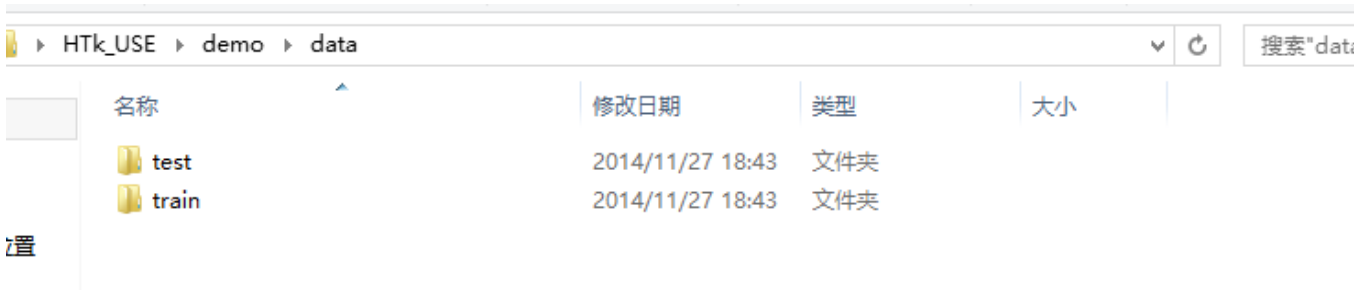


一文件目录布置（我认为这是最麻烦的一步）

1.创建一个 data 文件夹

2.\data\文件夹下放两个文件夹，分别是\train\和\test\



5.\data\train\下有两个文件夹分别是\wav\和\lab\



6.\data\train\wav\和\data\train\lab 下的目录名是一样的，都必须以训练用到的词汇做目录名如 caozuo , changge

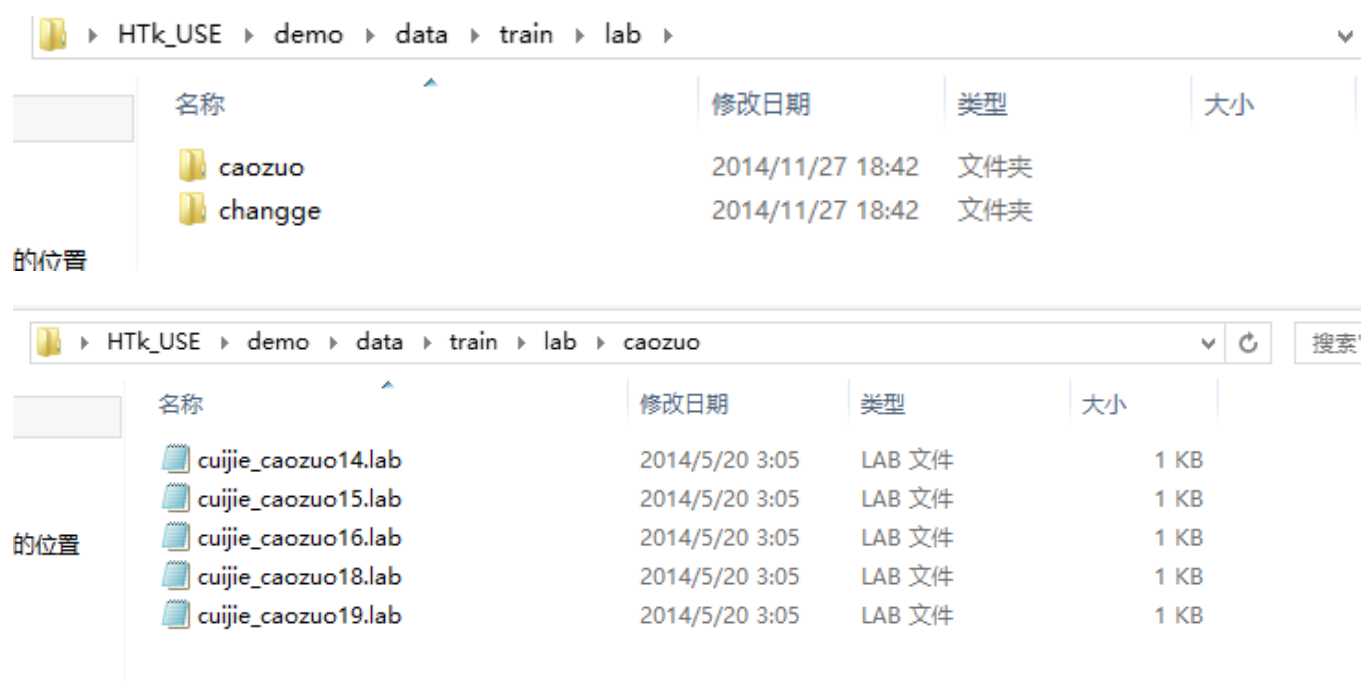


以词汇 caozuo 为例

7.\data\train\wav\caozuo\下放所有的分好的训练集中的 caozuo 的所有.wav 文件



8.\data\train\lab\caozuo\下放与上面.wav 文件对应好的.lab 文件

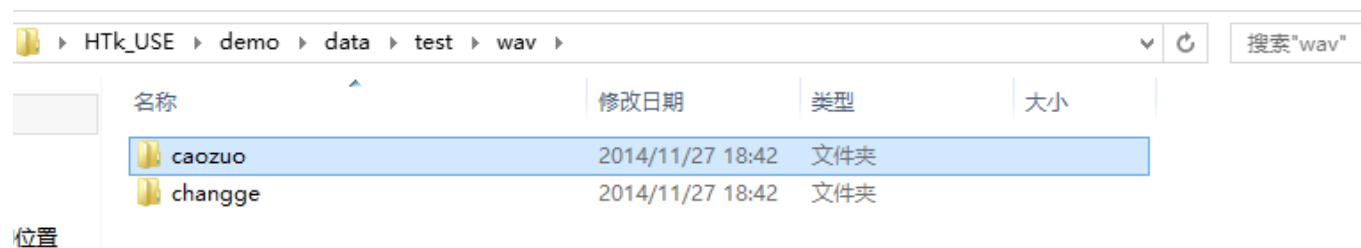


至此训练集文件放好了

测试集也是相同放置，不过要求 lab 文件夹下的目录和目录下的文件
必须与 wav 文件夹下对应的目录和目录文件完全一样（除了尾缀名）

比如

9.\data\test\wav\caozuo\cuijie_caozuo1.wav



HTk_USE > demo > data > test > wav > caozuo				
名称	#	标题	参与创作的艺术家人	
cuijie_caozuo1.wav				
cuijie_caozuo2.wav				

对应的 wav 文件路径一定是

10.\data\test\lab\caozuo\cuijie_caozuo1.lab

HTk_USE > demo > data > test > lab				
名称	修改日期	类型	大小	
caozuo	2014/11/27 18:42	文件夹		
changge	2014/11/27 18:42	文件夹		

HTk_USE > demo > data > test > lab > caozuo				
名称	修改日期	类型	大小	
cuijie_caozuo1.lab	2014/5/20 3:05	LAB 文件	1 KB	
cuijie_caozuo2.lab	2014/5/20 3:05	LAB 文件	1 KB	

位置

二、

1、data 文件夹替换与 myfile 同目录下的 data 文件夹

2、如果需要手写网络文件的话,把写好的网络文件复制到与 myfile 同目录下,并命名为 net.slf

HTk_USE > demo				
名称	修改日期	类型	大小	
data	2014/11/27 18:42	文件夹		
myfile	2014/11/27 19:01	文件夹		
data文件夹布置.doc	2014/11/27 20:18	Microsoft Word ...	14	
net.slf	2014/11/24 17:02	SLF 文件		
run.bat	2014/11/24 18:23	Windows 批处理...		

3、双击 run.bat,看到如下界面,开始输入各个词汇的音素个数,回车进入下一个词汇,按“B”或“b”返回上一个词汇

设置完最后一个词汇 sil 时,就不用进行任何其他的人为操作了

```

C:\WINDOWS\system32\cmd.exe
子目录或文件 ..\..\data 已经存在。
子目录或文件 ..\..\data\train 已经存在。
子目录或文件 ..\..\data\test 已经存在。
子目录或文件 ..\..\data\train\wav 已经存在。
子目录或文件 ..\..\data\train\lab 已经存在。
子目录或文件 ..\..\data\test\wav 已经存在。
子目录或文件 ..\..\data\test\lab 已经存在。
input the segment number(not state),like caozuo : 5
print 'b' or 'B' to backspace to last word .
caozuo : _

```

注：如果是第二次或两次以上运行它，会把上次设置的词汇音素设置出来，并会有一个是是否重写的提示，输入“Y”或‘y’表示重写：

```

C:\WINDOWS\system32\cmd.exe
子目录或文件 ..\..\data 已经存在。
子目录或文件 ..\..\data\train 已经存在。
子目录或文件 ..\..\data\test 已经存在。
子目录或文件 ..\..\data\train\wav 已经存在。
子目录或文件 ..\..\data\train\lab 已经存在。
子目录或文件 ..\..\data\test\wav 已经存在。
子目录或文件 ..\..\data\test\lab 已经存在。
子目录或文件 ..\bate 已经存在。
子目录或文件 ..\pro 已经存在。
the word segment is like that :
caozuo : 5
changge : 4
sil : 1
Do you want to rewrite hmmState ? (Y/N)_

```

三. 生成的文件

名称	修改日期	类型	大小
data	2014/11/27 18:42	文件夹	
model	2014/11/27 20:30	文件夹	
myfile	2014/11/27 19:01	文件夹	
result	2014/11/27 20:31	文件夹	
training	2014/11/27 20:30	文件夹	
data文件夹布置.doc	2014/11/27 20:18	Microsoft Word ...	149 KB
run.bat	2014/11/24 18:23	Windows 批处理...	1 KB

1. /mode：训练生成的 hmm 模型文件

HTk_USE > demo > model >

名称	修改日期	类型	大小
hmm0	2014/11/27 20:31	文件夹	
hmm1	2014/11/27 20:31	文件夹	
hmm2	2014/11/27 20:31	文件夹	
hmm3	2014/11/27 20:31	文件夹	
proto	2014/11/27 20:30	文件夹	

proto : 一直初始化的个文件 hmm 模型
hmm0 : Hinit 命令之后的 hmm 模型
hmm1 : 第一次 HRest 之后的 hmm 模型
hmm2 : 第二次 HRest 之后的 hmm 模型
hmm3 : 第三次 HRest 之后的 hmm 模型

2、\result

HTk_USE > demo > result

名称	修改日期	类型	大小
 reco.mlf	2014/11/27 20:37	MLF 文件	1 KB
 result.txt	2014/11/27 20:37	文本文档	1 KB

位置

reco.mlf : 测试数据的识别结果 (Hvite 命令后的结果)

result.txt : 测是结果识别率的分析文件 (HResult 命令后的结果)

3.\train 里是数据训练时的记录各词汇 mfcc 文件路径的文本

其他文件参见详细文件.docx