

男女声识别

周无寒

北京大学信息科学技术学院

2024 年 12 月

1 引言

我使用 Matlab 实现了一个男女声识别器。该识别器主要包含了预处理（提取有声段、去噪、分帧等）、特征提取，特征处理，结果判定等内容。

2 方法

2.1 预处理

提取有声片段的做法是，设置一个阈值，当音频的能量大于该阈值时，认为是有声片段。

```
1 threshold = 0.01;
2 voiced_indices = find(abs(x) > threshold);
3 x_voiced =
4     x(min(voiced_indices):max(voiced_indices), :);
```

降噪的做法是，对音频减去其直流分量。

```
1 x_denoised = x_voiced - mean(x_voiced);
```

分帧的做法是，设置好帧长和帧移，然后对音频进行分帧。使用一个二维矩阵来存储分帧之后的数据。

```
1 frame_size = 0.025;
2 frame_shift = 0.05;
3 frame_length = round(frame_size * fs);
```

```

4     frame_step = round(frame_shift * fs);
5     num_frames = floor((length(x_denoised)
6         - frame_length) / frame_step) + 1;
7
8     frames = zeros(num_frames, frame_length);
9     for i = 1:num_frames
10         start_index = (i-1) * frame_step + 1;
11         end_index = start_index + frame_length - 1;
12         frames(i, :) =
13             x_denoised(start_index:end_index);
14     end

```

2.2 特征提取

基频是区分男女声音的一个有效的特征。一般而言，男性的频率较低，女性的频率较高。

我们使用自相关方法提取基频 [1]。自相关方法通过计算信号的自相关函数，找到自相关函数的最大值位置，从而估计信号的周期，并计算基频。

```

1     [~,index]=max(data);
2     timewin=floor(0.015*fs);
3     xwin=data(index-timewin:index+timewin);
4     [y,~]=xcov(xwin);
5     ylen=length(y);
6     halflen=(ylen+1)/2 +30;
7     yy=y(halflen: ylen);
8     [~,maxindex] = max(yy);
9     fmax=fs/(maxindex+30);

```

梅尔频率倒谱系数是另一个区分男女声音的有效特征。梅尔刻度是一种基于频率定义的非线性刻度单位，它在对数频率尺度上进行了非线性变换，使得低频区域有更高的分辨率。由于男女声音的基频差异主要体现在低频区域，使用梅尔刻度可以更好地捕捉这种差异。Matlab 里有现成的函数可以计算梅尔频率倒谱系数。这里我们直接调用该函数，然后取平均值作为特征。

```
1     coeffs = mfcc(x, fs);
2     mean_coeffs = mean(coeffs);
```

3 结果

在给定的数据集中，使用基频判断时除了“f2.wav”以外，其余的音频文件都被正确识别；使用梅尔频率倒谱系数判断时，判断错误率较高。

4 讨论

在这个识别器中，分帧操作会大大降低识别的准确率，因此最后并未使用分帧之后的数据进行分析。对于该现象出现的原因，我认为可能是因为分帧之后的数据丢失了一部分信息，导致识别器无法准确识别。而降噪等其他预处理手段也可能会丢失一部分信息，因而“f2.wav”并未被正确识别。

梅尔频率倒谱系数的判断较为复杂。本程序中直接取平均值对结果影响较大，因此判断的准确性较差。

5 结论

总体上，该男女声识别器的准确率较高，但是仍有出错的情况。后续的改进可能包括使用不同的特征，比如共振峰、短时能量等。同时针对梅尔频率倒谱系数，考虑使用更加精细的判断方法，而非直接取平均值。对于判断标准的常数，也可以考虑在大数据集上使用机器学习的方法训练出一个优化的参数，提高准确率。

参考文献

[1] 郑君里，杨为理，应启珩. 信号与系统. 高等教育出版社, 2011.