

江西理工大学

语音信号处理 实验报告

实验名称 实验四 语音识别实践

日期 2023.03 专业 智能科学与技术 班级 智科 201 班

实验人 吴佳龙 学号 2420203236 同组人

一、实验目的

- 1、掌握语音识别的模板匹配法的原理和过程。
- 2、掌握 DTW 技术和梅尔频率倒谱系数 MFCC 特征提取。
- 3、应用 Python 实现基于 DTW 的 10 个阿拉伯数字的识别。

二、实验原理

下图为利用模板匹配法进行语音识别的原理框图。在训练阶段，用户将词汇表中的每一个词依次说一遍，并且将其特征矢量时间序列作为模板（Template）存入模板库；在识别阶段，将输入语音的特征矢量时间序列依次与模板库中的每个模板进行相似度比较，将相似度最高者作为识别结果输出。

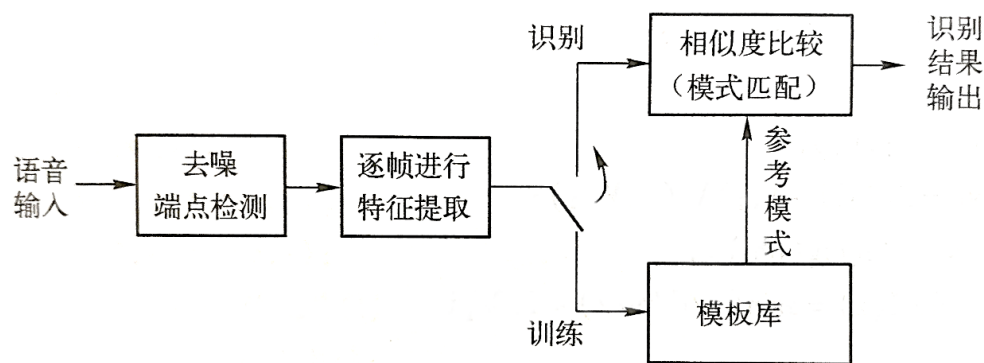


图 4-1 模板匹配法语音识别系统原理图

在特征提取阶段，本实验选用美尔频率倒谱系数（MFCC）作为识别特征；在识别阶段，实验选用动态时间规整（DTW）技术来进行模式匹配。下面将简单介绍这两个技术的原理和实现过程。本实验采用美尔频率倒谱系数（Mel - Frequency Cepstral Coefficients, MFCC）及其一阶和二阶差分作为特征参数。美尔频率倒谱系数不同于线性预测系数，它是将人耳的听觉特性和语音产生相结合的

一种特征参数。与普通倒谱分析不同，美尔倒谱系数着眼于分析人耳的听觉特性。人耳所听到的声音的高低与声音的频率并不呈线性正比关系，而用美尔频率尺度则更符合人耳的听觉特性。Mel 频率与实际频率的具体关系可用如下式表示：

$$Mel(f) = 2595 \lg(1 + f/700)$$

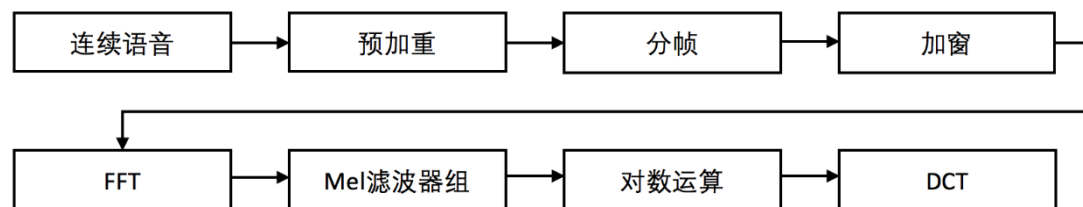


图 4-2 MFCC 语音特征的提取流程

三、实验设备及环境

- 1、设备：笔记本、台式电脑
- 2、实验环境：Windows 7、Windows10 或 Linux 操作系统

MATLAB/Python3.6+, Anaconda3, Pycharm2017+

四、实验内容与步骤

1、实验内容：

1) 匹配函数 DTW 的程序设计：

输入：参数 F 为模板的 MFCC 特征；R 为当前待识别语音的 MFCC；

输出：F 与 R 之间的匹配距离；

2) 对实验结果进行分析。

2、实验步骤

1) 实验语音来自 3 个人，每人说一遍 0-9 这 10 个数字，记录下相应的语音并保存。

2) 对语音进行预处理，端点检测，然后利用 MFCC 函数对有效语音段进行 MFCC 特征提取。

3) 以相同方式处理其他语音并保存。

4) 采用 DTW 模型对输入待识别与参考模板进行匹配，完成识别过程。

五、实验注意事项

1、实验中应遵守实验室实验规则，爱护实验室设备，保持卫生清洁；各种实验仪器设备在观察使用后放归原处，不得损坏或随意放置。

六、思考题

1、在 DTW 中，若采用欧氏距离将对结果造成何种影响，为什么？

如果在 DTW 中采用欧氏距离作为局部距离度量，那么会对结果造成以下影响：

优点：欧氏距离计算简单，直观，易于实现 3。

缺点：欧氏距离对噪声敏感，不能反映两个序列的形状变化和趋势变化

2、除 MFCC 外，语音信号的基本特征还有哪些？

能量：时域上每帧的能量，幅度的平方 1。能量可以用于判断语音和噪声，清音和浊音等。

短时过零率：计算信号跨越零点的次数 2。短时过零率可以用于判断语音和噪声，清音和浊音等。

基频：声带振动的频率，代表声音的音高 3。基频可以用于检测语音噪声，特殊声音检测，男女判别，说话人识别等。

谐波：基频对应的整数次频率成分，由声带发声带动空气共振形成的。谐波可以反映声音的音色，用于声纹识别，语音转换等。

七、实验程序及结果展示

```
1) #coding:utf-8-*  
%设置模板  
ncoeff=12; %mfcc 系数的个数  
fMatrix1 = cell(1,10);  
fMatrix2 = cell(1,10);  
fMatrix3 = cell(1,10);  
  
for i = 1:10  
    q = ['C:\Users\user\Desktop\4.基于动态时间规整（DTW）的孤立字语音识别实验(matlab)\4.基于动态时间规整（DTW）的孤立字语音识别实验(matlab)\SpeechData\p1\'num2str(i-1)'.wav'];  
    [speechIn1,FS1] = audioread(q);  
    speechIn1 = my_vad(speechIn1);  
    fMatrix1(1,i) = {mfccf(ncoeff,speechIn1,FS1)};  
  
end  
  
for j = 1:10  
    q = ['C:\Users\user\Desktop\4.基于动态时间规整（DTW）的孤立字语音识别实验(matlab)\4.基于动态时间规整（DTW）的孤立字语音识别实验(matlab)\SpeechData\p2\'num2str(j-1)'.wav'];  
    [speechIn2,FS2] = audioread(q);  
    speechIn2 = my_vad(speechIn2);
```

```

        fMatrix2(1,j) = {mfccf(ncoeff,speechIn2,FS2));
    end

    for k = 1:10
        q = ['C:\Users\user\Desktop\4.基于动态时间规整（DTW）的孤立字语音识别实验(matlab)\4.基于动态时间规整（DTW）的孤立字语音识别实验(matlab)\SpeechData\p3\'
        num2str(k-1)'.wav'];
        [speechIn3,FS3] = audioread(q);
        speechIn3 = my_vad(speechIn3);
        fMatrix3(1,k) = {mfccf(ncoeff,speechIn3,FS3));
    end

    %将数据保存为 mat 文件
    fields = {'zero','One','Two','Three','Four','Five','Six','Seven','Eight','nine'};
    s1 = cell2struct(fMatrix1, fields, 2);          %fields 项作为行
    save Vectors1.mat -struct s1;
    s2 = cell2struct(fMatrix2, fields, 2);
    save Vectors2.mat -struct s2;
    s3 = cell2struct(fMatrix3, fields, 2);
    save Vectors3.mat -struct s3;

```

结果图

```

>> matchTemplates
Press any key to start 2 seconds of speech recording...
Recording speech...
Finished recording.
System is trying to recognize what you have spoken...

Nbr =

     5     9
    10     2
     4     6

No microphone connected or you have not said anything.
|

```

实验结果分析

- DTW 模型对于孤立词语音识别有一定的效果，但是也存在一些缺点和局限性。例如，DTW 模型对于噪声敏感，对于发音不清晰或者口音不同的语音识别效果较差；DTW 模型计算量较大，需要对每个模板进行匹配，随着模板数量的增加，识别速度会降低；DTW 模型没有考虑语音信号的统计特性和上下文信息，只是基于局部

距离的最优化，可能会导致一些错误匹配。

- MFCC 特征对于语音信号有一定的表征能力，但是也存在一些缺点和局限性。例如，MFCC 特征只考虑了频域信息，忽略了时域信息和相位信息；MFCC 特征对于高频部分的信息损失较大，可能会影响语音信号的细节和差异；MFCC 特征没有考虑语音信号的动态变化和非线性变化，只是基于线性预测和倒谱分析。