音频分析及处理作业要求

多媒体（2）课程

2017.9.28

清华大学 软件学院

1. 作业概述

多媒体（2）课程的音频实验为一次小组实验，每个小组人数不得超过 3 人，也可 1 人或 2 人一组。作业成绩将根据完成难度、完整度、创新性等方面来给定。

2. 问题描述

2.1概述

同学们在本次实验中将实现一个全自动的多人会话记录器，该记录器的功能是：在一个多人进行对话的场合中（假设环境噪音很小），自动记录所有谈话内容，并且识别出该场景中一共有多少人参与谈话，分析出每个人说话的音频片段，并将说话的音频内容自动转成文本内容（中文或英文皆可），最后将所有谈话记录输出成文本对话。例如，输出内容可能如下：

TalkerA: Have you watched the NBA match last night? It was incredible. What a big win!

TalkerB: Yeah, best of the year.

TalkerC: I missed that match, but I will watch the replay today.

TalkerB: You must. It was awesome.

TalkerA: No wonder they could win the champion last year.

TalkerC: Alright. I will do it.

由于该实验涉及多个部分，包括音频自动分段、话者人数估计、话者识别、音频转文本等。本实验要求是离线完成以上各项操作，即会话片段已提前

录制完成并作为输入。有兴趣的同学可以选择在线完成以上操作，即随着谈话的进行而实时的识别出话者并将其谈话内容实时的转成文本，谈话结束后，所有谈话内容即生成完毕。

本次实验的输入可由同学们自己根据选择的离线会话或在线会话而录制，最后提交时将录制的音频一同提交。

注：若有同学希望完成自选题目或对本实验内容有任何疑问，请联系助教：钱珺，jy8239778@126.com，18651144571。

2.2音频自动分段

输入一段包含语音和空白的音频片段，将该音频片段分段，使得同一话者所说的连续的一段音频独立成段，无谈话的段落不用保留。例如针对前文中的例子，分段之后得到的结果为 6 个音频小片段，按先后顺序为：TalkerA、TalkerB、TalkerC、TalkerB、TalkerA、TalkerC。

一种可能的执行方式是，先找出静音段和语音段，以静音段作为分段依据，保留语音段。

2.3话者人数估计

根据给出的音频片段，估计参与谈话的人数，无需判断每个话者具体是什么人，在最后输出到文本时，可以用话者 1、话者 A 等来替代。

2.4话者识别

将音频自动分段和话者人数估计的结果作为输入，判断每个分段是哪位话者发出，例如分段 1 对应话者 2 等。

2.5音频转文本

将各语音分段作为输入，输出各语音段中的内容到文本。并根据话者识别的结果，将谈话的所有内容组织成前文例子中的谈话记录。

2.6实时会话记录

实时会话记录是本次实验的提高要求,随着谈话的进行而实时的识别出话者并将其谈话内容实时的转成文本，谈话结束后，所有谈话内容即生成完 毕。
注:为避免作业任务过重,以上各部分内容均可以使用现有的库,参考的库链接在本文末尾。同学们也可以自行搜索查找相关的可用库，并自行学习使用。

3. 作业提交

作业请于 2017 年 11 月 6 日晚 24 点前提交至网络学堂。

内容包括以下几个方面：

1） 作业报告：较详细的介绍实现的算法，突出介绍作业的创新点，同时需要说明小组分工情况；

2） 源代码：包括同学自己的源代码，以及依赖的外部库文件，若外部库文件过大，请给出相关链接；

3） 仅实现离线会话的同学需提交录制的音频文件。

4） 可执行程序及运行说明；

4. 评分标准

实验基础分为 90 分，要求如下：

1） 音频分段、话者估计与识别准确，占 50 分

2） 音频转文本，输出的文字正确无误，占 20 分

3）实验报告内容充分且不冗余，占 10 分

4）课堂展示，条理清晰，内容全面，占 10 分，时间为 11 月 9 日

额外分 10 分。一切合理的额外或者创新工作，都可以获得一定量的额外分数，比如将上文提到的将音频输入由离线提高至在线等。请在实验报告中写明你认为可以获得额外分的点。

5. 参考资料

1）Marsyas（http://marsyas.info/），音频特征提取、分析、合成、处理库，也含有部分机器学习算法的实现，C++语言；

2）jAudio（http://sourceforge.net/projects/jaudio/），功能与 Marsyas 类似，Java 语言；

3）Lame（http://lame.sourceforge.net/），一个高质量的 MP3 文件编码、解码器，可用其将 MP3 文件转为无损的 WAV 文件；

4）Essentia（http://essentia.upf.edu/），基于 C++的音频分析处理包，含 Python 调用接口；