**P1：Lab 2: Automatic Speech Recognition**

**P2：这是一个语音识别系统的框架示意图**

**P3：现在语音交互VUI触发的常见类型有：**

（1）语音触发：例如iphone配置的siri

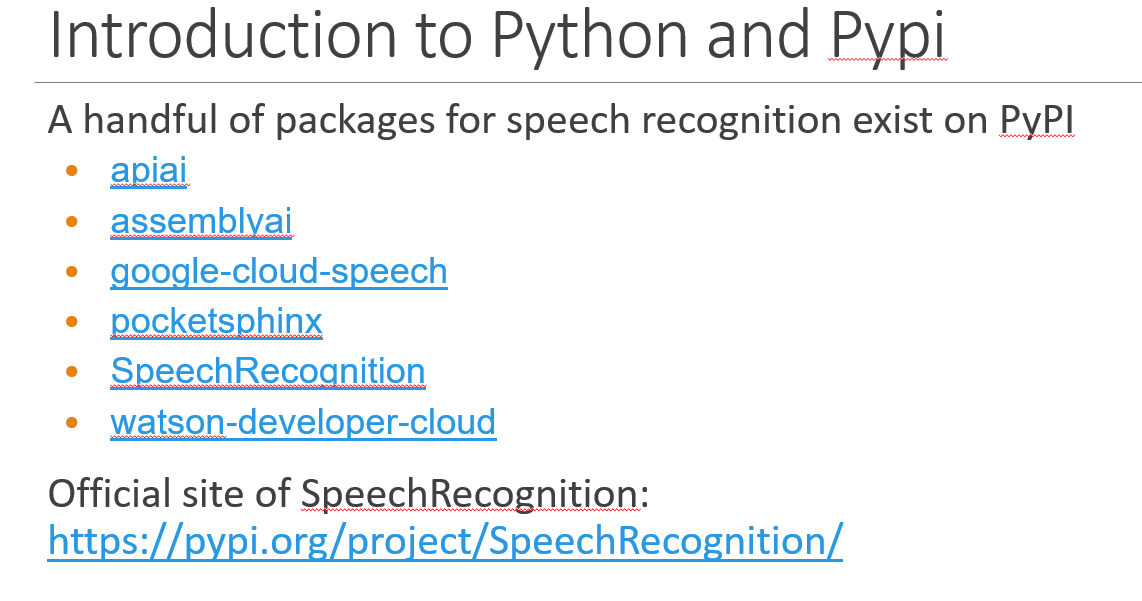
（2）触觉触发：按下按钮或者切换空间

（3）动作触发：在传感器前完成某种手势动作

（4）设备自己触发：事先预定好的设置触发设备

**P4：这个作业，推荐用Anaconda版python完成**

**P5：同时需要安装以下软件包**



PyPI是Python的正式第三方软件包的软件存储库，它类似于CPAN。一些软件包管理器例如pip，就是默认从PyPI下载软件包。用户通过PyPI可以下载超过235,000个Python软件包。

Python apiai是一个第三方库语音识别和自然语言处理的(模块包)。

可以利用AssemblyAI（https://www.assemblyai.com/）的高精度语音转文本Web API，从MP3录音中提取文本（也支持许多其他格式）。

https://jishuin.proginn.com/p/763bfbd2aca2

google-cloud-speech基于神经网络模型，可从80多种语言的音频文件中进行语音到文本的转录。

https://cloud.google.com/speech-to-text

SpeechRecognition是python自带的语音识别函数库，无需构建访问麦克风和从头开始处理音频文件的脚本，只需几分钟即可自动完成音频输入、检索并运行。因此易用性很高。

<https://blog.csdn.net/qq_32643313/article/details/99936268?spm=1001.2101.3001.6650.1&utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1.pc_relevant_antiscanv2&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7ECTRLIST%7ERate-1.pc_relevant_antiscanv2&utm_relevant_index=2>

pocketsphinx是由美国卡内基梅隆大学开发的一个计算量和体积都很小的嵌入式语音识别引擎。适合用于移动机器人的语音交互系统开发。

https://blog.csdn.net/zouxy09/article/details/7942784

IBM Watson是机器学习和认知计算最著名的使用平台之一。 IBM Watson Developer Cloud于2013年十一月推出，并提供了一套完整的API（常用功能，测试以及实验），允许开发人员利用机器学习技术，如自然语言处理、计算机视觉以和预测功能，来构建应用程序。IBM Watson Developer Cloud的API套件包括：语音到文本、文本到语音、权衡分析、独特见解、提问和回答、语气分析器以及视觉识别。

**P6：windows里**

这是windows里配置python环境的方法。

1、打开 Anaconda Prompt

2、切换到项目目录下(以D盘为例)：

切换盘符：D:

切换目录：cd 目录名称

3、创建ananconda环境：

激活conda：conda activate

创建python3.7虚拟环境：conda create -n speech python=3.7

激活虚拟环境： conda activate speech

4、pip配置清华镜像源：

pip config set global.index-url https://pypi.tuna.tsinghua.edu.cn/simple

python自带的语音识别函数库SpeechRecognition的安装方法

**P7：然后安装SpeechRecognition：pip install SpeechRecognition**

**P8：然后在windows里安装PocketSphinx**

1、下载二进制包：https://www.lfd.uci.edu/~gohlke/pythonlibs/#pocketsphinx

以windows python3.7版本为例：https://download.lfd.uci.edu/pythonlibs/archived/cp37/pocketsphinx-0.1.15-cp37-cp37m-win\_amd64.whl

2、将whl文件放在项目目录下

3、安装二进制包：pip install pocketsphinx-0.1.15-cp37-cp37m-win\_amd64.whl

**P9：**

然后安装PyAudio：pip install PyAudio

**P10:**

然后安装PyQT5

**P11：**

然后可以运行示例程序。

**P12：**

个人项目2的参考用户界面，在提供的代码里

同学们的报告里，要说明对界面和代码的修改

界面要能够接收用户的声音指令，识别内容，执行任务

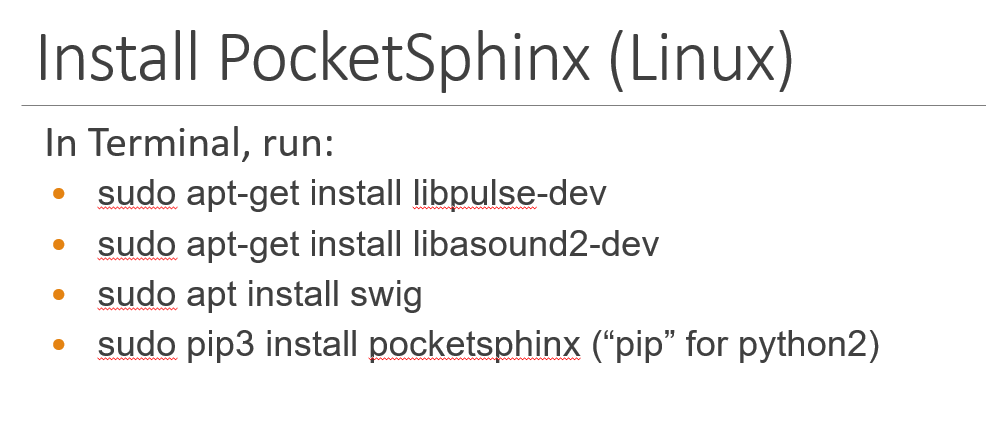
输入为：用户的声音指令

输出为：执行指令任务

一些其他参考

https://blog.csdn.net/zxy13826134783/article/details/103642266

**P13：Linux里**



可以安装卡内基梅隆大学的PocketSphinx，在terminal里跑以下指令

先读取软件包列表

安装语音开发包libasound2-dev

安装SWIG这个接口生成器，为脚本语言(tcl，perl，python等)提供C和C++的接口，

安装PocketSphinx

运行 pocketsphinx进行语音识别需要指定三个文件：声学模型、语言模型和字典文件（Sphinx是xxx）

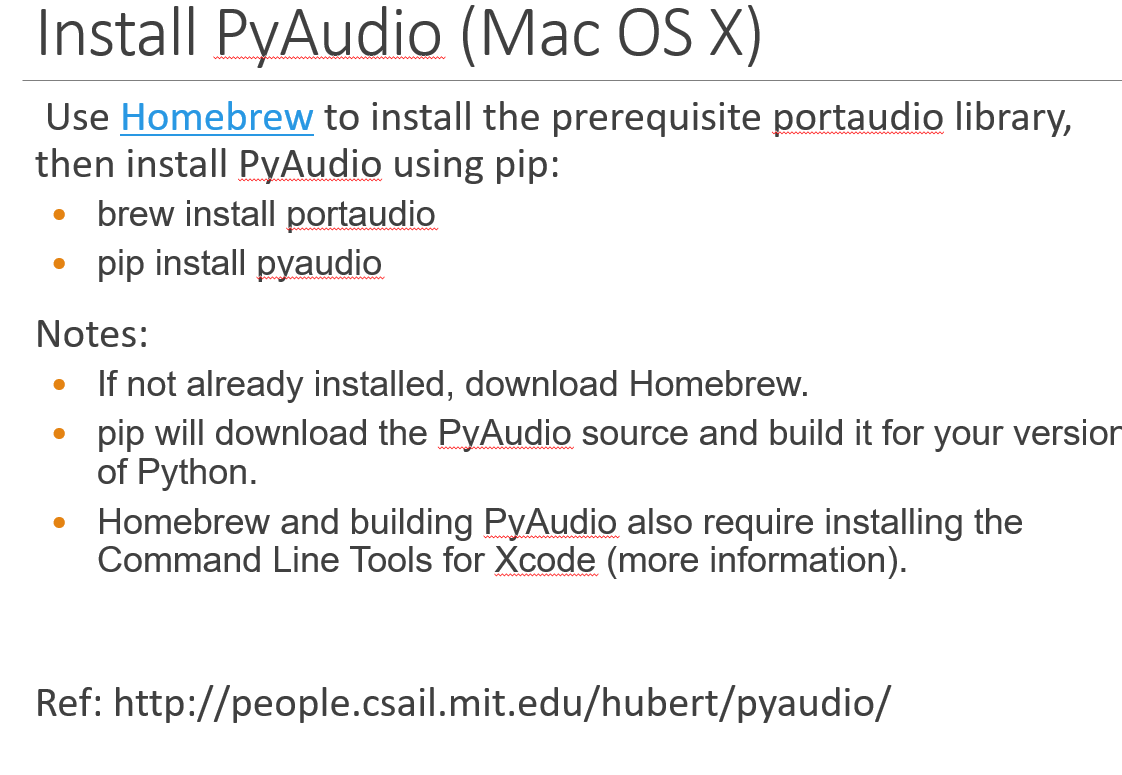
**P14：**

安装PyAudio语音操作工具包，它是一个跨平台的音频处理工具包，使用该工具包可以在Python程序中播放和录制音频，也可以产生wav文件等

一些其他参考

https://blog.csdn.net/qq\_35385687/article/details/119026887

**P15：Mac系统里**



可以用开源软件包管理系统Homebrew安装portaudio库函数，PortAudio是一个跨平台、开源的音频I/O库

然后安装语音操作工具包PyAudio

Homebrew，Xcode的命令行工具也请提前安装

具体参见：

http://people.csail.mit.edu/hubert/pyaudio/

**P16：作业内容**

Xxx

**P17：作业要求**

Xxx