# 语音识别调研报告

目录

[语音识别调研报告 1](#_Toc71555933)

[调研目标 1](#_Toc71555934)

[调研内容 1](#_Toc71555935)

[其他服务提供商概要 4](#_Toc71555936)

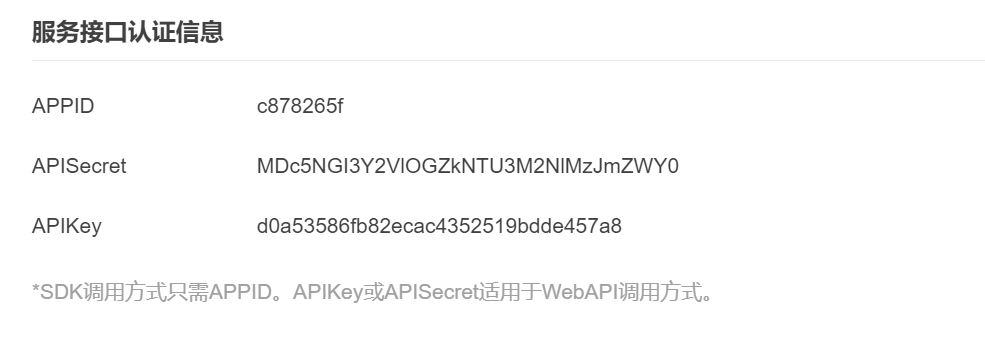
[DEMO实现 7](#_Toc71555937)

[调研总结 12](#_Toc71555938)

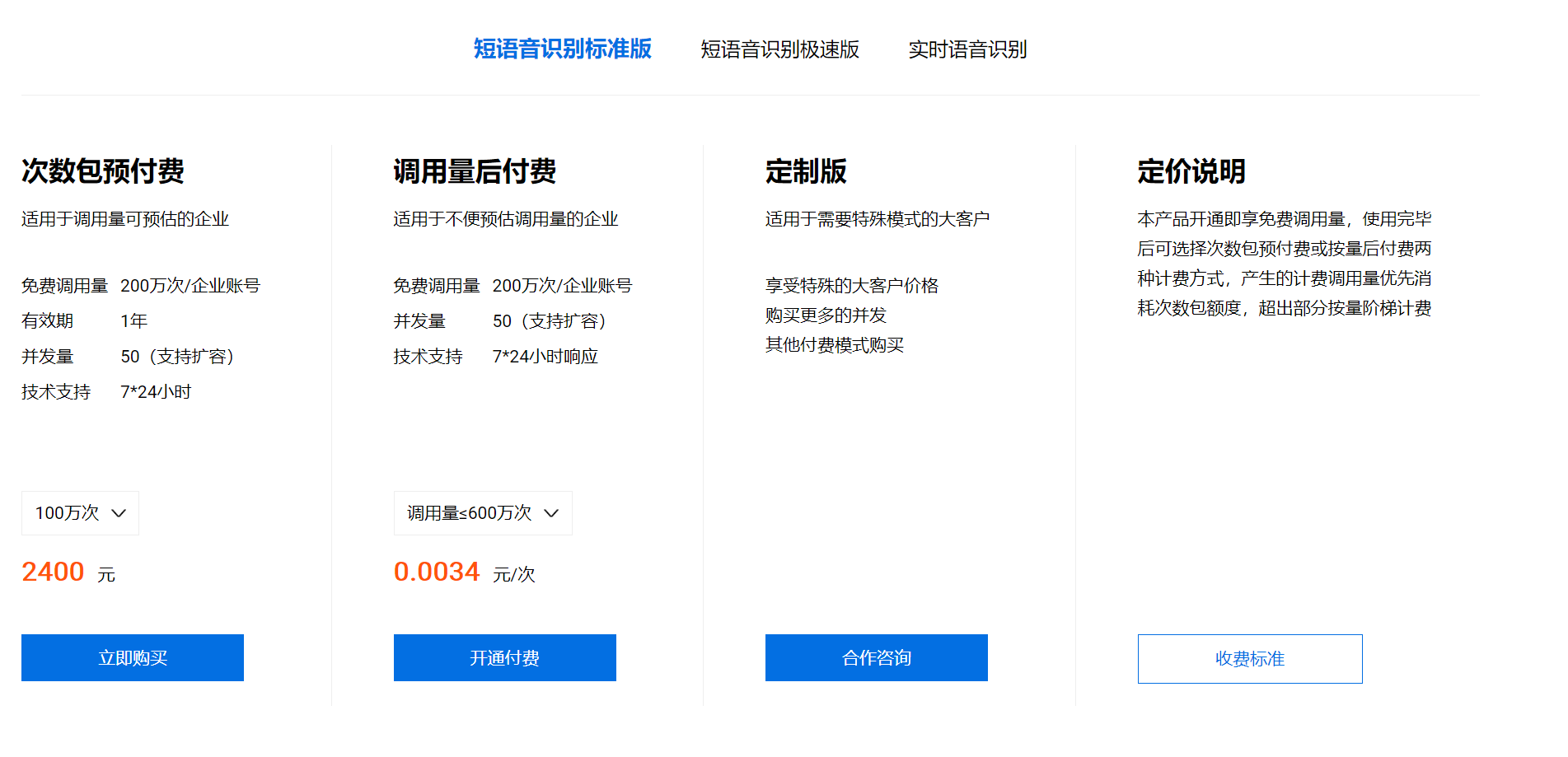
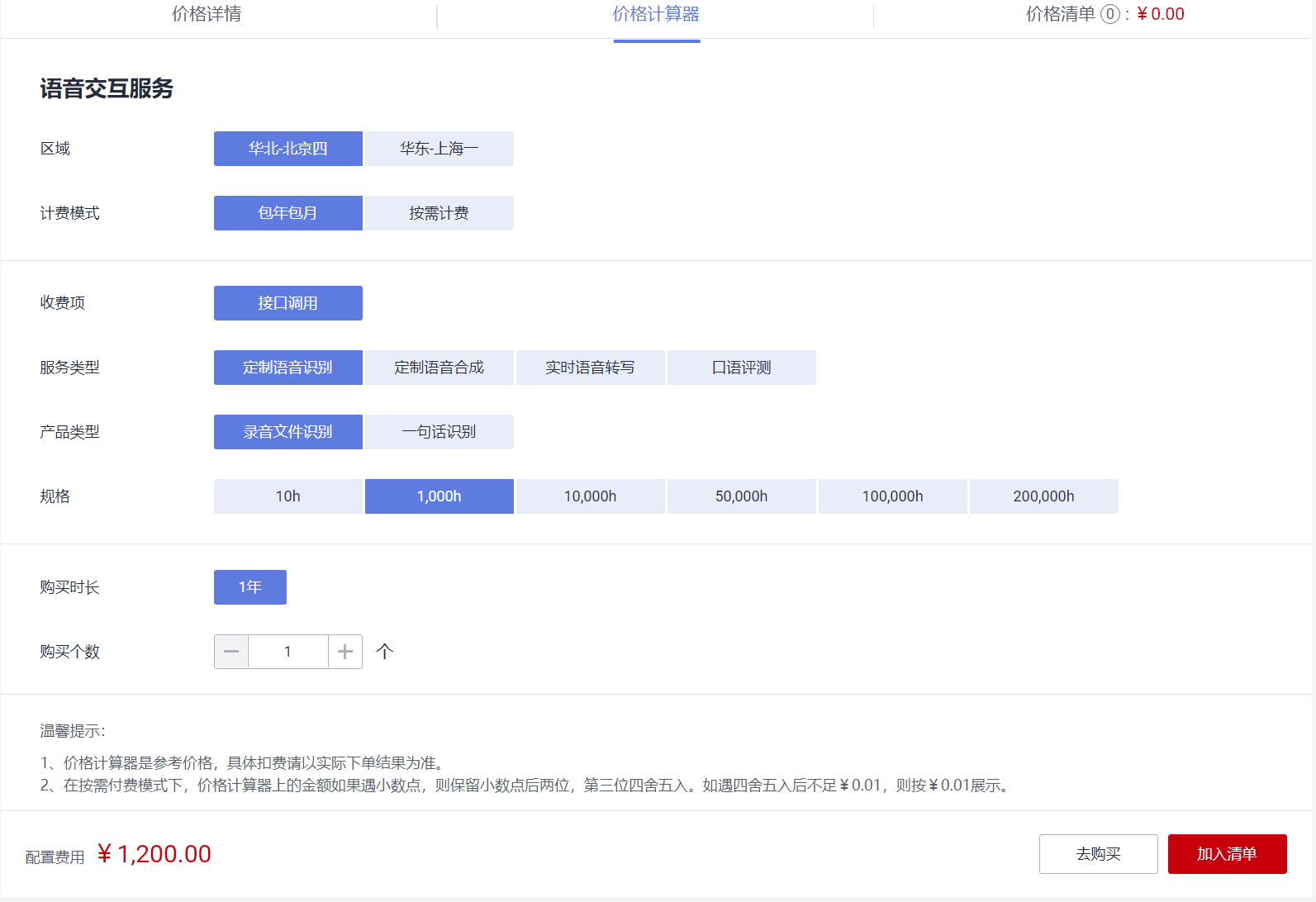
### 调研目标

1. 调研对象：
   1. 科大讯飞
   2. 其他提供方（百度微软腾讯）
2. 调研方向：（以科大讯飞为主要调研对象）
   1. 概要
   2. 适用范围
   3. 现有产品服务
   4. 调用方式
   5. 服务计价
   6. 杂务

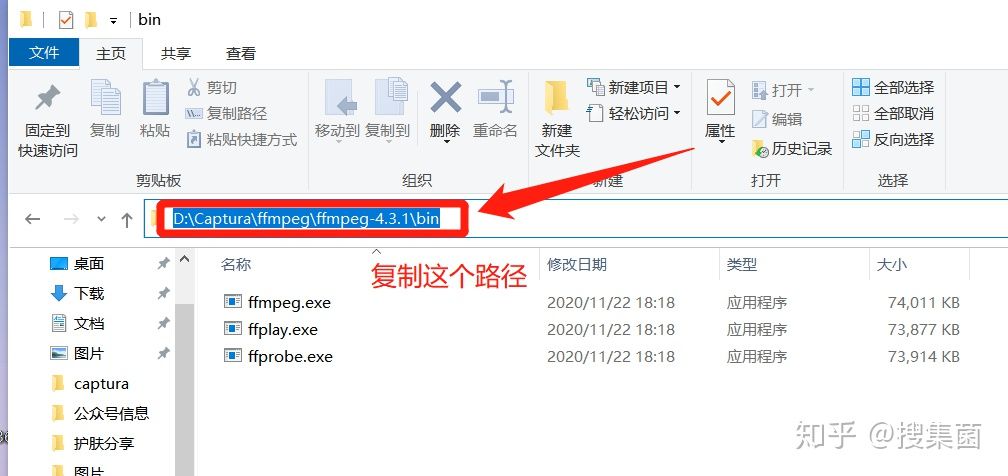
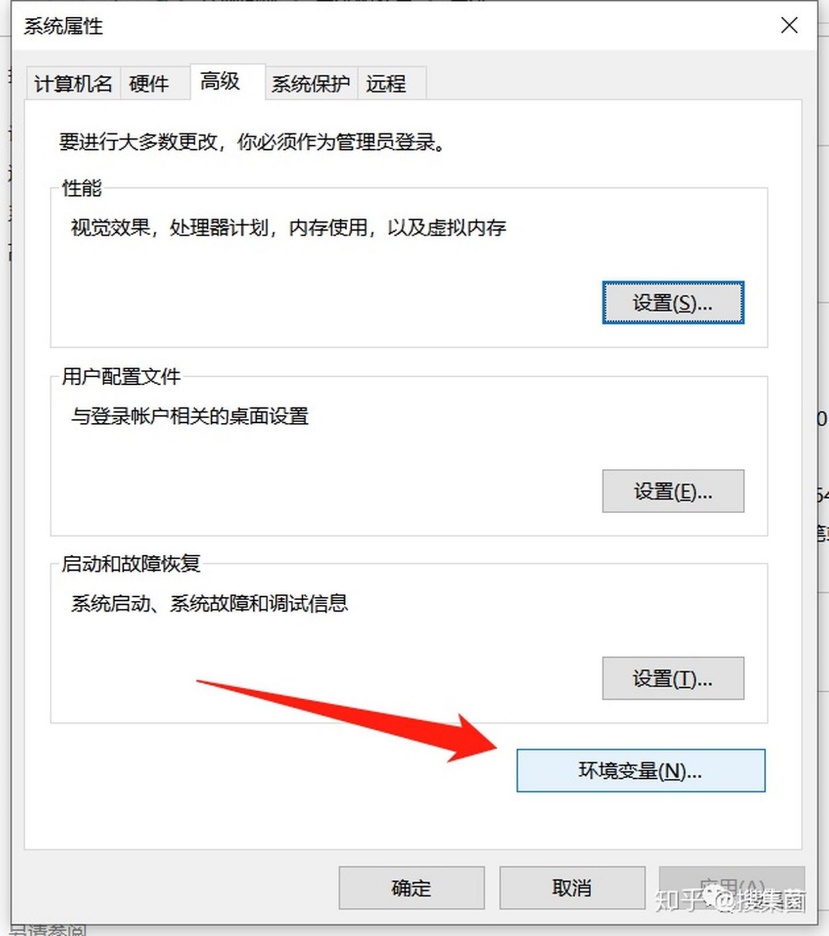
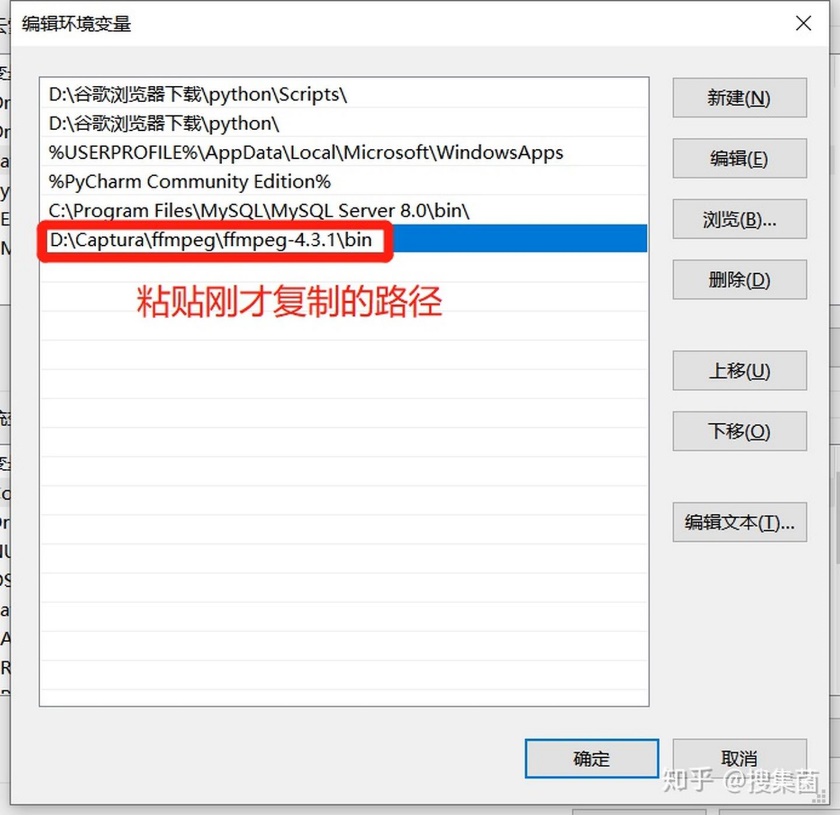
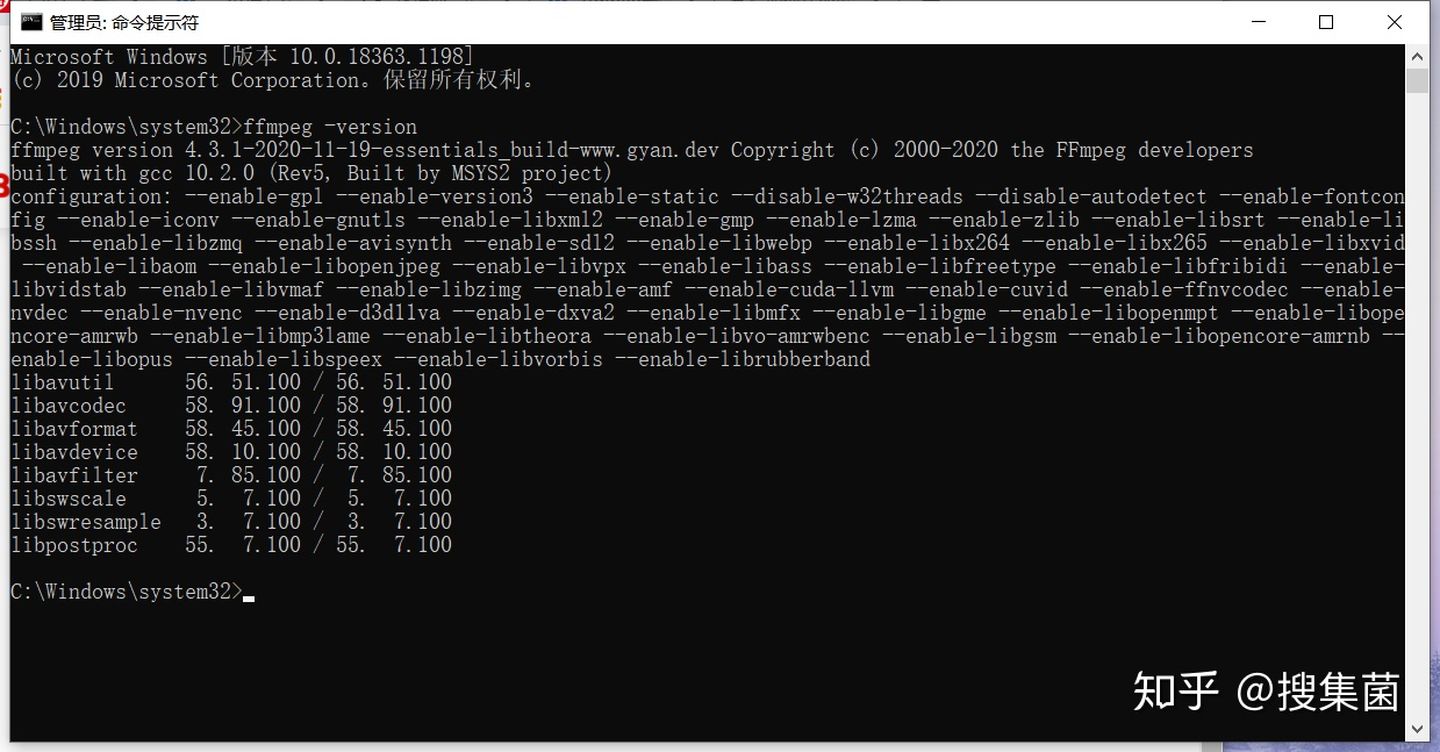
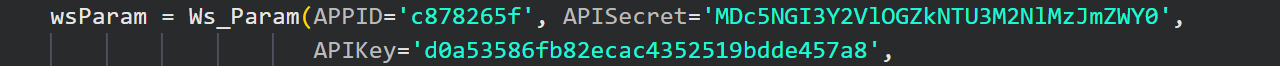
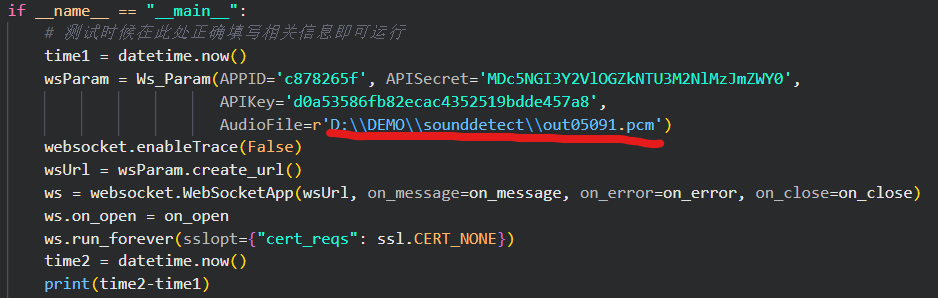
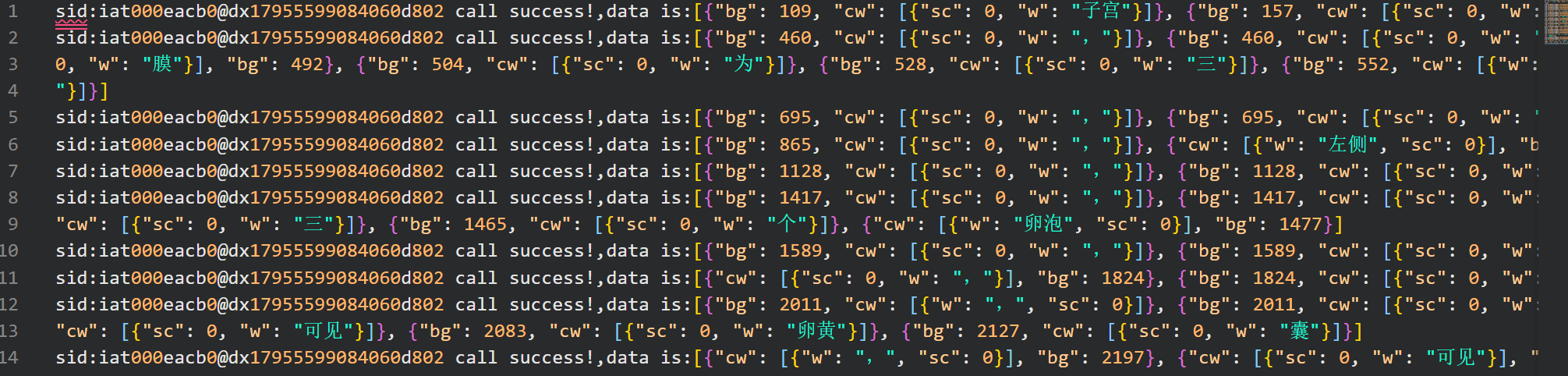
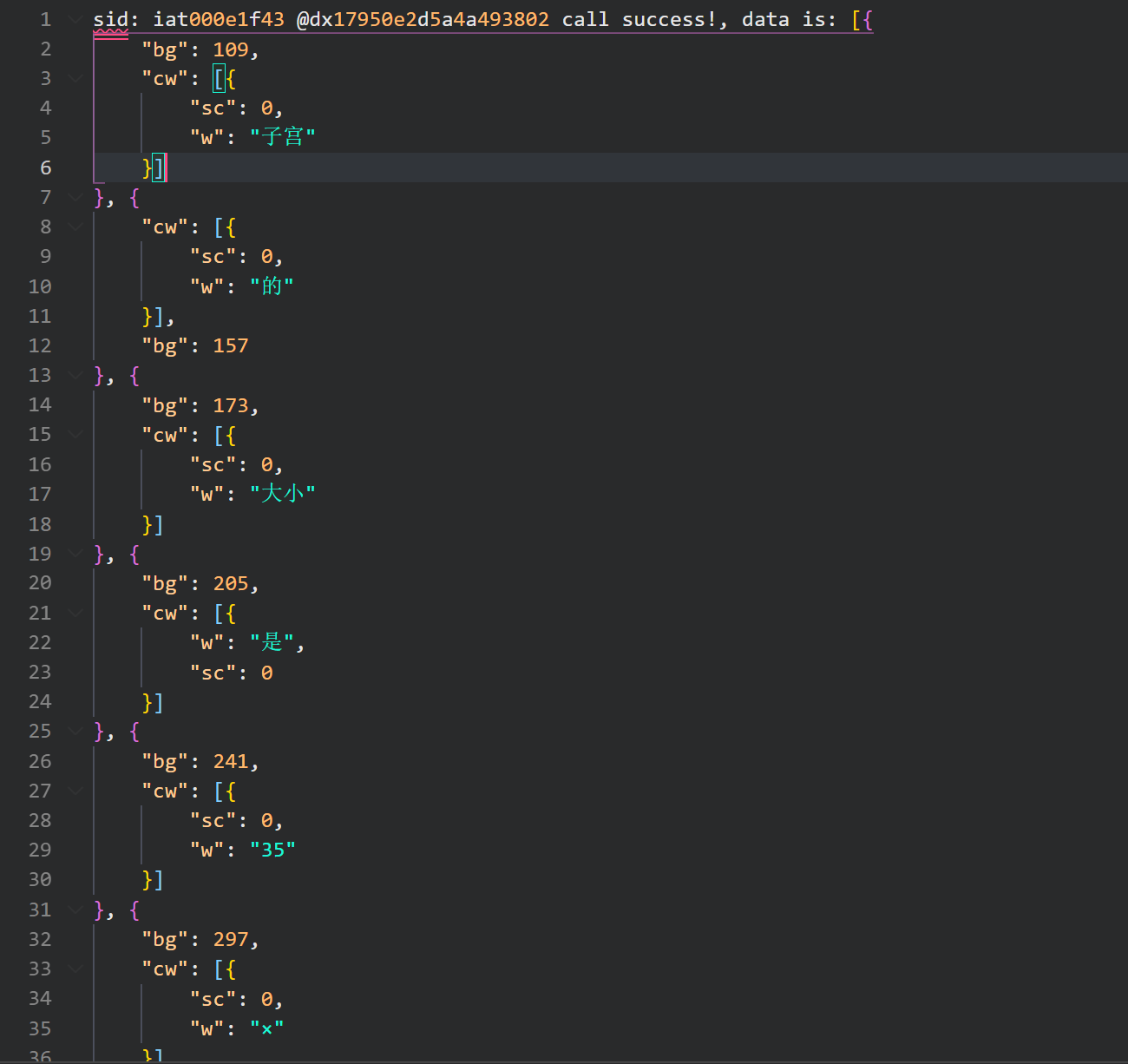
## 调研内容

1. 语音听写概要：语音听写 可以将60秒内的短语音转换成对应的文字信息，让机器能够“听懂”人类语言。适合 输入法、社交聊天、人机交互等60秒内即时短语音的场景；支持实时返回识别结果，达到一边上传音频一边获得识别文本的效果。
2. 适用范围  
   普通话识别率极高，经测试提前录入热词（常用词）后识别率可以达到满意水平。支持广东话的识别，同时不进行额外收费。  
   
3. 现有产品服务：
   1. 使用可能性较大的是动态修正功能。  
      
   2. 提供个性化热词功能  
      经测试，添加热词前后识别率有所提高  
      未添加识别结果：开心好  
      添加热词后识别结果：胎心好  
      
4. 调用方式：  
   提供WEB流式开发接口、WINDOWS端、LINUX端、Android端SDK包。  
   WEB调用时使用APPID、APISecret、APIKey进行调用。  
   其他方式只使用APPID即可调用。  
   
5. 计价 计算方式次数计价，套餐付费  
   
6. 其他服务内容  
   语音助手  
   语音合成  
   语音唤醒  
   硬件配套服务  
   

## 其他服务提供商概要

1. 共性  
   都支持多平台调用方式同时对编程语言没有过多限制。  
   
2. 区别：计价方式有所不同  
   总结：  
   讯飞按次使用，套餐形式购买，一次性购买一定的服务量。  
   百度与讯飞类似，套餐价格稍低，同时支持用多少扣多少，不使用套餐方式购买。  
   腾讯与讯飞类似，价格稍低。  
   华为，价格最高。  
   **讯飞市场占有率较高，价格偏高。百度收费灵活，腾讯价格最低，华为价格最高。**
   1. 讯飞  
      
   2. 百度  
      
   3. 腾讯  
      
   4. 华为  
      

## DEMO实现

1. 环境配置  
   本机环境：WINDOWS10  
   编译环境：Python3.6.6  
   开发环境：VScode
2. 外部工具
   1. FFmpeg  
      FFmpeg官方网址——下载安装包  
      https://www.gyan.dev/ffmpeg/builds/  
      
   2. 解压后配置环境变量  
        
      将解压后的bin目录复制到环境变量中  
        
      
   3. 在CMD中执行“ffmpeg –version”出现下图即成功。  
      
3. 文件格式转换  
   使用FFmpeg进行转换  
   **下面以MP3文件为例**  
   首先将MP3格式转换为wav文件  
   ffmpeg -i input.mp3 -acodec pcm\_s16le -ac 1 -ar 8000 output.wav  
   再将wav文件转换为pcm文件  
   ffmpeg -i output.wav -f s16le -ar 16000 -acodec pcm\_s16le output.pcm
4. 配置APPID、APISecret、APIKey  
   
5. 修改文件地址  
   
6. 输出json字串  
   
7. 格式化文档  
   
8. 进行字串的提取

## 调研总结

本次调研以科大讯飞为调研主要对象，了解的语音识别行业的大体情况。在经费充足的情况下推荐使用讯飞进行开发，因为市场占有率高同时有开发的DEMO。

同时必须要提到的是，语音识别在市场上基本上形成了行业标准，所以其他的供应商也会提供相应服务，差距不会很大。

在硬件方面讯飞提供了接口——linux端C语言开发等，本次使用 web流式进行开发主要是因为操作简单。