

# 在线语音合成接口说明

---

在线语音合成接口可以将文字信息转换为声音信息，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。

## 合成结果获取方式

---

在线语音合成有两种方式可以获取到合成的结果：

- 调用者主动查询，调用者将文本数据发送给服务器后，需要定时的去服务器查询合成结果。
- 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，当服务器有合成结果后，会把合成结果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合成结果。**

## 接口概述

---

在线语音合成接口分为put、get两个HTTP API。开发者通过put API将文本信息传递给服务器。如果采用主动查询的方式获取合成结果时，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。

## 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

---

将使用utf-8编码的文本信息传递给服务器，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。每次合成请求文本数据的长度需小于4K。

### 调用地址

---

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

## 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	合成结果回调的地址（后面会有详细介绍）
input_mode	string	仅支持"once"

request\_id 说明

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

audio\_format 说明

audio\_format表示合成的音频格式，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）。

callback\_url 说明

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合成结果；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合成结果。

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。

根据语音合成处理的状态，响应分为两种请情况。合成成功有音频数据时，回调报文的主体内为合成的音频数据，一些相关的状态信息位于回调报文的头部；合成失败时，回调报文的主体内为状

态信息。可以通过回调报文头部中的Content-Type来判断是否成功，具体介绍如下。

- 当Content-Type的值为**audio/mpeg**时，表示有音频数据，音频数据位于回调的主体内，此时的一些状态信息位于回调报文的头部中，字段如下

关键字	类型	说明
B-Code	int	处理的结果码
B-Message	string	描述信息
B-Request-Id	string	请求的唯一标示（与业务参数中的request_id相同）
B-Is-End	int	0-语音合成的结果未回调完毕；1-语音合成的结果已全部回调完毕，

- 当Content-Type的值为非**audio/mpeg**时，表示合成失败，此时的一些状态信息位于回调报文的主体中，是json格式的，具体的参数如下：

关键字	类型	说明
code	int	处理的结果码
message	string	描述信息
request_id	string	请求的唯一标示（与业务参数中的request_id相同）

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

需要进行语音合成的文本信息，文本信息需要使用utf-8编码。

## 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
request_id	string	请求的唯一标示（与业务参数中的request_id相同） ;当认为请求不合法时，没有该字段

## 相应报文示例

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

## get 接口要求

### 功能说明

从服务器获取合成的语音数据，业务参数中的request\_id需和put 接口中的request\_id一致。

put接口请求完成后，需要连续请求get接口，直到所有合成的音频数据获取完成。两次get请求之间可以间隔150~200毫秒，避免浪费CPU。

### 调用地址

[GET] http://host:port/v1/service/private/v1/tts

### 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用来标示一次语音合成的事务，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

根据get请求是否处理成功，响应分为两种请情况。get请求成功时，响应报文的主体内为合成的音频数据，一些相关的状态信息位于响应报文的头部；get请求失败时，响应报文的主体内为状态信息。可以通过响应报文头部中的Content-Type来判断是否成功，具体介绍如下。

### get接口成功时

当Content-Type的值为**audio/mpeg**时，表示请求成功，此时的一些状态信息位于响应报文的头部中，字段如下

关键字	类型	说明
B-Code	int	请求处理的结果码
B-Mess age	str in g	描述信息
B-Requ est-Id	str in g	请求的唯一标示（与业务参数中的request_id相同）
B-Is-En d	int	0-语音合成的结果未获取完毕，需继续发送get请求；1-语音合成的结果已全部获取完毕，不需再次发送get请求

### get接口失败时

当Content-Type的值为非**audio/mpeg**时，表示请求失败，此时的一些状态信息位于响应报文的主题中，是json格式的，具体的参数如下：

关键字	类型	说明
code	int	请求处理的结果码
messag e	strin g	描述信息
request _id	strin g	请求的唯一标示（与业务参数中的request_id相同） ;当认为请求不合法时，没有该字段

## 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_form at
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