

语音信息处理实验lab1

姓名：雷贺奥

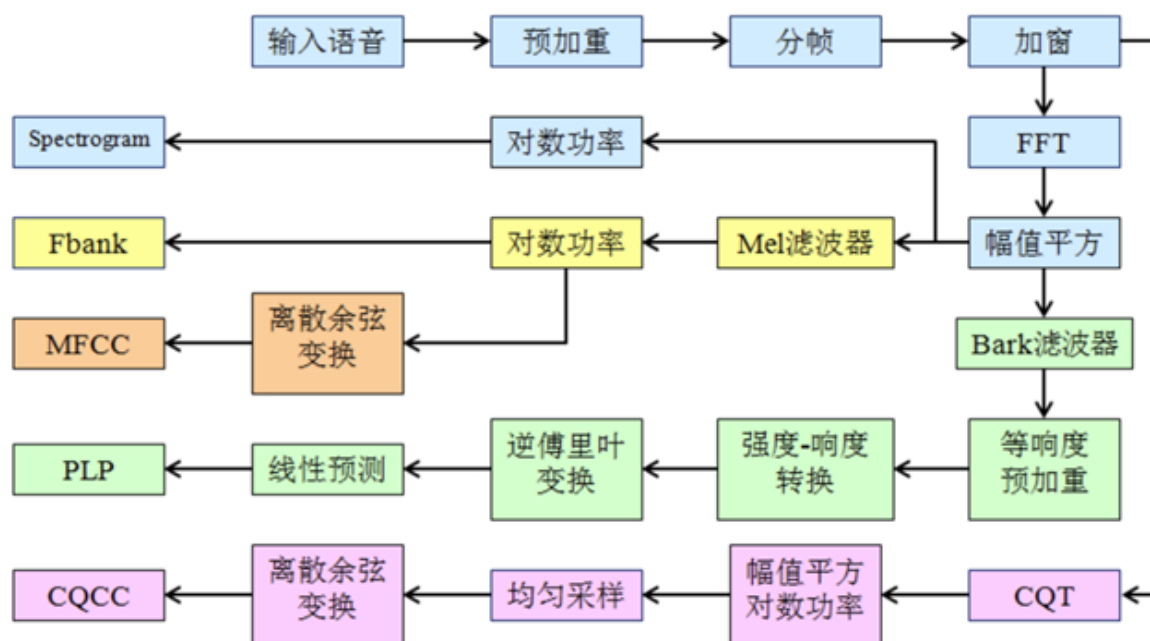
学号：2013551

专业：计算机科学与技术

实验描述

1. 给定一段语音信号（16KHZ Wav PCM），提取80维Log Mel Spectrogram（Fbank）特征，并画图。
2. 根据上课内容和代码实践回答雨课堂中的问题。
3. 抽取spectrogram特征, 并可视化
4. 抽取MFCC特征, 并可视化
5. 抽取PLP特征, 并可视化

实验过程



实验代码

• Fbank特征

```
#读取音频文件
import soundfile as sf
audio, fs = sf.read("wav/我爱南开.wav")
```

```

print(fs)
print(type(audio))
plot_time(audio, fs, "原图像时域图")
plot_freq(audio, fs, "原图像频域图")

#进行语音预加重量，加重高频
audio=pre_emphasis(audio)
#分帧
frame_sig=framing(audio, fs)
#Todo:: frame_sig为一个二维数组
plot_spectrogram(frame_sig.T, "维度", "分帧二维数组")
#加窗
frame_sig=add_window(frame_sig, fs)
#fft+幅值平方
frame_pow=stft(frame_sig)
#Mel滤波器
filter_banks=mel_filter(frame_pow, fs)
#对数功率
filter_banks=log_pow(filter_banks)
plot_spectrogram(filter_banks.T, "Dimension", "Fbank")

```

- spectrogram特征

```

import soundfile as sf
audio, fs = sf.read("wav/我爱南开.wav")
print(fs)
print(type(audio))
#进行语音预加重量，加重高频
audio=pre_emphasis(audio)
#分帧
frame_sig=framing(audio, fs)
#Todo:: frame_sig为一个二维数组
plot_spectrogram(frame_sig.T, "Dimension", "分帧二维数组")
#加窗
frame_sig=add_window(frame_sig, fs)
#fft+幅值平方
frame_pow=stft(frame_sig)
#spectrogram特征
spectrogram=log_pow(frame_pow)
plot_spectrogram(spectrogram.T, "Dimension", "spectrogram")

```

- MFCC特征

```

# 只需要在Fbank的基础上使用离散余弦变换即可
#MFCC
mfcc=discrete_cosine_transform(filter_banks)
plot_spectrogram(mfcc.T, "Dimension", "MFCC")

```

- PLP特征

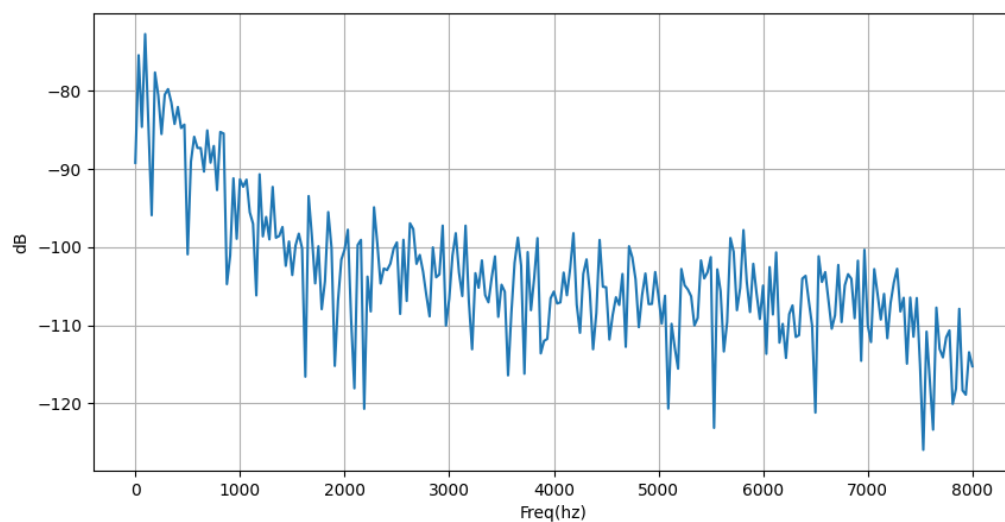
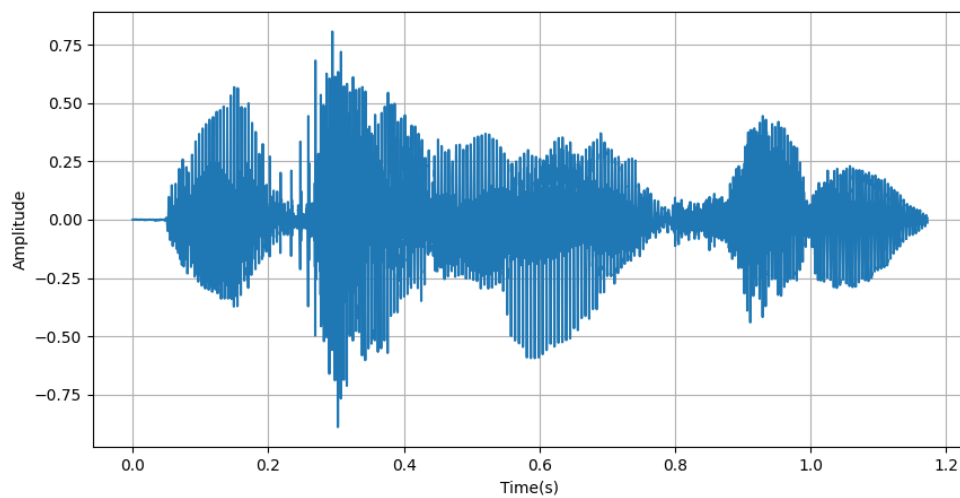
```

#PLP，调用spafe标准库中的plp函数即可
from spafe.features.rplp import plp
plp_result=plp(audio, fs,)
plot_spectrogram(plp_result.T, "Dimension", "PLP")

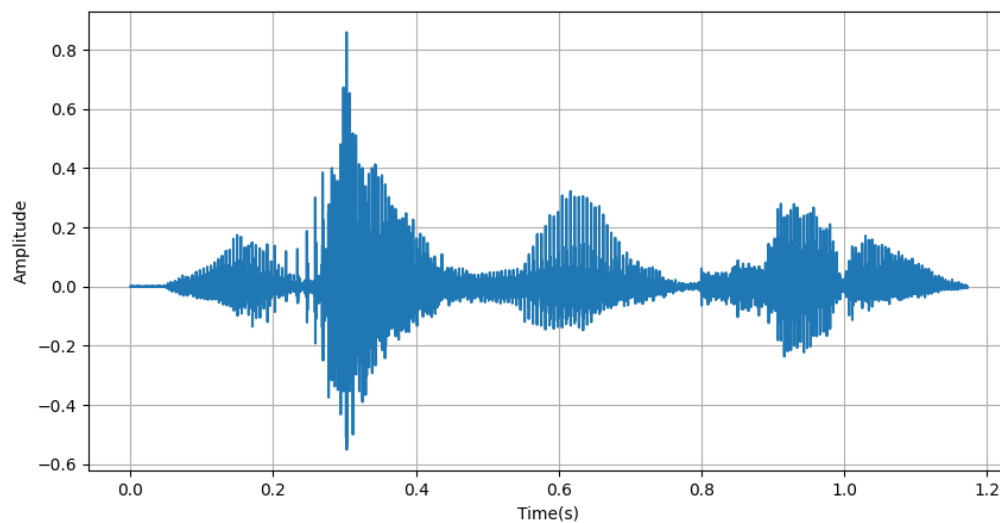
```

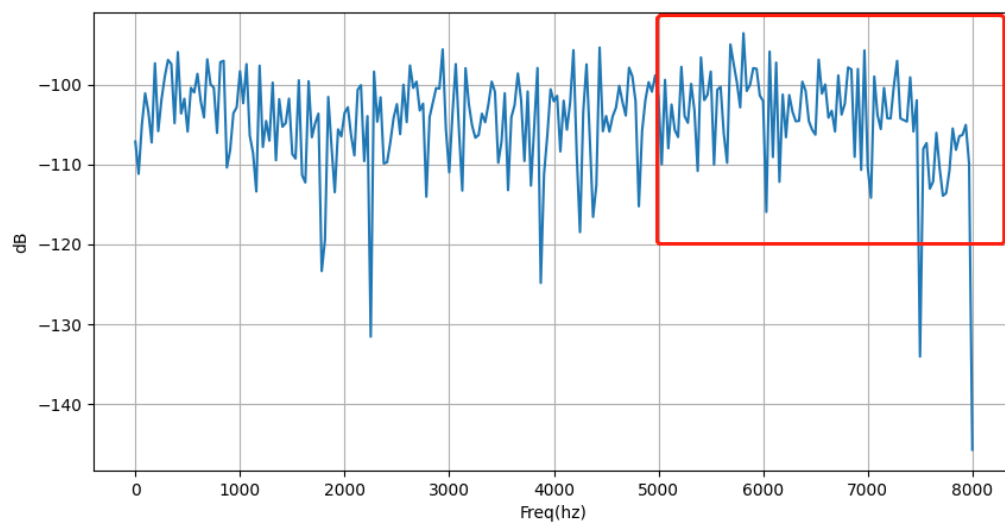
实验结果

- 原文件的时域图、频率图如下所示：

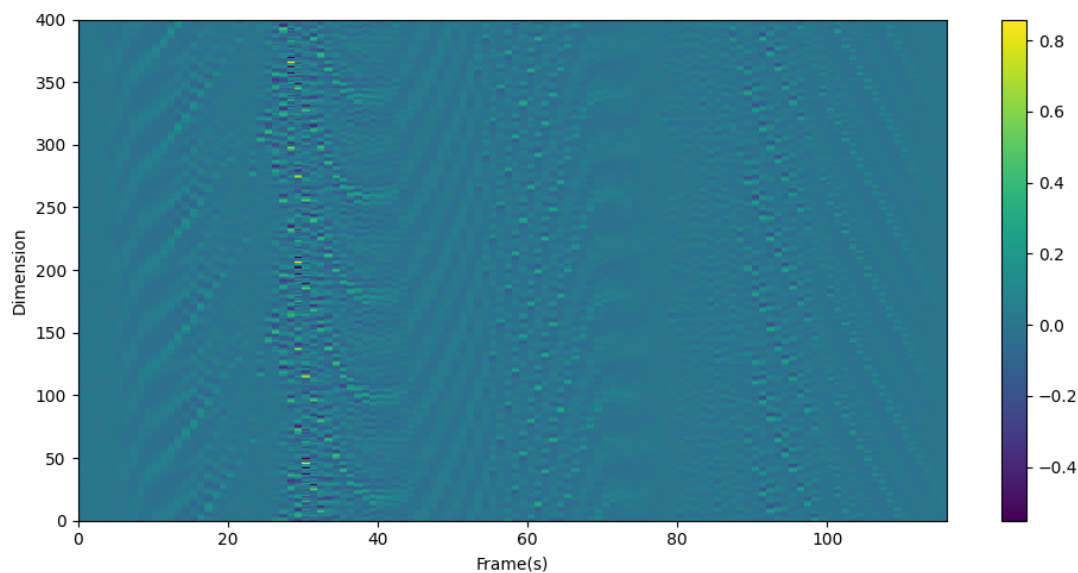


- **预加重处理**，可以发现高频部分的能量显著提高

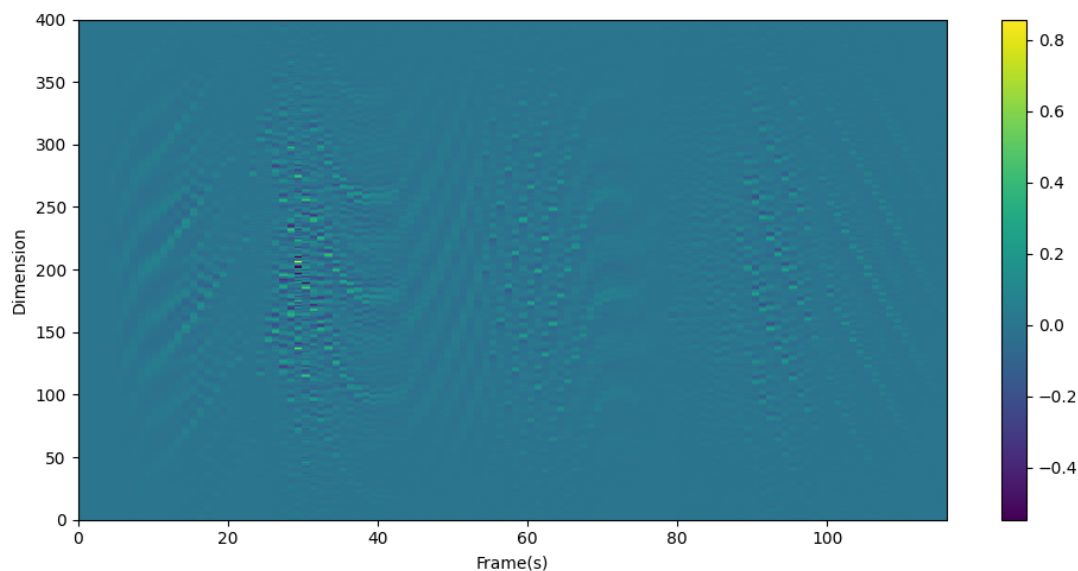




- **分帧操作**, 使用`framing(sig,fs,frame_len_s=0.025,frame_shift_s=0.01)`, 默认帧长 (len) 为 25ms, 偏移 (shift) 为 10ms, fs即为原视频的采样率。最后返回的, 是一个二维list, 一个元素是一帧信号。



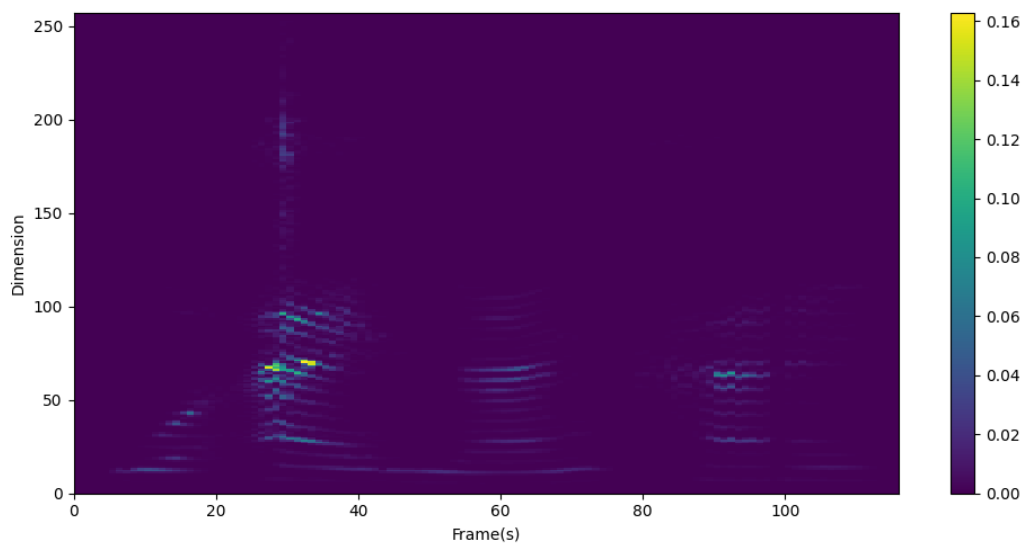
- **加窗操作**, `add_window(frame_sig,fs,frame_len_s=0.025)`



(1) 使全局更加连续。

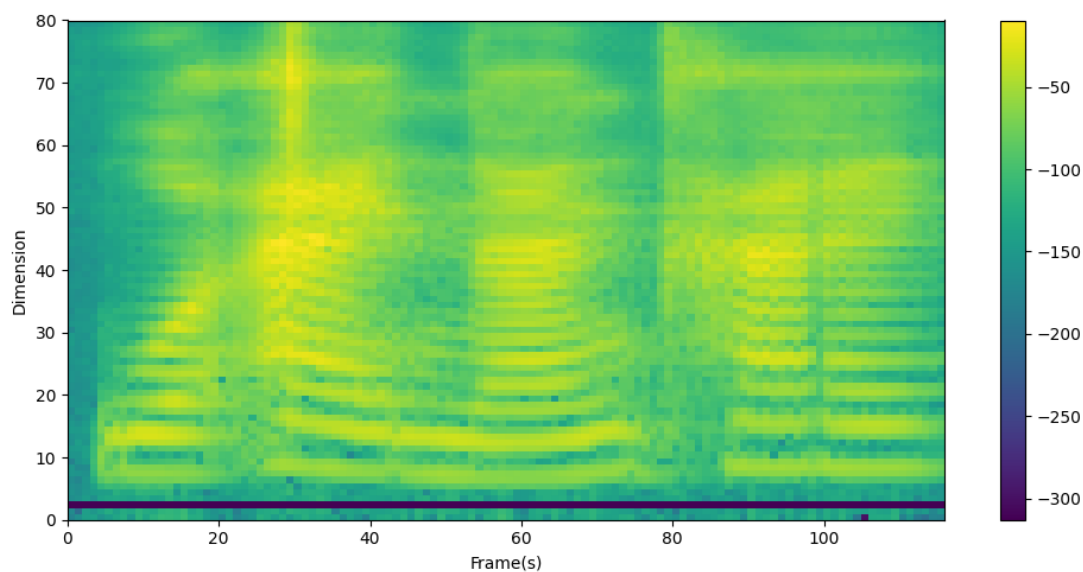
(2) 加窗之后，出现了周期函数的部分特征。

- **FFT+幅值平方**，调用`stft(frame_sig, nfft=512)`，将短时傅里叶变换将帧信号变为帧功率，返回分帧信号的功率谱。



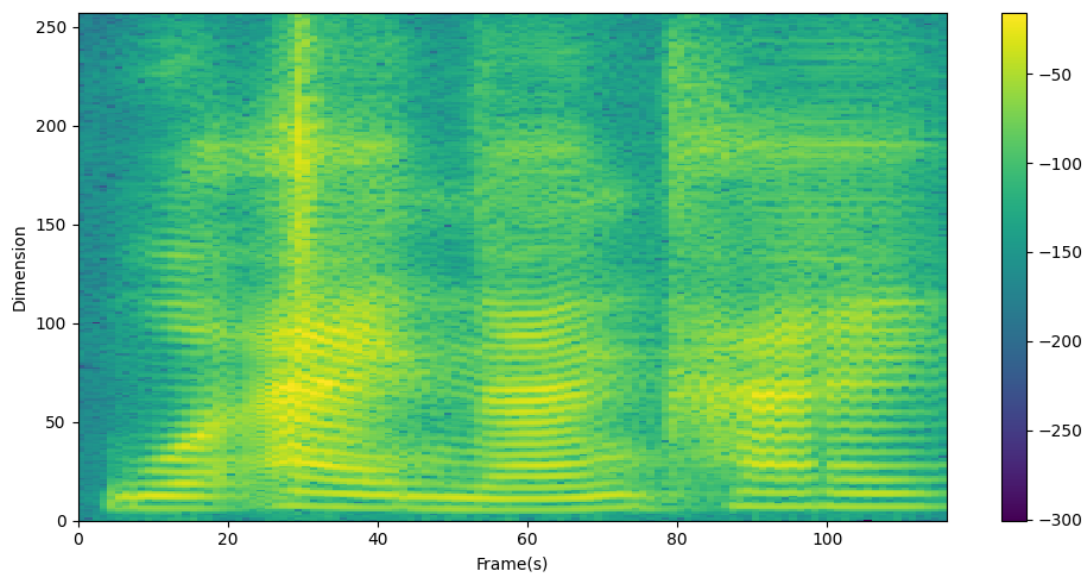
Fbank

使用mel滤波器，再取对数，即可得到80维特征。



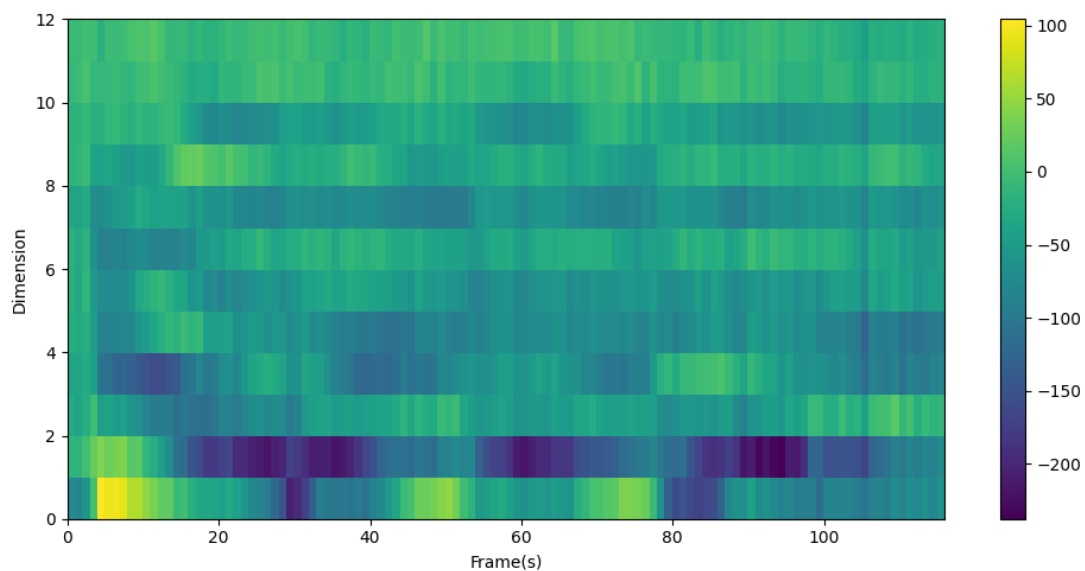
spectrogram

不使用mel滤波器，直接取对数，即可得到spectrogram特征。



MFCC

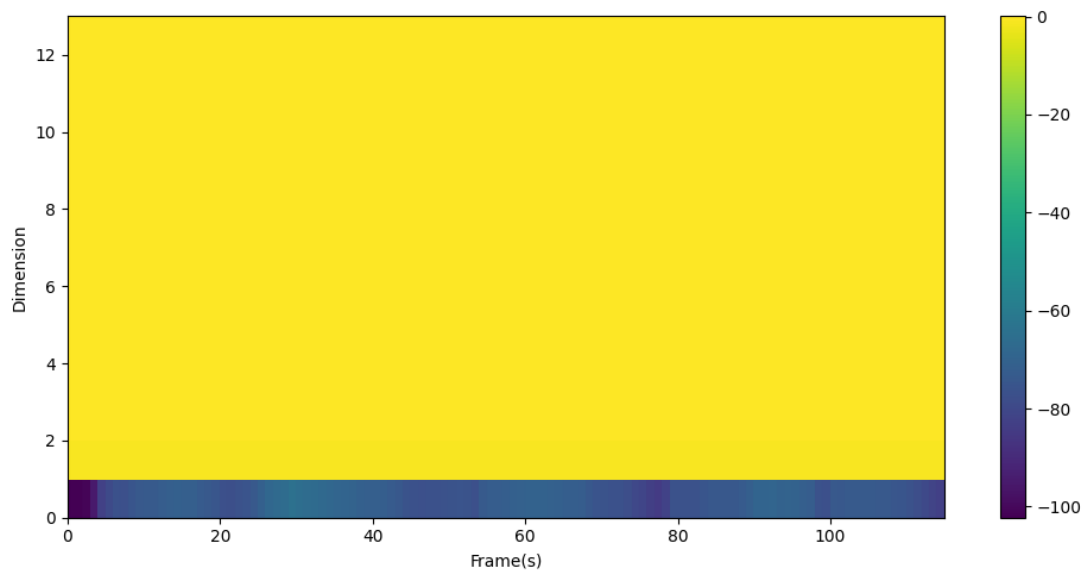
在Fbank基础上，使用离散余弦变换即可，结果如下：



通过阅读函数代码，我们可以知道，MCFF抽取的特征是2-13（共12维），然后MFCC被用于去做语音任务（如语音合成）时，会再加上energy，则一共是13维。

PLP

阅读spafe库函数的代码，可以找到需要使用的plp位于**spafe/features/rplp.py**中，即直接使用位于其中plp的API函数即可，结果如下：



遇到的问题与解决方式

1. 阅读MFCC函数代码时，不知道最后分出来的数组维数，最后向助教咨询解决。
2. 不知道如何提取音频文件，通过csdn查询，得知使用soundfile，以及相应的函数调用接口。

参考资料

- <https://blog.csdn.net/wudibaba21/article/details/108863431>