

浙江大学实验报告

课程名称：____数字视音频处理____ 实验类型：____综合____
实验项目名称：____语音差异分析与说话人识别____
学生姓名：____* *____ 专业：____数字媒体技术____ 学号：____
同组学生姓名：____无____ 指导老师：____杨莹春____
实验地点：____ 实验日期：____2019 年 1 月 5 日____

一、 实验目的和要求

1. 15 次以上诗歌朗诵录音（每周 1 次以上）
2. VOICEBOX 作 MFCC, GMM 训练与测试，记录结果。
3. 找出识别得分偏低的语句，用 PRAAT 分析其与模板语音之间的听感、特征等差异
4. 通过数据处理和创意设计，展示语音的时间变化趋势

二、 实验环境

编程工具：Matlab R2018b

录音工具：Windows10 录音机

可视化工具：echarts

Praat：语音分析

三、 实验步骤

1. 语音采集使用：

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，共六句话。

每周录制 1 次，采集 15 周共 15 次以上语音，每次分别用正常、快速、慢速三种方式录制。保持同一电脑和麦克风，安静的环境。第一周录音的正常语速部分语音用于训练模拟，其余各次语音按句用于测试模型。

每周的录音存在一个文件夹中，命名格式为 ID-W1、ID-W2.....其中录音文件命名格式 N1、N2.....N6 为正常，F1、F2.....F6 为快速，S1、S2.....S6 为慢速。



2. 使用 VOICEBOX 作 MFCC, GMM 训练与测试，记录结果：

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` 和要对比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 `IY`。

2.1 参数设置：

```
MFCC_size=12;%mfcc的维数
GMM_component=16;%GMM component 个数
mu_model=zeros(MFCC_size,GMM_component);%高斯模型 分量 均值
sigma_model=zeros(MFCC_size,GMM_component);%高斯模型 分量 方差
weight_model=zeros(GMM_component);%高斯模型 分量 权重
```

2.2 训练模型：

```
%train model
%使用1_1~1_6训练
for i=1:num_train
    train_file=[train_file_path 'N' num2str(i) '.wav'];
    [wav_data ,fs]=audioread(train_file);
    train_feature=melcepst(wav_data , fs);
    all_train_feature=[all_train_feature;train_feature];
end
[mu_model, sigma_model, weight_model]=gmm_estimate(all_train_feature', GMM_component);
```

2.3 测试：

```

for i=1:3
    for j=1:6
        test_file=[test_file_path file(i) num2str(j) '.wav'];
        [wav_data ,fs]=audioread(test_file);
        test_feature=melcepst(wav_data ,fs);
        [lYM, lY] = lmultigauss(test_feature', mu_model, sigma_model, weight_model);
        score((i-1)*6+j) = mean(lY);
        fprintf(' Test: %s_%d score:%f\n', file(i), j, score((i-1)*6+j));
    end
end

```

2.4 结果打印:

```

%result
[max_score, max_id]=max(score);
[min_score, min_id]=min(score);
fprintf(' Max score:%f\nMin score:%f\n', max_score, min_score);
fprintf(' Max-id:%d\nMin-id:%d\n', max_id, min_id);

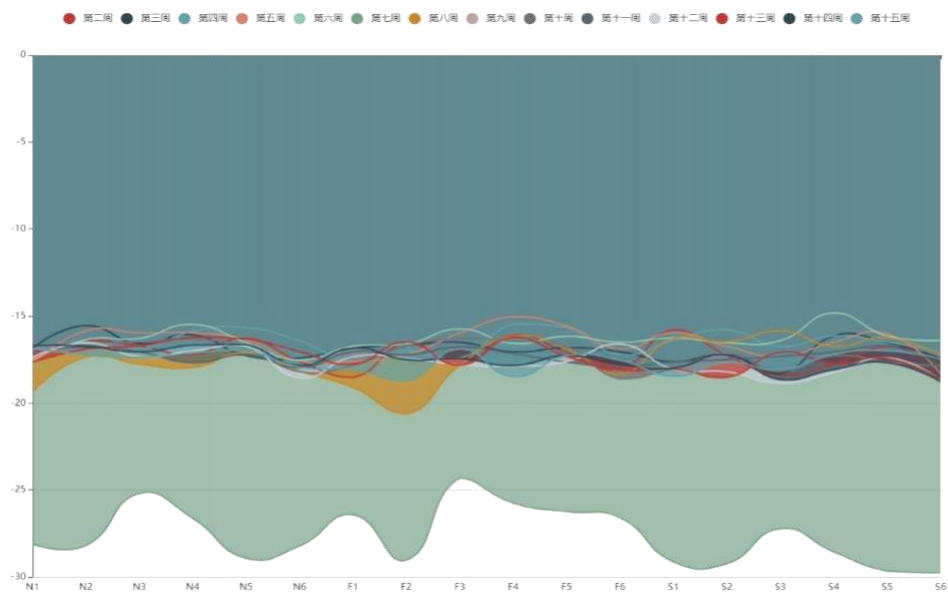
```

将每周的结果存入 txt 文件中, 并在 matlab 中 load 为 18(每周 18 句)*13(15 周除去第 1、7 周)的矩阵:

18x13 double

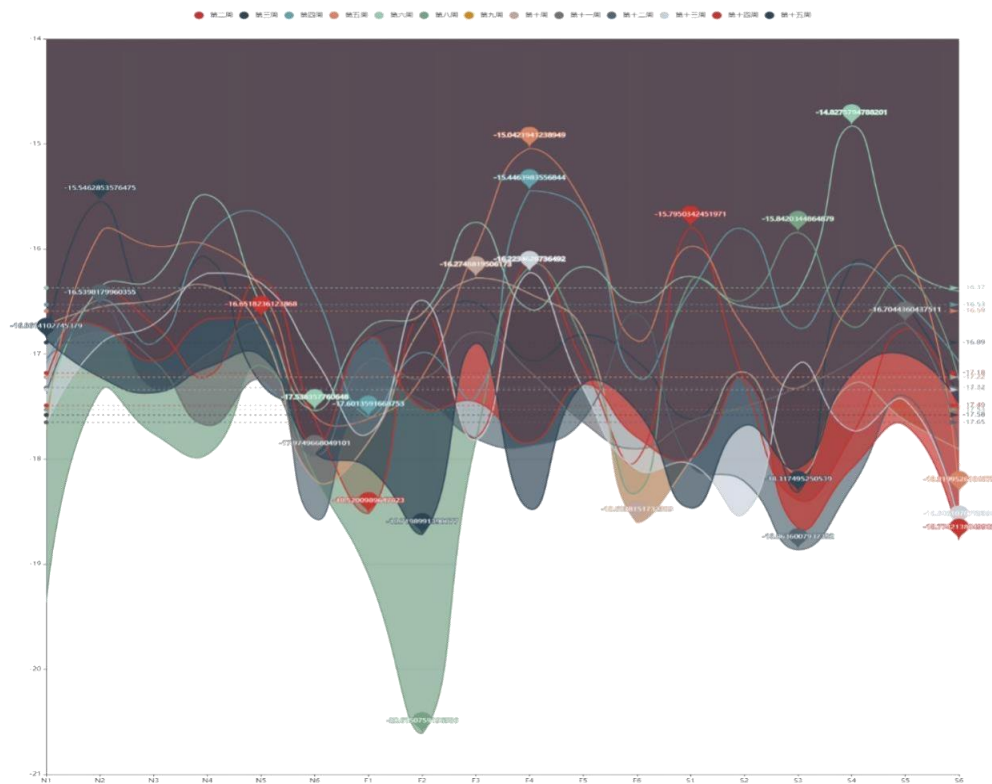
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-17.2419	-16.8547	-17.0917	-17.7169	-17.5280	-19.3668	-16.7219	-16.7219	-17.0401	-17.3763	-17.6679	-16.7117	-16.8614
2	-16.4356	-15.5463	-16.4840	-15.8504	-16.3654	-17.3589	-16.5517	-16.5517	-16.7930	-16.5398	-16.9067	-16.7350	-17.2197
3	-16.7248	-16.5896	-16.8976	-15.9732	-16.2855	-17.7589	-16.4894	-16.4894	-17.0623	-17.3334	-16.7162	-17.0642	-17.3687
4	-17.2154	-16.0825	-15.8821	-15.9623	-15.4852	-17.9563	-16.3475	-16.3475	-17.6783	-17.0009	-16.2302	-16.6664	-17.1278
5	-16.2887	-17.2812	-15.6651	-16.3869	-16.5321	-17.1115	-16.7684	-16.7684	-17.2452	-16.7515	-16.3243	-16.6518	-17.0145
6	-17.6172	-17.4139	-16.4265	-17.5854	-17.5384	-18.1622	-18.1962	-18.1962	-17.9750	-18.5673	-17.0442	-17.8531	-17.9345
7	-18.5201	-16.8526	-17.6014	-17.6174	-16.7450	-19.1014	-17.8372	-17.8372	-17.0625	-17.3520	-17.7598	-16.8421	-18.0782
8	-16.6078	-16.6802	-16.9847	-17.2174	-16.6168	-20.6151	-16.7338	-16.7338	-17.2339	-17.4388	-16.4883	-17.5090	-18.7199
9	-17.7956	-16.5148	-17.4126	-15.9952	-15.7461	-17.8295	-16.2749	-16.2749	-16.7921	-17.8048	-17.8179	-17.4534	-16.8991
10	-16.1078	-17.0599	-15.4464	-15.0422	-16.5730	-16.1680	-16.4245	-16.4245	-17.1641	-17.8762	-16.2235	-17.8434	-18.4735
11	-17.0830	-16.8171	-15.6559	-15.5716	-16.1691	-16.8833	-16.7780	-16.7780	-17.6495	-17.6225	-17.3304	-17.2883	-17.2853
12	-18.0121	-17.0640	-17.2192	-16.8733	-16.5101	-18.3187	-18.6038	-18.6038	-17.9017	-16.6175	-18.0981	-17.7908	-17.5391
13	-15.7950	-17.6191	-16.4036	-15.9753	-16.2613	-16.3077	-18.0074	-18.0074	-17.6384	-18.0552	-18.0454	-17.9899	-18.4654
14	-17.2612	-17.2301	-15.8118	-16.7486	-16.5465	-16.4842	-17.1719	-17.1719	-17.5018	-18.2118	-18.5232	-17.2340	-17.6807
15	-18.3881	-18.3175	-16.7509	-17.3233	-16.4130	-15.8420	-18.1482	-18.1482	-17.3191	-18.8636	-17.0954	-18.6549	-18.0869
16	-17.7788	-16.1520	-16.1497	-16.5131	-14.8276	-16.7379	-17.2851	-17.2851	-17.0738	-18.2669	-17.7254	-18.0875	-17.1946
17	-16.7643	-16.5282	-16.4950	-16.0158	-16.1412	-16.2582	-17.6313	-17.6313	-16.7044	-17.3987	-17.4435	-17.6784	-17.0037
18	-17.5320	-17.4317	-17.0729	-18.3200	-16.4018	-17.2510	-17.9044	-17.9044	-17.8702	-18.6656	-18.6461	-18.7742	-17.4690

3. 找出得分偏低的语句, 进行差异分析:



上图为每周每句音频的评分分布图，中可以看出，除了第七周，其他各周评分差异不大。对于第七周的反常情况，原因是当时耳机出了问题，没有意识到，所以在做分析时，去除第七周的样本。

除去第七周的样本，放大 y 轴，标识出每一周的最高分、最低分及平均分的结果：



分数最高前六:

W3,N2:-15.5462853576475

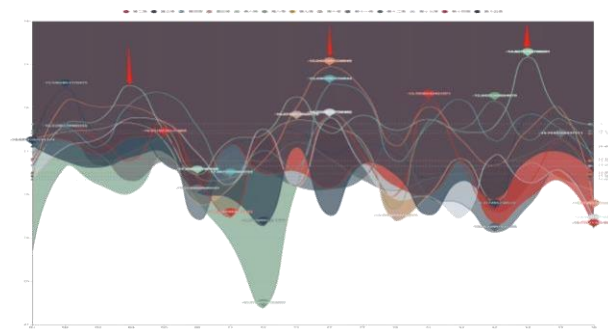
W4,F4:-15.4463983556844

W5,F4:-15.0421941238949

W5,F5:-15.5716246051991

W6,N4:-15.4852268923935

W6,S4:-14.8275794788201



统计分数最高的前六句, 结合两幅图, 我们可以看出第四句和其他几句相比分数较高。

分数最低前六:

w8,N1:-19.3668079307752

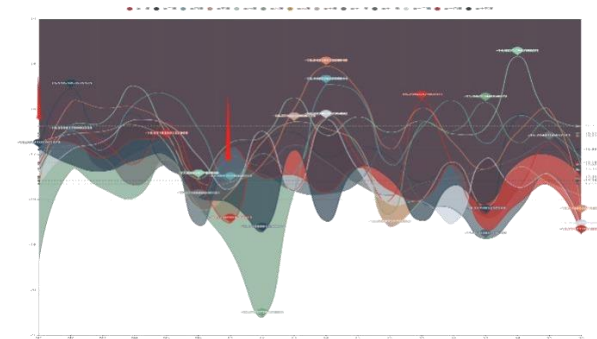
w8,F1:-19.1013568108366

w8,F2:-20.6150759196586

w12,S3:-18.8636007937392

w14,S6:-18.7742138049902

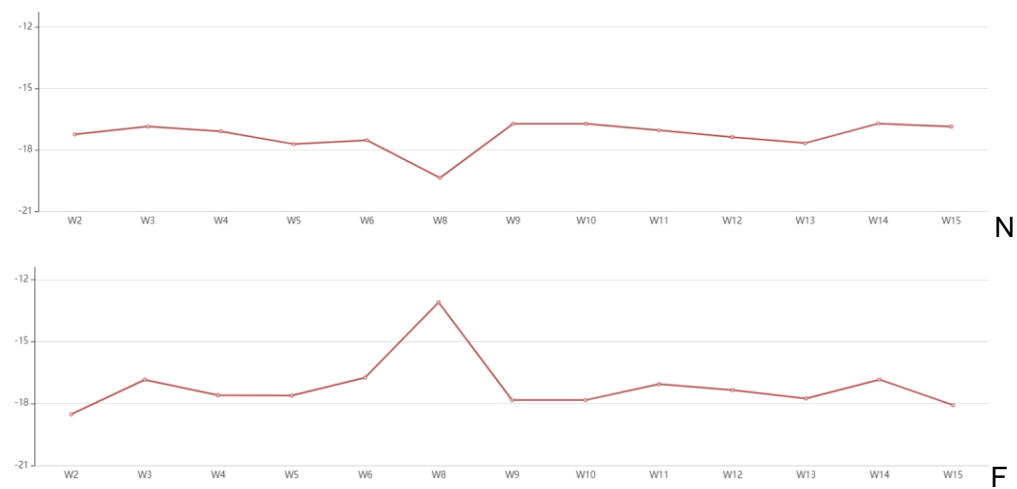
w15,F2:-18.7198991390077

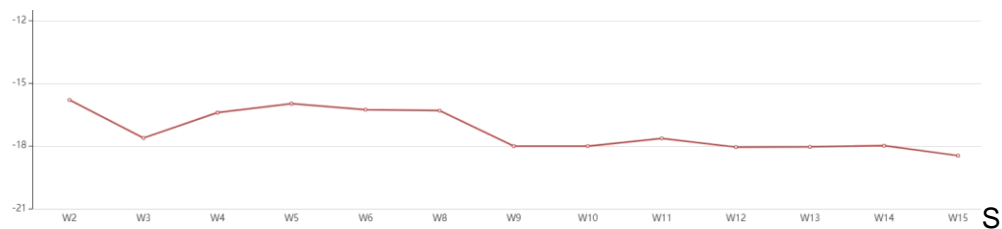


统计分数最低的前六句, 结合两图, 我们可以看出, 第一句的评分相对较低, 但是差别不是特别明显。

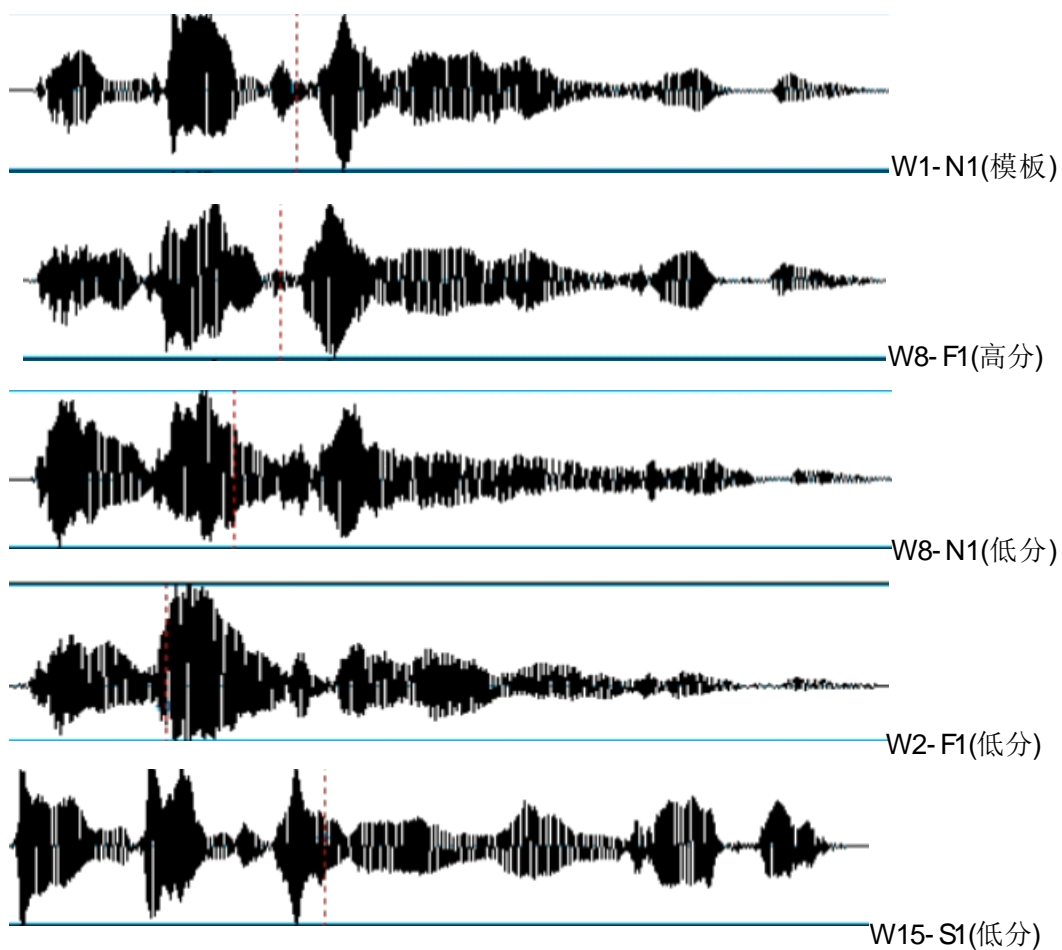
接下来对评分较低的语句进行分析, 以第一句为例:

不同速度语句的分数随时间的变化图



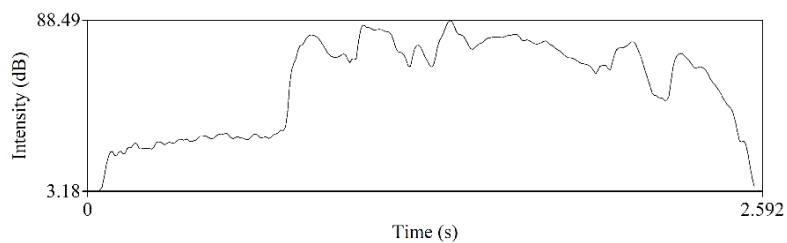


时域分析:

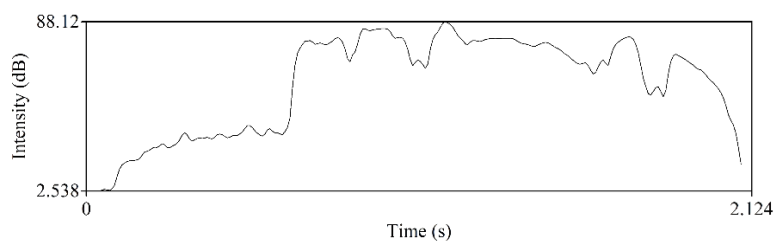


由上图可看出，分数较高的语句和模板语句的形状比较相似，而分数较低的语句由于重音，吐词清晰度，连贯性等不同，与模板语句相差较大。W8- N1 和模板语音相比前几个音发的比较重，后面几个音一掠而过，没有强调。W2- F1 的“登”和“鹤”字发音较重较长。W15- S1 每一个字都太过强调，与模板语音不同。

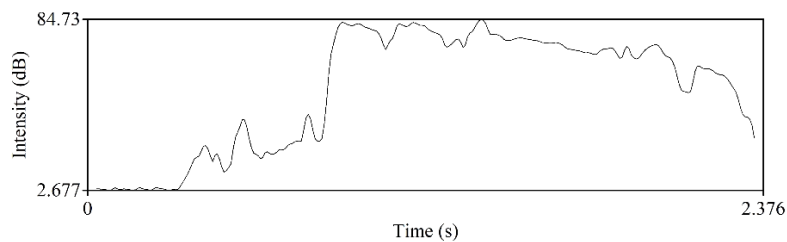
强度分析:



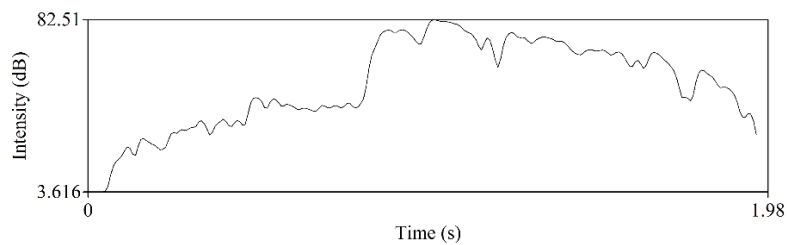
W1- N1(模板)



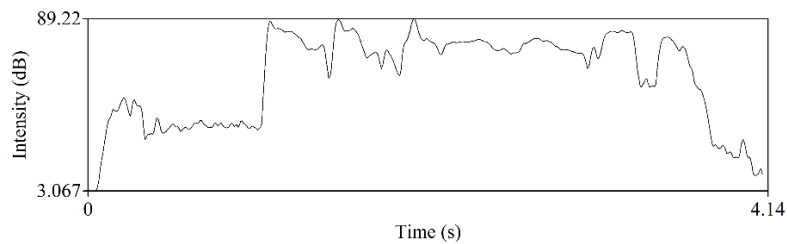
W8- F1(高分)



W8- N1(低分)



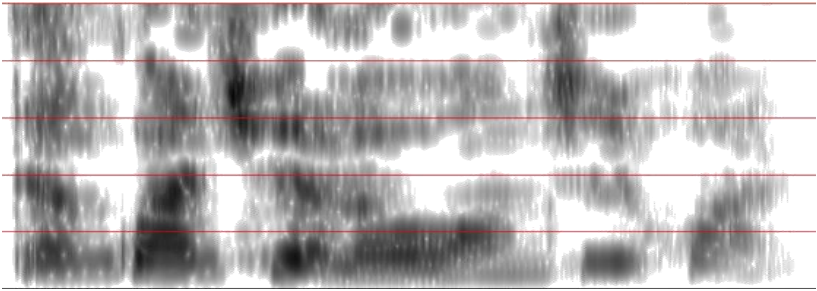
W2- F1(低分)



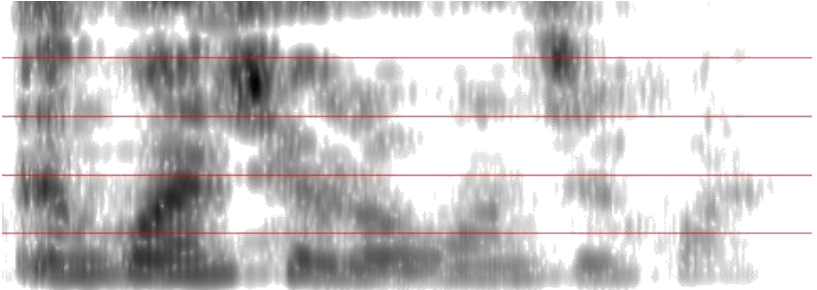
W15- S1(低分)

从强度分析可以看出,得高分的语句曲线形状和模板语句几乎相同, W8- N1(低分)的前段强度较低, W2- F1(低分)的峰值起点较后, W15- S1(低分)的形状与模板语句相差的较多。

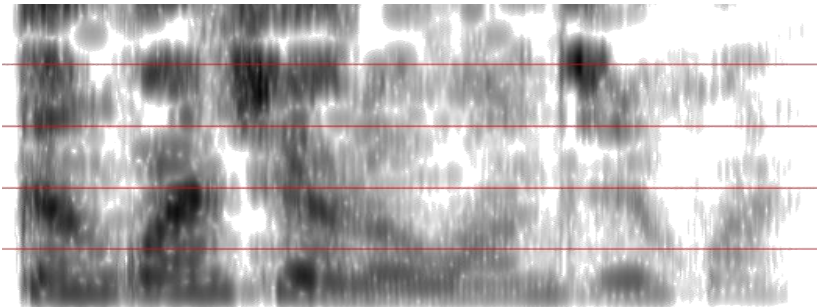
频谱分析：



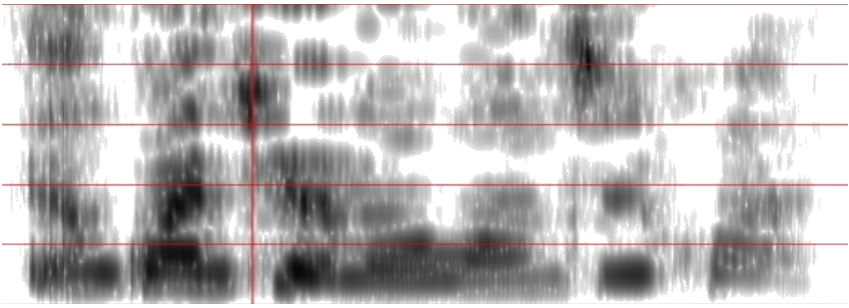
W1- N1(模板)



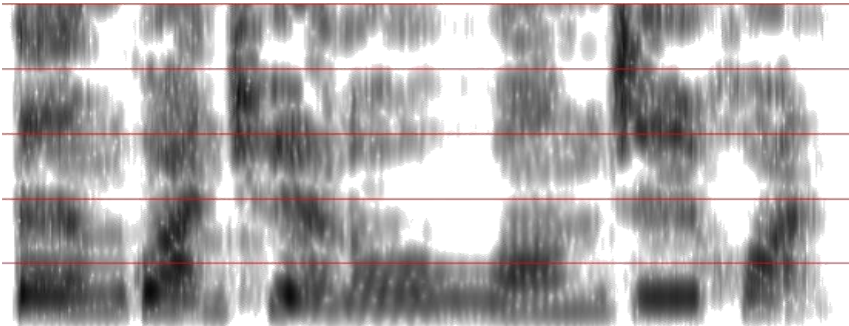
W8- F1(高分)



W8- N1(低分)



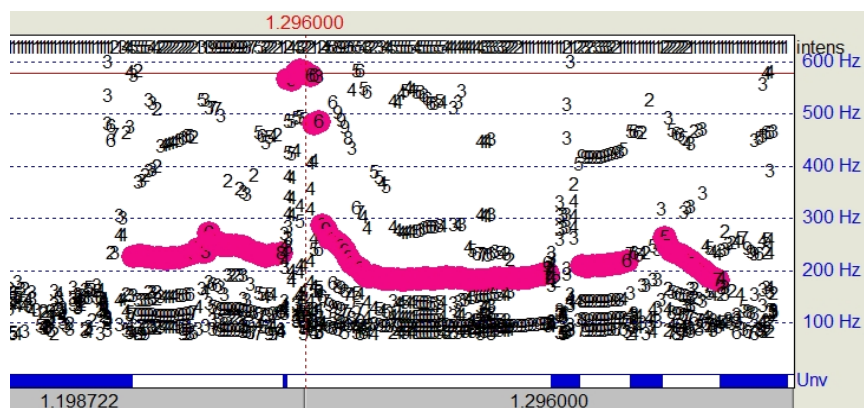
W2- F1(低分)



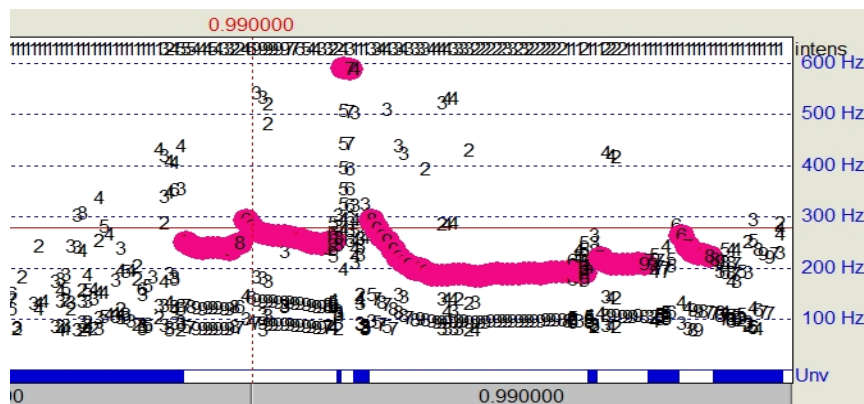
W15- S1(低分)

从频谱分析可以看出，W8- F1(高分)与模板频谱形状及浓淡都比较相似，最后轻音的几个字比较淡。W8- N1(低分)与模板的前几个字差异较大，且字与字之间较难分辨。W2- F1(低分)中间几个字的发音较模板偏重。W15- S1(低分)相比模板整体的发音较重，每一个字都说的比较清楚。

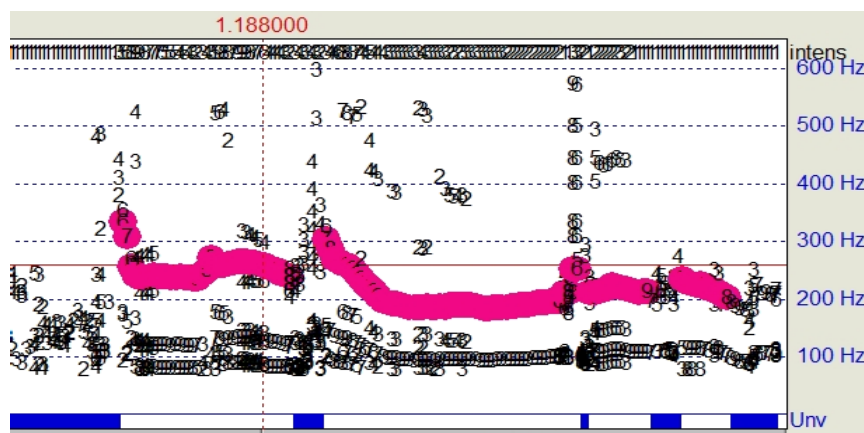
基频分析:



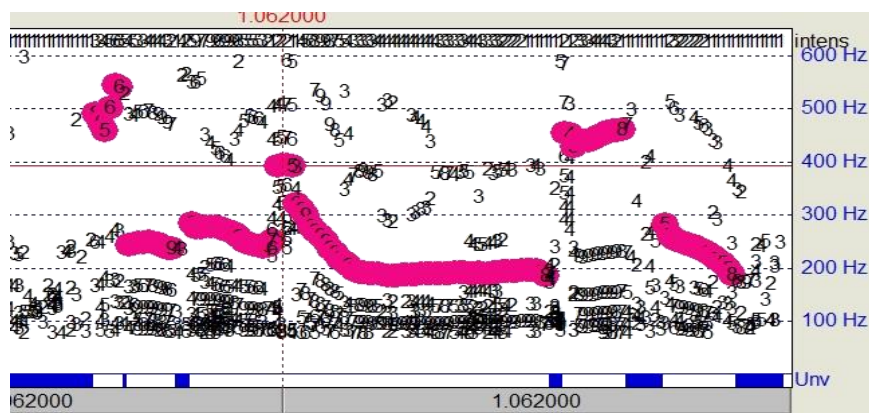
W1-N1(模板)



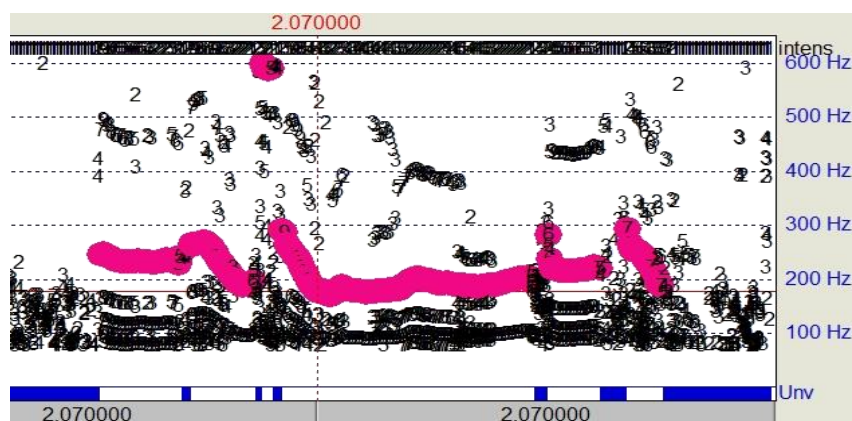
W8-F1(高分)



W8-N1(低分)



W2- F1(低分)

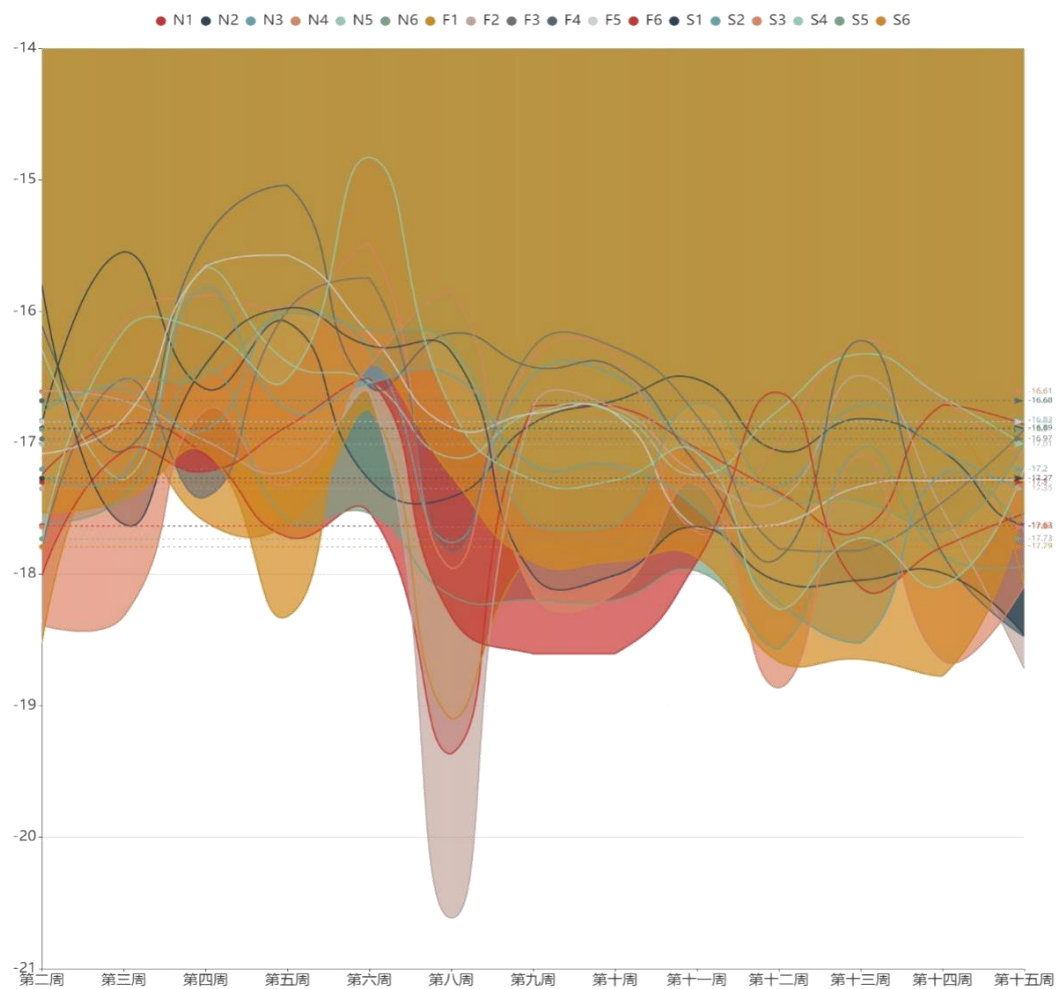


W15- S1(低分)

从基频分析可以看出,W8- F1(高分)的基频与模板相似,每个字的发音频率都十分相似。W8- N1(低分)的“雀”字发音与模板相比不够高。W2- F1(低分)的“登”字和“之”字与模板相比比较高。W15- S1(低分)的“雀”字和模板比较高,“楼”字和“王”字直接的区别较小。

4. 音频的时间变化分析

下图为N1-S6随时间变化的分布图,从中可以看出,音频的识别分数总体呈现微弱的下降的趋势,说明人的声音是随时间而发生缓慢的变化的,但是不是很明显。其中第八周分数总体明显的偏低,可能是因为感冒或者当时嗓子的状态不太好导致。



5. 总结:

在本次实验中,我学会了 **praat** 软件的基本操作,利用该软件对语音进行分析。其可视化的将语音的各个信息展示出来,让我对语音的不同属性有了更好的理解,明白了语音之间的差异主要体现在哪些地方。希望以后有时间再对这方面的知识进行更深入的了解,尝试实现不同的语音识别的算法。