

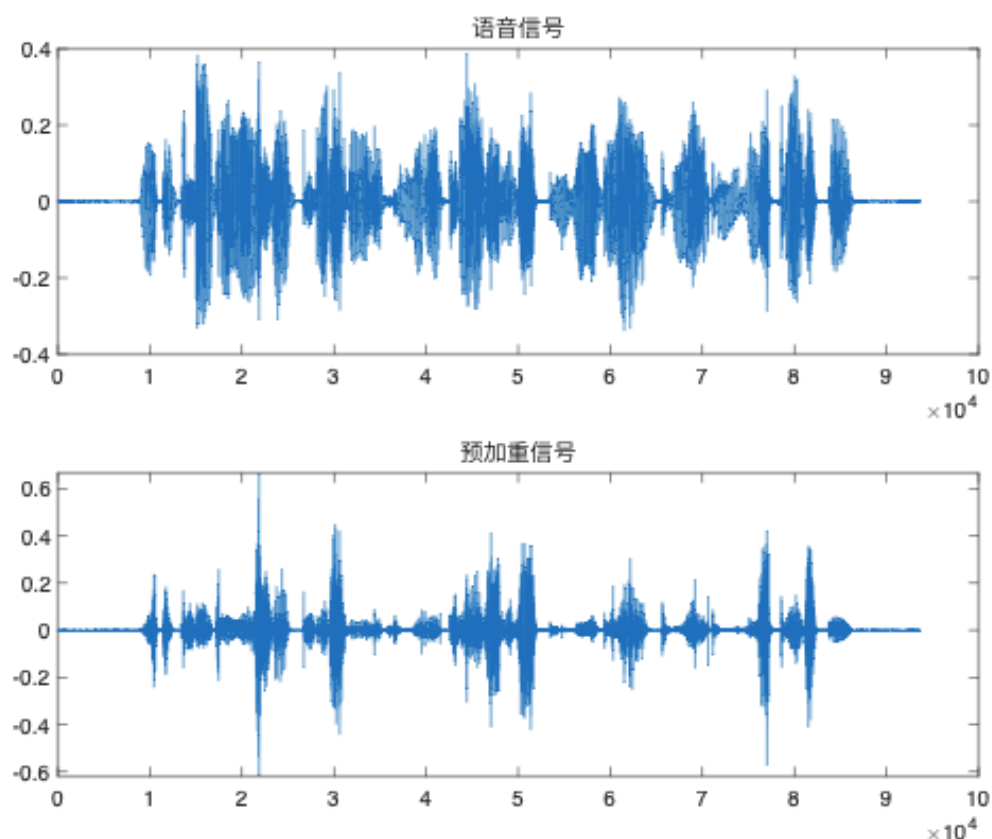
# 实验一 语音信号 MFCC 特征提取

## 1.wav实验各阶段结果分析

### 1. 信号提取及预加重

```
1 %提取信号、预加重
2 [x,Fs]=audioread('1.wav');
3 x=filter([1 -0.97],1,x);%通过预加重滤波器
```

o



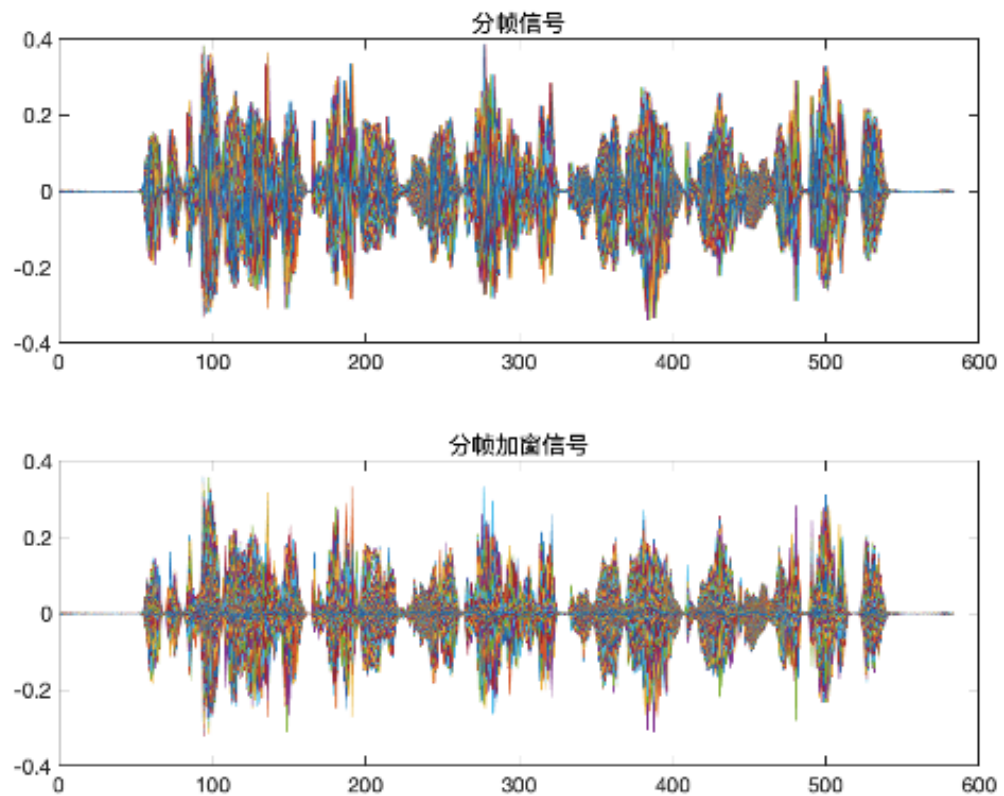
o 从图中可以看出预加重明显地放大了高频部分预，得以

- 平衡频谱，因为高频通常比低频小
- 避免傅立叶变换操作期间出现数值问题
- 还可以提高信噪比（SNR）

### 2. 分帧与加窗

```
1 fs=Fs/1000;%1ms的采样数
2 wlen=25*fs;%帧长
3 inc=10*fs;%帧移
4 overlap=wlen-inc;%每帧重叠部分
5 nf=fix((length(x)-overlap)/inc);%帧数
6 win=0.54-0.46*cos(2*pi*(1:wlen)/(wlen-1));%窗函数
7 y=enframe(x,win,inc);%分帧并加窗
```

○

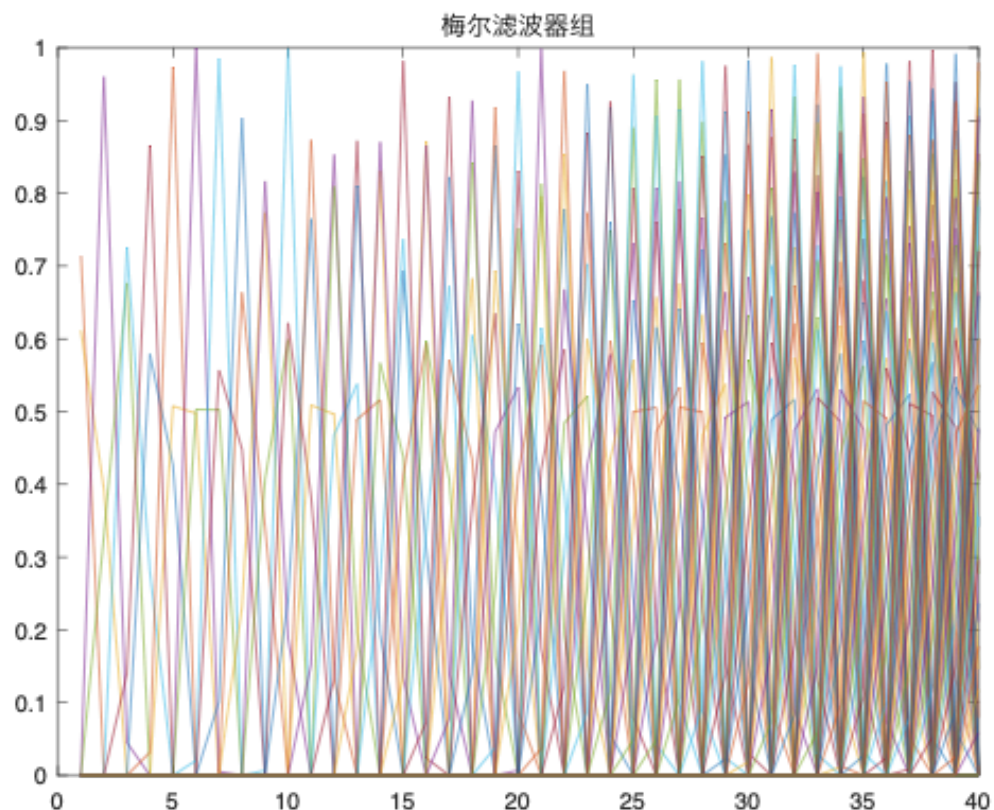


○ 可以看出加窗后分帧信号的高频分量有明显的衰减

### 3. 梅尔滤波器、DCT

```
1 bank=melbankm(40,512,Fs,0,0.5,'t'); %梅尔滤波器的阶数为40，FFT变换的长度为512
2 bank=full(bank); %稀疏矩阵转为正常矩阵
3 bank=bank/max(bank(:)); %归一化
4 dctcoef=zeros(13,40); %余弦变换系数矩阵
5 for k=1:13
6     n=0:39;
7     dctcoef(k,:)=cos((2*n-1)*k*pi/(2*40));
8 end
```

○



- 用voicebox中的melbankm函数设置梅尔滤波器组，阶数设为40，最高频率设为采样频率的一半，采用三角窗。归一化并变换滤波器组的存储形式，以便后续进行矩阵计算。
- 余弦变换系数矩阵直接套用公式两层for循环可得。

#### 4. 频谱、MFCC参数

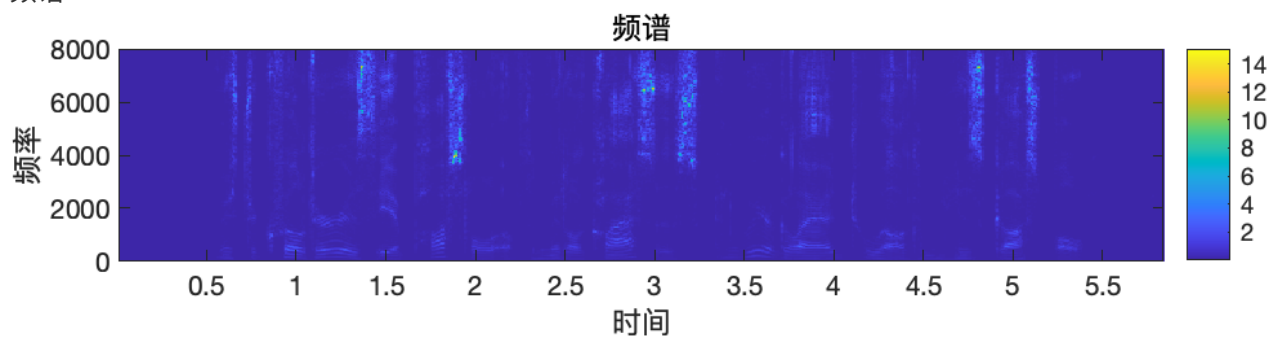
```

1  fourier=zeros(512,nf);
2  fbank=zeros(nf,40);
3  mfcc=zeros(13,nf);
4  for i=1:nf
5      y1=y(i,:);%取出一帧信号
6      Y=abs(fft(y1,512));%FFT快速傅里叶变换
7      fourier(:,i)=Y;
8      P=Y'./2/512;%取平方求功率
9      fbank(i,:)=bank*P(1:257);%频谱通过梅尔滤波器
10     mfcc(:,i)=dctcoef*10*log(fbank(i,:));%对数运算和DCT
11 end

```

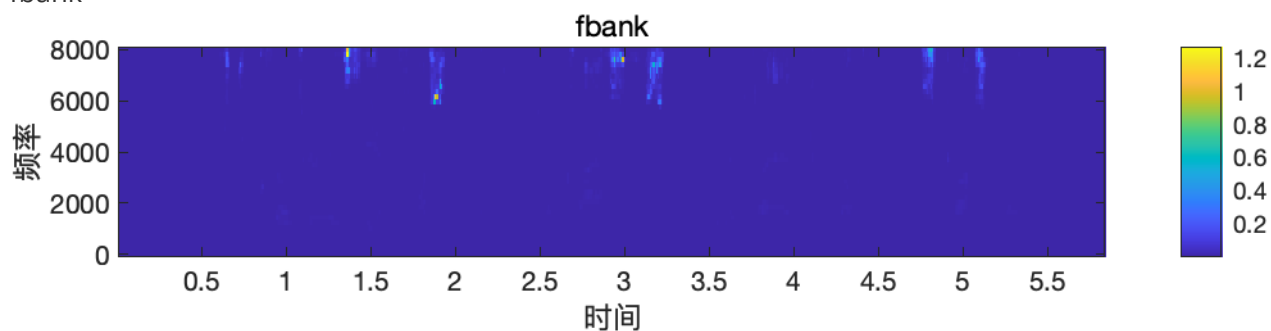
- 先初始化fourier、fbank、mfcc三个矩阵，分别用于存放信号频谱，梅尔滤波后的功率谱及MFCC参数。而后进行分帧帧数nf=584次for循环，每次取出之前加窗后的一帧，依次进行fft快速傅里叶变换、求功率谱、过梅尔滤波器、求MFCC参数，并依次存放。由于语音信号为实矩阵，所以fft结果共轭对称，功率谱及频谱只取一半即可。

- 频谱



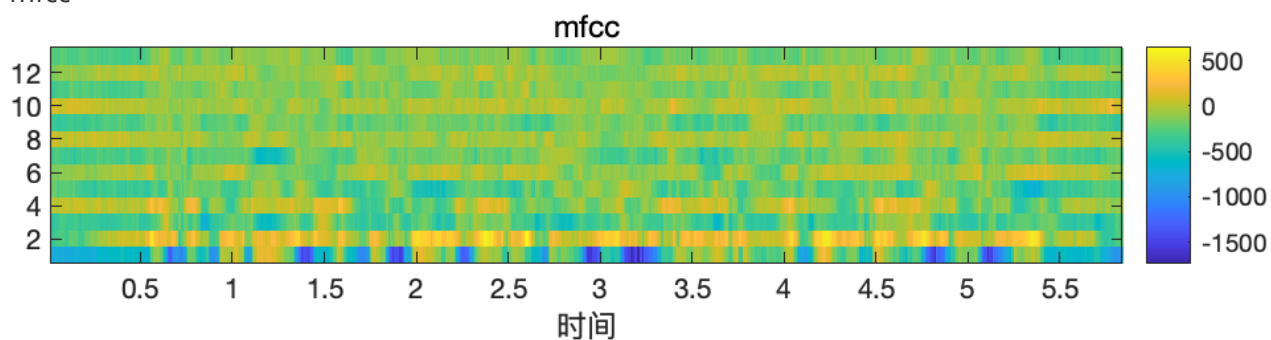
- 可以看出频谱图中颜色较亮的部分与初始的语音信号高频部分对应良好。

- fbank

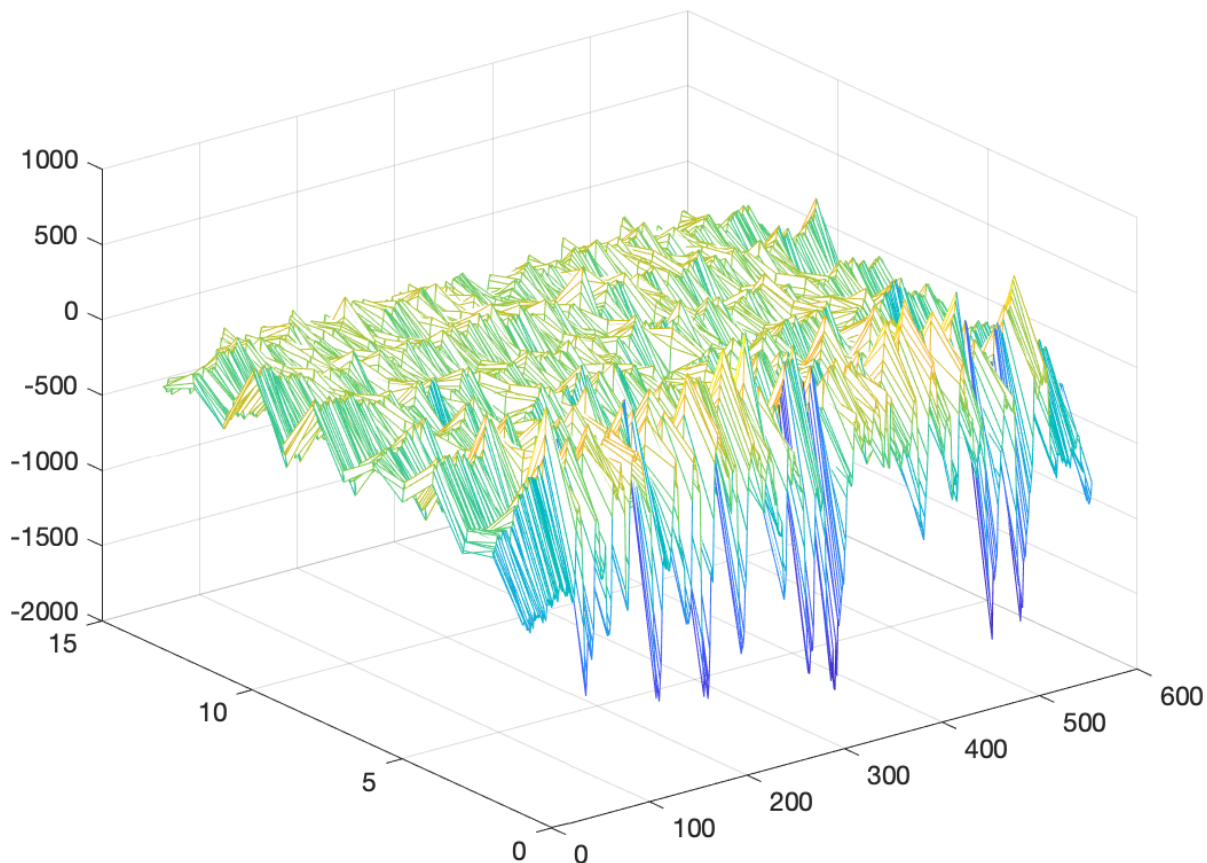


- 经过梅尔滤波器后高频分量明显减少

- mfcc



- 可以看出mfcc参数在频率很低时值很小，频率增大一点后达到峰值，而后振荡式分布
- 三维图



## 2.wav运行耗时

```

编辑器 - /Users/jinqiao/Documents/MATLAB/jianmo/exp1/mycode.m
example.m  mycode.m  untitled.m  +
1      clear;clc;close all;
2
3      tic %开始计时
4
5      %提取信号、预加重
6      [x,Fs]=audioread('2.wav');
7      x=filter([1 -0.97],1,x);%通过预加重滤波器
8      %figure();subplot(211);plot(x);title('语音信号');subplot(212);plot(x1);title('预加重')
9
10     %分帧、加窗
11     fs=Fs/1000;%1ms的采样数
12     wlen=25*fs;%帧长
13     inc=10*fs;%帧移
14     overlap=wlen-inc;%每帧重叠部分
15     nf=fix((length(x)-overlap)/inc);%帧数
16     win=0.54-0.46*cos(2*pi*(1:wlen)/(wlen-1));%窗函数
17
命令窗口
历时 4.088287 秒。
fx >>

```

- 使用MATLAB自带的tic/toc计时，用时4s。

## fbank与mfcc比较

- fbank特征更多是希望符合声音信号的本质，拟合人耳的接收特性。DCT是线性变换，会丢失语音信号中原本的一些高度非线性成分。
- 在深度学习之前，受限于算法，mfcc配GMMs-HMMs是ASR的主流做法。当深度学习方法出来之后，由

于神经网络对高度相关的信息不敏感，mfcc不是最优选择，经过实际验证，其在神经网络中的表现也明显不如fbank。