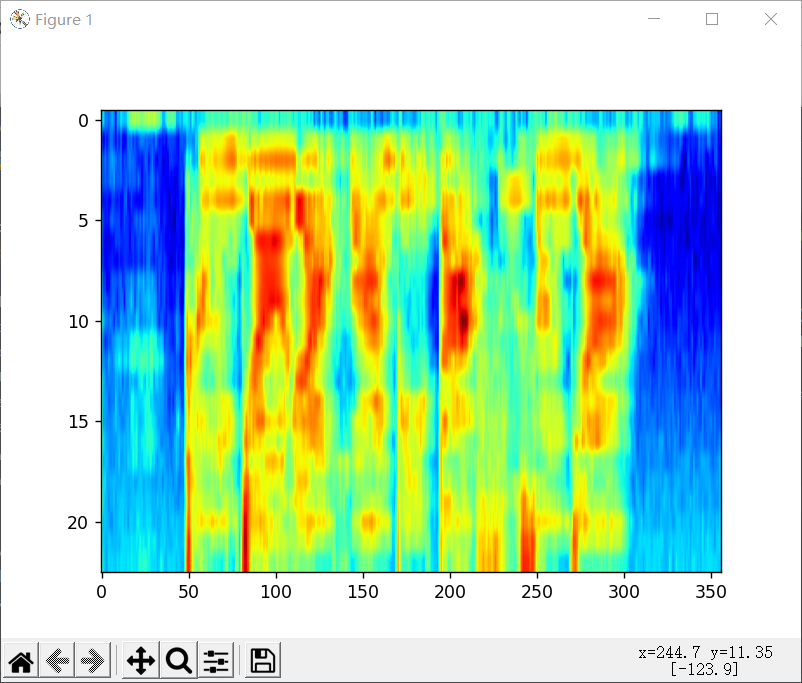
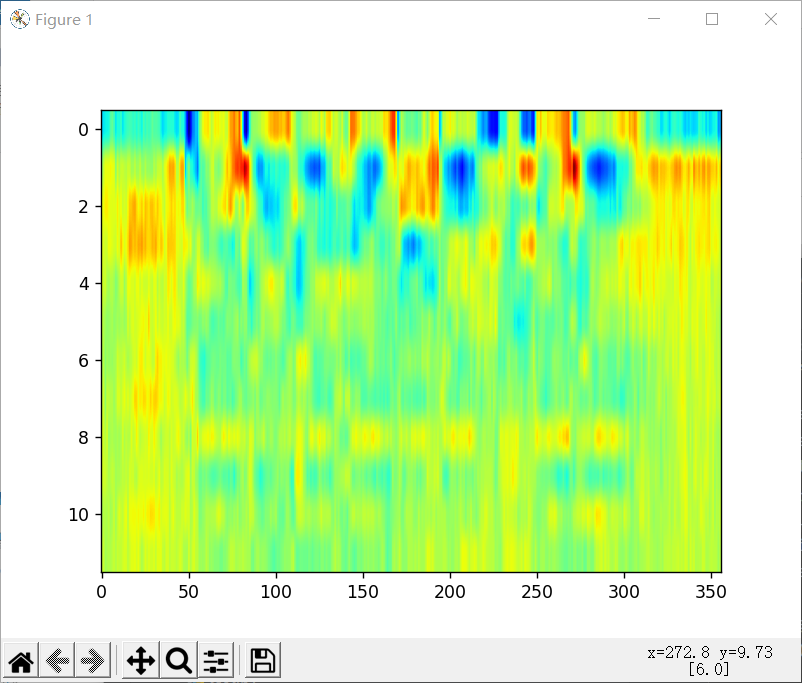
Q： 给定一段音频，请提取12维MFCC特征，阅读代码预加重、分帧、加窗部分，完善作业代码中fbank和mfcc部分，并给出最终的Fbank和MFCC特征

A：

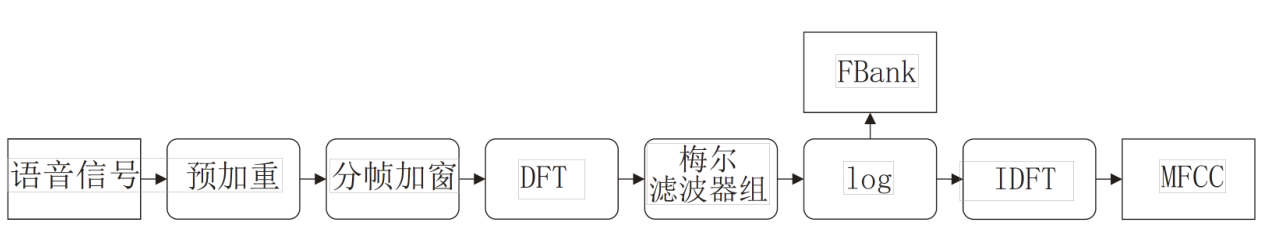
Fbank\_beats：



MFCC\_beats：



Fbank和MFCC（Mel-Frequency Cepstral Coefficients）提取流程：



Fbank和MFCC特征目前仍是主要使用的特征，虽然有工作尝试直接使用波形建模，但是效果并没有超越基于频域的特征。

* Step1：预加重(emphasis)

为什么要预加重？

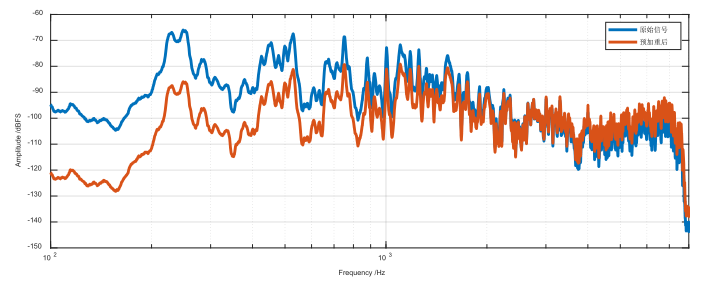
——提高信号高频部分的能量，高频信号在传递过程中，衰减较快，但是高频部分又蕴含很多对语音识别有利的特征，因此，在特征提取部分，需要提高高频部分的能量。

预加重滤波器是一个一阶高通滤波器，给定时域输入信号x[n],预加重之后的信号为:

}O%360IJ{B8_2EME)CV{I)R

其中，0.9<=a<=1.0

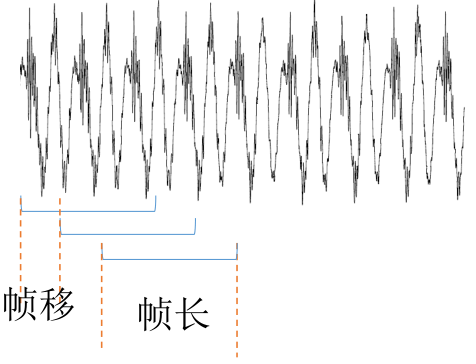
直观的解释：预加重是一个高通滤波器，因此，低频信号(即时域上信号变换慢)将被抑制；从公式可以看出来：如果x[n]是低频信号，变化缓慢，那么x[n]与x[n-1]相近，当a接近1的时候，x[n]-ax[n-1]约等于0,此信号的幅度将被大大抑制。如果x[n]是高频信号，变化快，那么x[n]与x[n-1]相差很大，x[n]-ax[n-1]的值不会趋于0，此信号的幅度还能保持，可以通过滤波器。



* Step2：加窗分帧

为什么要分帧？

——语音信号为非平稳信号，其统计属性是随着时间变化的，以汉语为例，一句话中包含很多声母和韵母，不同的拼音，它的发音特点很明显不一样。但是，语音信号又具有短时平稳的属性，比如汉语中的一个声母或者韵母，往往只会持续几十到几百毫秒，在这一个发音单元里，语音信号表现出明显的稳定性，规律性。所以在进行语音识别的时候，对于一句话，识别的过程也是以较小的发生单元(音素、字、字节)为单位进行识别，因而用滑动窗来提取短时片段。



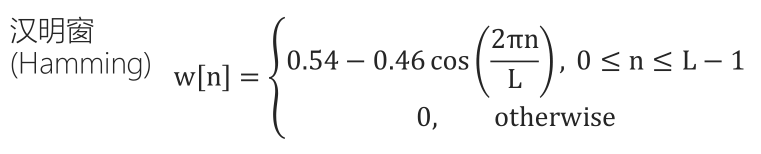
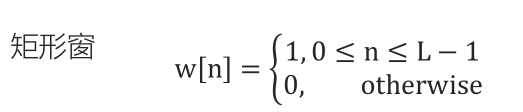
帧长、帧移、窗函数的概念，对于采样率为16kHz的信号，帧长、帧移一般为25ms、10ms，即400和160个采样点。

分帧的过程，在时域上，即用一个窗函数和原始信号进行相乘：

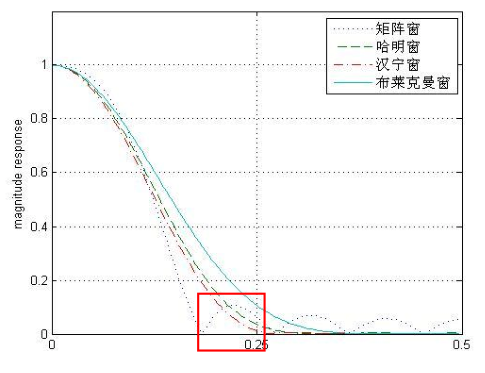
IMG_256

w[n]称为窗函数，常用的窗函数有：矩形窗、汉明窗.....

Dsp书已经详细讲过，这里就列一下公式：



而一般不直接使用矩形窗的原因是，矩形窗主瓣窄，但是旁瓣较大，将其与原信号的频谱式进行卷积，就会导致频谱泄露。



* Step3. 傅里叶变换

将上一步分帧之后的语音帧，由时域变换到频域，取DFT系数的模，得到谱特征。

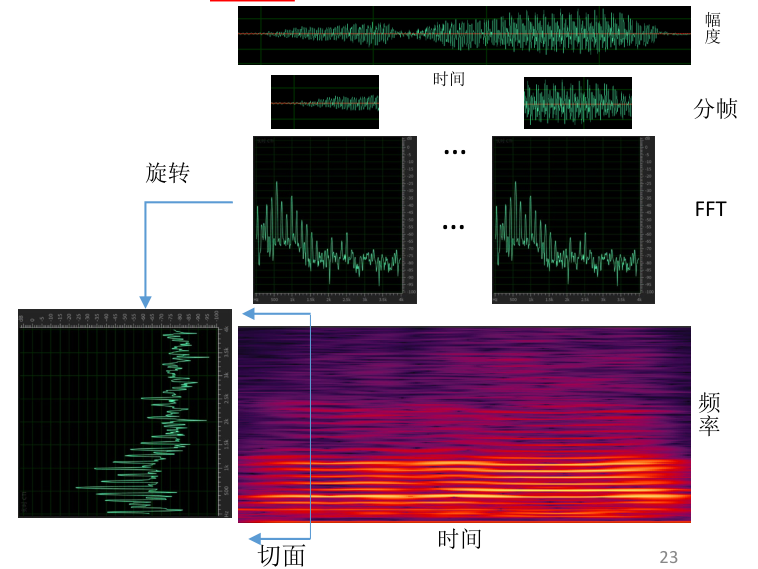
右图展示了语谱图的生成过程：

1. 加窗分帧

2. 将每一帧信号进行DFT（FFT），如第t帧信号，DFT系数为Xt(m) ,m = 0,1,…,N

3. 将每一帧DFT的系数按时间顺序排列，得到一个矩阵Y ∈ C^ T∗N ，且Y[t,m] = Xt(m)

4. 语谱图是一个三维图，横轴表示时间(t)，纵轴表示频率，颜色的深浅表示当前时频点上幅度的大小|Y[t,m]|



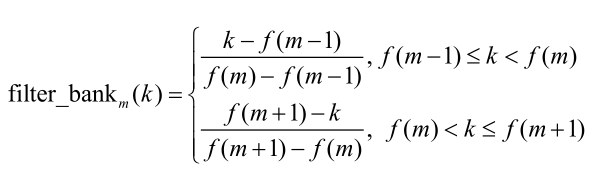
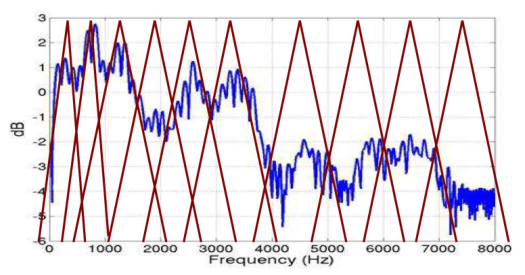
* Step4. 梅尔滤波器组和对数操作

• DFT得到了每个频带上信号的能量，但是人耳对频率的感知不是等间隔的，近似于对数函数

• 将线性频率转换为梅尔频率，梅尔频率和线性频率转换关系

mel f = 2595log 10 (1 +f/700 )

• 梅尔三角滤波器组：根据起始频率、中间频率和截止频率，确定各滤波器系数



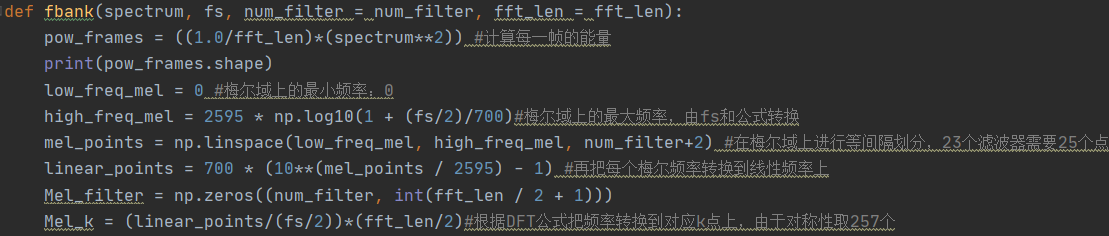
滤波器公式其实是一个分段的线性直线，比如f(m+1),f(m),f(m-1) = 30,20,10，那么10<k<20时，结果为(k-10)/10；20<k<30时，结果为(30-k)/10；k<10和k>20，结果均为0；可以看到，对于每个k，画出来的就是两条对称的线，他们的顶点重合。也就是说该滤波器在中心频率点响应值为1，在两边的滤波器中心点衰减到0。其中k表示滤波器个数。

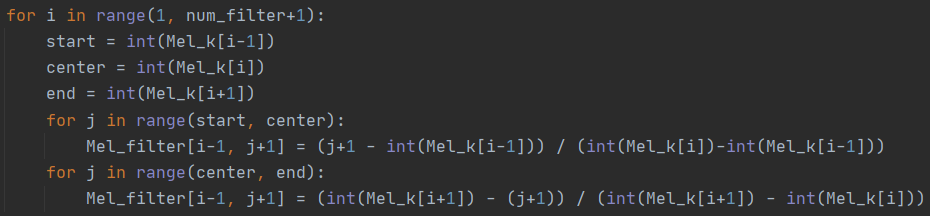
现在提取一个wav音频的12维mfcc:

①确定滤波器组个数P，这里取23

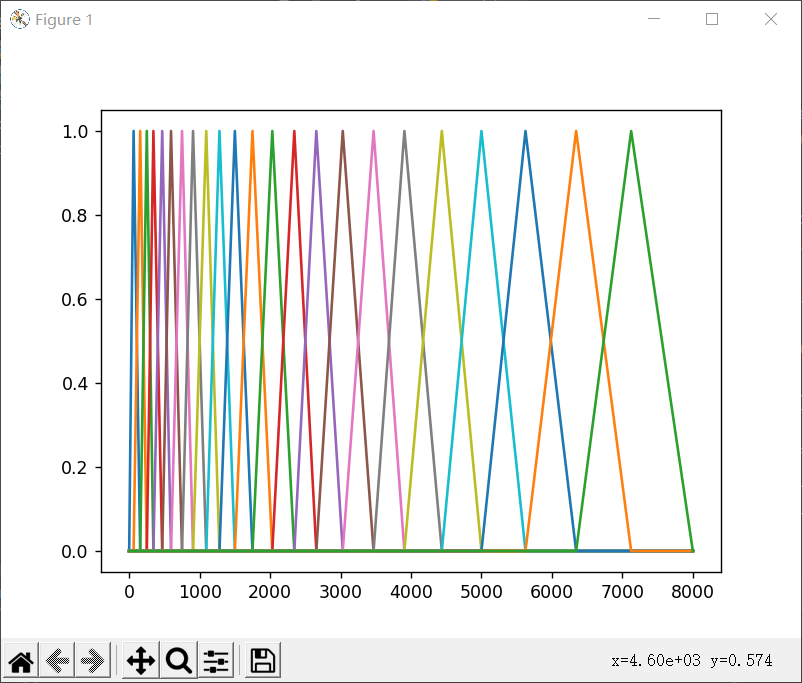
②根据采样率fs，DFT点数N，滤波器个数P，在梅尔域上等间隔地产生每个滤波器的起始频率、中间频率和截止频率。这里上一个滤波器的中间频率为下一个滤波器的起始频率。

③将梅尔域上每个三角滤波器的起始、中间和截止频率转换为线性频率，并对DFT之后的谱特征进行滤波，得到P个滤波器组的能量，得到Fbank特征。

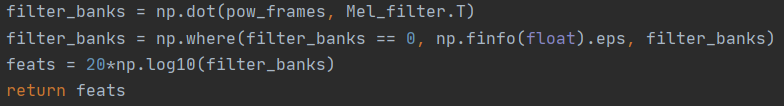




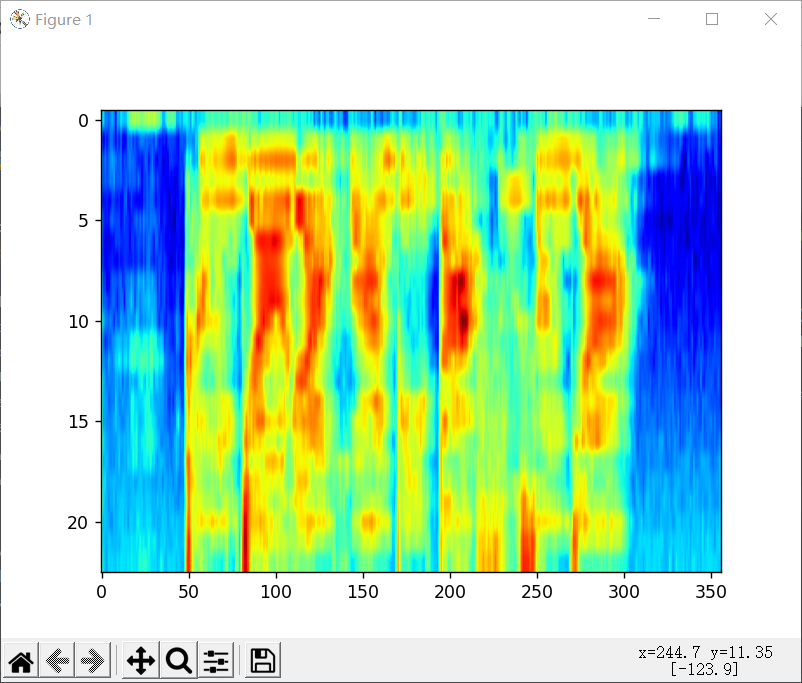
这样滤波器就能画出来了：



把每一帧能量送入梅尔滤波器，并进行dB转换，就可以得到Fbank特征。

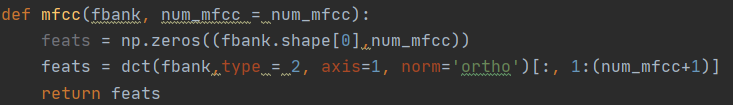


Fbank\_beats：

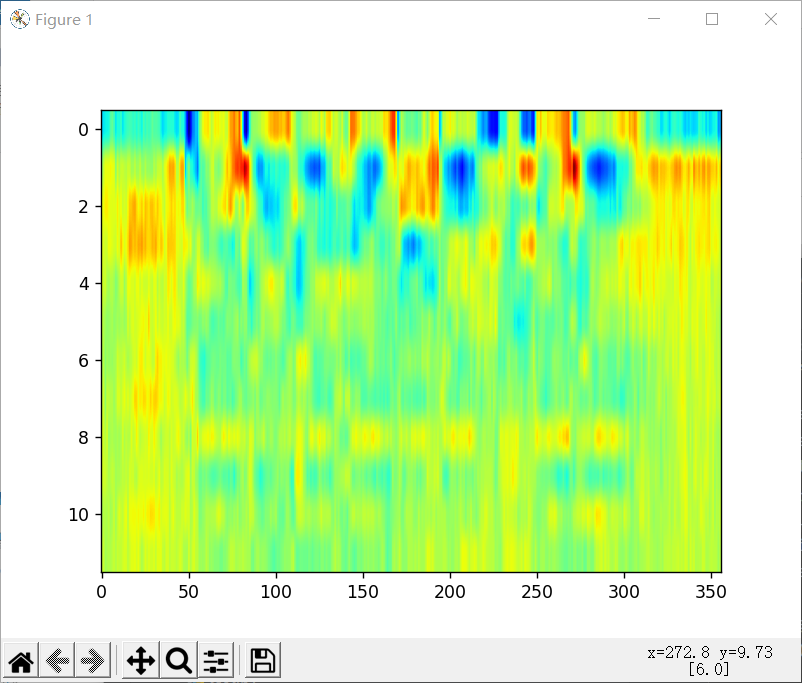


* Step5

MFCC特征在Fbank特征基础上继续进行IDFT变换等操作即可，准确的叫做DCT变换。DCT其实是IDFT的等价替代，所以MFCC名字里面有倒谱。

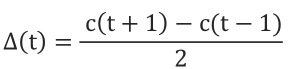


MFCC\_beats：

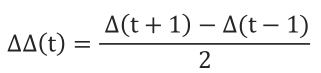


此外，再对这种静态特征做一阶和二阶差分，得到相应的动态特征。

①一阶差分(Delta)，类比速度，最简单的一阶差分计算方法：



②二阶差分(Delta,delta),类比加速度，简单计算方法：



一般常用的MFCC特征维39维，包括12维原始MFCC，12维一阶差分，12维二阶差分，1维能量，1维能量一阶差分，1维能量二阶差分。

MFCC特征一般用于对角GMM训练，各维度之间相关性小。

Fbank特征相关性更强，一般用于DNN训练。