

科大讯飞语音转写测试记录表

📅 DATE	@Feb 19, 2020
🏠 Property	
📄 issue	科大讯飞语音转写api
📌 类型	语音识别

科大讯飞语音转写官方文档地址：
<https://www.xfyun.cn/doc/asr/lfasr/API.html#接口说明>

语音转写介绍

语音转写

语音转写（Long Form ASR）基于深度全序列卷积神经网络，将长段音频（5小时以内）数据转换成文本数据，为信息处理和数据挖掘提供基础

免费试用

服务管理

个人用户

语音转写针对是有编程基础的开发者用户。如果您是个人用户，不想通过编程方式直接实现语音转写功能，可以去[讯飞听见官网](#)



{ 技术规格 }



输入

形式：已录制音频

声道：单声道&多声道

支持格式：
wav,flac,opus,m4a,mp3

支持语种：中文普通话、英语

采样率：8KHz,16KHz

采样精度：8bits,16bits



输出

形式：json格式字符串

编码：UTF-8

结果：分词形式&完整句子形式，词&句置信度
词&句时间戳，词属性，多候选词，文法格式智能转换，多发言人分离

- 注：
- 1、文法格式智能转换包括电话号码、车牌、日期、时间以及量词的规整，如“五点三十”引擎会识别规整为“5:30”。
 - 2、词属性包括：普通词；语气犹豫词，如“嗯”、“啊”、“呃”或者无意义的重复词等；标点符号，如“，”等。
 - 3、语音转写的识别文字结果为异步获取，最长12小时之内获取。
 - 4、语音转写引擎是适用于5小时以内的长音频识别，若您有大量1分钟以内的短音频需要识别，为了提高识别效率，建议使用[语音听写](#)服务。
 - 5、使用本服务前请参阅[实践指南](#)，以获得更高的转写准确率。

注意：这里使用的是音频文件，如果是视频的话，可能要把音频拆分出来单独处理，上传的时候会把音频切片拆分编码，返回json格式数据。

应用场景



将坐席通话转成文字，帮助电话质检和信息同步，也为数据挖掘提供原料基础



将会议和访谈的音频转换成文字存稿，让后期的信息检索和整理更方便快捷



将视频中音频文件进行语音转写，
轻松生成与视频相对应的字幕文件



可以从转写出的文字结果中搜索匹配相关词类，对黄暴/涉政内容进行高效鉴别

产品价格

{ 产品价格 }

套餐	体验包	20小时套餐	200小时套餐	500小时套餐	1000小时套餐
价格	免费	¥198.00	¥1180.00	¥2700.00	¥4900.00
时长量	5小时	20小时	200小时	500小时	1000小时
有效期	30天	一年	一年	一年	一年
单小时成本	—	¥9.90	¥5.90	¥5.40	¥4.90
使用产品	立即领取	立即购买	立即购买	立即购买	立即购买

温馨提示:

- 1、平台免费赠送的时长为5小时，供测试使用，每个账户限领取1次；
- 2、后台计量系统根据您的实际使用情况，以秒为单位向上取整，进行使用量扣除；
- 3、订单和套餐使用情况可在“控制台”查看；
- 4、本产品一经购买，不支持退货；
- 5、扣量优先级：批量购买+免费试用，即在“免费试用”的套餐额度剩余的情况下，又购买了“批量购买”套餐，则批量购买的套餐立即生效，并被设定为当前扣量套餐。而之前生成的“免费试用”的额度到期日不变。

测试过程

测试材料：官方demo+idea+易中天品三国14天生奇才.mp3（38:33）

APPID

5e4cb2a5

SecretKey

fb3e810f1cbfb391594241d6ddfd1c6e

测试结果（部分log）：

[illegible]

```
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务处理中：{"status":3,"desc":"音频转写中"}
sleep a while Zzz
任务完成！

转写结果： [{"bg": "0", "ed": "2360", "onebest": "嗯。", "speaker": "0"},
{"bg": "2360", "ed": "8980", "onebest": "公元207，一直副作用的刘备终于从笼中请出了诸葛亮，", "speaker": "0"},
{"bg": "8980", "ed": "13290", "onebest": "诸葛亮出山帮助刘备取得了用武之地。", "speaker": "0"},
{"bg": "13440", "ed": "17780", "onebest": "刘备不但有了自己的地盘，而且地盘还越做越大，", "speaker": "0"},
{"bg": "18000", "ed": "23170", "onebest": "最后和曹操、孙权形成了三国鼎足而立的局面。", "speaker": "0"}, {"bg": "23170", "ed": "26760", "onebest": "而曹操集
```

测试结果分析

1.返回参数官方文档

转写结果字段说明

字段名	说明
bg	句子相对于本音频的起始时间，单位为ms
ed	句子相对于本音频的终止时间，单位为ms
onebest	句子内容
speaker	说话人编号，从1开始，未开启说话人分离时speaker都为0
si	句子标识，相同si表示同一句话，从0开始 注：仅开启分词或者多候选时返回
wordsResultList	分词列表 注：仅开启分词或者多候选时返回
alternativeList	多候选列表，按置信度排名 注：仅开启分词或者多候选时返回
wordBg	词相对于本句子的起始帧，其中一帧是10ms 注：仅开启分词或者多候选时返回
wordEd	词相对于本句子的终止帧，其中一帧是10ms 注：仅开启分词或者多候选时返回
wordsName	词内容 注：仅开启分词或者多候选时返回
wc	句子置信度，范围为[0,1] 注：仅开启分词或者多候选时返回
wp	词属性，n代表普通词，r代表人名，d代表数字，m代表量词，s代表顺滑词（语气词），t代表地名&多音字，p代表标点，g代表分段标识 注：仅开启分词或者多候选时返回

2.分析

{"bg":"2360","ed":"8980","onebest":"公元207，一直副作用的刘备终于从笼中请出了诸葛亮，","speaker":"0"}

结论：

- 1.可以得到内容的开始时间结束时间，在这里2360表示第2秒，有时间数据就可以得到时间轴生成字幕。
- 2.这里调用的是本地的音频文件，不是在线实时转换的，从log可以看出，当文件内容全部转换成功后才会返回内容。
- 3.每次处理的音频时长在5小时以内，预处理文件大小限制在500m内，接口访问频率受限（默认1秒内不得超过20次）。