

科大讯飞语音api

以下是科大所有语音转文字api,之所以还调研科大讯飞,是因为科大讯飞的准确率是最高的最稳定的,他有自动纠正,还对音频的背景音乐和人声做了分离处理,即使在有背景声音情况下准确率还是很高,如果不是url 限制,我还是推荐用科大讯飞,成熟稳定准确率高调用方便! 就算这个项目没有用到,以后有语音识别的项目也可以当作参考。在上面三个版本api测试结果中可以体现。

科大讯飞

1.语音听写

(之前用过jsdemo,得用谷歌浏览器,ie不兼容。读取电脑麦克风,准确度很高,如果是数字也可以识别,但是有限制,如果五秒还是十秒不说话会停止识别(可以修改限制参数),有一分钟限制(无法改变,只能多次调用),websocket通讯方式)

```
public class WebIATWS extends WebSocketListener {  
    private static final String hostUrl = "https://iat-api.xfyun.cn/v2/iat"; //中英文, http url 不支持解析 ws/wss schema  
    // private static final String hostUrl = "https://iat-niche-api.xfyun.cn/v2/iat"; //小语种  
    private static final String apiKey = "xxxxxxx"; //在控制台-我的应用-语音听写(流式版)获取  
    private static final String apiSecret = "xxxxxxx"; //在控制台-我的应用-语音听写(流式版)获取  
    private static final String appid = "xxxxxxx"; //在控制台-我的应用获取  
    private static final String file = "resource\\iat\\16k_10.pcm"; // 中文
```

输入格式: 本地pcm音频文件

音频时长: 不超过一分钟

文档地址: <https://www.xfyun.cn/doc/asr/voicedictation/API.html>

2.语音转写

(已经测试过,切片上传本地文件,科大服务端进行重组。必须用文件方式,提过工单确认了)

输入格式: 本地音频文件

格式支持: wav/flac/opus/m4a/mp3

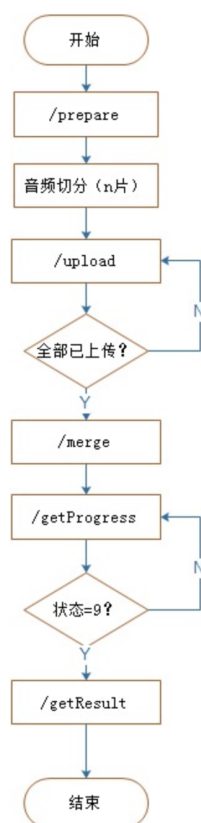
音频大小: 不超过500M

音频时长: 不超过5小时,建议5分钟以上

文档地址: <https://www.xfyun.cn/doc/asr/lfasr/API.html#接口说明>

接口调用流程: 转写 API 包括以下接口: 预处理、文件分片上传、合并文件、查询处理进度、获取结果。

- 预处理 /prepare:
- 文件分片上传 /upload:
- 合并文件 /merge:
- 查询处理进度 /getProgress:
- 获取结果 /getResult:



对于歌曲识别准确度不高，但是比腾讯的要好很多，因为伴奏也会对识别效果有影响。所以测试文件我选择易中天品三国，比较接近真实场景。

3.实时语音转写

(必须用文件方式，提过工单确认了)

支持的音频格式：采样率为16K，采样深度为16bits的pcm_s16le音频（本地，这个没有时间限制）

文档地址：<https://www.xfyun.cn/doc/asr/rtasr/API.html>

要求:调用的时候必须设置白名单IP，否则会报错，返回10105错误码

<http://rtasr.xfyun.cn/v1/ws>

2020-02-26 12:04:20.448 连接中

2020-02-26 12:04:20.736 连接建立成功!

2020-02-26 12:04:20.825 握手成功! sid: rta000b1ae1@dx4ae711b13a44a12100

2020-02-26 12:04:21.453 开始发送音频数据

2020-02-26 12:04:23.666 result: 床

2020-02-26 12:04:23.781 result: 床前

2020-02-26 12:04:23.912 result: 床前明月

2020-02-26 12:04:24.038 result: 床前明月光疑

2020-02-26 12:04:24.562 result: 床前明月光! 疑是地上

2020-02-26 12:04:26.046 result: 床前明月光, 疑是地上霜

2020-02-26 12:04:26.183 result: 床前明月光! 疑是地上霜举头

2020-02-26 12:04:26.213 result: 床前明月光! 疑是地上霜! 举头到

2020-02-26 12:04:26.608 result: 床前明月光, 疑是地上霜! 举头望明月, 低头

2020-02-26 12:04:26.663 发送结束标识完成

2020-02-26 12:04:26.851 result: 床前明月光, 疑是地上霜, 举头, 望明月, 低头思故乡。

2020-02-26 12:04:27.866 链接关闭

科大工单回复:

待确认

工单号: 2020022519094 | 提交时间: 2020-02-25 15:11:40



客户服务专员

实时转写也不支持, 这还真的非常抱歉, 后期我们会极力优化的, 您的需求就是我们的方向, 非常感谢您对贵公司的产品大力支持!

2020-02-25 17:22:45



您

因为我们这个可能需求量挺大的, 我们还是优先考虑讯飞平台

2020-02-25 17:18:53



您

那你们的语音实时转写 可以支持这样吗?

2020-02-25 17:18:26



客户服务专员

您好, 非常抱歉, 非常理解您的需求, 现在的我们的场景都是从本地获取音频来转写, 还不支持url模式来获取音频, 因为引擎端暂不支持直接获取url里面的数据流或者内容, 中和的方式就是在用户的服务端可以把url对应的音频下载下来, 然后上传到我们引擎端, 暂且我们把您的需求列为我们的产品需求, 后期在转写优化上优化以及实现该需求。感谢您的支持和理解。

2020-02-25 15:38:04



您

语音转写api, 有没有可能改成url的方式传入?

我们准备做一个教育平台的网页应用, 用户上传音频或者视频文件, 然后我们进行语音转换, 保存结果, 我测试过科大讯飞语音转写的api, 对本地文件可用, 不知道能不能把传入的参数改成url模式, 我们这个需求量还很大的, 希望有技术支持! 谢谢!

2020-02-25 15:11:40

上面内容都是基于官方demo给出的结果, 官方demo里都是本地的pcm音频文件, 之前没有做过这方面, 不知道是否有其他方式可以传入实时音频流, 像语音听写 (js模块, 直接传入电脑麦克风数据), 我测试的都是java版本的demo。