

哈尔滨工业大学

实验报告

实 验（三）

题 目 命令词识别实验报告

专 业 视听觉信息处理

学 号 1170300511

班 级 1703105

学 生 易 亚 玲

指 导 教 师 郑 铁 然

实 验 地 点 G 7 0 9

实 验 日 期 2019.11.4

计算机科学与技术学院

一、 设计命令词识别任务

1.1 介绍你的分组情况

小组成员：张亚博、束魏琦、陈鋆、强文杰、易亚玲

1.2 描述所设计的命令词识别任务

收集 5 个电影名字的发音，每个人说每个电影的名字十遍，然后进行训练。程序能够语音提示用户说出想要搜索的电影名字，程序能自动完成识别任务，并且打开一个窗口，显示关于该电影的影评信息。

1.3 列出词表

- 1.西游记
- 2.十面埋伏
- 3.赌神
- 4.阿甘正传
- 5.冰雪奇缘

1.4 介绍语料采集方法和规模，

语料采集：cooledit

规模：一共五句话，每个人每句话采集 10 次

二、特征提取

2.1 详细描述你所采用的特征和提取算法

特征提取由小组成员束魏琦完成，下面是运行截图

2.2 给出特征提取部分运行结果的截图

```

选择C:\Windows\System32\cmd.exe
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/smmf_10.mfc [sampSize=156,nSamples=118] with CRC
./QWJ_wav/smmf_10.wav -> ./QWJ_mfc/smmf_10.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_1.mfc [sampSize=156,nSamples=129] with CRC
./QWJ_wav/xyj_1.wav -> ./QWJ_mfc/xyj_1.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_2.mfc [sampSize=156,nSamples=97] with CRC
./QWJ_wav/xyj_2.wav -> ./QWJ_mfc/xyj_2.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_3.mfc [sampSize=156,nSamples=113] with CRC
./QWJ_wav/xyj_3.wav -> ./QWJ_mfc/xyj_3.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_4.mfc [sampSize=156,nSamples=128] with CRC
./QWJ_wav/xyj_4.wav -> ./QWJ_mfc/xyj_4.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_5.mfc [sampSize=156,nSamples=120] with CRC
./QWJ_wav/xyj_5.wav -> ./QWJ_mfc/xyj_5.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_6.mfc [sampSize=156,nSamples=105] with CRC
./QWJ_wav/xyj_6.wav -> ./QWJ_mfc/xyj_6.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_7.mfc [sampSize=156,nSamples=113] with CRC
./QWJ_wav/xyj_7.wav -> ./QWJ_mfc/xyj_7.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_8.mfc [sampSize=156,nSamples=124] with CRC
./QWJ_wav/xyj_8.wav -> ./QWJ_mfc/xyj_8.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_9.mfc [sampSize=156,nSamples=112] with CRC
./QWJ_wav/xyj_9.wav -> ./QWJ_mfc/xyj_9.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/xyj_10.mfc [sampSize=156,nSamples=100] with CRC
./QWJ_wav/xyj_10.wav -> ./QWJ_mfc/xyj_10.mfc

HTK Configuration Parameters[21]
Module/Tool      Parameter      Value
# HREC           FORCEOUT       TRUE
# HNET           TRACE         1
# HLABEL         TRACE         8
# HPARM          TRACE         65
# HSHELL         TRACE         2
#                FORCECTEXP     TRUE
#                ALLOWXWRDEXP  TRUE
#                ENORMALIZE   TRUE
#                NUMCEPS      12
#                CEPLIFTER    22
#                NUMCHANS     26
#                PREEMCOEF     0.970000
#                USEHAMMING    TRUE
#                WINDOWSIZE    250000.000000
#                SAVEWITHCRC   TRUE
#                SAVECOMPRESSED TRUE
#                TARGETRATE    100000.000000
#                TARGETKIND     MFCC_E_D_A_Z
#                ZMEANSOURCE    FALSE
#                SOURCEFORMAT   WAV
#                SOURCEKIND     WAVEFORM

C:\Users\15650\Desktop\MFCC提取工具>

```

2.3 给出特征文件内容的截图

```

C:\Windows\System32\cmd.exe
Microsoft Windows [版本 10.0.18362.476]
(c) 2019 Microsoft Corporation. 保留所有权利。

C:\Users\15650\Desktop\MFCC提取工具>hcopy -A -D -T 1 -C tr_wav.cfg -S list.scp
hcopy -A -D -T 1 -C tr_wav.cfg -S list.scp

HTK Configuration Parameters[21]
Module/Tool      Parameter      Value
# HREC           FORCEOUT       TRUE
# HNET           TRACE        1
# HLABEL         TRACE        8
# HPARM          TRACE       65
# HSHELL         TRACE        2
#               FORCECXTEXP    TRUE
#               ALLOWXWRDEXP  TRUE
#               ENORMALIZE   TRUE
#               NUMCEPS     12
#               CEPLIFTER    22
#               NUMCHANS    26
#               PREEMCOEF    0.970000
#               USEHAMMING   TRUE
#               WINDOWSIZE   250000.000000
#               SAVEWITHCRC   TRUE
#               SAVECOMPRESSED TRUE
#               TARGETRATE    100000.000000
#               TARGETKIND    MFCC_E_D_A_Z
#               ZMEANSOURCE   FALSE
#               SOURCEFORMAT   WAV
#               SOURCEKIND     WAVEFORM

HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_1.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_1.wav -> ./QWJ_mfc/agzz_1.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_2.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_2.wav -> ./QWJ_mfc/agzz_2.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_3.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_3.wav -> ./QWJ_mfc/agzz_3.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_4.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_4.wav -> ./QWJ_mfc/agzz_4.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_5.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_5.wav -> ./QWJ_mfc/agzz_5.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_6.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_6.wav -> ./QWJ_mfc/agzz_6.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_7.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_7.wav -> ./QWJ_mfc/agzz_7.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_8.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_8.wav -> ./QWJ_mfc/agzz_8.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_9.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_9.wav -> ./QWJ_mfc/agzz_9.mfc
HParm: Parm tab type MFCC E D A K Z saved to ./QWJ_mfc/agzz_10.mfc [sampSize=156,nSampl
./QWJ_wav/agzz_10.wav -> ./QWJ_mfc/agzz_10.mfc

```


选择C:\Windows\System32\cmd.exe

```
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/agzz_10.mfc [sampSize=156,nSam
./QWJ_wav/agzz_10.wav -> ./QWJ_mfc/agzz_10.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_1.mfc [sampSize=156,nSampl
./QWJ_wav/ds_1.wav -> ./QWJ_mfc/ds_1.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_2.mfc [sampSize=156,nSampl
./QWJ_wav/ds_2.wav -> ./QWJ_mfc/ds_2.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_3.mfc [sampSize=156,nSampl
./QWJ_wav/ds_3.wav -> ./QWJ_mfc/ds_3.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_4.mfc [sampSize=156,nSampl
./QWJ_wav/ds_4.wav -> ./QWJ_mfc/ds_4.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_5.mfc [sampSize=156,nSampl
./QWJ_wav/ds_5.wav -> ./QWJ_mfc/ds_5.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_6.mfc [sampSize=156,nSampl
./QWJ_wav/ds_6.wav -> ./QWJ_mfc/ds_6.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_7.mfc [sampSize=156,nSampl
./QWJ_wav/ds_7.wav -> ./QWJ_mfc/ds_7.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_8.mfc [sampSize=156,nSampl
./QWJ_wav/ds_8.wav -> ./QWJ_mfc/ds_8.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_9.mfc [sampSize=156,nSampl
./QWJ_wav/ds_9.wav -> ./QWJ_mfc/ds_9.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/ds_10.mfc [sampSize=156,nSampl
./QWJ_wav/ds_10.wav -> ./QWJ_mfc/ds_10.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_1.mfc [sampSize=156,nS
./QWJ_wav/frozen_1.wav -> ./QWJ_mfc/frozen_1.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_2.mfc [sampSize=156,nS
./QWJ_wav/frozen_2.wav -> ./QWJ_mfc/frozen_2.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_3.mfc [sampSize=156,nS
./QWJ_wav/frozen_3.wav -> ./QWJ_mfc/frozen_3.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_4.mfc [sampSize=156,nS
./QWJ_wav/frozen_4.wav -> ./QWJ_mfc/frozen_4.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_5.mfc [sampSize=156,nS
./QWJ_wav/frozen_5.wav -> ./QWJ_mfc/frozen_5.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_6.mfc [sampSize=156,nS
./QWJ_wav/frozen_6.wav -> ./QWJ_mfc/frozen_6.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_7.mfc [sampSize=156,nS
./QWJ_wav/frozen_7.wav -> ./QWJ_mfc/frozen_7.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_8.mfc [sampSize=156,nS
./QWJ_wav/frozen_8.wav -> ./QWJ_mfc/frozen_8.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_9.mfc [sampSize=156,nS
./QWJ_wav/frozen_9.wav -> ./QWJ_mfc/frozen_9.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/frozen_10.mfc [sampSize=156,n
./QWJ_wav/frozen_10.wav -> ./QWJ_mfc/frozen_10.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/snmf_1.mfc [sampSize=156,nSam
./QWJ_wav/snmf_1.wav -> ./QWJ_mfc/snmf_1.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/snmf_2.mfc [sampSize=156,nSam
./QWJ_wav/snmf_2.wav -> ./QWJ_mfc/snmf_2.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/snmf_3.mfc [sampSize=156,nSam
./QWJ_wav/snmf_3.wav -> ./QWJ_mfc/snmf_3.mfc
HParm: Parm tab type MFCC_E_D_A_K_Z saved to ./QWJ_mfc/snmf_4.mfc [sampSize=156,nSam
./QWJ_wav/snmf_4.wav -> ./QWJ_mfc/snmf_4.mfc
```

三、 基于 DTW 的命令词识别

3.1 介绍你所设计 DTW 算法，标明所采用的开发工具

开发工具：Pycharm 2019.1

度量方式：欧式距离

算法描述：首先初始化 (0, 0) 位置为 0，然后初始化 $cost[0,:]$ 和 $cost[:,0]$ 接下来进行动态规划，分别比较 $cost[i-1][j]+dis[i][j]$ ， $cost[i][j-1]+dis[i][j]$ 和 $cost[i-1][j-1]+dis[i][j]*2$ 的大小，取最小的一个，然后依次向下直到(0,0)

```
# 如果两个向量等长，不需要对齐
if len(x) == len(y):
    rel = 0
    for i in range(len(x)):
        rel += cal_distance(x[i], y[i])
    return rel
else:
    len_x = len(x)
    len_y = len(y)
    cost = [[0 for i in range(len_y)] for i in range(len_x)]
    # 初始化 dis 数组
    dis = []
    for i in range(len_x):
        dis_row = []
        for j in range(len_y):
            dis_row.append(cal_distance(x[i],
y[j]))
        dis.append(dis_row)
    # 初始化 cost 的第 0 行和第 0 列
    cost[0][0] = dis[0][0]
    for i in range(1, len_x):
        cost[i][0] = cost[i - 1][0] + dis[i][0]
    for j in range(1, len_y):
        cost[0][j] = cost[0][j - 1] + dis[0][j]
    # 开始动态规划
    for i in range(1, len_x):
        for j in range(1, len_y):
            cost[i][j] = min(cost[i - 1][j] +
dis[i][j] * 1,
                            cost[i - 1][j - 1] +
```

```
dis[i][j] * 2,
                                cost[i][j - 1] + dis[i][j]
* 1)
return cost[len_x - 1][len_y - 1]
```

3.2 正确率:

阿甘正传: 1.0

冰雪奇缘: 1.0

赌神: 1.0

西游记: 1.0

十面埋伏: 1.0

四、 基于 HMM 的命令词识别

4.1 分工

训练语料采集与处理: 束魏琦, 易亚玲, 陈鋆, 强文杰, 张亚博

电影内容爬虫: 强文杰, 陈鋆

hmm 算法实现: 陈鋆, 强文杰, 张亚博

viterbi 算法实现: 束魏琦, 易亚玲

实用系统语音采集和与处理: 易亚玲, 束魏琦, 张亚博

实用系统框架设计及代码整合: 张亚博, 陈鋆, 强文杰

4.2 正确率:

阿甘正传: $13/15 = 86.7\%$

冰雪奇缘: $13/15 = 86.7\%$

赌神: $14/15 = 93.3\%$

西游记: $14/15 = 93.3\%$

十面埋伏: $14/15 = 93.3\%$

五、 实用系统

5.1 介绍在搭建实用系统过程中，你所在组的具体分工

训练语料采集与处理：束魏琦，易亚玲，陈鋆，强文杰，张亚博

电影内容爬虫：强文杰，陈鋆

hmm 算法实现：陈鋆，强文杰，张亚博

viterbi 算法实现：束魏琦，易亚玲

实用系统语音采集和与处理：易亚玲，束魏琦，张亚博

实用系统框架设计及代码整合：张亚博，陈鋆，强文杰

5.2 介绍你承担部分的完成情况

我负责的是训练语料采集和处理部分的视频剪辑部分的工作，还有 viterbi 部分算法参与推导，前期的语音采集。

5.3 实用系统是否搭建成功，若是，请描述之

我们的能够自动完成采集实时训练语音，并且利用 python 程序完成特征提取，然后将提取后的特征作为 hmm 的输入，完成模板匹配，然后我们还进行了以后个扩充，根据匹配类型，我们从网上抓取相应电影的影评，并显示在屏幕上。由于是随机人声的识别，受语速和噪声的影响，所以正确率不能达到 1.0，但已经超过百分之 80，是具有一定的实用价值的。

六、 总结

6.1 请总结本次实验的收获

对 DTW, virtebi 和 HMM 算法的理解更加深刻了, 能够做一个简单的听觉方面的小应用。

6.2 请给出对本次实验内容的建议