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WebSocket 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	ws://api.baller-tech.com/v1/service/ws/v1/tts
字符编码	UTF-8
WebSocket版本	13 (RFC 6455)
响应格式	统一采用JSON格式

调用流程

1. 通过hmac-sha256计算签名，向服务器端发送WebSocket协议握手请求。
2. 握手成功之后，通过WebSocket连接上传和接收数据。
3. 请求方接收到服务器端推送的结果返回结束标记后断开WebSocket连接

握手和接口鉴权

在WebSocket的握手阶段，请求方需要对请求进行签名，服务端会根据签名检查请求的合法性。握手时请求方将签名相关的参数经过url编码后加到请求地址的后面，具体的参数和示例如下：

```
ws://api.baller-tech.com/v1/service/ws/v1/tts?
authorization=xxxx&host=xxxx&date=xxx
```

参数	类型	说明	示例
host	string	请求的主机	api.baller-tech.com
date	string	当前GMT格式的时间	Fri, 10 Jan 2020 07:31:50 GMT
authorization	string	鉴权信息Base64编码后的数据	-

握手和鉴权参数详细介绍

date介绍

1. date必须是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
2. 服务端允许date的最大偏差为300秒，超出此偏差请求会被拒绝

authorization介绍

authorization使用base64编码前的格式如下json格式

```
{
    "app_id": "1172448516240310275",
    "signature": "qaIpgE3Ecs78g6GRFxQBJKgdna28b7ronAcsDCsO+Zw="
}
```

app_id介绍

1. 由北京大牛儿科技发展有限公司统一分配。

signature介绍

1. signautre 是使用hmac-sha256对参数进行签名后并base64编码的字符串。
2. signautre 使用hmac-sha256签名前的原始字段由三部分构成，分别为app_id、date、host。每一部分使用换行符(\n)进行分割，“:”号前后无空格。

```
app_id:1172448516240310275
date:Fri, 10 Jan 2020 07:31:50 GMT
host:api.baller-tech.com
```

3. 使用hmac-sha256算法，结合app_key（由北京大牛儿科技发展有限公司统一分配）对signautre的原始字段进行签名。
4. 对签名数据进行base64编码，生成signature的字段值。

握手和鉴权消息响应

1. 接口鉴权成功时，WebSocket握手回复报文的状态码为101。
2. 接口鉴权失败时，WebSocket握手回复报文的状态码为403，可以通过响应行的原因短语查看接口鉴权失败原因。
3. 接口鉴权失败时，响应报文的主体中会返回json格式的数据，包含了以下信息

参数	类型	说明
task_id	string	本次任务的标识，如果对请求有疑问，可以将task_id提供给我公司进行排查
message	string	接口鉴权失败的原因，与响应行中的原因短语相同

数据的发送和接收

握手成功之后，请求方和服务器会建立WebSocket的连接，请求方将数据通过WebSocket发送给服务器，服务器有合成结果的时候，会通过WebSocket连接推送合成结果到请求方。请求方和服务器通过json的格式交换数据。

请求方发送数据时使用的参数

参数名	类型	是否每帧必须	描述
business	obj	否	业务参数，仅在握手成功后首帧中上传
data	obj	是	数据流参数，握手成功后所有帧中都需要上传

业务参数(business)

参数名	类型	是否必须	默认值	描述
language	string	是	无	音频的语种；参见支持的语种和采样格式
sample_format	string	否	audio/L16;rate=16000	音频采样格式；参见支持的语种和采样格式
audio_encode	string	否	raw	音频编码格式；参见支持的音频编码
voice_name	string	是	无	音频发言人；参见支持的发音人
speed	float	是	无	音频输出的语速；参见语速的取值范围
tempo	float	否	0	音频输出的节奏；参见节奏的取值范围
pitch	float	否	0	音频输出的音调；参见音调的取值范围
volume	float	是	1.0	音频输出的音量；参见音量的取值范围

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表示音频数据为16000hz，16bit的pcm音频数据

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。

数据流参数（data）

参数名	类型	是否必须	描述
txt	string	是	经过base64编码后的文本数据

- **拼音处理**：文本中包含人名等的汉语拼音，希望按照拼音发音时，需要添加指定的标签`[rp1]`、`[rp0]`
    - ```
      My name is [rp1]xiǎo péng you[rp0].
      你好啊，[rp1]xiǎo péng you[rp0]。
      ```

```
{
    "data": {
        "txt":
"AAAFAAoADwAXAB0AJgA0AEIATABPAE8AUQBRAEgAOwA0AC8AJwAcABUAEQAJAAIAAgADAAAA+P="
    },
    "business": {
        "language": "mon_i",
        "sample_format": "audio/L16;rate=16000",
    }
}
```

服务器推送结果的参数

参数名	类型	描述
task_id	string	本次任务的id，仅在第一帧中返回，如果对请求有疑问，可以将task_id提供给我公司进行排查
code	int	请求处理的结果码
message	string	错误提示
is_end	int	结果返回是否结束（0-未结束; 1-结束），当为1时，请求方需关闭WebSocket
data	string	base64编码后的合成音频数据

```
{
    "code": 0,
    "message": "success",
    "is_end": 0,
    "data": "xxxxxx",
    "task_id": "1172448516240310275-2903dc7e3ab65879b4fc66055720ec09"
}
```
```

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（传统）	kaz_i	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
蒙语（传统）	mon_i	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
蒙语（西里尔）	mon_o	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
朝鲜语	kor	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/

audio_encode	编码说明
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0.5到2.0，0.5最慢，1.0为正常，2.0最快。

节奏的取值范围

1. 节奏取值范围为-50到50，-50最慢，0为正常，50最快。

音调的取值范围

1. 音调取值范围为-10到10，-10最低，0为正常，10最高。

音量的取值范围

1. 音量的取值范围为0.0到1.0，0.0音量最低，1.0音量最高，默认1.0。
2. 目前仅中文、英文支持音量设置，其他语种仅支持音量为1.0的值。

支持的发音人

发音人	语种	备注
yiyi	中文	支持
runrun	中文	支持
ruirui	中文	支持
nana	中文	支持
lili	中文	支持
mingxuan	中文	支持
yueni	中文	支持
muze	中文	支持
tingyan	中文	支持
mary	英语（英音）	支持
elise	英语（美音）	支持
regina	英语（美音）	支持
chana	蒙语（传统）	支持
gerile	蒙语（传统）	支持
danba	蒙语（传统）	支持

发音人	语种	备注
tana	蒙语（西里尔）	支持
suolangcuomu	藏语（卫藏）	支持
gesangwangmu	藏语（卫藏）	支持
renyang	藏语（安多）	支持
yangla	藏语（安多）	支持
cangla	藏语（康巴）	支持
guli	维语	支持
amina	维语	支持
ailinna	哈萨克语（传统）	支持
mayila	哈萨克语（传统）	支持
minzhen	朝鲜语	支持
hailaiyousuo	彝语	支持
dafei	壮语	支持
yinan	壮语	支持