

建模以及实验

生理特征

I 年龄感

I 性别感

I 清脆度

- 数据准备
利用自己采集的860个短音频进行训练
对每个音频进行最大值为7的label标注
- 特征提取
MFCC特征
生理特征主要与音高、谐波以及音高抖动有关，所以采用MFCC可以对音高、谐波等有具体的表现
- 模型选择
采用卷积神经网络，由于MFCC特征提取后特征有较强的横纵向、局部相关性所以采用卷积核对其进行处理。
经过卷积池化降维后，利用全连接层对提取到的特征进行权重结合，最终预测出结果
- 最终结果

年龄感误差：0.70065402681 / 7.0

性别感误差：1.13084420650 / 7.0

清脆度误差：1.15585780881 / 7.0

发音调控

I 咬字阶段：咬字清晰度、相邻字发音黏着程度；

I 韵律阶段：平均音高、音高变化、语调的平顺性、停顿频度、语速控制、节奏的稳定性

- 数据准备
利用自己采集的860个短音频进行训练
对每个音频进行最大值为7的label标注
利用腾讯智聆api对860个音频做了音节边界标记，用于音节分割训练
- 特征提取
 - pitch
音高是能直接提取得到的，所以通过直接提取音频的音高来获得平均音高、音高变化
 - MFCC
黏着度与包络连贯性有关、咬字干净与谐波组成有关、不同音节在频谱上的表现也有所不同。所以本模块也用了mfcc作为发音清晰度、相邻字发音黏着度、语调平顺性。

- Fbank

和MFCC相比缺少了最后的离散余弦变化一步，更多的是希望符合声音信号的本质，拟合人耳接收的特性。不同的音节能在Fbank特征上有较好的体现

- 音节提取

语速、节奏稳定性、停顿等都需要用到音节边界这一特征

- 模型选择

- 发音清晰度、黏着度、平顺性三个直接利用MFCC特征训练的采用的是和生理特征模块相同的模型——卷积神经网络

最终结果为

```
发音清晰度平均误差:    0.7744751781850437  /   7.0
发音黏着程度的平均误差:    1.0075982218   /   7.0
语调的平顺性的平均误差:    0.92381925050   /   7.0
```

- 音节分割

采用vad方法检测出silence

将fbank特征与pitch按照时域拼接

利用卷积神经网络对特征进行处理训练，最终结果为

	F1	HR	OS	R
count	860.000000	860.000000	860.000000	860.000000
mean	0.812844	79.654628	-3.560585	0.822803

R-value采用动态region方法进行测评

基本满足需求

- 平均音高、音高变化

平均音高采用 pitch的（中位数+均值）/ 2

音高变化采用pitch的方差

- 停顿频度

音节之间间隔次数/时间

- 语速

音节发音时间的均值

- 节奏的稳定性

各音节发音时间的方差

宏观（抽象）评价

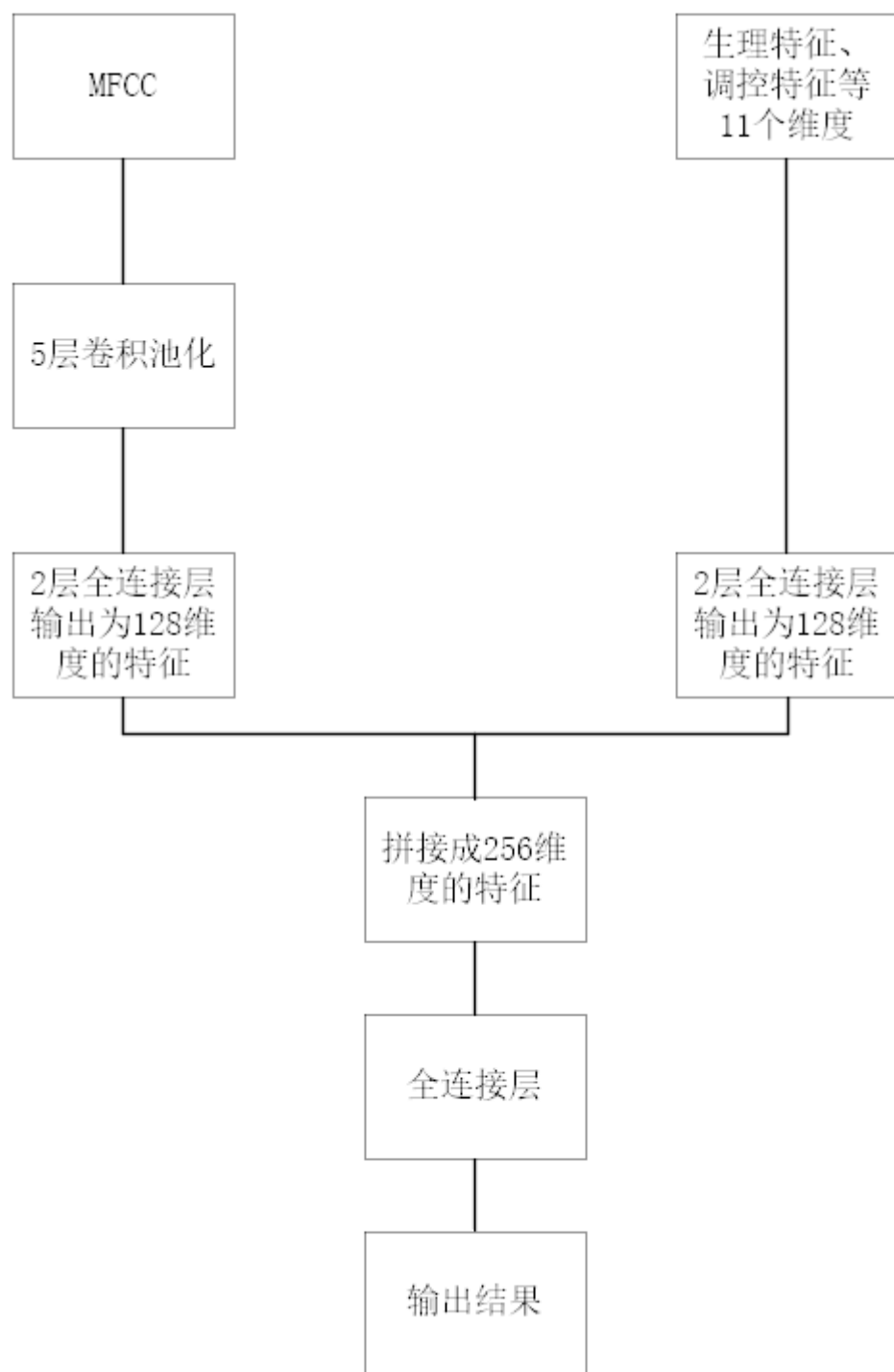
温柔、单纯、孱弱、有力、成熟、高傲、清爽、活力、激情、轻佻、尖细、滑稽、戏谑、冷艳、高冷、沧桑、神圣、正义、低缓、木楞

- 数据准备

利用自己采集的860个短音频进行训练

对每个音频进行最大值为7的label标注

- 特征提取
 - 生理特征
 - 年龄感
 - 性别感
 - 清脆度
 - 发音调控特征
 - 咬字阶段：咬字清晰度、相邻字发音黏着程度；
 - 韵律阶段：平均音高、音高变化、语调的平顺性、停顿频度、语速控制、节奏的稳定性
 - MFCC
- 模型选择
 - 采用多特征结合思想，先使用卷积对MFCC进行特征再提取，然后利用全连接将生理特征、调控特征组合提取，最后利用逻辑回归将拼接起来的特征进行组合得出结果



○ 最终结果

整体平均误差 : 1.036830073627828

低缓	loss : 1.7395625419453187
高冷	loss : 1.7860195660912903
冷艳	loss : 1.014093239664892
单纯	loss : 1.2696723516554393
孱弱	loss : 1.6063810307783024
尖细	loss : 0.9161219021014009
戏谑	loss : 0.6128652089974802
有力	loss : 1.110721804090578

木楞	loss : 1.5987329244573887
正义	loss : 1.4404893855803216
沧桑	loss : 1.360141946073011
活力	loss : 0.7689765134076854
清爽	loss : 1.535077632810883
温柔	loss : 1.932533746203615
滑稽	loss : 0.8968183388750208
激情	loss : 0.5312550728594192
神圣	loss : 1.196321327394942
轻佻	loss : 1.5537195912105972
高傲	loss : 0.7374732748721218
成熟	loss : 1.3160406374418765