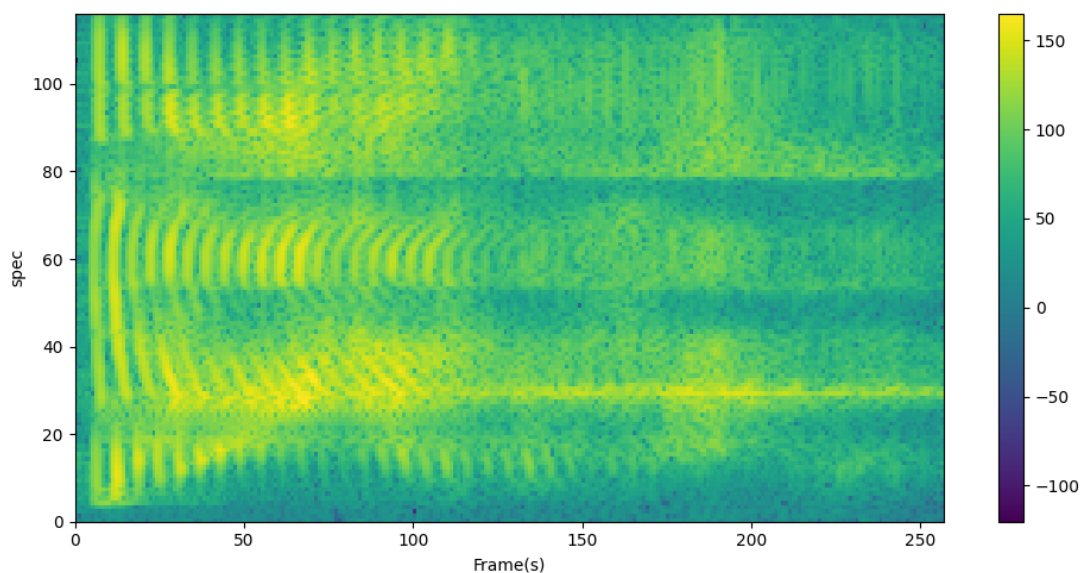


spectrogram

```

signal = bo.pre_emphasis(signal) # 预加重
signal = bo.framing(signal,sample_rate) # 分帧
signal = bo.add_window(signal, sample_rate) # 加窗
signal = bo.stft(signal) # FFT+幅值平方
signal = bo.log_pow(signal) # 对数功率
bo.plot_spectrogram(signal,"spec","spec")

```

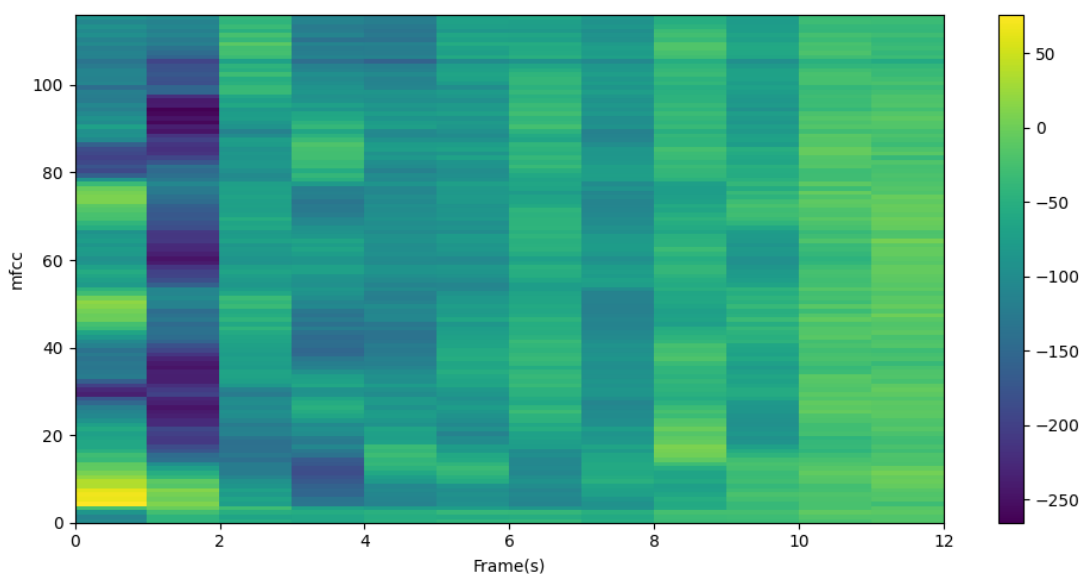


MFCC

```

signal = bo.pre_emphasis(signal) # 预加重
signal = bo.framing(signal,sample_rate) # 分帧
signal = bo.add_window(signal, sample_rate) # 加窗
signal = bo.stft(signal) # FFT+幅值平方
signal = bo.mel_filter(signal,sample_rate) # Mel滤波器
signal = bo.log_pow(signal) # 对数功率
signal=bo.discrete_cosine_transform(signal) #离散余弦变换
bo.plot_spectrogram(signal,"mfcc","mfcc")

```



遇到的问题及解决方式

- 不知道如何读取语音文件

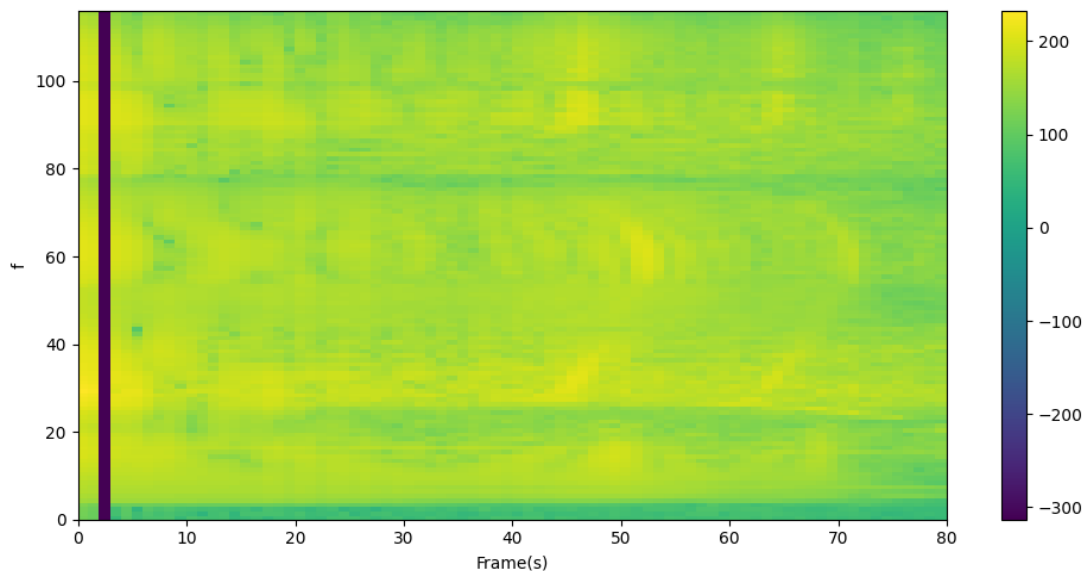
由于是调用已经给定的函数进行语音处理，所以有给定的函数参数限制，起初并不知道如何把.wav文件以数组形式读取进来，后来通过查阅资料，找到了读取方式。

- 生成图像有误

第一次还没有认真研读代码，直接按照图像进行函数调用，生成Fbank，代码具体为

```
import basic_operator as bo
import wave
import scipy.io.wavfile
sample_rate, signal = scipy.io.wavfile.read('./wav/我爱南开.wav')
signal = bo.pre_emphasis(signal) # 预加重
signal = bo.framing(signal, sample_rate) # 分帧
signal = bo.add_window(signal, sample_rate) # 加窗
signal = bo.my_fft(signal) # FFT
signal = bo.stft(signal) # FFT+幅值平方
signal = bo.mel_filter(signal, sample_rate) # Mel滤波器
signal = bo.log_pow(signal) # 对数功率
bo.plot_spectrogram(signal, "fbank", "fbank")
```

导致生成图像如下，后来发现是stft函数其实已经包含了my_fft函数功能，相当于多做了一次傅里叶变换，把这个冗余调用删除就得到了正确的图像。



参考资料

[语音识别第4讲：语音特征参数MFCC - 知乎 \(zhihu.com\)](#)

[\(超详细\) 语音信号处理之特征提取 语音特征提取 BarbaraChow 的博客 - CSDN 博客](#)

[python做语音信号处理 python 音频处理 凌逆战的 博客 - CSDN 博客](#)