

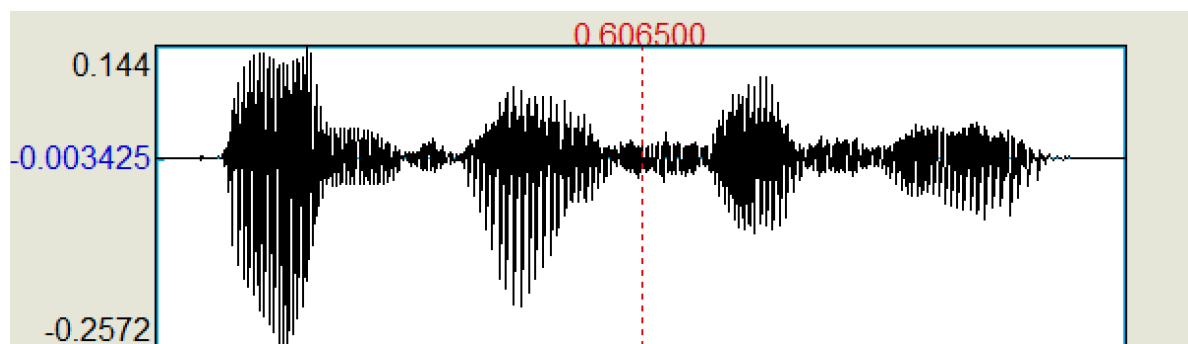
Praat课下实验

任务1

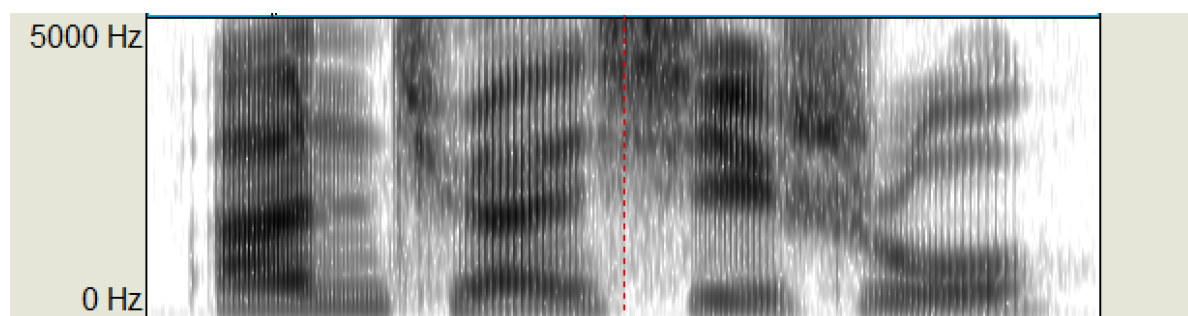
open->read from file 读取文件“GuoL/40004.wav”

1) View & Edit查看

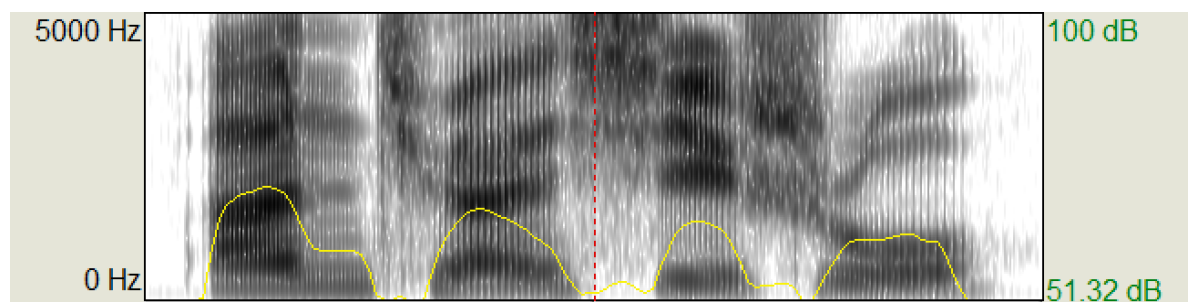
波形:



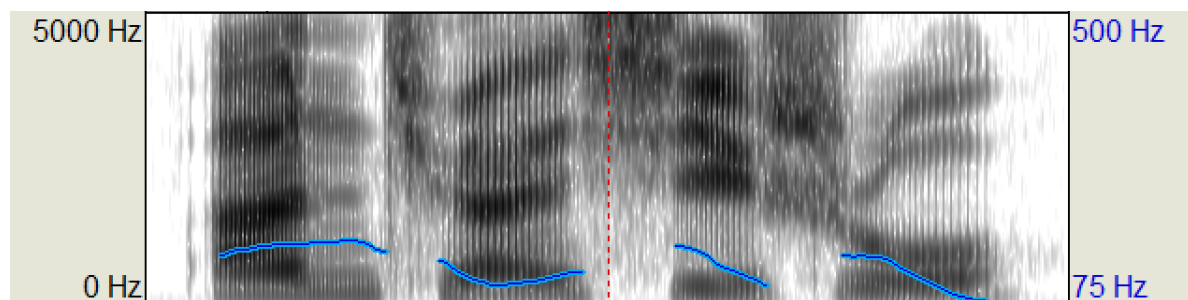
语谱图:



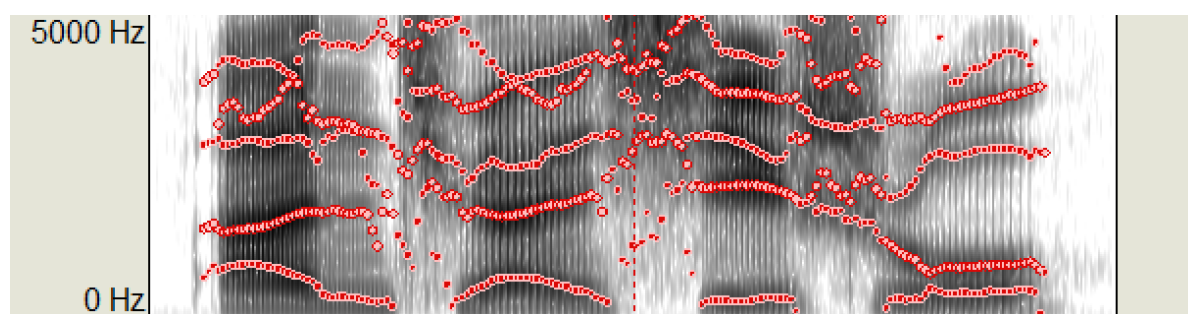
音强:



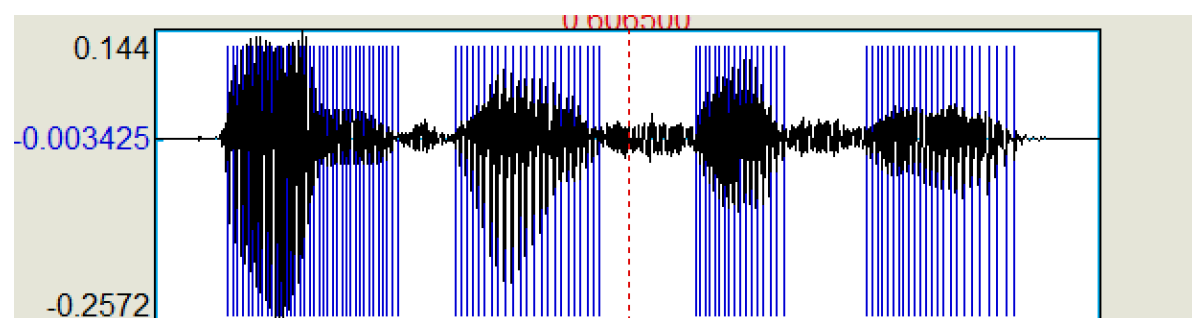
基音轮廓:



共振峰:



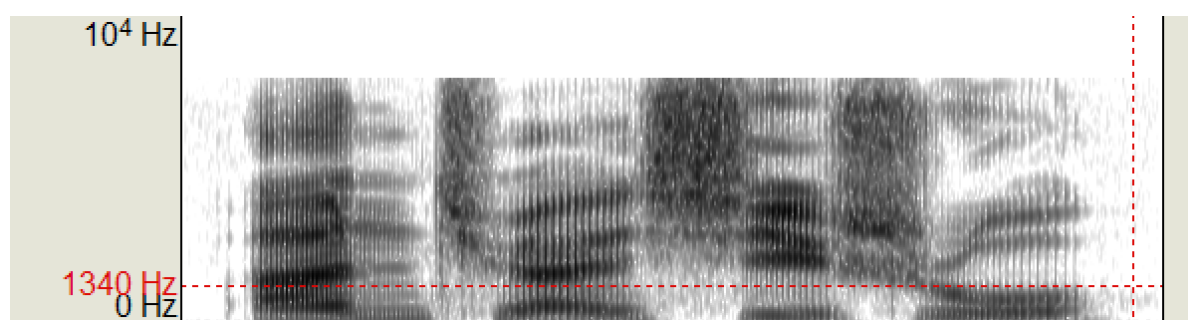
脉冲:



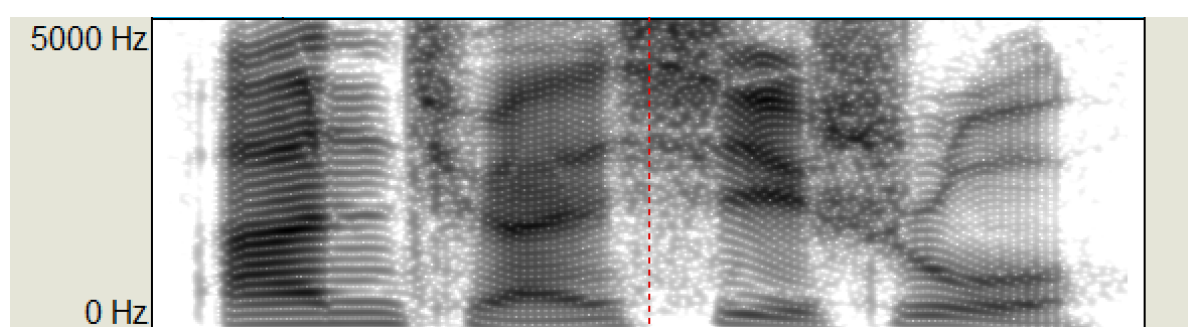
2) 调参

语谱图

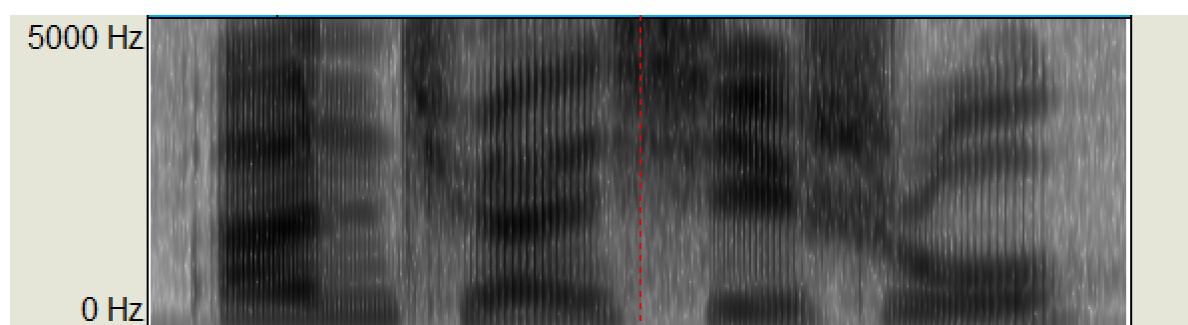
view range: 0-5000 -> 0-10000 (Hz) 频率显示范围变大



window length: 0.005->0.02(s) 黑色条带变细

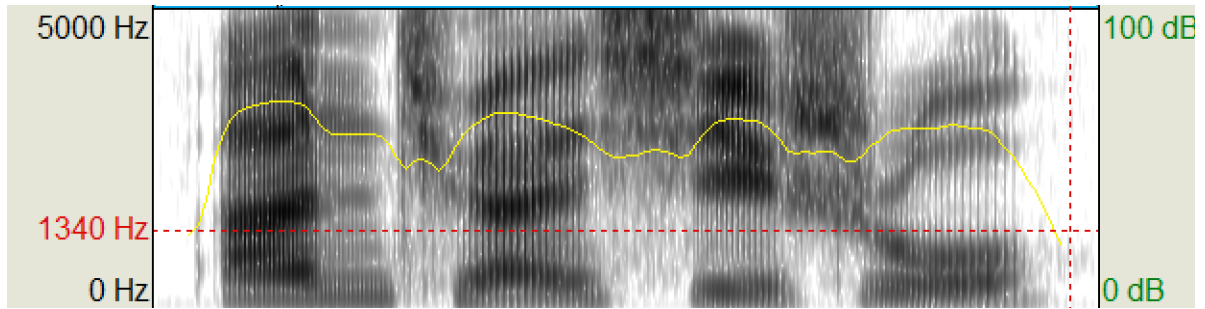


dynamic range: 70->140(dB) 整体颜色变暗



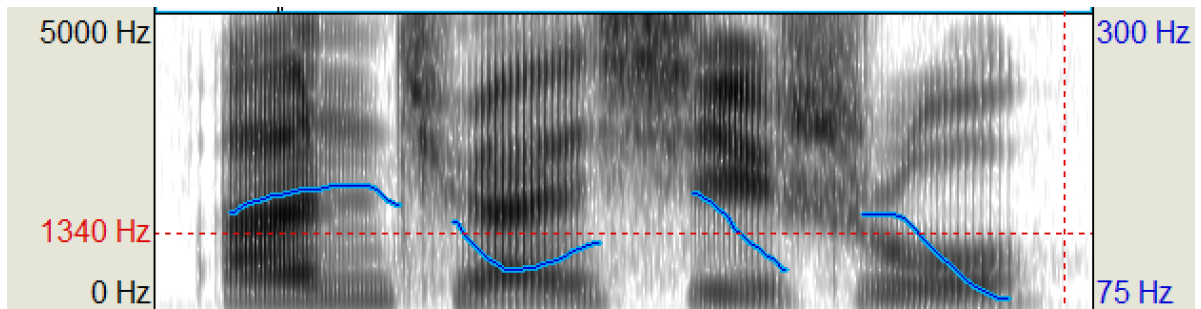
音强

view range: 50-100 -> 0-100 (dB) 观测音强范围变大



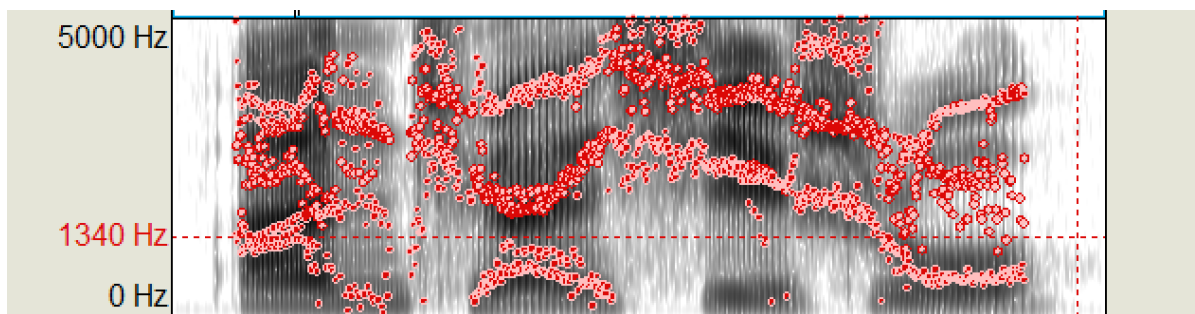
音高

view range : 75-500 ->75-300 (Hz) 观测频率范围变大



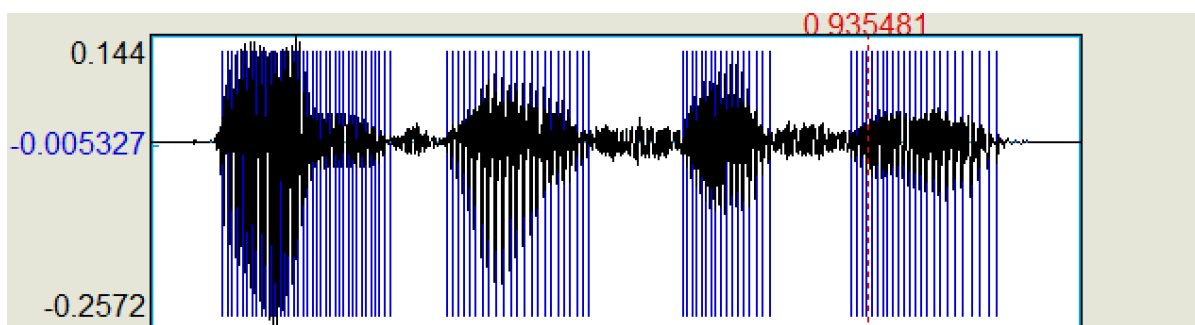
共振峰

改变共振峰数量 5->3 改变窗长0.025->0.005(s)



脉冲

maximum period factor1.3->3.0 maximum amplitude factor1.6->5.0



3)

语谱图 view range参数为频率显示范围；window length为窗长，即短时傅里叶变换的窗口大小；dynamic range是一个等比映射

音强是人主观的音量大小，改变view range即改变了音强的观测范围

基因轮廓: pitch指声音的高低, 由声音频率决定, 声源振动中有一个频率最低的振动, 由它发出的音就是基音; 改变pitch的view range即改变了音高的观测范围

共振峰: formant ceiling为共振峰搜索范围的最大频率上限; number of formants为共振峰的数目; window length为分析窗口长度

脉冲: maximum period factor为抖动计算中, 连续区间之间允许的最大抖动; maximum amplitude factor是最大振幅因子

4)

基于语音信号短时平稳的假设, 使用对语音加窗的方法进行短时分析

音强:

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)^* w(n-m)]^2 = \sum_{m=n}^{n+N-1} [x(m)^* w(n-m)]^2$$

音高:

使用短时自相关法进行音高特征提取:

$$R_n(k) = \sum_{m=-\infty}^{\infty} [x(m) \cdot w(n-m) \cdot x(m+k) \cdot w(n-(m+k))]$$

对函数值进行归一化

$$R'_n(k) = \frac{R_n(k)}{R_n(0)}$$

自0之后, 是上公式值结果取到峰值的k即代表波形的短时周期, 对其取倒数可以得到pitch

语谱图:

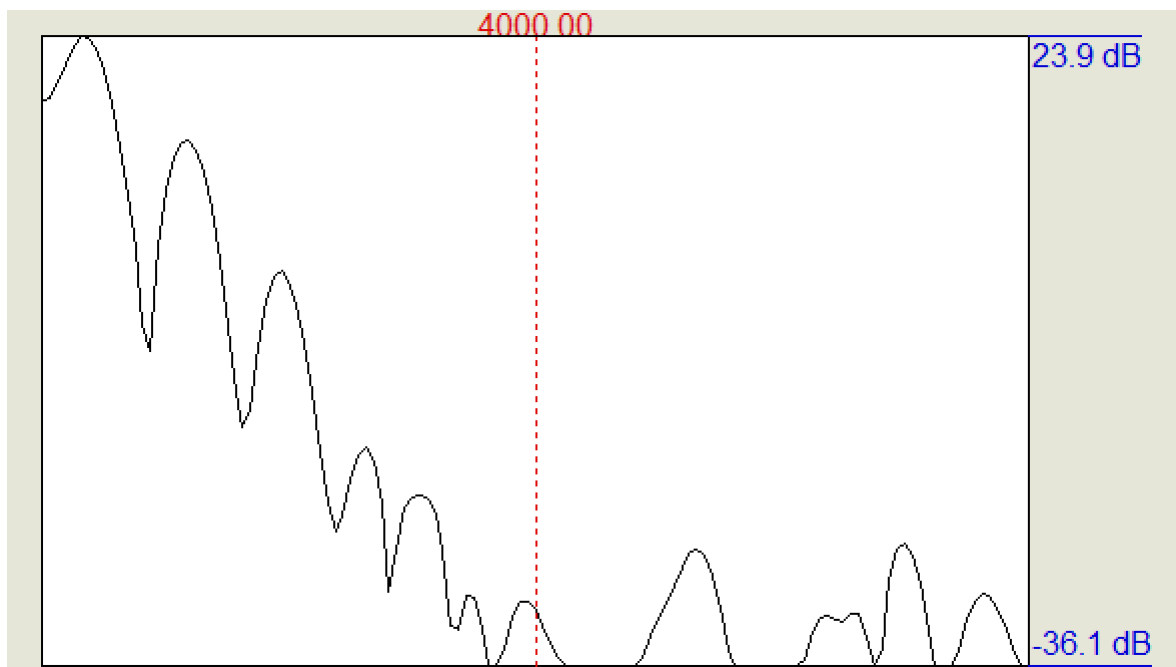
使用FFT方法

$$\text{STFT}\{x(n)\} \equiv X(n, w) = X_n(e^{jw}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-jwm}$$

5)

共振峰是声音的频谱中能量相对集中的区域, 共振峰作用在原始的频谱图上得到了最终的频谱图, 根据源-滤波器模型, 基音和其谐波由声源振动产生, 而共振峰体现了声道的特性, 共振峰的作用是使得源频谱图上某些频率的声音信号得到强度上的增加, 而某些频率的声音信号削弱, 最终形成输出的声音的频谱图。

获取spectral slice: Spectrum->view spectral slice



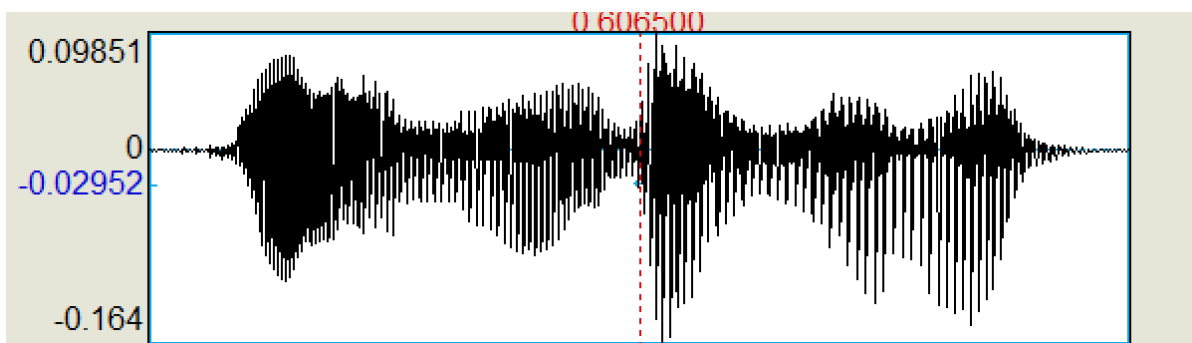
6)

脉冲：常是指电子技术中经常运用的一种像脉搏似的短暂起伏的电冲击。主要特性有波形、幅度、宽度和重复频率。脉冲是相对于连续信号在整个信号周期内短时间发生的信号，大部分信号周期内没有信号。发语音时声带的不断开启和关闭将产生间歇的脉冲波

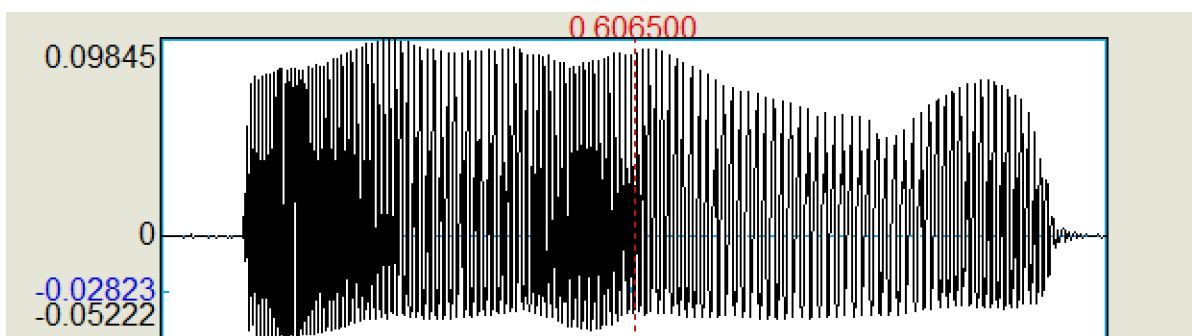
和脉冲有关的参数有maximum period factor和maximum amplitude factor

任务2

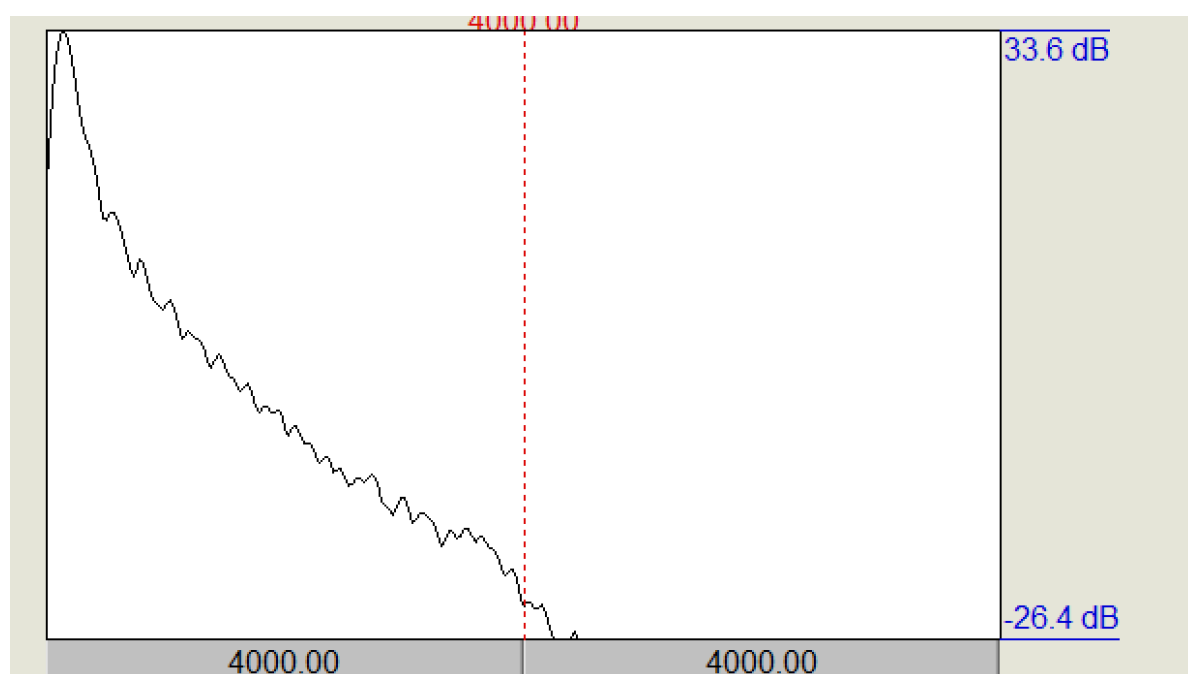
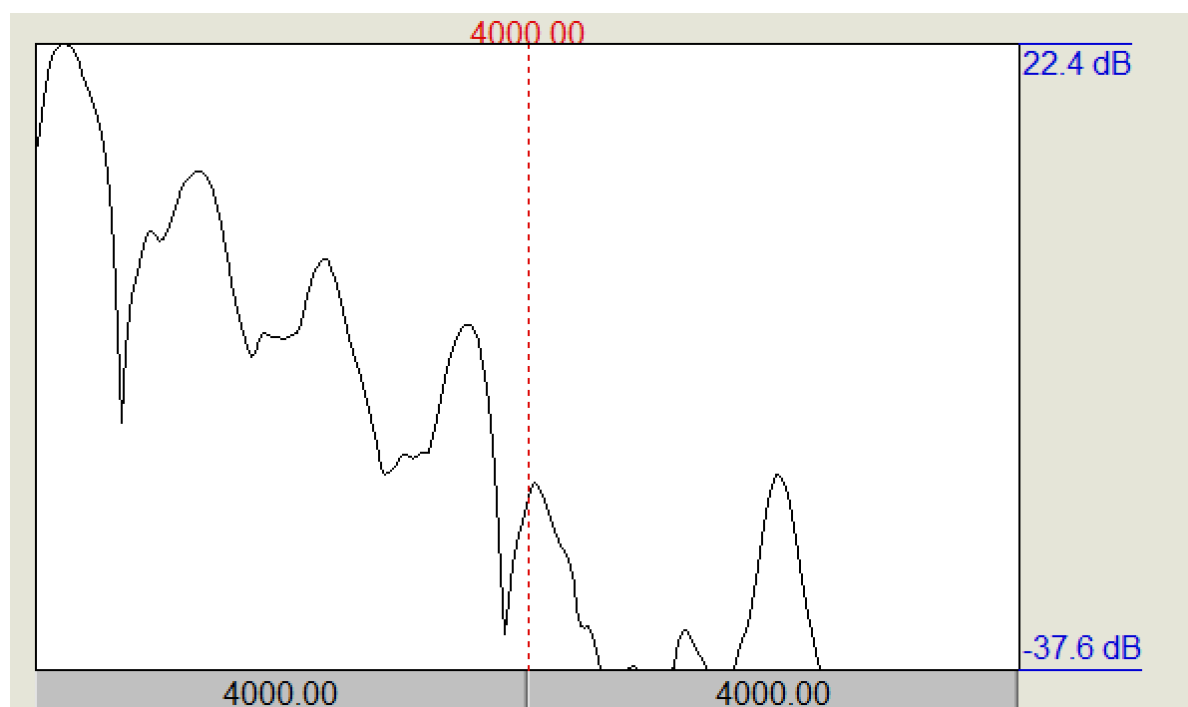
1) 打开40001.wav



打开40001.egg.wav



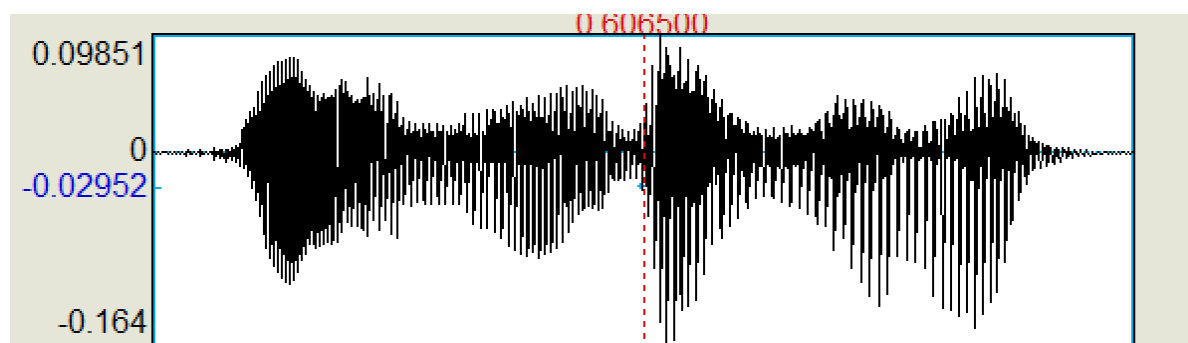
两者同时刻切片

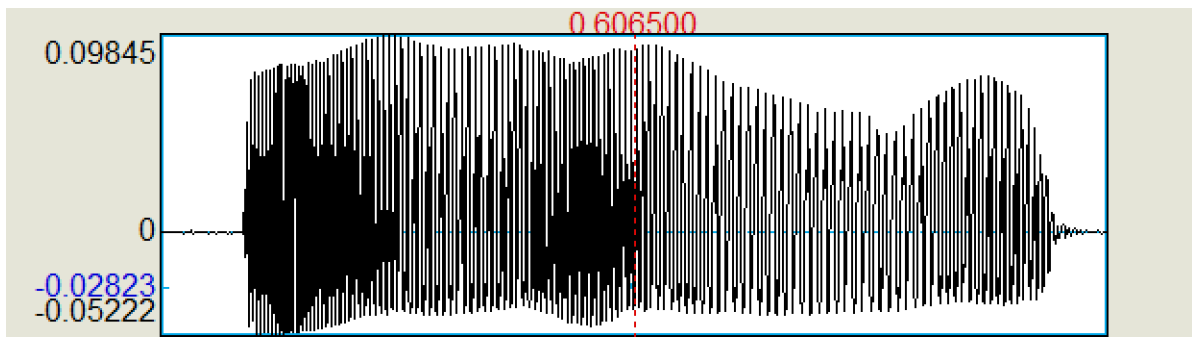


谐波是由基音衍生而来的，频率是基音频率整数倍的声波

后者更接近声带振动后未经调制得到的音频，从切片可以看出

2) 打开40001.wav和40001.egg.wav进行对比

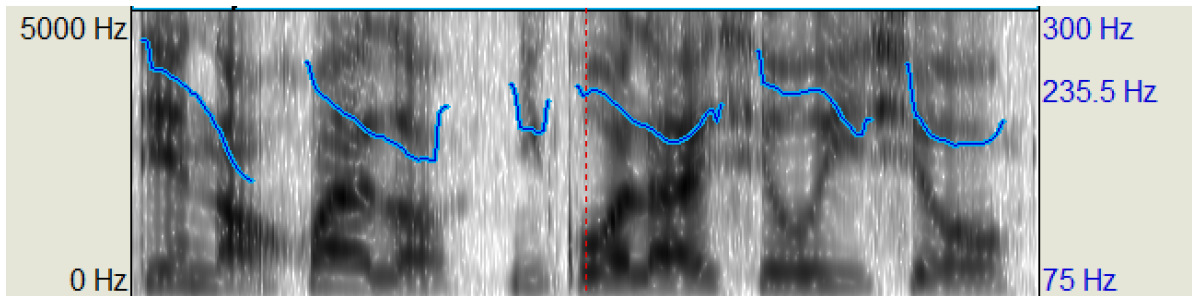




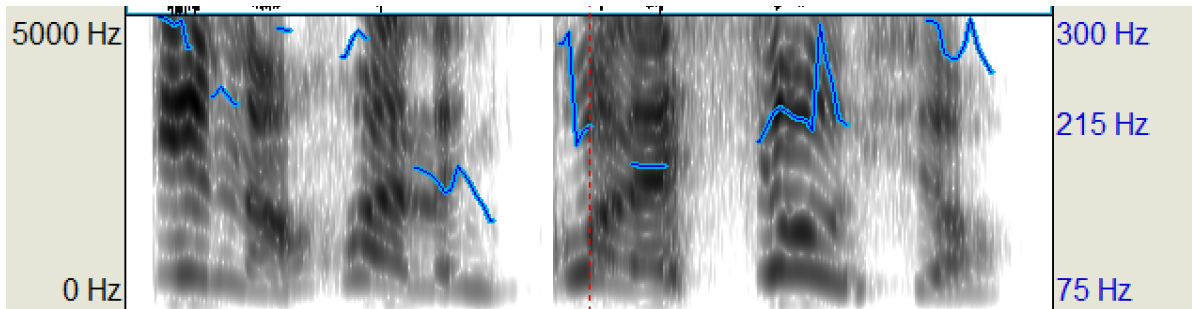
前者波形频谱变化丰富，音强浮动大；后者频率变化少，音强变化小

3)

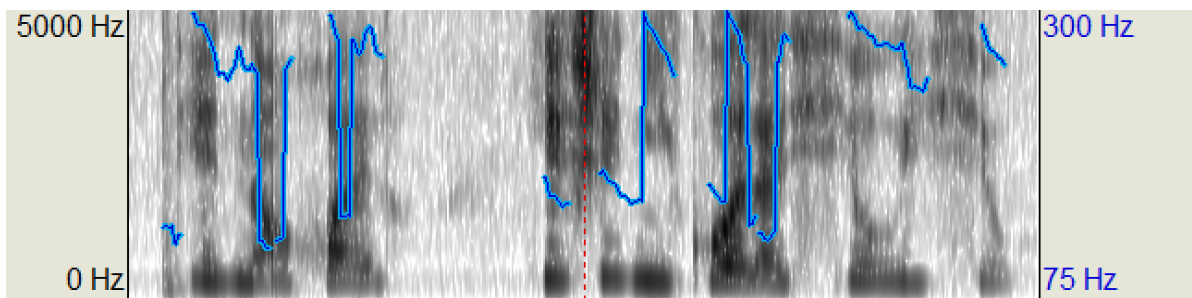
打开exp-0.wav



打开exp-1.wav

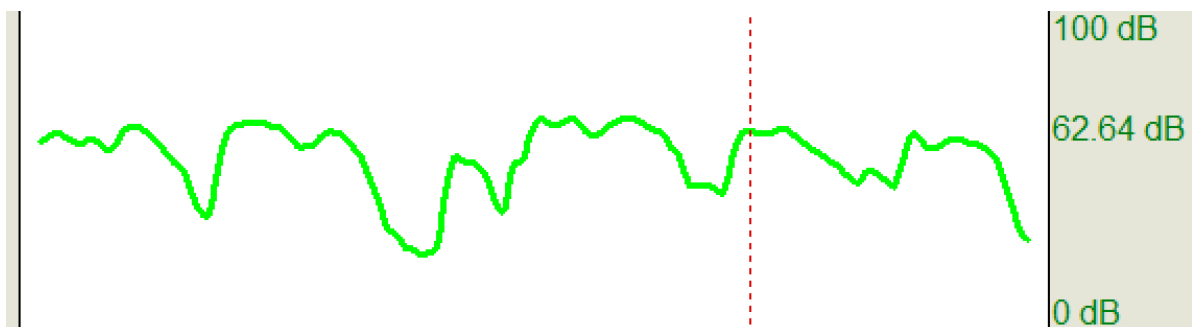


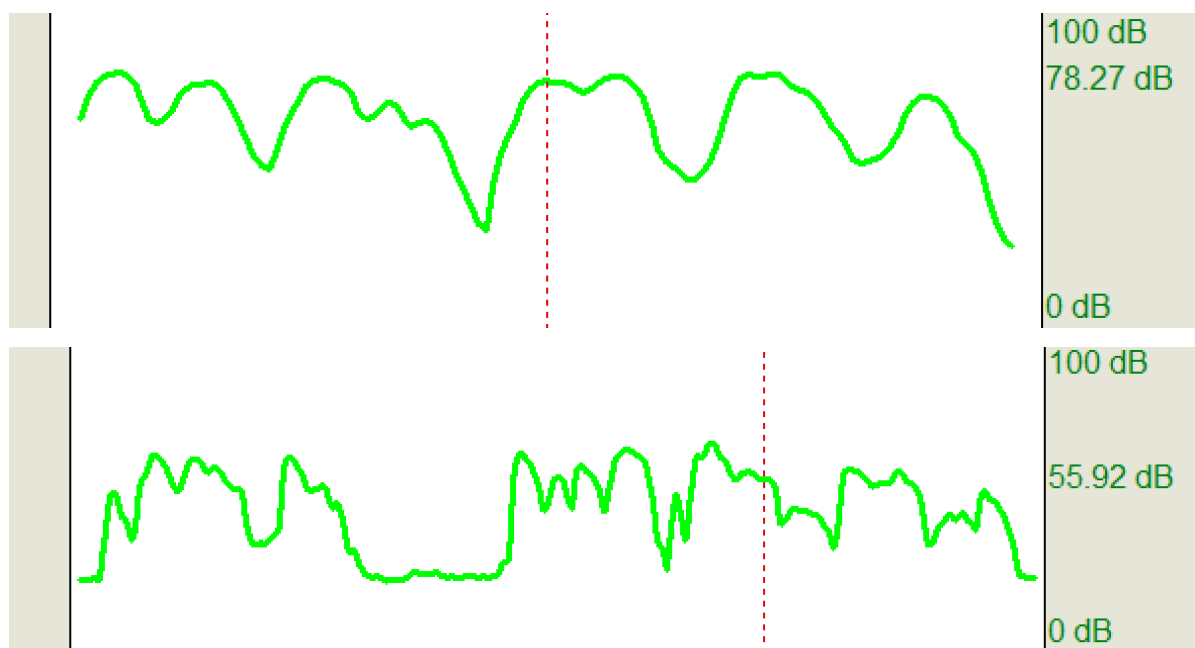
打开exp-4.wav



基音轮廓对比: 中性是连续且相对平滑、愤怒是不连续并忽高忽低、悲伤是连续但上下浮动大

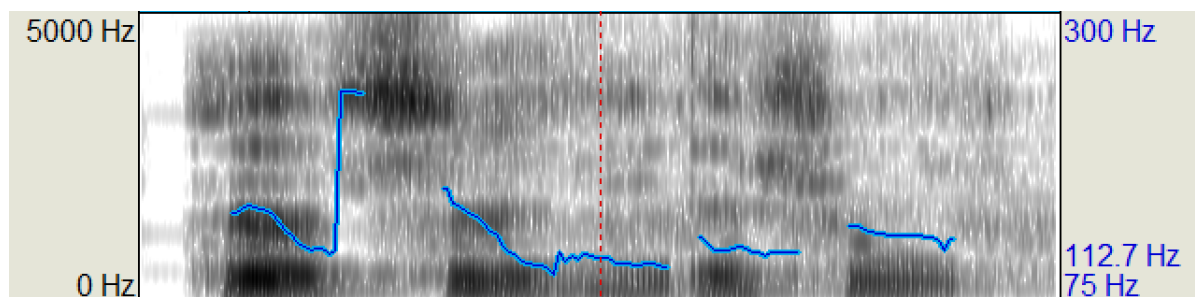
三者的音强也有区别: 中性音强中等，愤怒音强较大，悲伤音强较低



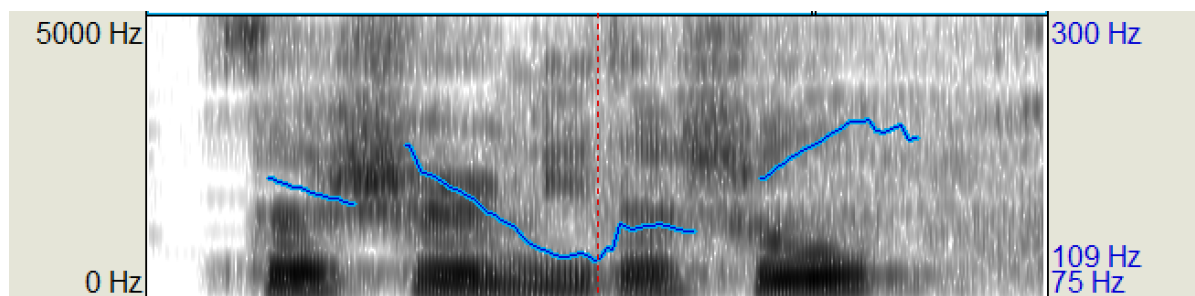


4)

打开book_declaration.wav



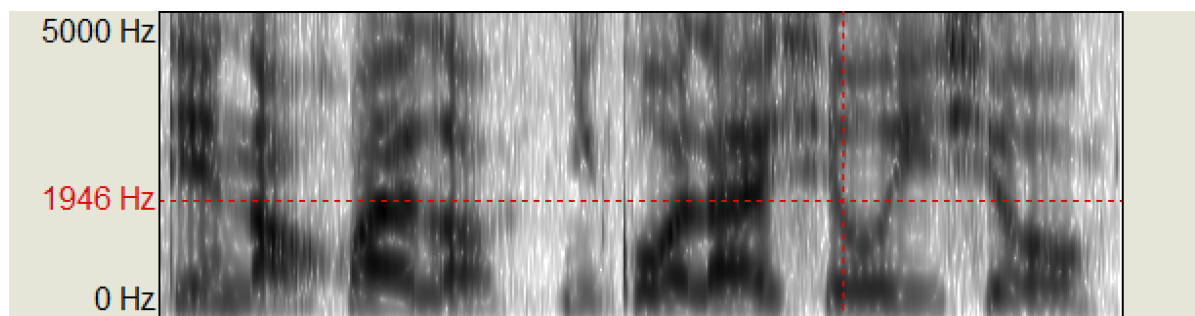
打开book_question.wav



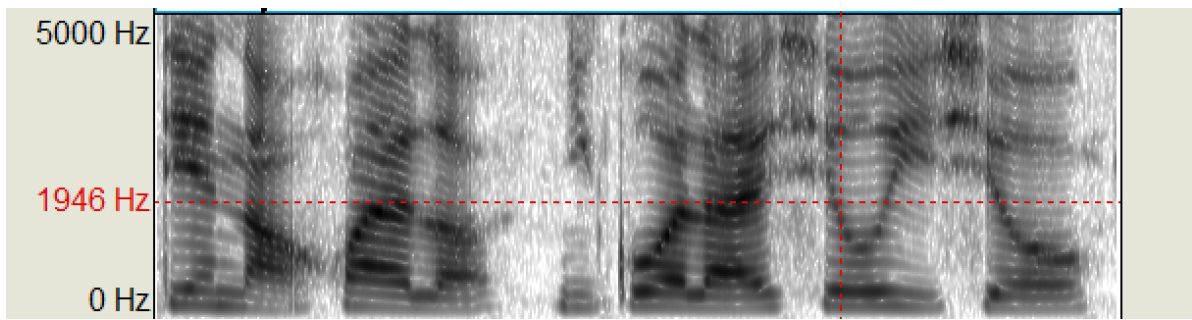
基音轮廓对比：declaration陈述句整体是随时间从高向低走；question疑问句先随时间从高往低走，后又转折上升

5)

绘制宽带语谱图，窗长：4ms



绘制窄带语谱图，窗长：10ms

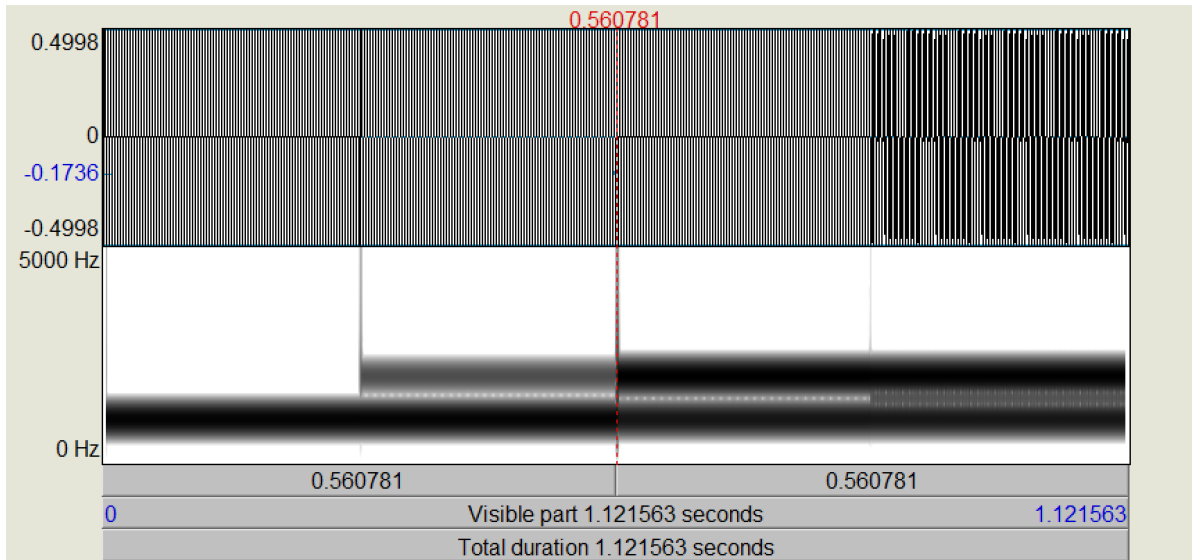


宽带语谱图：时间分辨率更高，频率分辨率更低，横向条纹更宽；

窄带语谱图：时间分辨率更低，频率分辨率更高，横向条纹更窄；

6)

打开abc.wav



一共有三种频率的声音，包括1000hz、1500hz和2000hz

0-0.28s只包含1000hz的声音

0.28-0.56s包含1000hz的声音和1500hz的声音

0.56-0.84s包含1000hz的声音和2000hz的声音

0.84-最后包含三种频率的声音

听到的情况：听到三种声音，第一部分是一种声音，第二部分是另一种声音，第三四部分是一种声音，听感区别不大

掩蔽效应是指高强度的强纯音掩蔽了其附近频率的弱纯音，使得弱纯音不被人耳感知

第三部分和第四部分听感相同是因为1000hz和2000hz的强纯音掩盖了1500hz的弱纯音