# 语音MP3转文字调研报告

# 1、调研目的

本次调研主要是针对最适合音频(MP3)转文字的实现方案

# 2、调研方向

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **类型** | **平台** | **地址** | **语言** | **协议** |
| 第三方服务 | 百度 | https://ai.baidu.com/ai-doc/SPEECH/Vk38lxily | java |  |
| 第三方服务 | 腾讯 | https://cloud.tencent.com/document/product/1278/46713 | java |  |
| 第三方服务 | 阿里 | 暂未对接，机器访问不了 |  |  |
| 开源引擎 | vosk | https://alphacephei.com/vosk/ <https://www.jb51.net/article/254571.htm>  http://www.mobange.com/nav/java/93720.html | java | Apache-2.0 |

# 3、调研分析

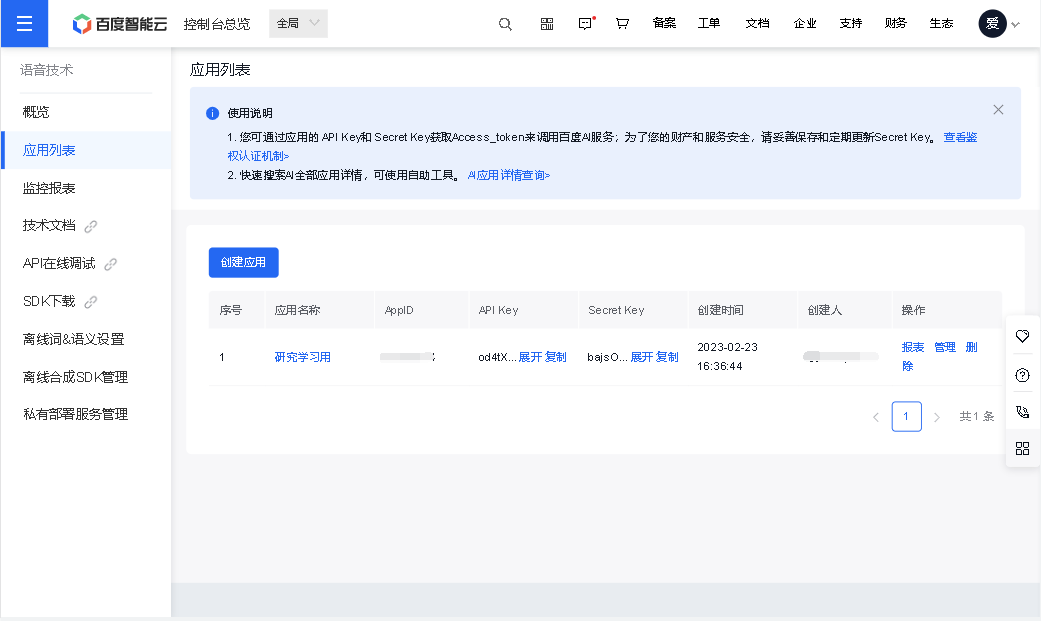
## 3.1、百度开放平台

### Ⅰ、百度智能云秘钥获取

地址：<https://console.bce.baidu.com/ai/#/ai/speech/app/list>

说明：这一步为了得到AppId，APIKey,SecretKey 调用SDK用

图示：

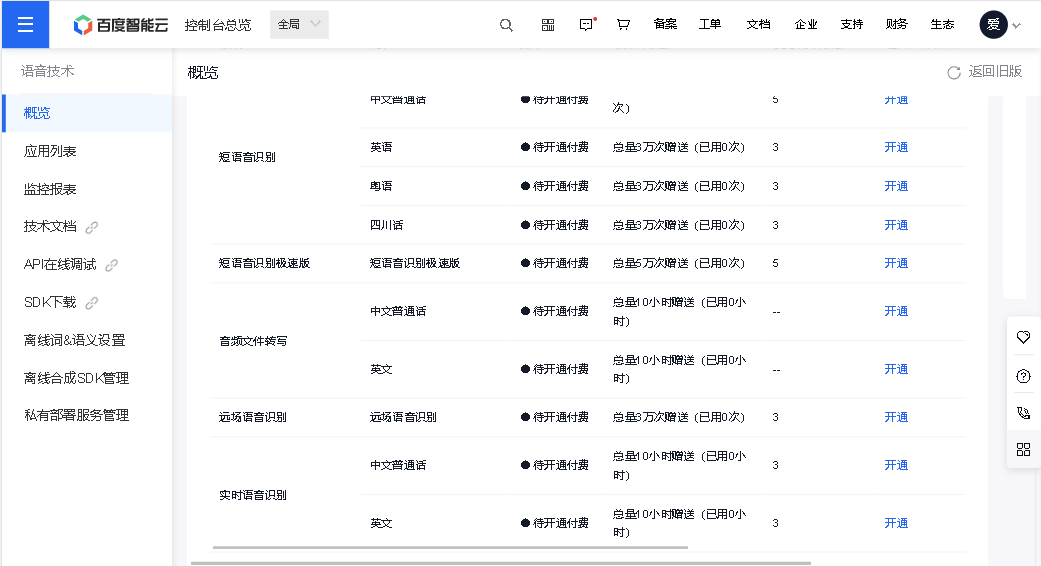


### Ⅱ、对接资料

地址：<https://ai.baidu.com/ai-doc/SPEECH/Vk38lxily>

图示：





## 3.2、腾讯云

### Ⅰ、腾讯云秘钥获取

地址：<https://console.cloud.tencent.com/cam/capi>

说明：这一步为了得到AppId，SecretId, SecretKey调用SDK用

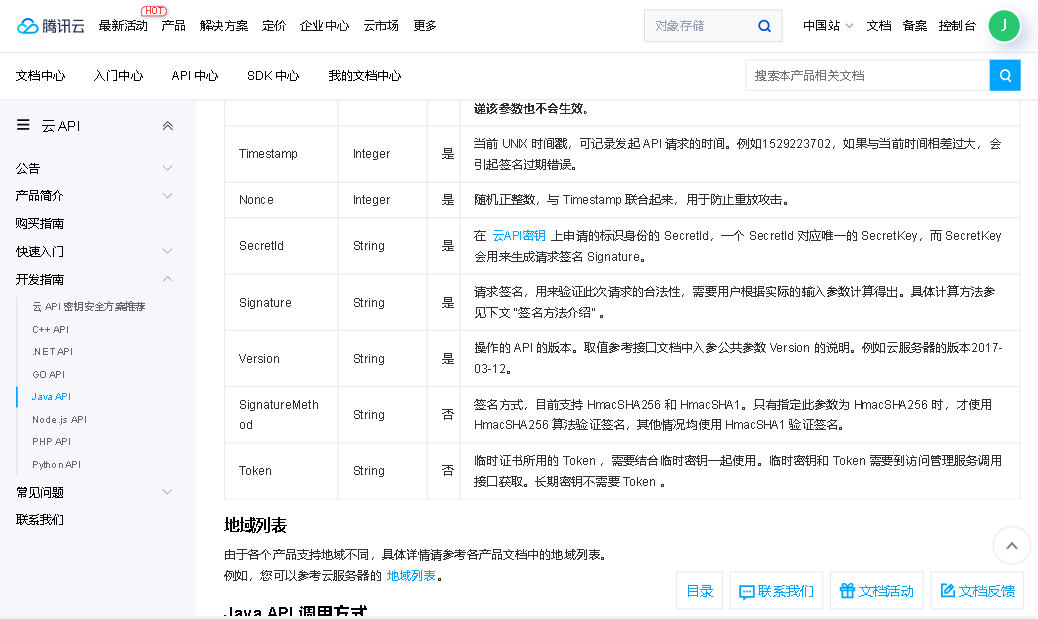
图示：



### Ⅱ、对接资料

地址：<https://cloud.tencent.com/document/product/1278/46713>

图示：



<https://cloud.tencent.com/document/product/1093/52097>



## 3.3、Vosk(开源引擎)

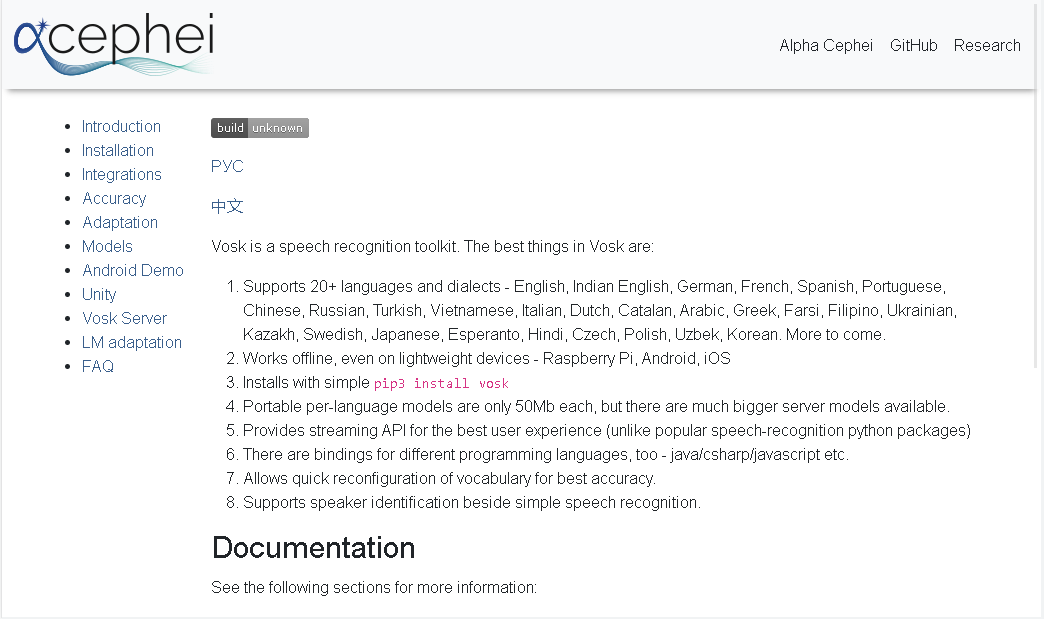
### Ⅰ、对接资料

地址：<https://alphacephei.com/vosk/>

项目地址：<https://github.com/alphacep/vosk-api>

模型地址：<https://alphacephei.com/vosk/models>

图示：



# 4、优缺点

## 4.1、百度智能云

优点：直接调用第三方服务，代码量少

缺点：1、目前格式仅仅支持 pcm，wav或amr，如填写 mp3 即会有此错误(需要代码中转)

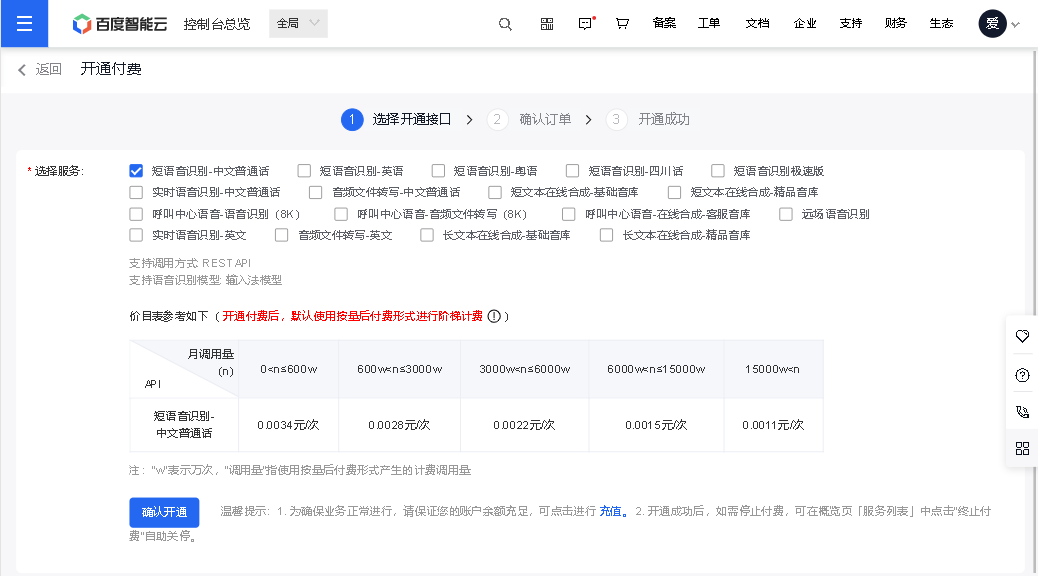
2、识别没有腾讯的准(一个音频百度识别的不准，腾讯可以)

3、需要付费(能试用3个月)

4、文档不是太清晰。

5、时长限制比较短(目前找到的文档是60S)

付费详情：<https://console.bce.baidu.com/ai/?fromai=1#/ai/speech/order/create~apiId=370>



## 4.2、腾讯云

优点：1、直接调用第三方服务，代码量少

2、支持的格式多：支持wav、mp3、m4a、flv、mp4、wma、3gp、amr、aac、ogg-opus、flac格式(比百度支持的多)

3、支持时长5小时以内的录音文件进行识别，异步返回识别全部结果。

4、支持各个时长（60S，30min…）的录音文件识别。

5、有免费额度(下图2)

缺点：暂无

付费详情：<https://cloud.tencent.com/document/product/1093/35686>

1、每月有免费额度，预付费的方式





## 4.3、Vosk(开源引擎)

优点：1、不要钱

2、没有时长限制(但是很慢)

3、可以训练模型(具体怎么实现暂未研究)

缺点：1、仅支持WAVE文件，需要手动转格式

2、完整模型(大型通用中文模型)使用比较耗费资源比较慢，轻量级模型使用解析不准确。

3、现阶段不支持断句(无标点符号和空格)

# 5、总结

## 5.1、解析结果

|  |  |
| --- | --- |
| **类型** | **转换值** |
| 原文 | 欢迎收听有声小说三体 刘慈欣原著 王明君演播 喜马拉雅三体宇宙联合荣誉出品 第89集 电子显微镜屏幕上所显示的 仍然是水滴 光滑的镜面 丁怡问道 现在放大了多少倍？西子指了指显微镜显示屏一角上的一个数字 现在是100倍 说着 他又把放大的倍数调到了一千倍 可放大后的表面 |
|  |  |
| 百度 | 解析不出来:content len too long（内容过长） |
|  |  |
| 腾讯 | 欢迎收听有声小说三体。刘慈欣远助。王明君演播。喜马拉雅三体宇宙联合荣誉出品。嗯。第89集。电子显微镜屏幕上所显示的。仍然是水滴。光滑的镜面。丁怡问道。现在放大了多少倍？西子指了指显微镜显示屏一角上的一个数字。现在是100倍。说着，他又把放大的倍数调到了一千零倍。可放大后的表面。 |
|  |  |
| Vosk | 欢迎收听有声小说三体刘慈欣原著党明军演播喜马拉雅身体宇宙联合荣誉出品的八十九级电子显微镜屏幕上所显示的仍然是谁的光滑的劲烈的疑问的现在发达了多少呗西子指了指显微镜显示屏一脚上的一个数字现在是一百倍说着他又把放大的被薯条到了以前北和放大后的表面 |

## 5.2、结论

1、如果是对接第三方服务的话都可以，按照计价模式来。但是经过对比还是腾讯的合适点。

2、如果想免费的话使用vosk(但是转换可能会不准)。

原因见：[优缺点](#_4、优缺点_1)