

# 在线语音识别接口说明

---

在线语音识别接口可以将声音信息转换为文本信息，通过HTTP API的方式给开发者提供一个通用的接口，相对于SDK，API具有轻量、跨语言的特点。

## 识别结果获取方式

---

在线语音识别有两种方式可以获取到识别结果：

- 调用者主动查询，调用者将音频数据发送给服务器后，需要定时的去服务器查询识别结果。
- 服务器主动推送，调用者发送数据给服务器时，需携带一个回调的http地址，当服务器有识别结果后，会把识别结果POST到回调地址，同一次请求事务有可能会POST多次，需根据回调中携带的信息判断是否结束。

**对于同一次的请求事务，只能使用其中的一种方式获取识别结果。**

## 接口概述

---

在线语音识别接口分为put、get两个HTTP API。开发者通过put API将声音信息传递给服务器。如果采用主动查询的方式获取识别结果时，需要定时的调用get接口获取数据，如果采用服务器主动推送的获取方式时，不需要调用get接口。

## 音频数据传输模式

---

put API支持once和continue两种音频数据的传输模式，前者支持将音频数据采集完成后一次性发送给服务器，后者支持将采集到的音频数据实时的发送给服务器。

不论是使用once还是continue，识别结果的返回都是分多次的。once时虽然只会向服务器发送一次音频数据，但识别结果的返回并不是一次，而是和continue一样，分多次返回。

## 需注意的地方

---

每一个HTTP API请求时，需要传递B-CurTime，值为当前的时间戳，服务器会检查该时间戳，并拒绝处理一个5分钟前发起的请求。

## put 接口要求

---

### 功能说明

---

将声音信息传递给服务器，每次请求事务开始时需要产生一个全局唯一的request\_id，用来标识本次请求事务。

使用once模式（既将采集好的音频数据一次性的传递给服务器），需设置业务参数的input\_mode为once；使用continue模式时（既边采集边将数据传递给服务器），如果不是最后一部分的音频数据，input\_mode需设置为continue；如果是最后一段音频数据，input\_mode需设置为end。

## 调用地址

```
[POST] http://host:port/v1/service/private/v1/asr
```

## HTTP 请求参数

### HTTP HEADER参数

关键字	说明
B-CurTime	当前时间戳
B-Param	相关业务参数BASE64编码后的字符串
Content-Type	application/octet-stream

### 业务参数

业务参数为json格式。

关键字	类型	说明
request_id	string	请求的唯一标示（后面会有详细介绍）
language	string	文本对应的语言（后面会有详细介绍）
audio_format	string	语音格式（后面会有详细介绍）
callback_url	string	识别结果回调的地址（后面会有详细介绍）
input_mode	string	支持"once"、"continue"、"end"

#### request\_id 说明

request\_id用来标示一次语音识别的事务，由开发者产生，全局唯一。建议使用uuid。

#### audio\_format 说明

audio\_format表示音频格式，根据RFC对MIME格式的定义，使用audio/L16;rate=16000来表明音频格式，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）。

调用put接口时，如果没有设置callback\_url或callback\_url的值为空，则认为调用者使用主动查询的方式获取识别结果；如果callback\_url不是http的请求地址，会返回对应的错误码，本次请求失败；其他情况服务器认为调用者使用了回调的方法获取识别结果。同一次语音识别任务多次put时，callback\_url的值不能改变。

目前回调的发送只支持POST的方式，每一次回调会设置5秒的超时，如果回调失败，会尝试重新发送3次。连续发送3次失败，就会停止发送，出现这种情况会导致该次请求事务收到的结果时不完整的。

回调的识别结果位于响应报文的主体内，是json格式，具体参数如下。

关键字	类型	说明
code	int	请求处理的结果码
request_id	string	请求的唯一标示（与业务参数中的request_id相同）
is_end	int	0-未结束；1-结束
data	string	本次请求事务的完整的识别结果，因为会对之前的结果进行调整，所以在拿到新的识别结果的时候需要对以前已经上屏显示的结果进行清除，然后将新获取的识别结果重新上屏显示（utf-8编码）

相应报文的示例

```
{
  "code": 0,
  "is_end": 0,
  "request_id": "cf6b746e-64f0-4319-9bd8-c0e2ad48053e",
  "data": "xxx"
}
```

调用者收到回调后，在回复的响应报文中，状态码200表示处理成功，其他状态码标示处理失败。服务器收到调用者回复的非200的错误码时，只会记录该信息，不会对相同的数据进行重新发

送。

## HTTP BODY参数

需进行识别的音频数据

## HTTP 响应说明

http响应数据为json格式，具体字段的含义如下

关键字	类型	说明
code	int	请求处理的结果码 【必有】
request_id	string	请求的唯一标示（与业务参数中的request_id相同）【code为0时必有，非0时不一定有】

## 相应报文示例

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

## get 接口要求

### 功能说明

从服务器获取识别的结果（识别结果的编码为utf-8编码），业务参数中的request\_id需和put 接口中的request\_id一致。

请求put接口和请求get接口可以成对出现，当没有put请求且识别结果没有获取完成时，可连续的调用get接口，此时两次get请求可以间隔150~200毫秒。

get请求每次获取的结果都是本次语音识别事务的完整结果。因为会对之前的结果进行调整，所以在拿到新的识别结果的时候需要对以前已经上屏显示的结果进行清除，然后将新获取的识别结果重新上屏显示。

## 调用地址

```
[GET] http://host:port/v1/service/private/v1/asr
```

## HTTP 请求参数

### HTTP HEADER参数

关键字	说明
B-CurTime	当前时间戳
B-Param	相关业务参数BASE64编码后的字符串

### 业务参数

业务参数为json格式。

关键字	类型	说明
request_id	string	请求的唯一标示（后面会有详细介绍）

#### request\_id 说明

request\_id用来标示一次语音识别的事务，get 接口中的request\_id需和put接口中的request\_id保持一致。

## HTTP 响应说明

识别结果位于响应报文的主体内，是json格式，具体参数如下。

关键字	类型	说明
code	int	请求处理的结果码 【必有】
request_id	string	请求的唯一标示（与业务参数中的request_id相同） 【code为0时必有，非0时不一定有】
is_end	int	0-未结束，需继续发送get请求；1-结束，不需再发送get请求 【code为0时必有，非0时不一定有】
data	string	识别的结果（utf-8编码） 【code为0时必有，非0时不一定有】

相应报文的示例

```
{
  "code": 0,
  "is_end": 0,
  "request_id": "cf6b746e-64f0-4319-9bd8-c0e2ad48053e",
  "data": "xxx"
}
```

## 支持的语种以及音频格式

语种	对应的language 字段	支持的音频格式	对应的audio_forma t
英语	eng	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
哈语	kaz	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
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
壮语	zha	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