

接口说明

语音合成（TTS）可以将文字信息转换为不同语种的声音信息。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。

使用时请求方通过HTTP协议的POST方法将文字信息一次性的发送到服务器，然后通过HTTP协议的GET方法去服务器获取合成结果。与一次性交互的方式（既将文字信息一次性发送到服务器，然后等服务器处理完成之后该请求才携带合成结果返回）相比，发送文字的请求会在服务器收到文字之后就返回，不会阻塞到服务器合成完成，减少调用等待的时间，应用处理起来更灵活。

接口要求

项目	说明
请求地址	http://api.baller-tech.com/v1/service/v1/tts
请求方式	发送文字数据时使用POST；获取合成结果时使用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。

接口调用模式

根据合成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方法获取合成结果，适用于直接在终端客户的设备上发起请求时。
2. 将合成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方法获取合成结果

1. 通过HTTP协议POST方法，将文本数据一次性的发送到服务器。
2. 通过HTTP协议GET方法，去服务器获取合成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方法请求合成结果；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）。

将合成结果推送到请求时指定的HTTP 地址上

1. 通过HTTP协议POST方法，将文字数据一次性的发送到服务器，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，分多次将合成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； 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； 需和接口签名时的请求时间一致； 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 07:31:50 GMT
B-Param	string	经过BASE64编码后的业务参数，参见 业务参数	
B-Checksum	string	参见 接口签名 。	
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	本次语音合成事务的请求ID； 获取该请求合成结果时需携带相同的请求ID； 调用者需保证请求ID的唯一性，建议使用UUID	6497c282-9371-4c68-a9f1-522212b5ac1d
sample_format	string	合成的采样格式，参见 支持的语种和采样格式	audio/L16;rate=16000
language	string	合成音频的语种，参见 支持的语种和采样格式	zho
audio_encode	string	音频编码格式；参见 支持的音频编码	raw
speed	int	音频输出的语速；参见 语速的取值范围	0
callback_url	string	合成结果推送的回调地址； 通过调用HTTP的GET方法获取合成结果时不需设置	http://192.168.1.234:18888/tts/callback

1.1.1.1 sample_format 介绍

根据RFC对MIME格式的定义，使用audio/Lxx;rate=xxxxx 表明采样格式，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表示合成的音频数据为16000hz，16bit的pcm音频数据

1.1.1.2 audio_encode 介绍

语音合成的原始数据是未经过压缩的采样数据，播放器可以直接播放，它的数据量比较大，以audio/L16;rate=16000为例，一秒的音频需要32000字节的数据来表示。如果对带宽比较敏感，希望减少传输的数据量，可以指定编码格式，对原始采样数据进行编码（压缩），编码（压缩）后的数据需解码后才能正常播放。

WebAPI返回的是编码后的裸流，不包含任何的封装信息。接口每次返回一帧或多帧完整的音频数据，不会将一帧音频数据分多次返回。

为了方便解码，当该参数指定为speex或opus时，在每帧数据前会添加4个字节，用来表示当前帧的字节数。

1.2 HTTP请求Body

待合成的文本数据。

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； 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； 需和接口签名时的请求时间一致； 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 07:31:50 GMT
B-Param	string	经过BASE64编码后的业务参数，参见 业务参数	
B-Checksum	string	参见 接口签名 。	

2.1.1 业务参数介绍

参数	类型	说明	举例
request_id	string	本次语音合成事务的请求ID； 需与POST时保持一致	6497c282-9371-4c68-a9f1-522212b5ac1d

2.2 响应报文

合成的语音数据位于响应报文的主体中，一些状态的控制信息位于响应报文的头部中，如下所示

参数	类型	说明
B-Code	string	请求处理的结果码 ("0": 成功; 其他: 失败)
B-Message	string	对B-Code字段的文本说明
B-Request-Id	string	请求时传入的request_id
B-Is-End	string	合成结果是否获取结束 ("1": 结束; "0": 未结束)

3. 推送合成结果的消息格式

采用服务器推送合成结果时，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响应的Header中添加B-Order参数，表示本次事务推送的次序，从0开始依次递增。

支持的语种以及采样格式

语种	对应的language 字段	支持的采样格式	对应的 sample_format
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
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
中文	zho	采样率: 16000hz 采样点大小: 16bits	audio/L16;rate=16000
英文	eng	采样率: 16000hz 采样点大小: 16bits	audio/L16;rate=16000

支持的音频编码

audio_encode	编码说明
raw	未压缩的原始音频采样数据
alaw	A-law编码, 详细介绍请参考: https://github.com/dystopiancode/pcm-g711
ulaw	μ-law编码, 详细介绍请参考: https://github.com/dystopiancode/pcm-g711
mp3	mp3编码, 详细介绍请参考: https://lame.sourceforge.io/
speex	speex编码 (会在每帧数据前添加4个字节, 表示当前帧的大小), 详细介绍请参考: https://www.speex.org/
opus	opus编码 (会在每帧数据前添加4个字节, 表示当前帧的大小), 详细介绍请参考: https://opus-codec.org/

语速的取值范围

1. 语速取值范围为-100到100, 取值越大语速越慢。
2. 目前仅有中文、英文两个语种支持调整语速。

