

浙江大学

本科实验报告

课程名称：数字视音频处理

姓 名：钱旭峰

学 院：计算机科学与技术学院

系：计算机科学与技术

专 业：计算机科学

学 号：3140102491

指导教师：杨莹春

2016 年 12 月 27 日

浙江大学实验报告

课程名称：数字视音频处理_____ 实验类型：_____ 设计实验

实验项目名称：_____ Project：语音差异分析

学生姓名：钱旭峰_____ 专业：计算机科学与技术_____ 学号：3140102491

同组学生姓名：无_____ 指导老师：杨莹春_____

Project：语音差异分析

一、 实验目的和要求

1. 掌握如何用 GMM（Gaussian Mixture Model）的方法来进行说话人辨识（Speaker Identification）；
2. 掌握如何利用 parrt 软件进行元音三角形的分析；
3. 掌握说话人识别的基本原理和过程；
4. 理解 MFCC 特征提取。

二、 主要仪器设备

1. matlab R2015a：用于编程
2. 录音机：用于语音的录制、采集
3. Praat：用于语音的元音三角形及基频图等分析

三、 操作方法与实验步骤

1. 语音采集使用
















两首诗歌的朗诵：

登鹳雀楼 王之涣

白日依山尽，黄河入海流。欲穷千里目，更上一层楼。八千里路云和月。

在这个实验中我们首先录下 15 周的录音，在以其他同学的第一周的录音作为训练数据，训练说话人模型的函数 `gmm_estimate`，以自己剩余 14 周的录音作为测试数据，检测识别的正确率，实验中一般会出现误识别的现象，我们将分析误识别的几段语音和我们自己的语音的特征差别，并设计检测容易误识别的语音样本的算法

首先将其他同学第一周的录音建立模型，自己的 2~15 周录音与建立模型的录音放在一起作为测试数据，并找出自己所有语句中测试得分比模型录音高（被误识别为模型录音）的那些语句。

 3140102491-W1	2016/9/25 15:21	文件夹
 3140102491-W2	2016/10/1 12:08	文件夹
 3140102491-W3	2016/10/11 22:14	文件夹
 3140102491-W4	2016/10/15 13:09	文件夹
 3140102491-W5	2016/10/24 11:15	文件夹
 3140102491-W6	2016/11/7 0:10	文件夹
 3140102491-W7	2016/12/16 19:14	文件夹
 3140102491-W8	2016/12/16 19:21	文件夹
 3140102491-W9	2016/12/16 19:31	文件夹
 3140102491-W10	2016/12/17 18:51	文件夹
 3140102491-W11	2016/12/17 18:56	文件夹
 3140102491-W12	2016/12/17 19:06	文件夹
 3140102491-W13	2016/12/18 12:14	文件夹
 3140102491-W14	2016/12/18 12:19	文件夹
 3140102491-W15	2016/12/18 12:24	文件夹

模型识别:

先用自己第一周的模型测试, 计算出最低分数 self score。再依次对其他每个同学的录音建模, 用自己的录音测试, 并筛选出自己的录音中测试得分最接近于模型的录音语句。由于每建立一个模型, 都能测试出最接近模型录音的一句自己的录音, 而我们一共可以建立 46 个模型, 这样一来就能统计出自己录音中最容易被误识别的那些语句。只要分析出这些语句的特征, 即可推测出容易被误识别的特征。

代码通过 `wavread` 以及 `melcepst` 读取 .wav 文件并提取特征 `train_feature` (12 维 MFCC), 然后使用 `gmm_estimate` 为说话人训练模型 (16 阶 GMM), 得到模型的 3 个参数 `[mu, sigma, c]`。最后将被测特征 `test_feature` 和要比对的说话人模型参数传入函数 `lmultigauss`, 即得到该被测特征与指定模型的比对得分 `lY`。

代码中的 `[YM, Y] = lmultigauss(x, mus, sigm, c)` 试图计算模型匹配得分, 得到 `YM, Y` 两个参数。对 `Y` 求 `mean`, 即该段测试语音得分与对应模型的比对得分。

代码训练了 1 个说话人的模型, 测试了 13 组特征, 最终 `score` 为得分数组, 得分数组的第 `i` 行第 `j` 列的值代表第 `i` 个测试语音与第 `j` 个说话人模型的得分, 分数高则表示更匹配。

四、 实验结果与分析

实验结果：

-17.6661	'3140102491-W1'	-17.6661		-20.6401	'3140104620-W1'	-29.8344
-13.8796	'3140102491-W1'	-13.8796		-19.0009	'3140104327-W1'	-29.7069
-14.3506	'3140102491-W1'	-14.3506		-21.9167	'3140104620-W1'	-32.3619
-16.9761	'3140102491-W1'	-16.9761		-20.489	'3140104620-W1'	-27.8008
-14.9119	'3140102491-W1'	-14.9119		-21.1029	'3140104620-W1'	-30.1907
-14.1885	'3140102491-W1'	-14.1885		-22.757	'3140104726-W1'	-34.7332
-19.8529	'3140104620-W1'	-20.6905		-20.4178	'3140104726-W1'	-29.6304
-16.8296	'3140104620-W1'	-18.1267		-19.6011	'3140104108-W1'	-31.9127
-21.4391	'3140104620-W1'	-22.0834		-21.2414	'3140104620-W1'	-30.0815
-15.7144	'3140102491-W1'	-15.7144		-19.9035	'3140104108-W1'	-28.4478
-17.2541	'3140102491-W1'	-17.2541		-21.6498	'3140104620-W1'	-32.0573
-10.7755	'3140102491-W1'	-10.7755		-21.9033	'3140104726-W1'	-32.4292
-18.1783	'3140102491-W1'	-18.1783		-23.9426	'3140104726-W1'	-33.1263
-15.9662	'3140102491-W1'	-15.9662		-24.6596	'3140104620-W1'	-34.6525
-19.7988	'3140102491-W1'	-19.7988		-25.6679	'3140104620-W1'	-35.1164
-18.088	'3140102491-W1'	-18.088		-23.0749	'3140104108-W1'	-32.8147
-16.6759	'3140102491-W1'	-16.6759		-23.1415	'3140104620-W1'	-32.3573
-15.2554	'3140102491-W1'	-15.2554		-26.0695	'3140104726-W1'	-35.4264

等共有 15 张表（详见 result.xlsx 文件）

匹配的正确率：



曲线很奇怪

之后我分析了对于第二周的 N3.wav 的误识别的情况：

在第二周的 N3.wav 3140104620 同学将近比我高 10 分，以下是具体分析：

听感特征：

第一周语音：

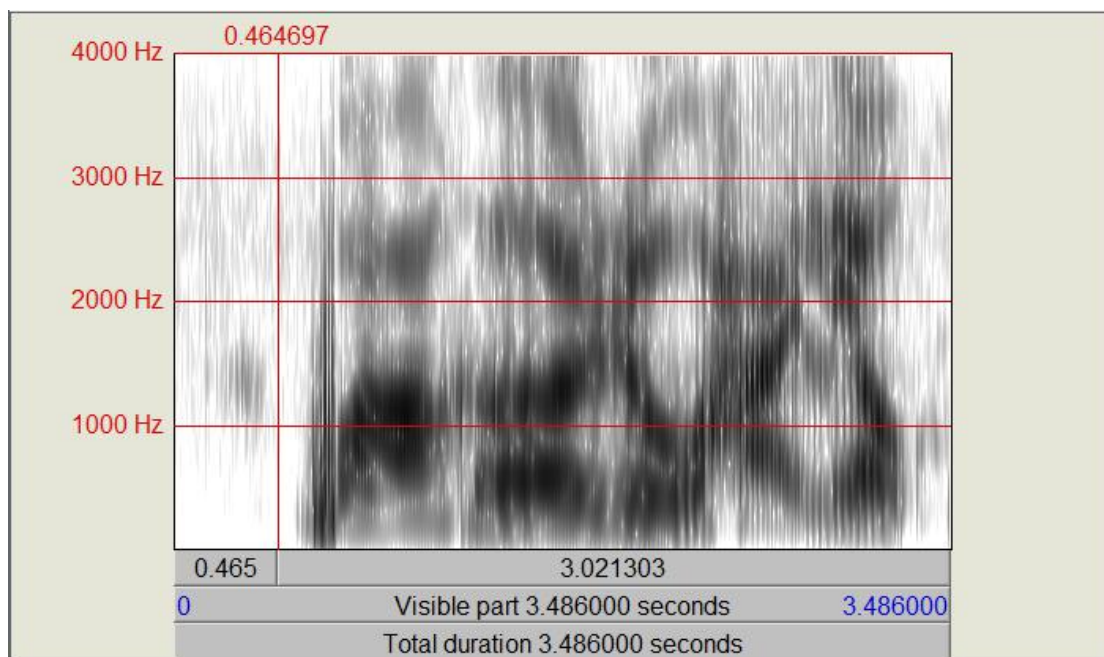
1. 噪音较多
2. 呼吸气的声音较重
3. 音色较为低沉
4. 前面慢后面快

误识别语音(3140104620-W1)：

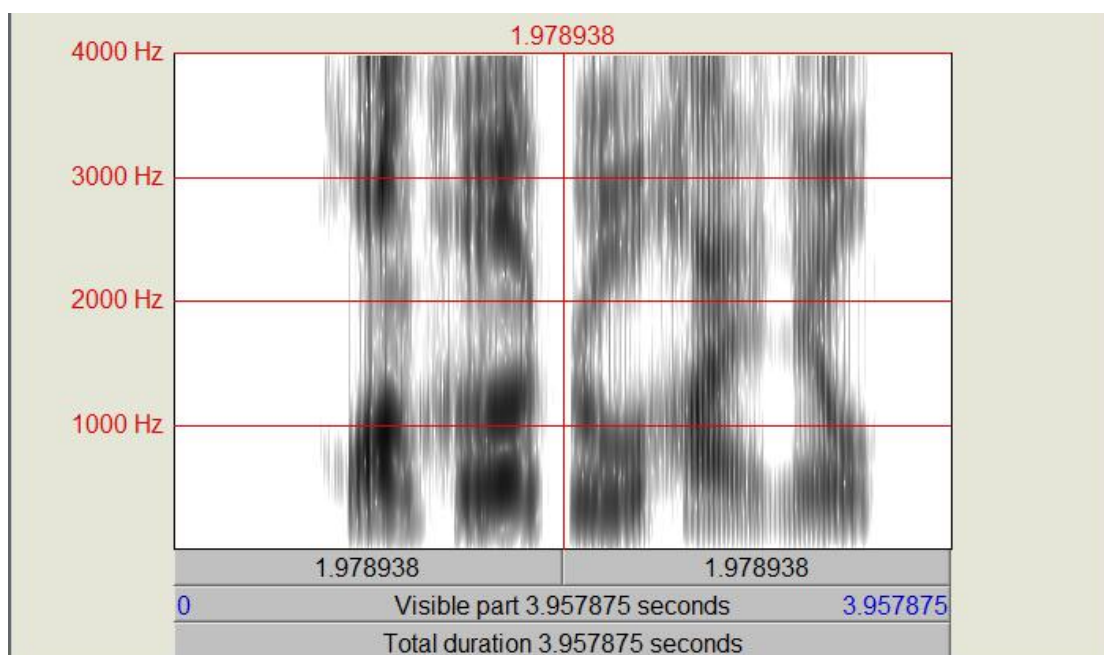
1. 噪音较少
2. 没有呼吸气的声音
3. 音较高
4. 速度均匀

频谱分析：

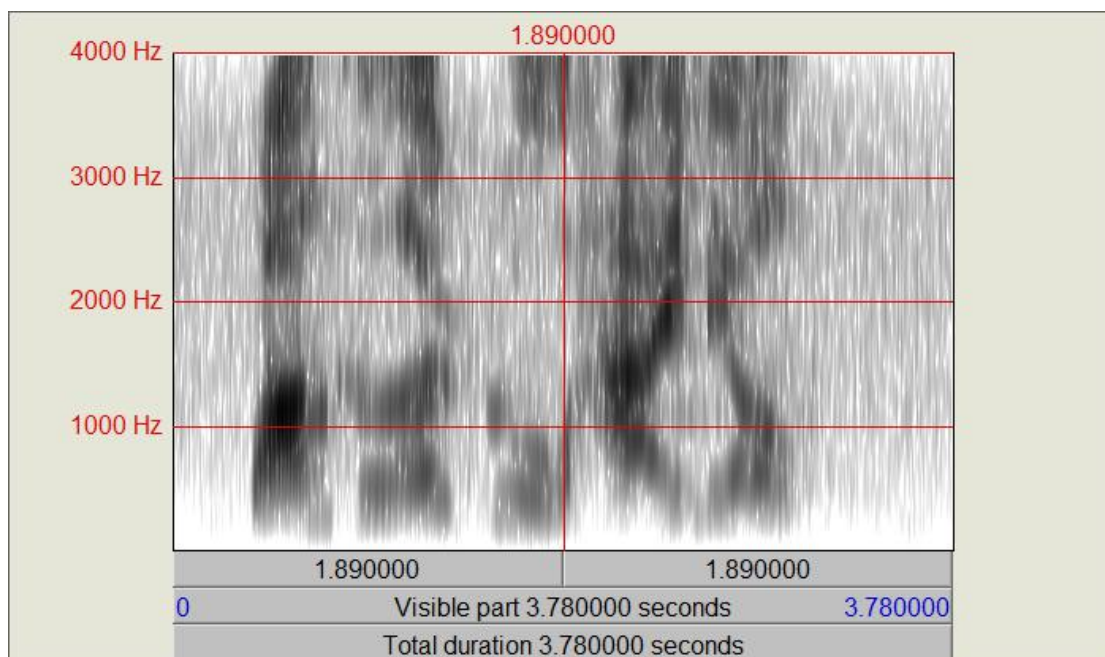
第一周录音：



误识别录音:

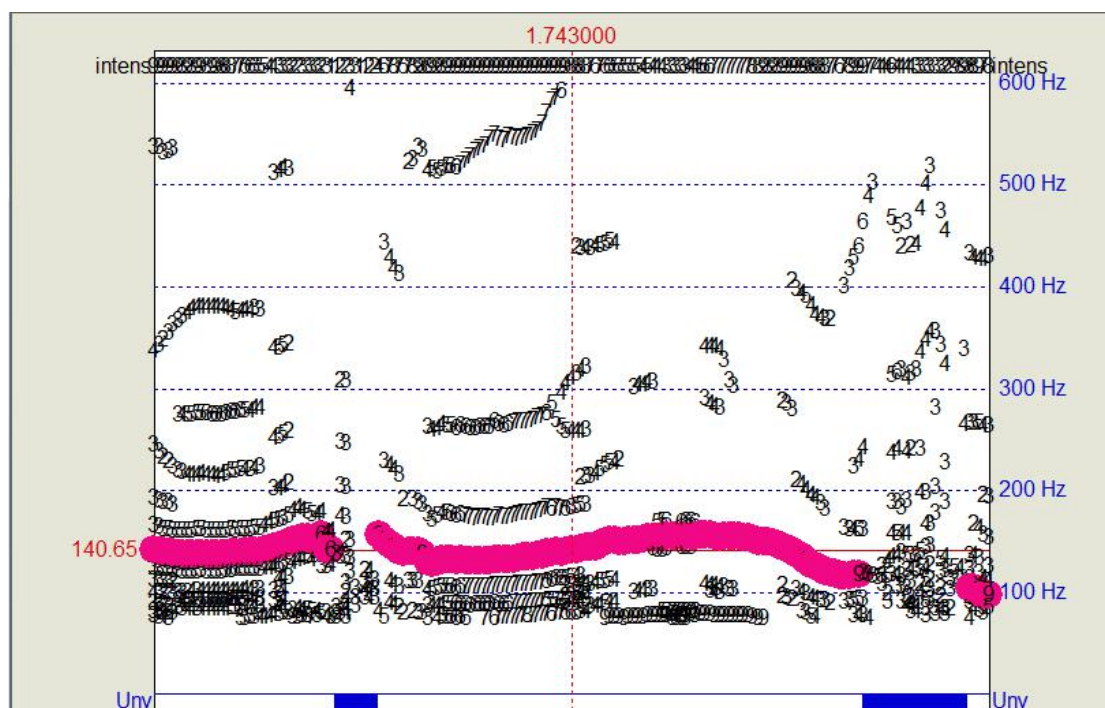


标准:

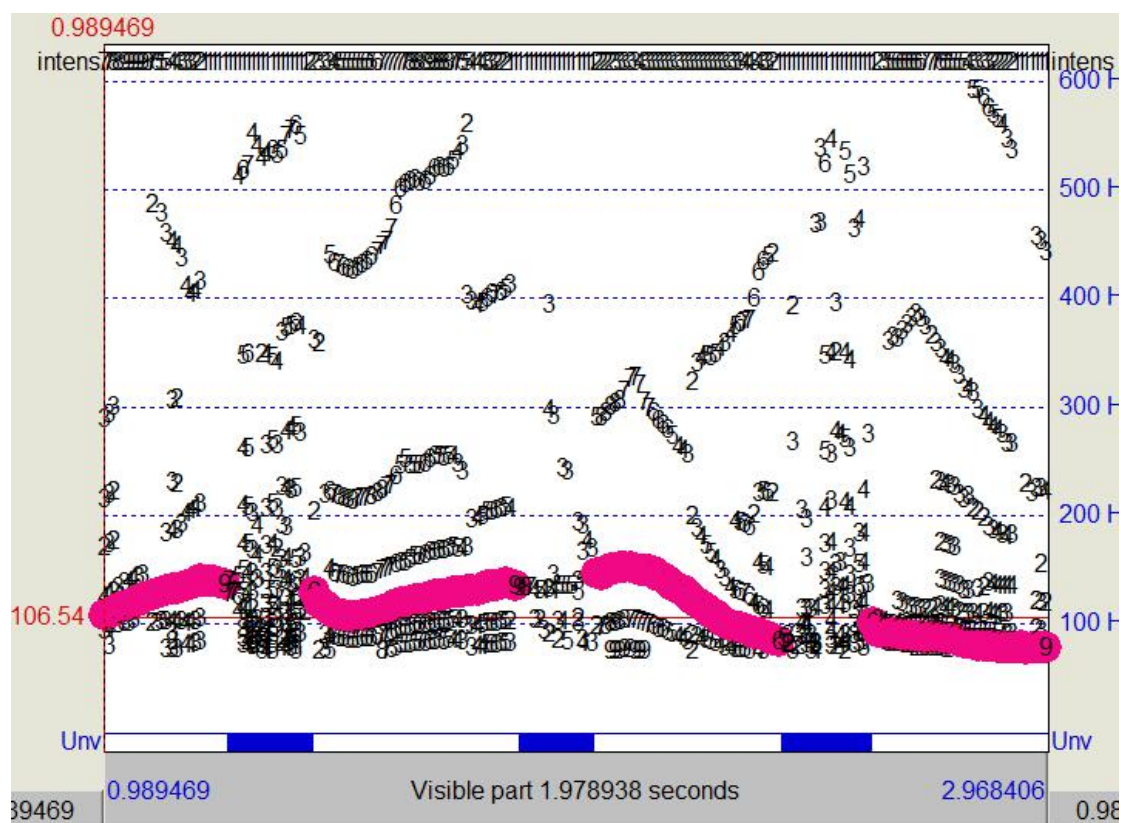


基频分析：

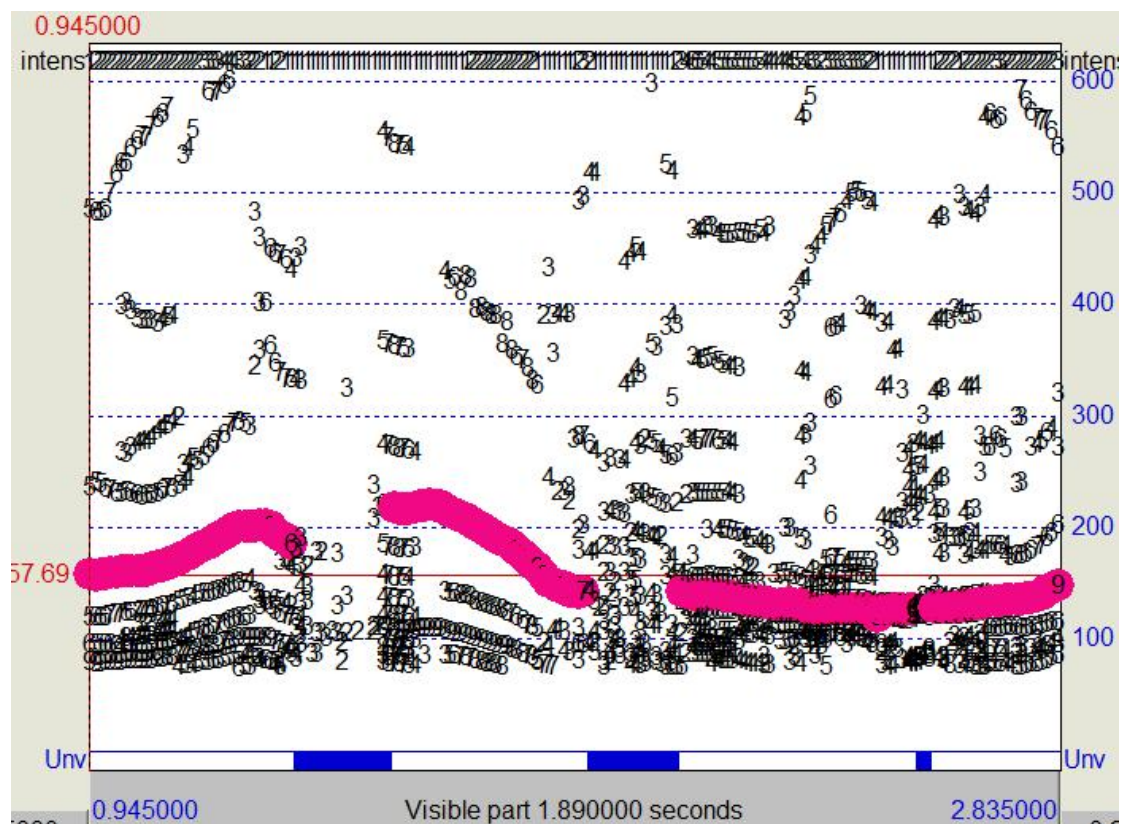
第一周录音：



误识别语音：

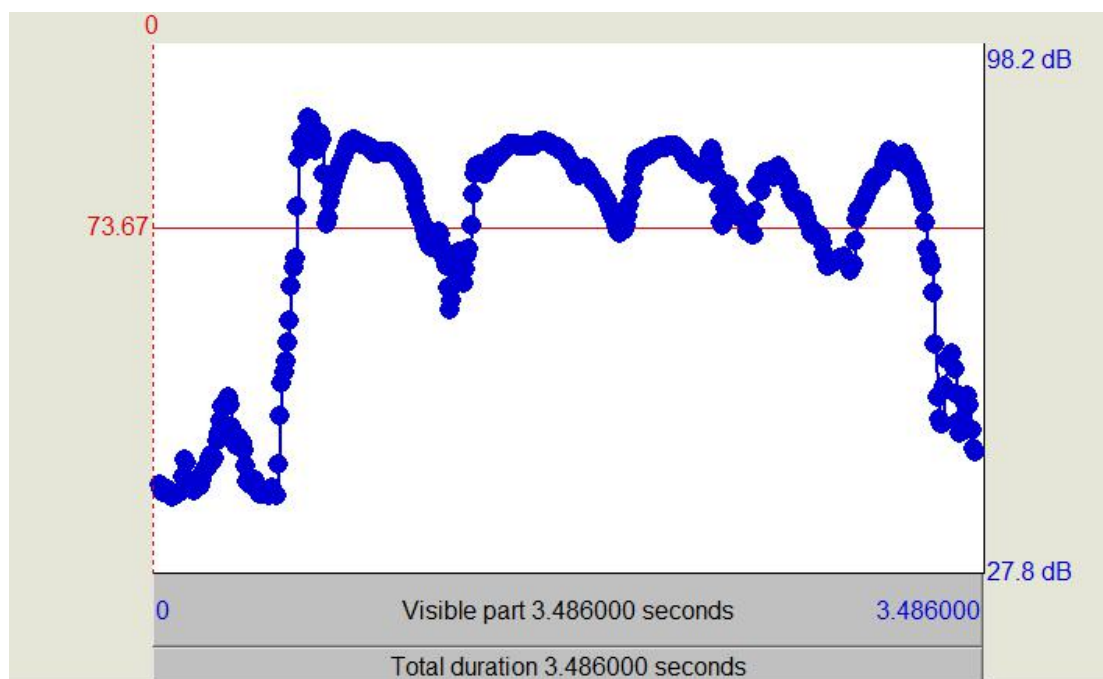


标准:

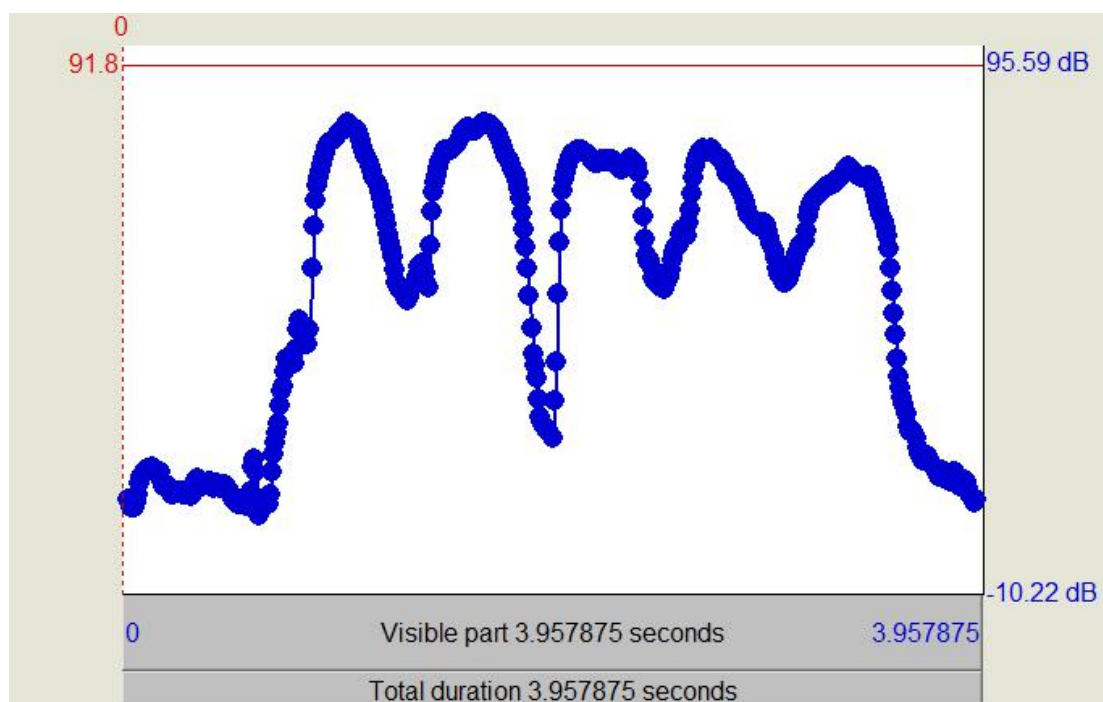


强度分析：

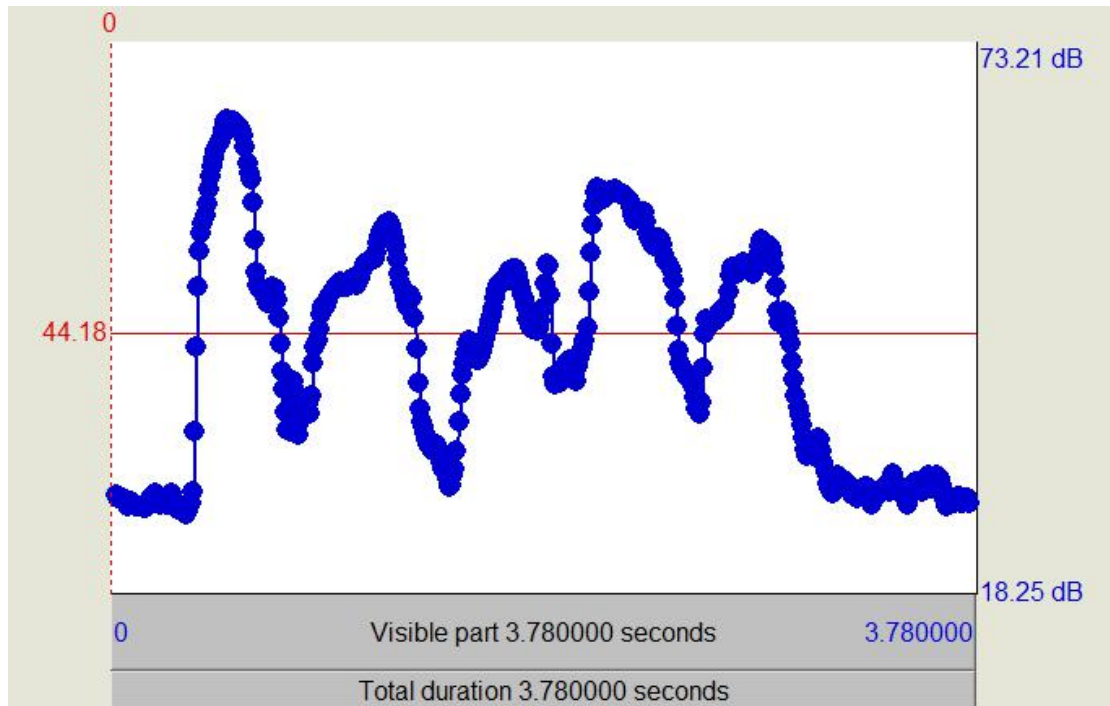
第一周录音：



误识别语音：

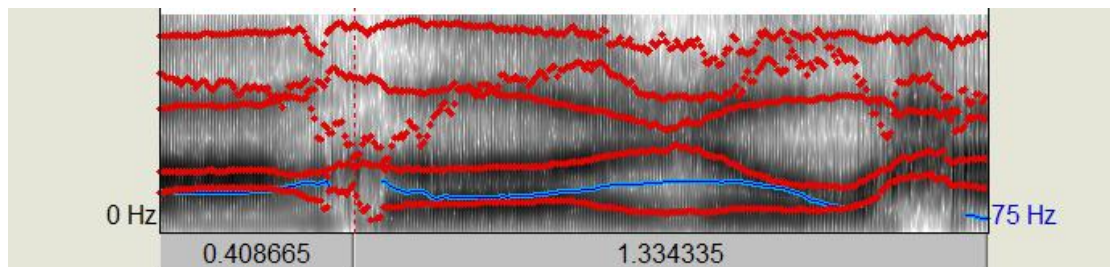


标准：

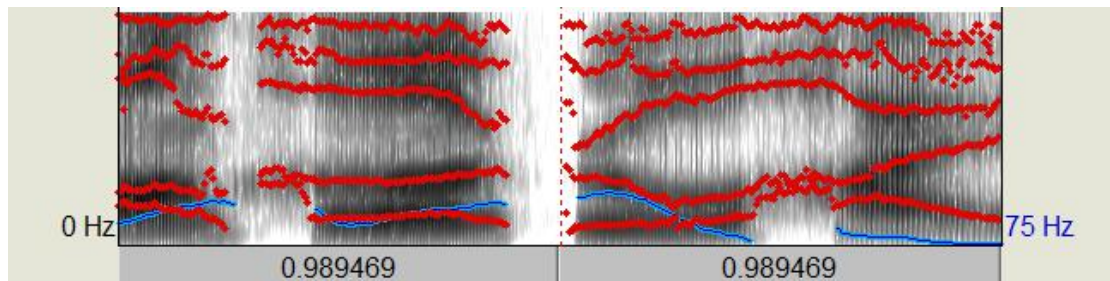


共振峰分析：

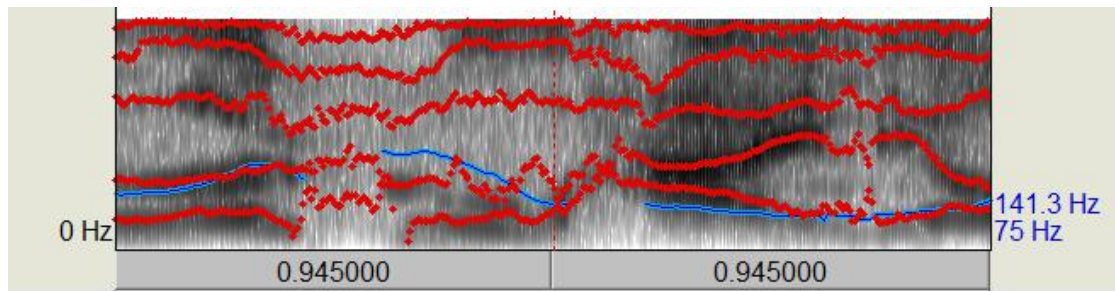
第一周语音：



误识别语音：



标准：



五、心得与体会

经过本次实验，我对于 Praat 这个软件已经可以做到熟练使用，了解了如何分析两段语音的差别，提取语图，基频，强度，共振峰。同时，我解了如何识别以及对这个算法的改进，我语音识别在将来的生活中会有很多的实际作用，将会对我未来的发展有很大的作用。