

倾听你的声音

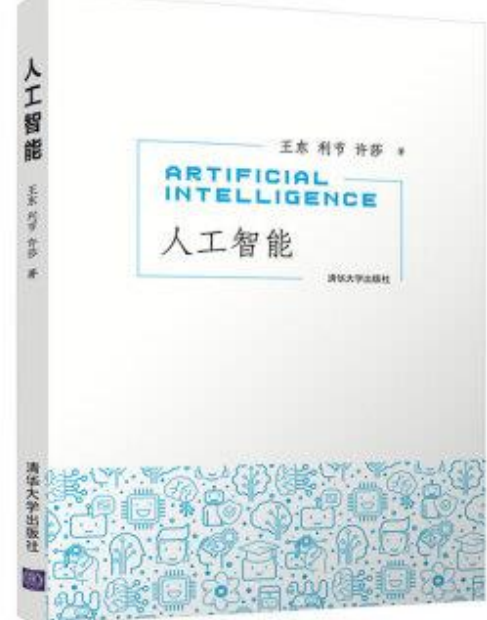


利节



AIDemo示例

<http://aibook.csit.org>



实验准备:

AIDemo 提供了一个轻量级的示例程序

SpeechSeparation, 读者可以通过这个例子加深对语音信号处理过程的理解。**SpeechSeparation** 是一个语音分离程序, 输入一个多音源混杂的语音, 该程序可以将每个音源的声音分离出来。



实验原理：

本实验将基于最近提出的深度聚类(**Deep Clustering**) 方法对混杂语音进行分离。这一方法的基本思路是将声音转化为频谱图，将该图上每个时-频块 (**TF**) 通过一个深度神经网络映射到一个特征空间，使得在该特征空间中各个**TF** 块的相邻矩阵与依实际标注得到的相邻矩阵尽可能相似。基于这一特征空间，对混合语音的**TF** 块进行聚类，再将同一类的**TF** 块选出来作为一个声源的频谱，即可实现语音分离。



实验一：运行缺省配置

打开终端窗口，进入`speech/speechSeparation/code` 目录，运行`run.sh`程序，即可观察到基于缺省参数的DeepCluster 语音分离效果。查看`run.sh`，可以发现该过程将调用如下Python 命令：

```
python3 main.py model/speech_model.h5
```

```
./data/speech_separate/01u/01ua010a.wav./data/speech_separate/20r/20ra010a.wav
```



实验一：运行缺省配置

上述命令将01ua010a.wav 和20ra010a.wav 两个声音文件进行混合，得到音源混合的语音文件mixed.wav，再对该混合语音利用DeepCluster 算法进行分离，得到分离后的两路声音，保存成输出文件out_1.wav 和out_2.wav。播放mixed.wav,out_1.wav 和out_2.wav，可以感受DeepCluster 的分离效果。观察spec.pdf 中所示的频谱，可以查看DeepCluster 对TF 的聚类情况，如图3.24所示。从图中可以看到，DeepCluster 算法可以实现准确度较高的TF 聚类，从而实现较好的音源分离。



实验二：体验不同场景下的分离效果

前一个实验中的两个声源文件，`01ua010a.wav` 和 `20ra010a.wav`，是两个性别不同的发音人发出的内容不同的语音，因此音源之间的差异性比较大，**DeepCluster** 比较容易实现对音源的分离。本实验将探讨在相对复杂的条件下，**DeepCluster** 方法的分离性能。下面的实验均需通过修改 `run.sh` 完成。



实验二：体验不同场景下的分离效果

(1) 观察说话内容对人声分离效果的影响

依两路音源的发音内容不同，分离的难度会有所不同。通过修改 `run.sh`，比较 `DeepCluster` 对内容相同的发音和内容不同的发音的分离效果。相同内容：

```
python3 main.py model/speech_model.h5  
../data/speech_separate/01u/01ua010a.wav  
../data/speech_separate/20r/20ra010a.wav
```

不同内容：

```
python3 main.py model/speech_model.h5  
../data/speech_separate/01u/01ua010a.wav  
../data/speech_separate/20r/20ra010b.wav
```




实验二：体验不同场景下的分离效果

(2) 观察发音人性别对人声分离效果的影响

依两路音源的发音人性别不同，分离的难度会有所不同。通过修改 `run.sh`，比较 `DeepCluster` 对同性别混合发音和不同性别混合发音的分离效果。相同性别：

```
python3 main.py model/speech_model.h5  
../data/speech_separate/01u/01ua010a.wav  
../data/speech_separate/027/027a010b.wav
```

不同性别：

```
python3 main.py model/speech_model.h5  
../data/speech_separate/01u/01ua010a.wav  
../data/speech_separate/20r/20ra010b.wav
```



实验二：体验不同场景下的分离效果

(3) 观察音调匹配性对人声-背景音乐分离效果的影响

如果两路音源属于不同类的声音，则分离会相对容易，如音乐和人声。然而，分离效果可能受到音调匹配度的影响。下面的实验考察人声和音乐在音调匹配和不匹配两种情况下的分离效果。

音调匹配的人声和音乐分离：

```
python3 main.py model/music_model.h5  
../data/music_separate/0_accompaniment.wav  
../data/music_separate/0_vocals.wav
```

音调不匹配的人声和音乐分离：

```
python3 main.py model/music_model.h5  
../data/music_separate/0_accompaniment.wav  
../data/music_separate/1_vocals.wav
```



实验三：分离自己的声音

录制你和朋友的混合声音，上传到**AIDemo** 虚拟机，对该混合声音进行音源分离。通过实验，可以发现你和哪位朋友的声音更相像（相像意味着更不容易被分离）。



The end !