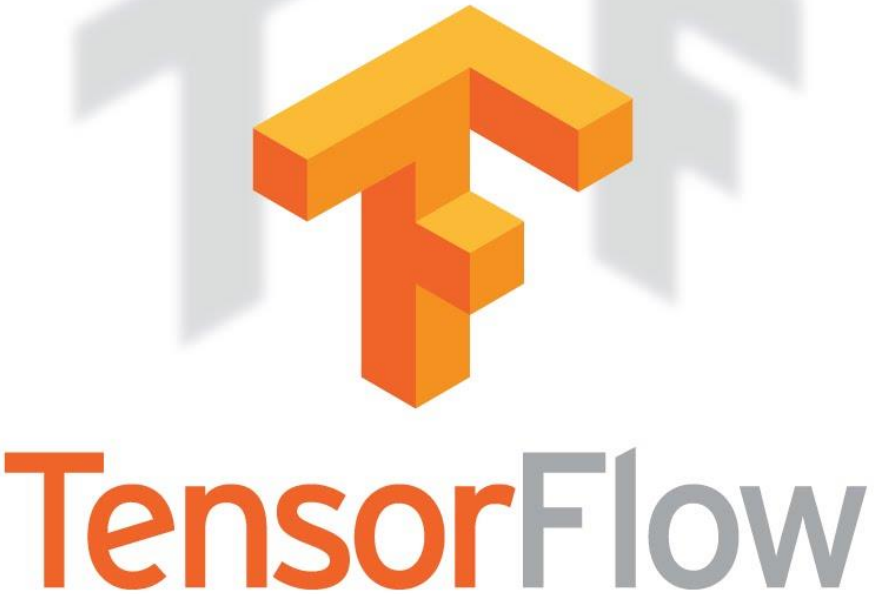
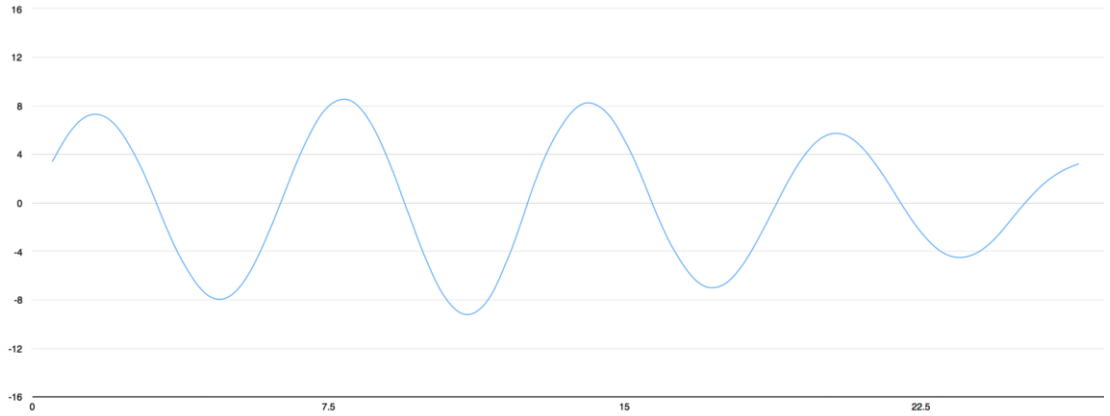


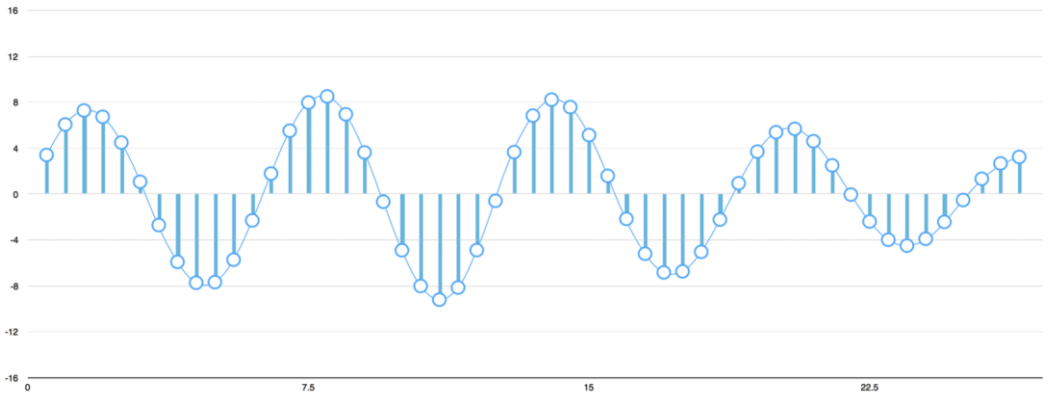
深度学习框架Tensorflow学习与应用 第12课

DATAGURU专业数据分析社区

声音信号

- 将N个采样点集合成一个观测单位，成为帧。通常N的值为256或512，覆盖的时间约为20-30ms左右。为了避免两帧之间变化过大，因此会让两相邻帧之间有一段重叠区域。通常语音识别所采用的语音信号的采样频率为8KHz或16KHz。



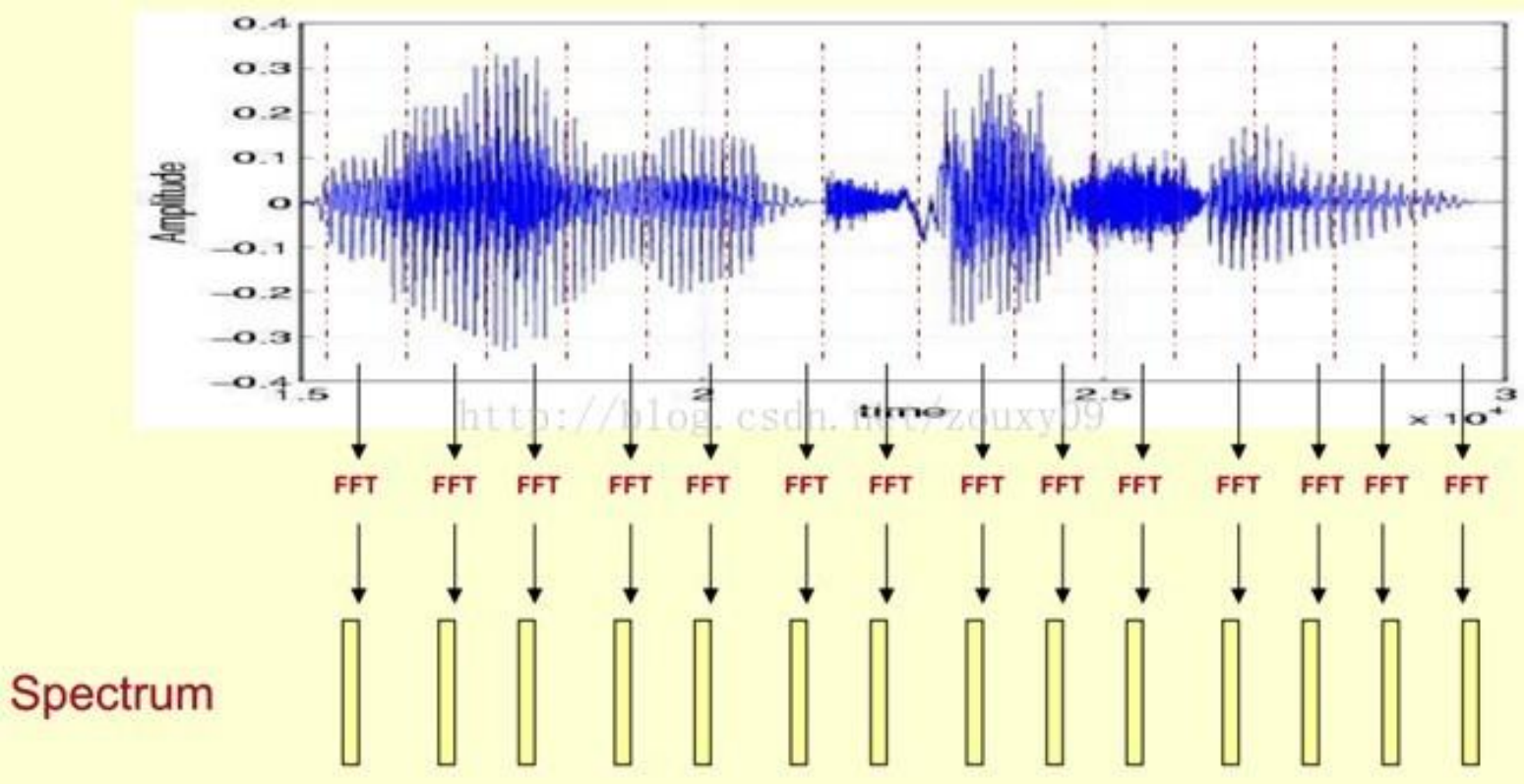


梅尔频率倒谱系数（MFCC）

- MFCC是一种广泛使用的语音特征，在1980年由Davis和Mermelstein研究出来。

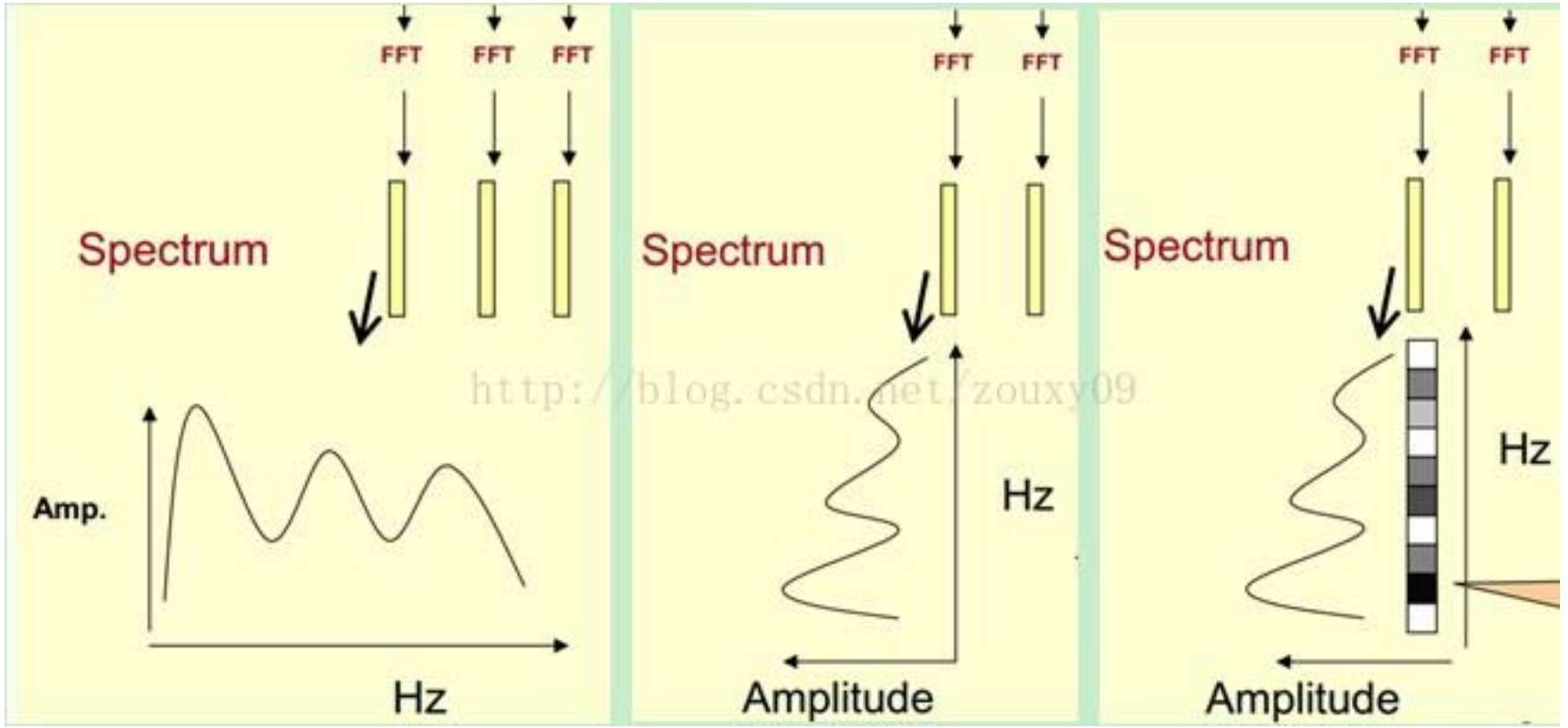
声谱图

- 语音被分为很多帧，每帧语音都对应于一个频谱（通过FFT计算得到），频谱表示频率与能量的关系。

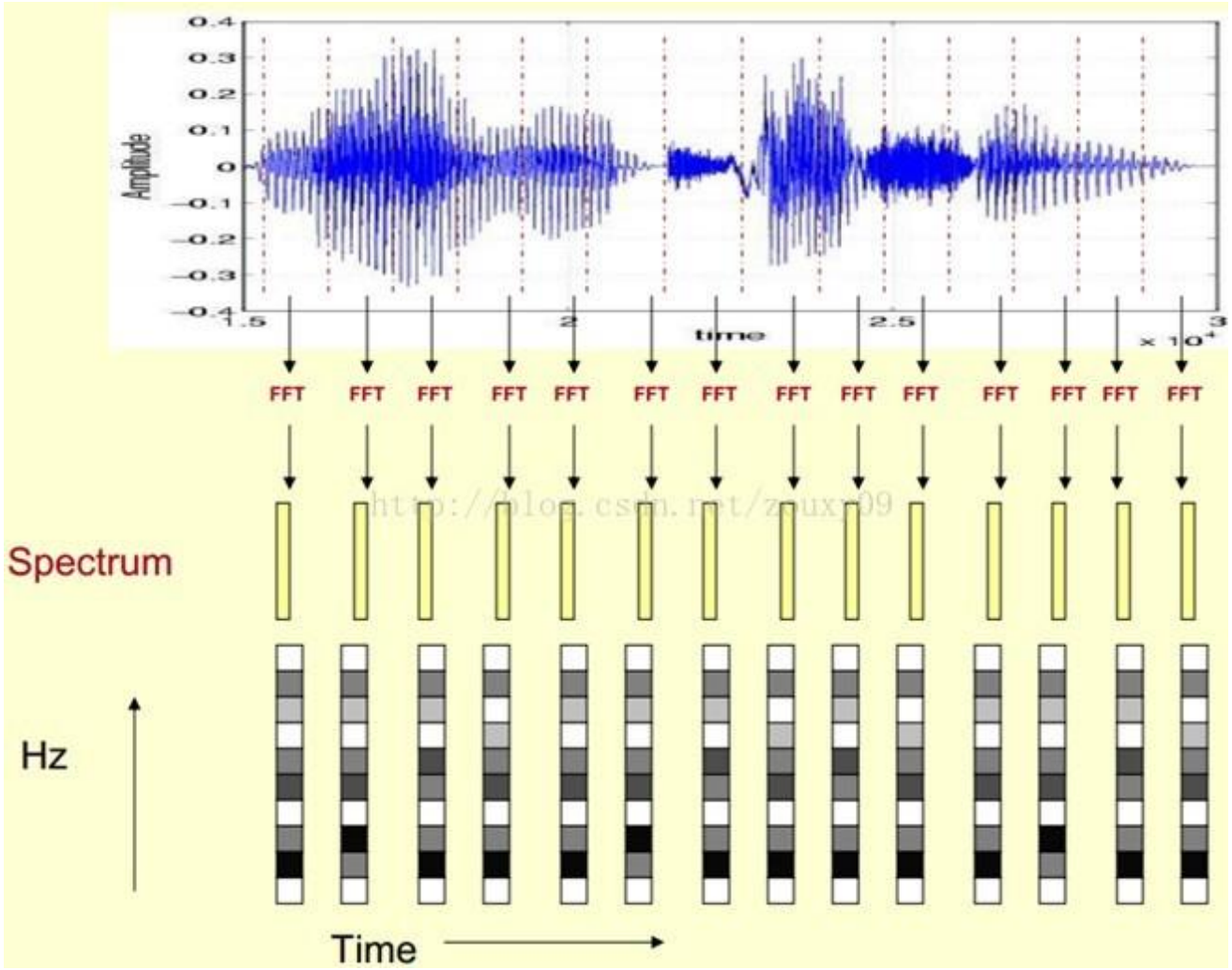


频谱图

- 我们先将一帧语音的频谱通过坐标表示出来，如左图。再将图旋转90度，如中间的图。然后把这些幅度映射到一个灰度级表示。

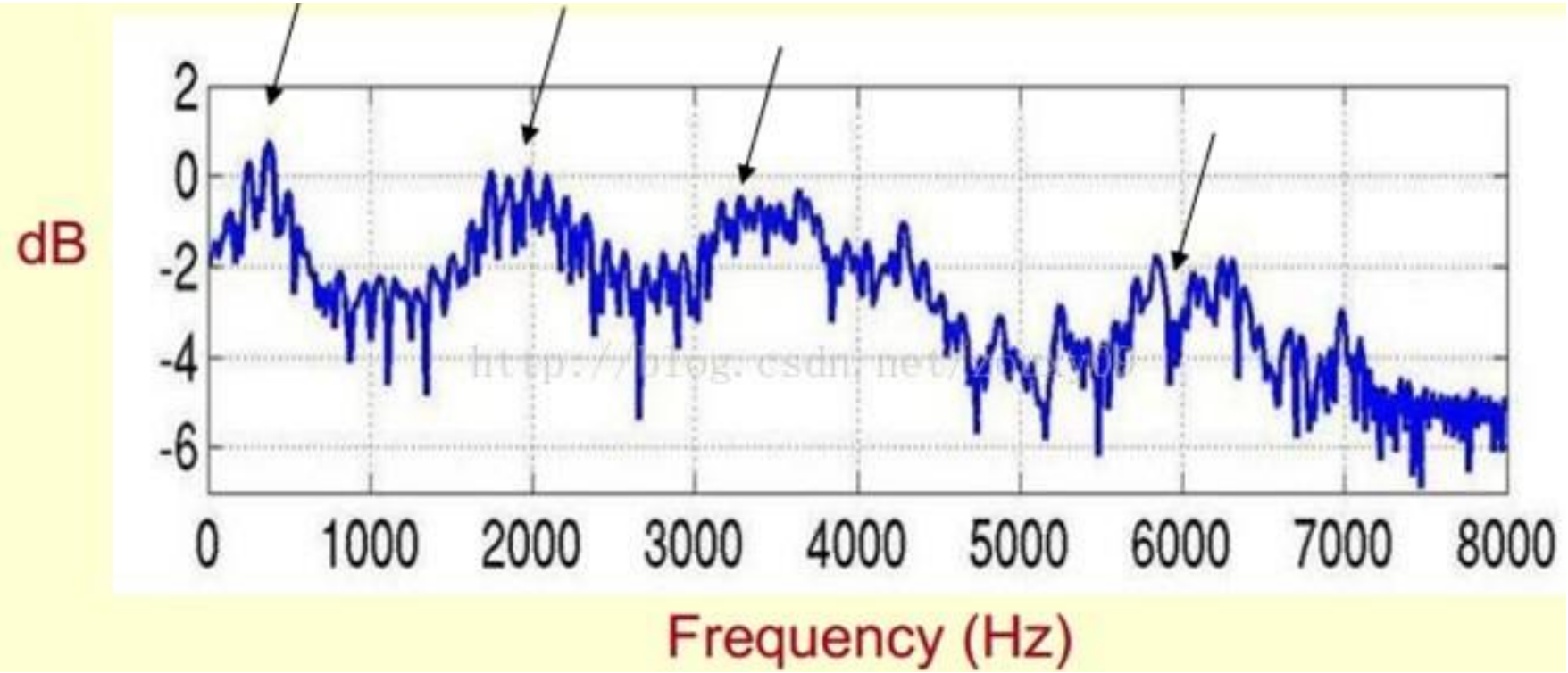


spectrogram声谱图



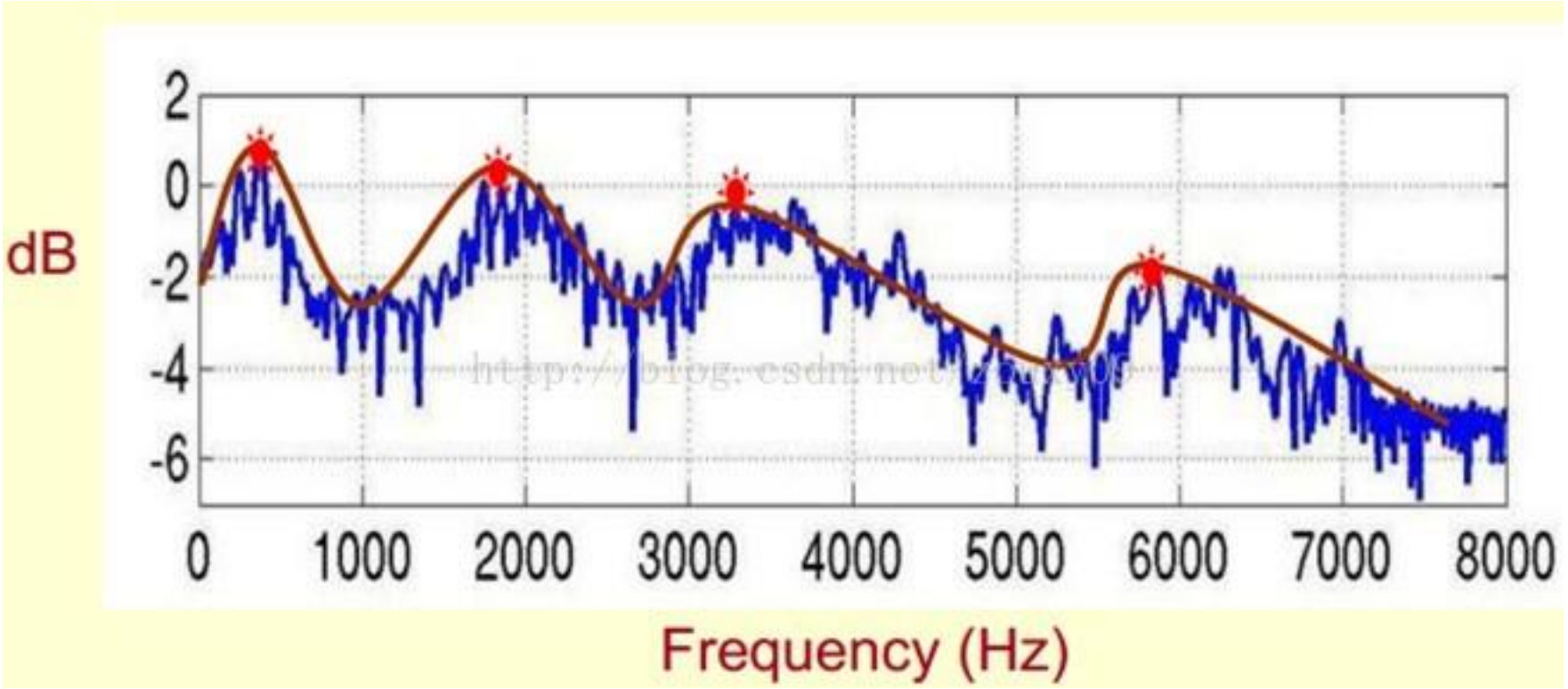
共振峰

- 下面是一个语音的频谱图。峰值就表示语音的主要频率成分，这些峰值成为共振峰。共振峰携带了声音的辨识属性。用它就可以识别不同的声音。

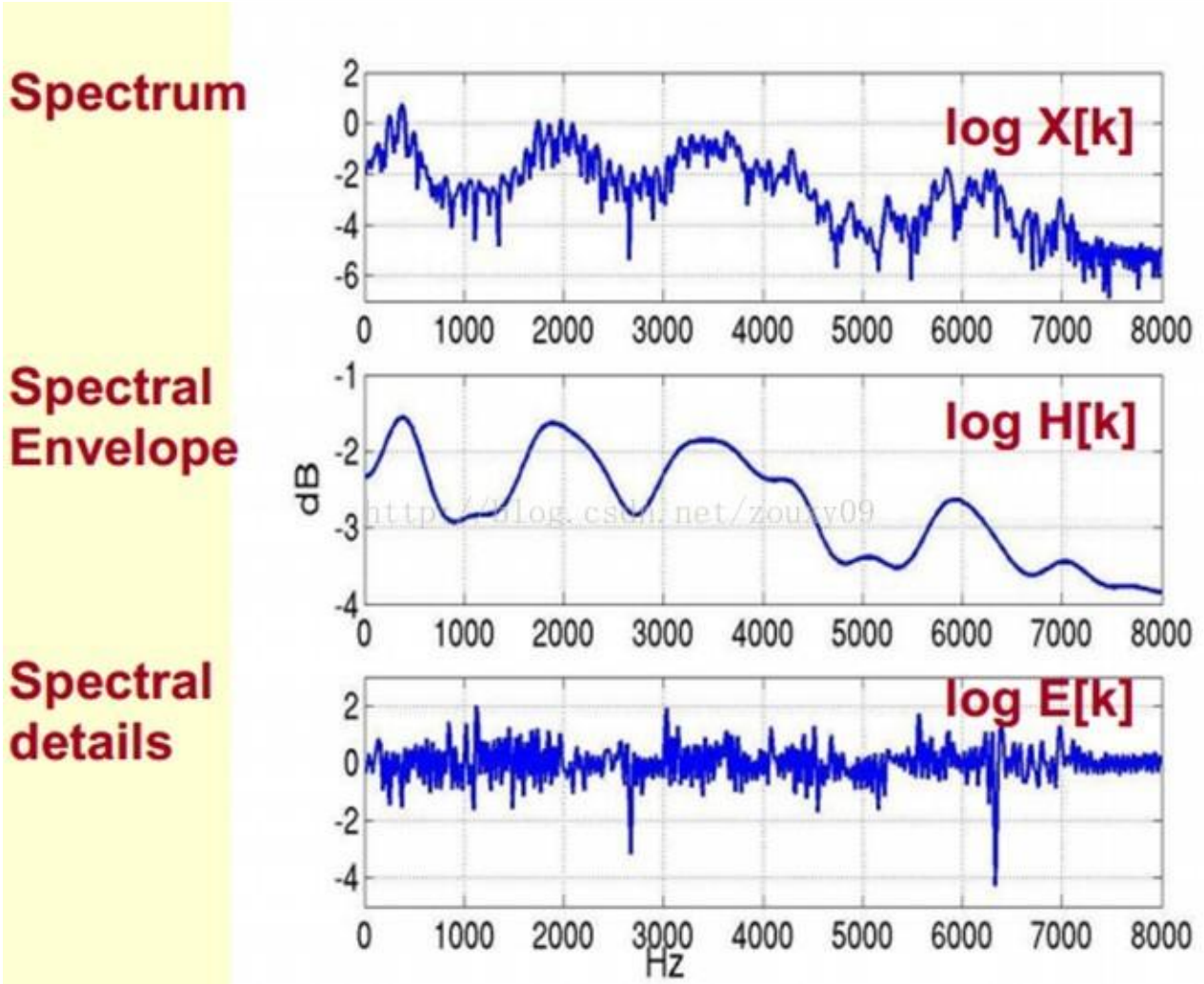


包络

- 我们需要把共振峰提取出来，不仅要提取共振峰的位置，还要提取它们转变的过程，也就是频谱的包络。包络就是一条连接这些共振峰点的平滑曲线。

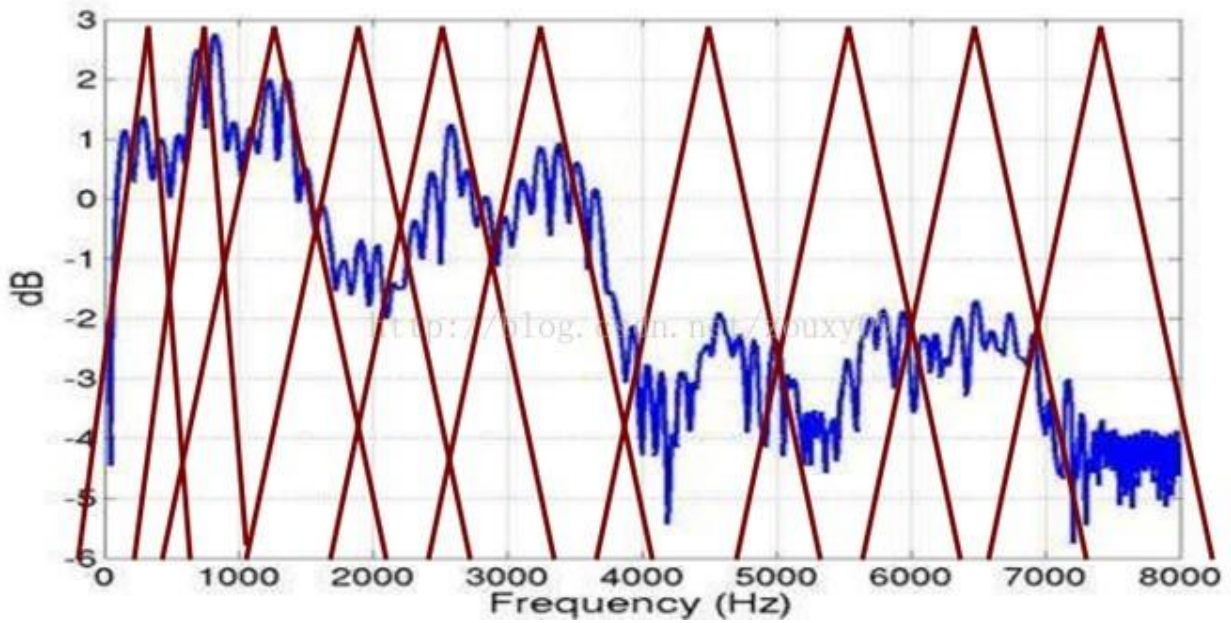


分离包络和频谱的细节


Spectrum
Spectral Envelope
Spectral details
log X[k]
log H[k]
log E[k]
dB
Hz
2
0
-2
-4
-6
-1
-3
1000
2000
3000
4000
5000
6000
7000
8000

Mel频率分析

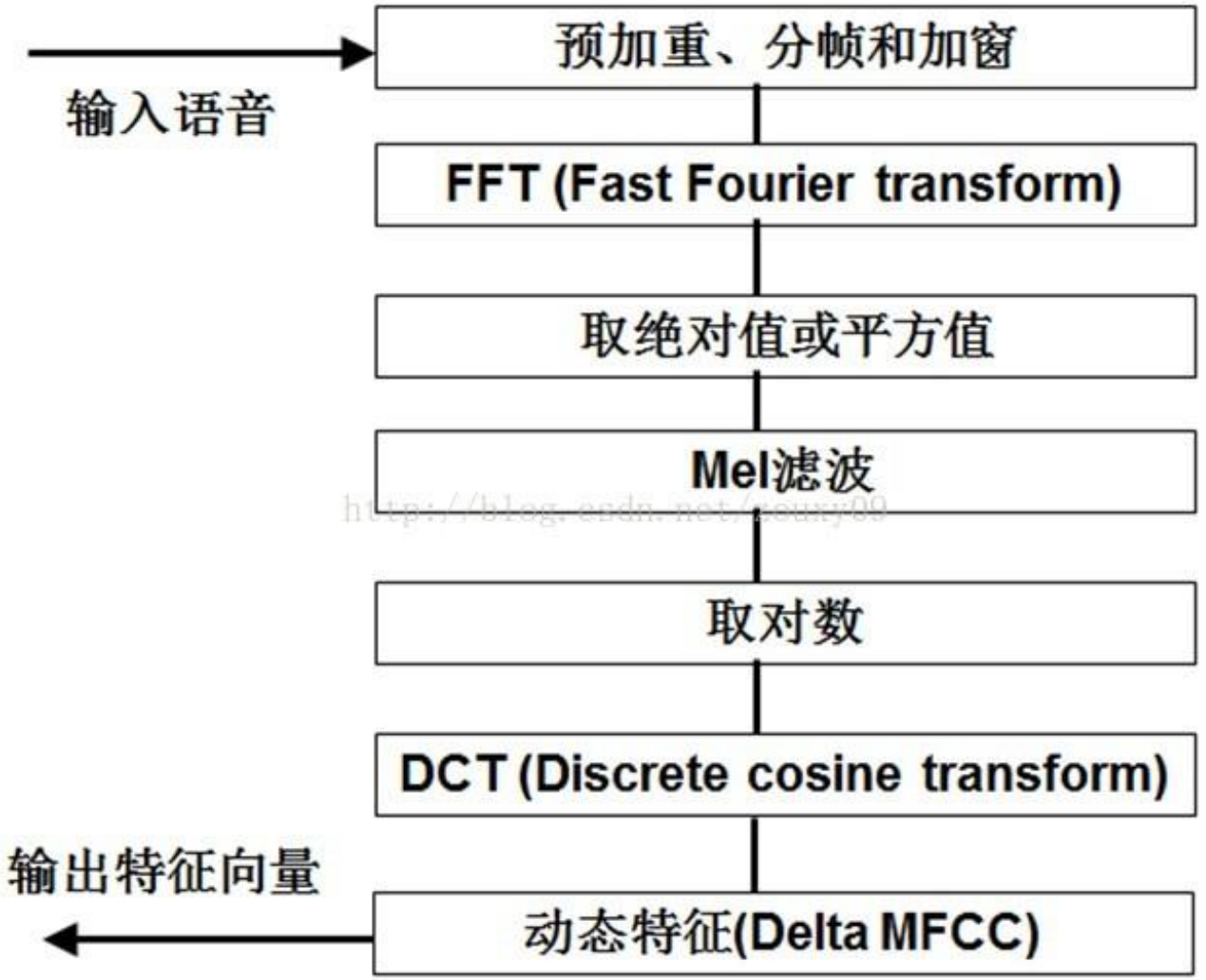
- 刚刚我们得到了频谱包络，不过人类听觉感知实验表明，人类的听觉的感知只聚焦在某些特定的区域，而不是整个频谱包络。Mel频率分析就是基于人类听觉感知实验的。人耳就像一个滤波器组，它只关注某些特定频率的分量，也就是说它只让某些频率的信号通过。并且在低频区域由很多的滤波器，分布比较密集，在高频区域，滤波器比较少，也比较稀疏。



Mel频率分析

- 人的听觉系统是一个特殊的非线性系统，它响应不同频率信号的灵明度是不同的。在语音特征的提取上，人类的听觉系统非常好，它不仅能提取出语义信息，而且能提取出说话人的个人特征。所以语音识别系统中能模拟人类听觉感知处理的特点，就有可能提高语音的识别率。
- MFCC考虑到了人类的听觉特征，将线性频谱映射到基于听觉感知的Mel非线性频谱中。

语音处理流程


输入语音
预加重、分帧和加窗
FFT (Fast Fourier transform)
取绝对值或平方值
Mel滤波
取对数
DCT (Discrete cosine transform)
动态特征(Delta MFCC)
输出特征向量

ffmpeg

- conda install -c conda-forge ffmpeg
- Windows安装方式：http://www.bubuko.com/infodetail-786878.html
- Ubuntu安装方式：http://blog.csdn.net/u012386199/article/details/51188988

法律声明

【声明】本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

http://edu.dataguru.cn

炼数成金逆向收费式网络课程

- Dataguru（炼数成金）是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情，请看我们的培训网站 http://edu.dataguru.cn

Thanks

FAQ时间