

捷通华声 灵云公有云能力平台 HTTP-ASR 接口 开发手册

灵云科技 源自清华 服务全球

北京捷通华声语音技术有限公司

Beijing InfoQuick SinoVoice Speech Technology Corp.

2014 年 6 月

目 录

1. 概述.....	1
2. HTTP 头.....	1
3. ASR 访问接口	2
3.1. ASR 识别请求接口	2
3.2. 访问	2
3.3. 服务器应答	2
3.4. 示例	2
3.5. x-task-config 参数说明.....	3

1. 概述

灵云HTTP-ASR接口，客户端通过HTTP-ASR接口使用POST方式与灵云平台进行通讯，返回数据为XML格式。灵云的HTTP-ASR接口支持实时识别方式，也就是在录音的同时可以将数据分为多次同步上传，这样后台就可以实时同步地进行识别过程，以提高响应时间。

2. HTTP 头

客户端POST请求时HTTP头信息，必须带上如下：

顺序	名称	值（例子）	含义	备注
1	x-app-key	ac5d5431	应用标识	必选，由捷通公司提供
2	x-sdk-version	3.6	sdk 版本号	必选，使用 http 接口时，可以指定为 3.6
3	x-request-date	2014-6-18 10:10:11	请求时间	必选，每个任务的请求时间不能重复
4	x-task-config	capkey=asr.cloud.freetalk, audioformat=alaw	任务参数信息	为 name=value 形式，多个参数以逗号隔开 capkey 为必选 splitsentence 为可选，默认为 yes
5	x-session-key	18EA9384AA34	请求数据签名	必选 x-session-key 生成算法说明： x-session-key =md5（x-request-date + devkey）
6	x-udid	101:1234567890	设备标识	必选，如使用设备取设备标识号，如不使用设备设置为例子中的默认值
7	x-tid		保留字段	保留字段，不用设置
8	x-eid		保留字段	保留字段，不用设置

注：头的内容必须为英文。

devkey是应用开发者密钥，在注册开发者用户时由开发者网站统一分配，用作请求数据签名；

3. ASR 访问接口

3.1. ASR 识别请求接口

本接口在提交识别请求时，同步在 http 请求的回应消息中返回识别状态或识别结果。

3.2. 访问

(POST) [ServiceUrl]/asr/Recognise

说明：

包体中传输语音数据

需要以下 HTTP 头：x-app-key 、 x-request-date、 x-task-config、 x-auth、 x-sdk-version

3.3. 服务器应答

应答格式：XML 文件

XML 的元素（Element）及含义

顺序	名称	值（例子）	含义	备注
1	ResCode	Success (IN_PROGRESS,FAILED)	成功（处理中，失败）	必选
2	ResMessage	成功（处理中，失败）	成功（处理中，失败）	必选
3	Result_Token	0E01123456	结果令牌	成功时必选
4	ResultCount	2	结果个数	可选
5	Result	<pre><Result> < ResultText>你好</ResultText> <Score>90</Score> </Result></pre>	识别结果，其中 ResultText 是结果文本； Score 为结果得分，最大 100，最小为 0；	成功时必选
6	ErrorNo	8	失败时错误码	失败时必选

3.4. 示例

识别完成返回结果示例：

```
<?xml version="1.0"?>
<ResponseInfo>
  <ResCode>Success</ResCode>
  <ResMessage>成功</ResMessage>
  <Result_Token>0E01123456</Result_Token>
  <ResultCount>2</ResultCount>
  <Result>
    <ResultText>012345678</ResultText>
    <Score>90</Score>
  </Result>
  <Result>
    <ResultText>012345679</ResultText>
    <Score>80</Score>
  </Result>
</ResponseInfo>
```

实时识别中间返回结果示例:

```
<?xml version="1.0"?>
<ResponseInfo>
  <ResCode>InProgress</ResCode>
  <ResMessage>expect more chunk</ResMessage>
  <Result_Token>1_8_20_21608_20131118192712_0</Result_Token>
</ResponseInfo>
```

识别失败返回结果示例:

```
<?xml version="1.0"?>
<ResponseInfo>
  <ResCode>FAILED</ResCode>
  <ResMessage>CheckSign failed</ResMessage>
  <ErrorNo>8</ErrorNo>
  <Result_Token>1_8_20_21608_20131118192712_0</Result_Token>
</ResponseInfo>
```

3.5. x-task-config 参数说明

使用 ASR 进行语音数据识别时需要指定使用的引擎，目前支持引擎有：

✧ capkey（语音合成能力名称，必选）

Capkey	备注
asr.cloud.freetalk	基于自由说引擎的语法识别，通用领域。

asr.cloud.freetalk.poi	基于自由说引擎的语法识别，poi 导航领域。
asr.cloud.freetalk.music	基于自由说引擎的语法识别，音乐领域。

✧ **audioformat**（语音格式），目前支持以下格式

pcm16k16bit	16k 16bit 录音数据（默认）
ulaw16k8bit	16k 8bit u-law 数据
alaw16k8bit	16k 8bit a-law 数据
speex	speex 压缩格式

✧ **addpunc**（是否需要为识别结果添加标点）：

yes	需要
no	不需要（默认）

✧ **domain**（指定识别领域，目前仅在mode=freetalk下支持）：

common	通用识别（默认）
poi	导航
music	音乐

注：只有当 capkey 为 asr.cloud.freetalk 时，domain 参数才有效。

✧ **identify**(会话标识)：

该字段为会话的唯一标识，该字段将和 x-udid 字段一起作为一次会话的唯一标识。在一个设备上每次也需要不一样。

✧ **index**(识别序号)：

某次会话的请求序号，从 1 开始，最后一次的序号以负数作为终结。如可能的序列号为：

- -1 此时请求数据只有一块
- 1, 2, -3 此时请求数据有三块，此时只有最后一块会返回识别结果，其余均返回 ResCode 为 InProgress（见实时识别中间块返回结果示例）

其他可选字段：

✧ **lon**（经度，目前保留，将来为导航识别时使用）

✧ **lat**（纬度，目前保留，将来为导航识别时使用）

✧ **areacode**（城市编码，目前保留，将来为导航识别时使用）