# 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 |  |
| 开源引擎 | vosk | https://alphacephei.com/vosk/ | java | Apache-2.0 |
| 第三方服务 | 讯飞 | <https://www.xfyun.cn/services/lfasr> | java |  |
| 第三方服务 | 阿里 | https://help.aliyun.com/document\_detail/432038.html | java |  |

# 3、调研分析

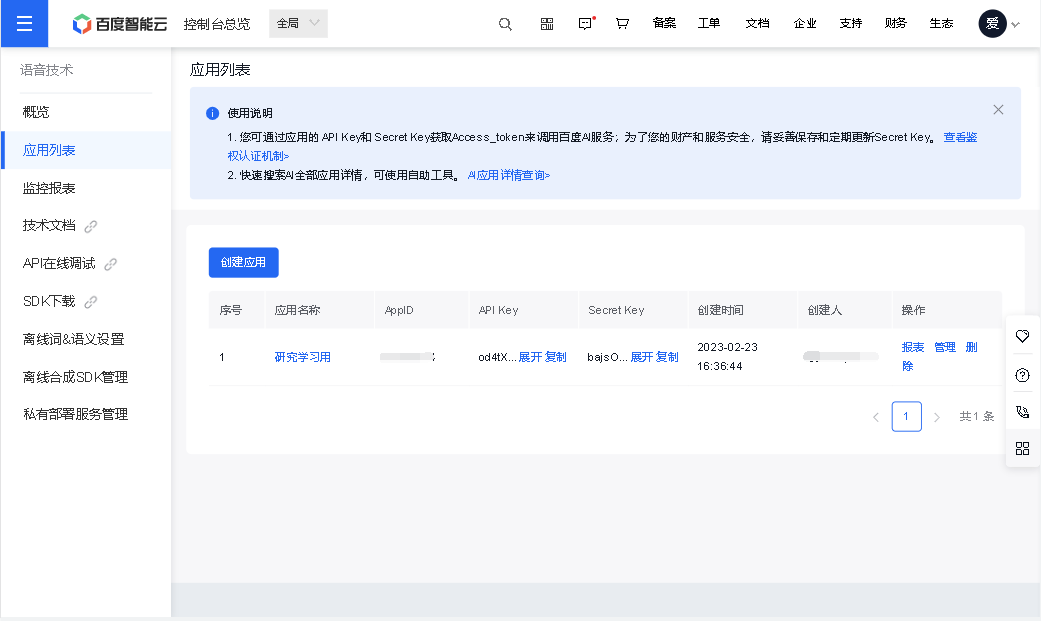
## 3.1、百度开放平台

### 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：

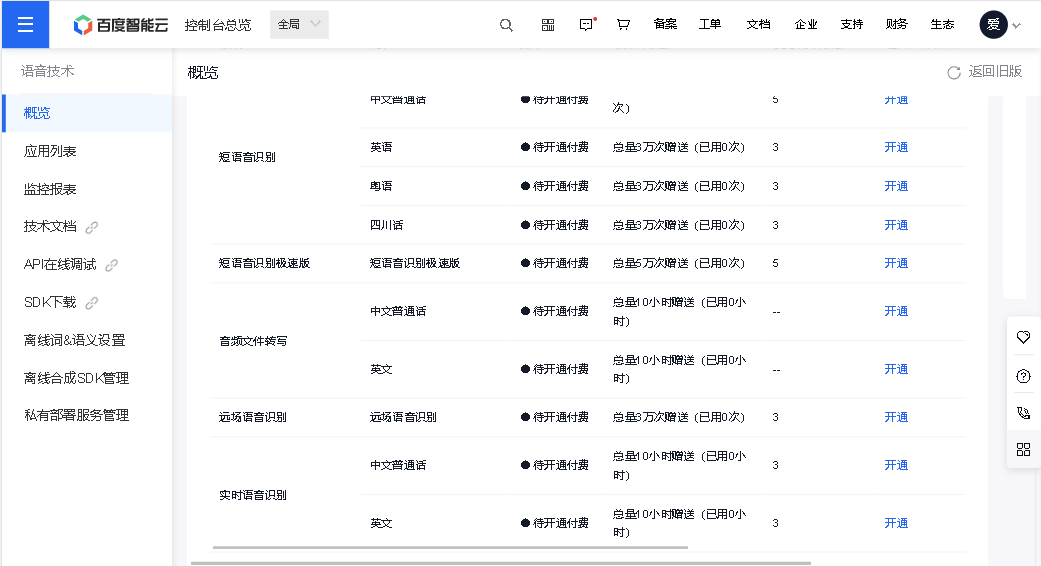


### Ⅱ、对接资料

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

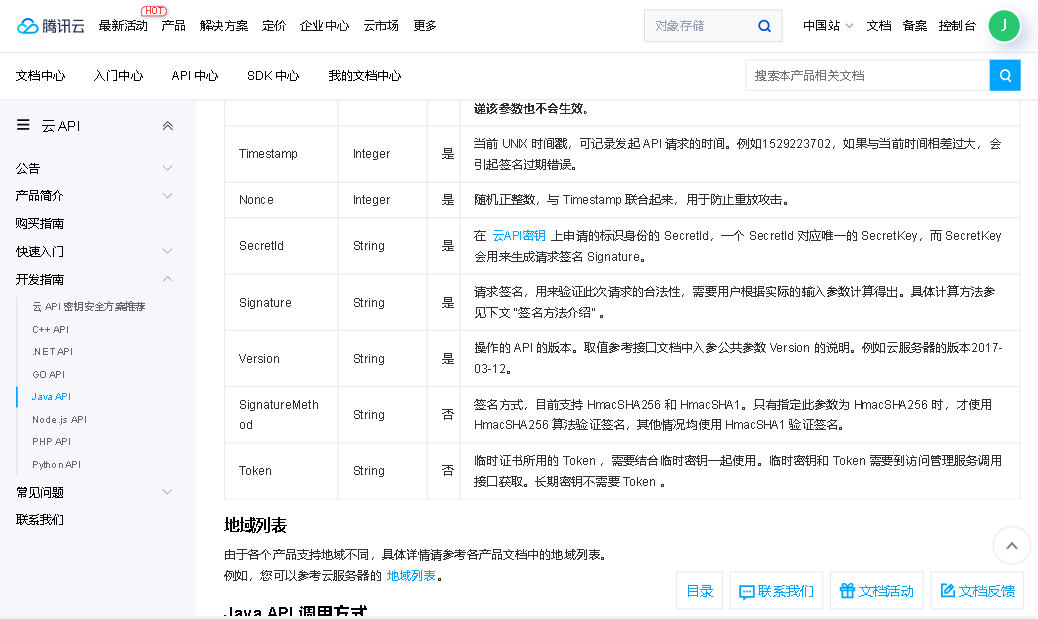
图示：



### 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)

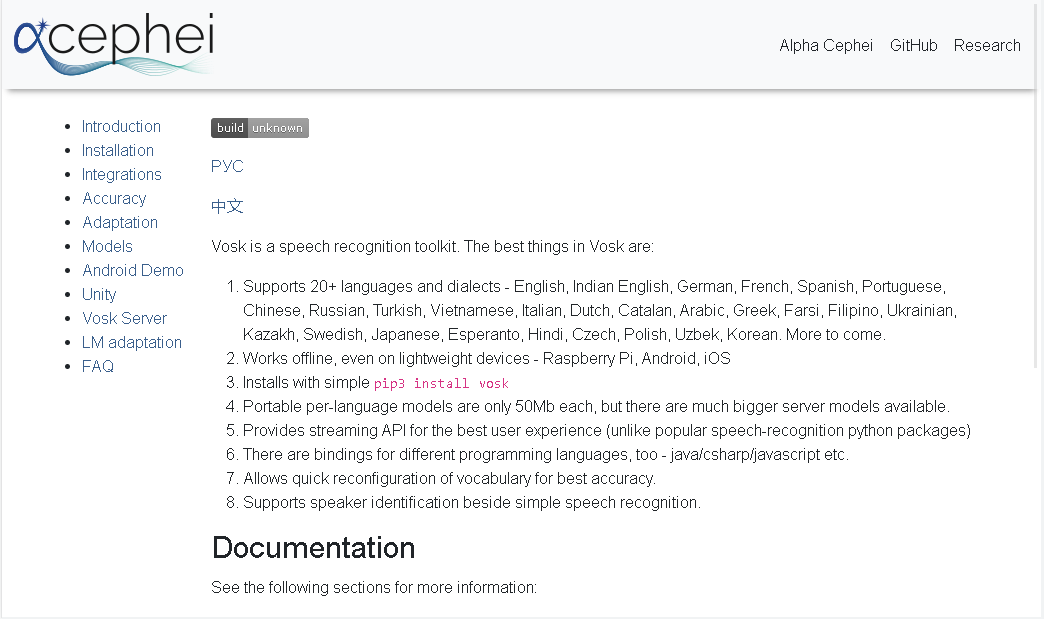
### Ⅰ、对接资料

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：



## 3.4、讯飞

### Ⅰ、讯飞云秘钥获取

地址：<https://console.xfyun.cn/services/lfasr>

说明：这一步为了得到AppId, SecretKey调用SDK用

图示：



### Ⅱ、对接资料

地址：<https://www.xfyun.cn/services/fast_lfasr>

图示：





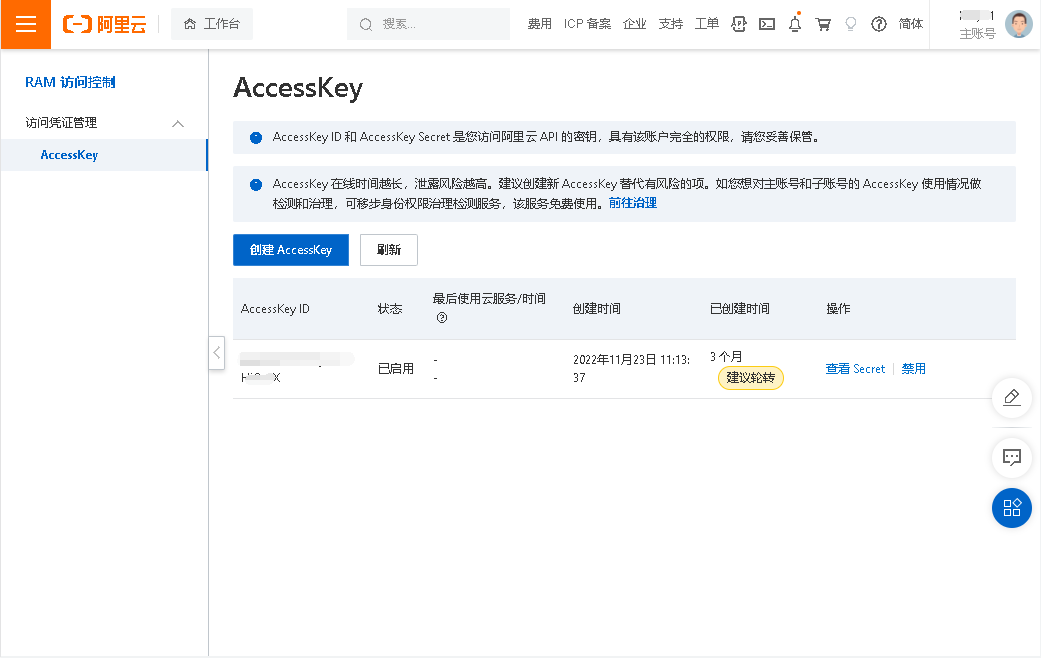
## 3.5、阿里

### Ⅰ、阿里云秘钥获取

地址：<https://ram.console.aliyun.com/manage/ak>

说明：这一步为了得到AppId, SecretKey调用SDK用

图示：



### Ⅱ、对接资料

地址：<https://help.aliyun.com/document_detail/72138.html?spm=5176.27813197.J_2883378880.6.213d50515tvY45>

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)

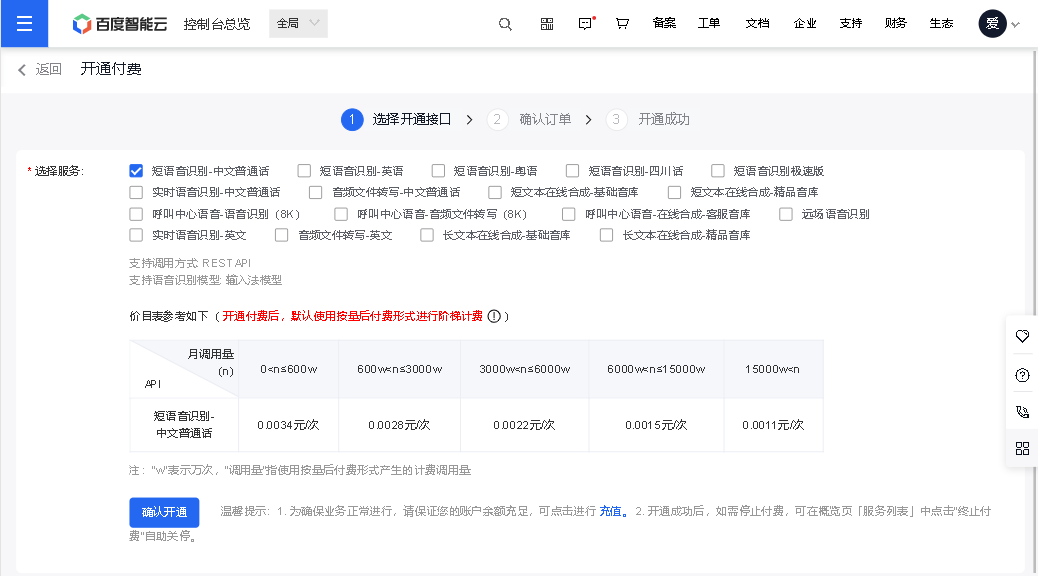
2、识别没有腾讯的准(一个音频百度识别的不准，腾讯可以)

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、有免费额度(下图2)

4、文档清晰

缺点：1、本地上传的文件识别范围为5M以内，否则需要公网文件地址

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)

## 4.4、讯飞

优点：1、时常限制没有

2、识别文件不超过5小时，建议5分钟以上，不超过500M

缺点：1、不支持maven，仅提供jar

2、集成方式复杂，对接文档很差

3、音频文件限制较大：pcm（pcm\_s16le），wav，speex(speex-wb)

采样率为16000 或者 8000.

推荐使用16000，比特率为16bit单声道

付费详情：<https://www.xfyun.cn/services/fast_lfasr#anchor4503211>

1、按时长套餐计费



## 4.5、阿里云

优点：1、直接调用第三方服务，代码量少

2、时间和文件大小限制宽松(支持单轨和双轨的WAV、MP3、M4A、WMA、ACC、OGG、AMR、FLAC格式录音文件识别音频文件大小不超过512 MB，视频文件大小不超过2 GB)

3、文档清晰

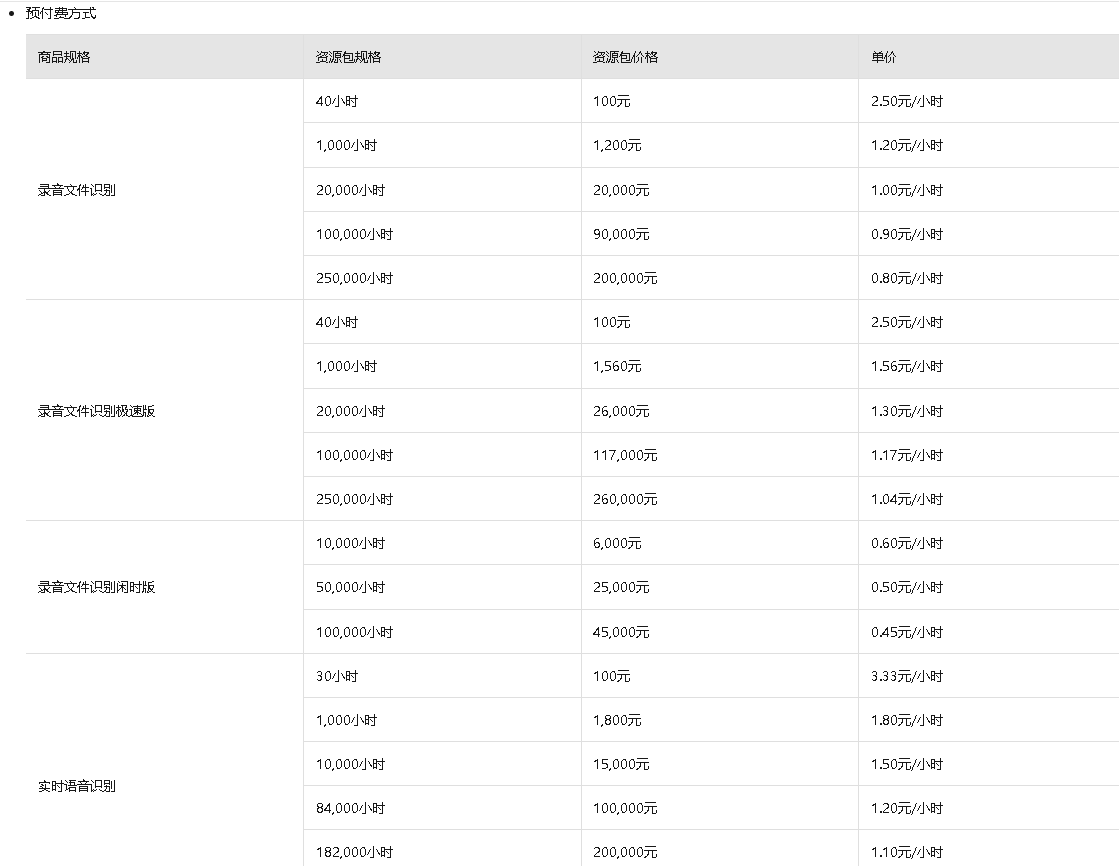
4、急速识别支持本地100M以内不超过2小时的文件

缺点：1、录音文件识别(非急速识别)必须为公网的环境

2、不支持MP3

付费详情：<https://help.aliyun.com/document_detail/207373.html>

1、预付费和后付费



# 5、总结

## 5.1、解析结果

40S的语音

|  |  |  |
| --- | --- | --- |
| **类型** | **转换值** | **耗时** |
| 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 | - |
|  |  |  |
| 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。刘慈欣援助。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。 | 3230毫秒 |
|  |  |  |
| 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 | 12700毫秒 |
|  |  |  |
| 讯飞 | 嗯嗯学生必须从医保是我去好好去唉呀我去我我我 | 4804毫秒 |
|  |  |  |
| 阿里 | 嗯是啊是是是嗯嗯嗯是可以我付出了钱你是我的一二三四五哎嗯谢谢好玩的嗯嗯嗯嗯 | 89254毫秒 |

4S的语音：

|  |  |  |
| --- | --- | --- |
| **类型** | **转换值** | **耗时** |
| 原文 | 北京科技馆。 | - |
|  |  |  |
| 百度 | 北京科技馆。 | 745毫秒 |
|  |  |  |
| 腾讯 | 北京科技馆。 | 910毫秒 |
|  |  |  |
| Vosk | 北京科技馆 | 7339毫秒 |
|  |  |  |
| 讯飞 | 北京科技馆。 | 2455毫秒 |
|  |  |  |
| 阿里 | 北京科技馆 | 5978毫秒 |

## 5.2、结论

1、如果是对接第三方服务暂时建议腾讯。1识别快，2识别准。

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)