语音识别实验报告

【实验目的】

1. 了解目前开放的各个语音识别API
2. 针对《脑健康系统研究》中的语音题设计测试录音并对所选API进行测试和评估
3. 决定《脑健康系统研究》所使用的语音识别API

【实验说明】

1. 根据调查，一共选择了三个备选API做测试（百度云、科大讯飞、Google），具体请见下一小节
2. Google API测试的方法为合成录音文件为FLAC格式并传输后得到结果，其余API测试为电脑播放录音，手机端得到测试结果
3. 针对《脑健康系统研究》中的语音题，设计测试录音如下：

* 录音A：中文单词测试：金属，哭泣，黑暗，北方，学校，玫瑰，向上，英尺，苹果，钢笔
* 录音B：中文连续测试：草原上有对狮子母子，小狮子问母狮子：“妈妈，幸福在哪里？”母狮子说：“幸福就在你的尾巴上呀”于是小狮子不断追着尾巴跑，但始终咬不到。母狮子笑道：“傻瓜！幸福不是这样得到的。只要你昂首向前走，幸福就会一直跟着你。”
* 录音C：英文单词测试，metal,cry,dark,south,school,rose,up,inch,apple,pen
* 录音D：英文连续测试：Anna, of South Boston, employed as a scrub woman in an office building, reported at the City Hall Station that she had been held up, on State Street, the night before, and robbed of fifteen dollars. She had four little children, the rent was due, and they had not eaten for two days. The officers, touched by the woman’s story, made up a purse for her.
* 每个录音会播放三遍：1遍正常音量、1遍音量较小、1遍正常音量＋噪音
* 录音是我自己录的，在英语上可能有一些词语发音不标准导致结果有偏差

【实验结果】

经过有如下开放的语音识别API：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 语言支持 | 在线 | 收费 | 其他 |
| 百度云语音 | 中英 | 在线 | 免费 | 集成录音功能、语义理解  上线后会有次数限制 |
| 科大讯飞 | 中英 | 在线 | 含免费部分 | 有语义解析功能、降噪  大量使用后会有次数限制 |
| Google语音识别 | 中英 | 在线 | 免费 | 非针对移动应用，必须FLAC格式文件 |
| 搜狗语音云 | 仅中文 | 离线 | 免费 | 服务次数上限：1000/一天  集成录音、降噪网络通讯和状态通知 |
| 微信语音平台 | 中英 | 在线 | 收费 |  |
| 云知声语音云 | 中文 | 在线 | 收费 |  |
| 语意果 | 中文 | 在线 | 收费 |  |

为了便于开发，我们的APP需要兼容中英两种语言，故选取了前三个（百度、讯飞、google）作为备选API。

通过接入SDK后（Google为合成录音并传送数据流）得到测试结果如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 中文单词 | 中文连续 | 英文单词 | 英文连续 |
| 百度云语音 | 正常音量和加噪下非常顺利通过测试  小音量下错误率比较高 5/10 | 正常音量加噪下错了一句话  大体识别正确  小音量下几乎不能用 | 正常音量：识别对了7/10  噪音：5/10  小音量：3/10 | 正常音量情况下文字大体长得差不多（语法基本能识别正确），关键词有错误情况  噪声情况下略差  小音量几乎不能用 |
| 科大讯飞 | 正常音量下顺利通过测试  噪声环境下有一个词语错误  小音量下错误率较高 5/10 | 正常音量和加噪情况下顺利通过测试  小音量下错误率很高 | 正常音量：8/10  噪音：7/10  小音量：3/10 | 和百度云的情况相近 |
| Google语音识别 | 正常10/10  噪声9/10  因为是文件传输，小音量没有很大影响8/10 | 加噪情况下错了2句话  小音量下错了1句 | 10/10  小音量：8/10 | 语法基本正确  关键词基本上都差不多 |

百度的sdk中有一些模版的接口，比较适用于android开发。

在中文处理上，百度和讯飞做得还算不错，正确率能达到90%以上。Google的稍逊。

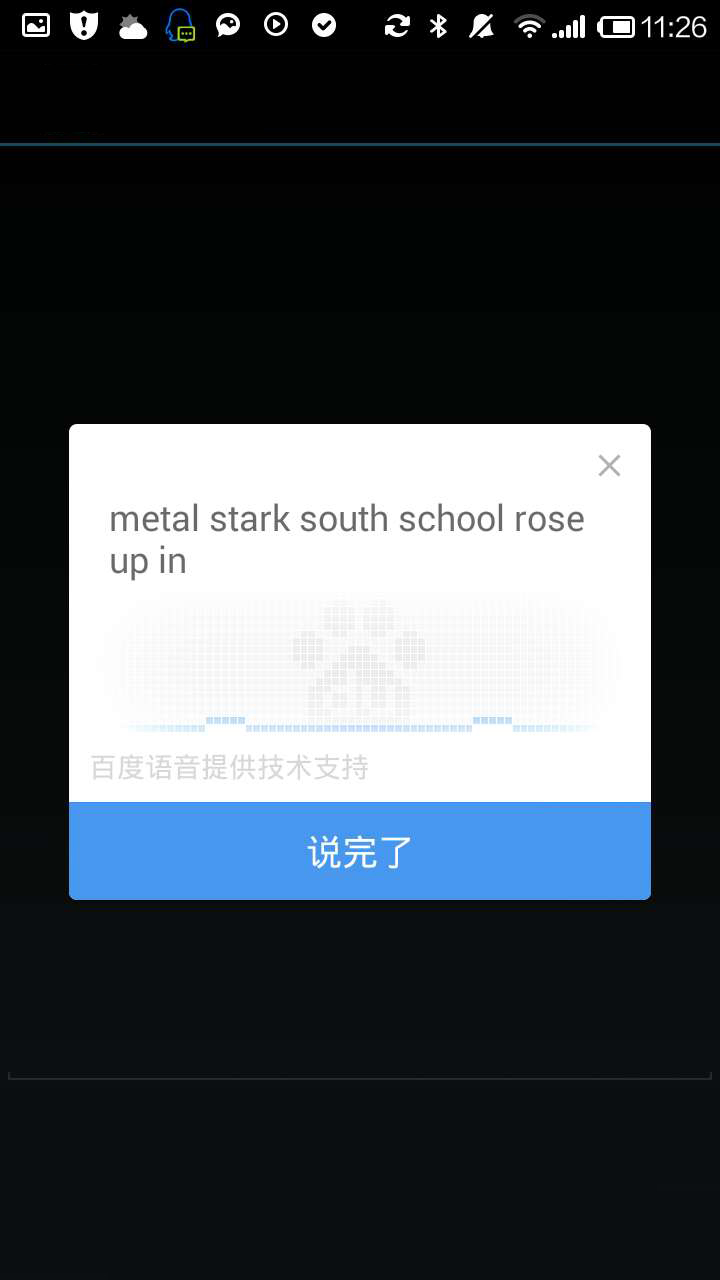
但是在英语处理上，百度和讯飞在语法的识别上还是可以，但是在关键词的识别上稍微差一些。

小音量的情况下，百度和讯飞处理都非常不好。

Google的中英文处理总体性能更好，但是由于它规定了接口FLAC文件，它不适合做手机应用上。

总体来说，百度和讯飞在测试结果上没有明显的差异。

附：实验截图（百度测试）



【实验结论】

1. 有三个比较合适（免费、支持中英文）备选API：百度、讯飞、google
2. 主要实验时间花费在接口SDK上
3. 中文处理上百度和讯飞功能强于google，加噪的条件下准确率90%以上
4. 英文处理上，百度和讯飞对语法的识别还算可用，但有关键词识别错误（可能测试的发音有问题）
5. 小音量下API的性能很差完全不能使用
6. Google的api非针对移动端，考虑到接到android开发上可能会非常不方便，百度和讯飞的总体上来说性能接近。由于百度的应用接口更容易使用，选择使用百度的语音识别sdk