

# 接口说明

语音识别（ASR）可以将语音信息转换为文字信息。该能力通过HTTP API的方式提供给开发者，相较于SDK，该方式具有轻量、跨平台、跨开发语言的特点。

## 接口要求

| 项目   | 说明  |
|------|---|
| 请求地址 | <a href="http://api.baller-tech.com/v1/service/v1/asr">http://api.baller-tech.com/v1/service/v1/asr</a> |
| 请求方式 | 发送语音数据时使用POST；获取识别结果时使用GET  |
| 字符编码 | UTF-8   |

## 接口签名

为了防止通信过程中发送的消息被他人窃取和修改，每一个HTTP协议接口都需要进行签名验证，服务器发现请求的签名不一致时会拒绝处理。

将**app\_key**（由北京大牛儿科技发展有限公司统一分配）、**请求时间**（GMT格式）、**base64编码后的业务参数**按照固定的顺序组成的字符串MD5后的结果作为签名，放到请求报文的Header的B-Checksum参数中。

## 音频数据发送模式

向服务器发发送音频数据时，可以一次性的将音频数据发送到的服务器,然后连续的获取识别结果，也可以将音频数据分多次发送到服务器，每发送一次音频数据，获取一次识别结果，当音频数据发送完毕后，再连续的获取识别结果，不论使用那种模式向服务器发送音频数据，识别结果的获取模式和方法是一样的，服务器会分多次将实时的识别结果返回。两种模式的适用场景如下：

1. 已经录制好的音频且时长较短（一般60秒内）时，即可以一次性将音频数据发送到服务器，也可以分多次发送到服务器；
2. 已经录制好的音频且时长较长时，分多次将音频数据发送到服务器；
3. 音频数据实时录取，实时识别时，分多次将音频数据发送到服务器。

## 识别结果获取模式

根据识别结果获取的方式不同，分为两种调用模式：

1. 连续调用HTTP的GET方法获取识别结果，适用于直接在终端客户的设备上发起请求时。
2. 将识别结果推送到请求时指定的HTTP 地址上，适用于在对接方公司服务器上发起请求时（终端客户与对接方公司服务器通信，对接方服务器调用本请求）。

## 连续调用HTTP的GET方法获取识别结果

1. 通过HTTP协议POST方法，将音频数据发送到服务器。
2. 通过HTTP协议GET方法，去服务器获取识别结果以及是否获取结束的状态；
3. 如果HTTP协议GET方法的响应中是否获取结束的状态为未结束，需要继续调用HTTP协议GET方法请求识别结果；为了避免频繁的交互浪费CPU和网络资源，两次HTTP协议GET方法的请求之间可

以间隔一段时间（具体值可以根据使用场景进行测试确定，建议150~200毫秒）。

## 将识别结果推送到请求时指定的HTTP 地址上

1. 通过HTTP协议POST方法，将音频数据发送到服务器，发送数据时携带结果推送的地址。
2. 服务器通过HTTP协议的POST方法，分多次将识别的结果推送到请求时指定的地址。

## 接口参数

### 1. POST方法请求参数

#### 1.1 HTTP请求Header中需设置参数

| 参数           | 类型     | 说明  | 举例                               |
|--------------|--------|---|----------------------------------|
| B-AppId      | string | 由北京大牛儿科技发展有限公司统一分配；<br>分配的值为64位的整型，此处需要转换为string  | 1176611429127553031              |
| B-CurTime    | string | GMT+0时区的符合RFC1123格式的日期和时间，星期和月份只能使用英文表示；<br>需和接口签名时的请求时间一致；<br>服务器会拒绝处理请求时间与当前时间相差300秒的请求 | Fri, 10 Jan 2020<br>07:31:50 GMT |
| B-Param      | string | 经过BASE64编码后的业务参数，参见 <a href="#">业务参数</a>  |                                  |
| B-Checksum   | string | 参见 <a href="#">接口签名</a> 。   |                                  |
| Content-Type | string | 传输数据的类型，此处使用固定值   | application/octet-stream         |

##### 1.1.1 业务参数介绍

| 参数                 | 类型     | 说明   | 举例  |
|--------------------|--------|--|---|
| request_id         | string | 本次语音识别事务的请求ID；<br>获取该请求识别结果时需携带相同的请求ID；<br>调用者需保证请求ID的唯一性，建议使用UUID     | 6497c282-9371-4c68-a9f1-522212b5ac1d  |
| audio_format       | string | 音频格式，参见 <a href="#">支持的语种和音频格式</a>                                     | audio/L16;rate=16000  |
| language           | string | 语种，参见 <a href="#">支持的语种和音频格式</a>                                       | zho   |
| input_mode         | string | 音频数据的发送模式，支持以下字段：<br>once<br>continue<br>end                           | once  |
| service_type       | string | 服务类型，支持以下字段：<br>sentence: 句子识别（默认值，任务有时长限制）<br>realtime: 实时识别（任务无时长限制） | sentence  |
| dynamic_correction | string | 是否启用动态纠正：<br>on：启用（暂不支持）<br>off: 不启用（默认值）                              | off   |
| vad                | string | 是否启用端点检测，支持以下字段：<br>on: 启用（默认值）<br>off: 不启用                            | on  |
| callback_url       | string | 识别结果推送的回调地址；<br>通过调用HTTP的GET方法获取识别结果时不需设置                              | <a href="http://192.168.1.234:18888/ocr/callback">http://192.168.1.234:18888/ocr/callback</a> |

#### 1.1.1.1 audio\_format 介绍

根据RFC对MIME格式的定义，使用audio/Lxx;rate=xxxxx 表明音频格式，audio/L后面的数字表示音频的采样点大小（单位bit），rate=后面的数字表示音频的采样率（单位hz）。

比如audio/L16;rate=16000表示音频数据为16000hz，16bit的pcm音频数据

#### 1.1.1.2 input\_mode 介绍

一次性将音频数据发送到服务器时，input\_mode应设置为once。当分多次将音频数据发送到服务器时，如果不是本次识别事务的最后一次，input\_mode应设置为continue；如果是本次识别事务的最后一次应设置为end。

### 1.2 HTTP请求Body

待识别的音频数据（二进制）。

## 1.3 响应报文

http响应数据为json格式，具体字段的含义如下

| 参数         | 类型     | 说明                       |
|------------|--------|--------------------------|
| code       | int    | 请求处理的结果码 (0: 成功; 其他: 失败) |
| message    | string | 对code字段的文本说明             |
| request_id | string | 请求时传入的request_id         |

```
{
  "code": 0,
  "message": "success",
  "request_id": "f7409982-dc05-4d19-80c9-6169dd70b247"
}
```

## 2. GET方法请求参数

### 2.1 HTTP请求Header中需设置参数

| 参数         | 类型     | 说明  | 举例                               |
|------------|--------|---|----------------------------------|
| B-AppId    | string | 由北京大牛儿科技发展有限公司统一分配;<br>分配的值为64位的整型，此处需要转换为string  | 1176611429127553031              |
| B-CurTime  | string | GMT+0时区的符合RFC1123格式的日期和时间，星期和月份只能使用英文表示;<br>需和接口签名时的请求时间一致;<br>服务器会拒绝处理请求时间与当前时间相差300秒的请求 | Fri, 10 Jan 2020<br>07:31:50 GMT |
| B-Param    | string | 经过BASE64编码后的业务参数，参见 <a href="#">业务参数</a>  |                                  |
| B-Checksum | string | 参见 <a href="#">接口签名</a> 。   |                                  |

#### 2.1.1 业务参数介绍

| 参数         | 类型     | 说明                            | 举例                                       |
|------------|--------|-------------------------------|--|
| request_id | string | 本次语音识别事务的请求ID;<br>需与POST时保持一致 | 6497c282-9371-4c68-a9f1-<br>522212b5ac1d |

## 2.2 响应报文

语音识别时，会将传入的音频分为不同的子句，每次GET请求返回的是一个子句的结果。子句的识别结果分为最终结果和非最终结果两种状态；最终状态表示结果为当前子句的最终结果，之后再获取到的结果为新子句的结果；非最终状态表示结果为当前子句的中间状态的结果，之后再获取到的结果还是该子句的识别结果。

一般我们只需关注最终状态的识别结果即可，如果需要更快速的让用户的看到部分识别结果，并动态的调整用户看到的识别结果时，才需要考虑非最终状态的结果。

响应数据为json格式，具体字段的含义如下

| 参数          | 类型     | 说明                             |
|-------------|--------|--------------------------------|
| code        | int    | 请求处理的结果码 (0: 成功; 其他: 失败)       |
| message     | string | 对code字段的文本说明                   |
| request_id  | string | 请求时传入的request_id               |
| is_end      | int    | 识别结果是否获取结束 (0: 未结束; 1: 结束)     |
| data        | string | 一个子句的识别结果                      |
| is_complete | int    | 子句结果是否是最终的 (0: 非最终结果; 1: 最终结果) |

```
{
  "code": 0,
  "message": "success",
  "is_end": 1,
  "request_id": "3488a4fa-5f7d-11ea-b739-4023431f608e",
  "data": "xxx",
  "is_complete": 1
}
```

### 3. 推送识别结果的消息格式

采用服务器推送识别结果时，推送的消息格式与GET请求的响应报文格式基本一致。不一样的地方是会在code参数同级添加一个order参数，表示本次事务推送的次序，从0开始依次递增。

```
{
  "code": 0,
  "message": "success",
  "is_end": 1,
  "request_id": "3488a4fa-5f7d-11ea-b739-4023431f608e",
  "data": "xxx",
  "is_complete": 1,
  "order": 0,
}
```

## 支持的语种以及音频格式

| 语种      | 对应的language<br>字段 | 支持的音频格式                    | 对应的audio_format      |
|---------|-------------------|----------------------------|----------------------|
| 哈语      | kaz               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 蒙语      | mon               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 藏语 (安多) | tib_ad            | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 藏语 (康巴) | tib_kb            | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 藏语 (卫藏) | tib_wz            | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 维语      | uig               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 汉语      | zho               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 壮语      | zha               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 彝语      | iii               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |
| 韩语      | kor               | 采样率: 16000hz 采样点大小: 16bits | audio/L16;rate=16000 |