

大模型语音合成API

最近更新时间：2025.07.09 18:24:57首次发布时间：2024.04.25 10:44:40

[我的收藏](#)

有用

无用

Websocket

使用账号申请部分申请到的 appid&access_token 进行调用
文本一次性送入，后端边合成边返回音频数据

1. 接口说明

接口地址为 wss://openspeech.bytedance.com/api/v1/tts/ws_binary

2. 身份认证

认证方式使用 Bearer Token，在请求的 header 中加上 `"Authorization": "Bearer; {token}"`，并在请求的 json 中填入对应的 appid。

注意

Bearer 和 token 使用分号 ; 分隔，替换时请勿保留 {}

AppID/Token/Cluster 等信息可参考 [控制台使用FAQ-Q1](#)

3. 请求方式

3.1 二进制协议

报文格式(Message format)

Byte \ Bit	7	6	5	4	3	2	1	0
0	Protocol version				Header size			
1	Message type				Message type specific flags			
2	Message serialization method				Message compression			
3	Reserved							
4	[Optional header extensions]							
5	[Payload, depending on the Message Type]							
6	...							

所有字段以 **Big Endian(大端序)** 的方式存储。

字段描述

消息类型详细说明

目前所有 TTS websocket 请求都使用 full client request 格式，无论"query"还是"submit"。

Full client request

- Header size为 **b0001** (即 4B，没有 header extension)。
- Message type为 **b0001** 。
- Message type specific flags 固定为 **b0000** 。
- Message serialization method为 **b0001** JSON。字段参考上方表格。
- 如果使用 gzip 压缩 payload，则 payload size 为压缩后的大小。

Audio-only server response

- Header size 应该为 **b0001** 。

- Message type为 `b1011` .
- Message type specific flags 可能的值有：
 - `b0000` - 没有 sequence number.
 - `b0001` - sequence number > 0.
 - `b0010` or `b0011` - sequence number < 0, 表示来自服务器的最后一条消息, 此时客户端应合并所有音频片段(如果有多条)。
- Message serialization method为 `b0000` (raw bytes).

4.注意事项

- 每次合成时reqid这个参数需要重新设置, 且要保证唯一性 (建议使用uuid.V4生成)
- websocket demo中单条链接仅支持单次合成, 若需要合成多次, 需自行实现。每次创建websocket连接后, 按顺序串行发送每一包。一次合成结束后, 可以发送新的合成请求。
- operation需要设置为submit才是流式返回
- 在 websocket 握手成功后, 会返回这些 Response header

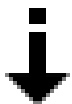
Demo

python



tts_websocket_demo.py

未知大小



Java



tts-demo-java.zip

7.01KB

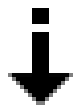


Go



tts_websocket_demo.go

7.68KB



HTTP

使用账号申请部分申请到的 appid&access_token 进行调用
文本全部合成完毕之后，一次性返回全部的音频数据

4. 接口说明

接口地址为 <https://openspeech.bytedance.com/api/v1/tts>

5. 身份认证

认证方式采用 Bearer Token.

1)需要在请求的 Header 中填入"Authorization":"Bearer;\${token}"

注意

Bearer 和 token 使用分号 ; 分隔，替换时请勿保留\${}

AppID/Token/Cluster 等信息可参考 [控制台使用FAQ-Q1](#)

6. 注意事项

- 使用 HTTP Post 方式进行请求，返回的结果为 JSON 格式，需要进行解析
- 因 json 格式无法直接携带二进制音频，音频经 base64 编码。使用 base64 解码后，即为二进制音频
- 每次合成时 reqid 这个参数需要重新设置，且要保证唯一性（建议使用 UUID/GUID 等生成）

参数列表

| Websocket 与 Http 调用参数相同

请求参数

- 2. ssml 能力已支持，详见 [SSML 标记语言--豆包语音-火山引擎 \(volcengine.com\)](https://volcengine.com)
- 3. 暂时不支持音高调节
- 4. 大模型音色语种支持中英混
- 5. 大模型非双向流式已支持latex公式
- 6. 在 websocket/http 握手成功后，会返回这些 Response header

请求示例：

代码块

```
1  {
2      "app": {
3          "appid": "appid123",
4          "token": "access_token",
5          "cluster": "volcano_tts",
6      },
7      "user": {
8          "uid": "uid123"
9      },
10     "audio": {
11         "voice_type": "zh_male_M392_conversation_wvae_big tts",
12         "encoding": "mp3",
13         "speed_ratio": 1.0,
14     },
15     "request": {
16         "reqid": "uuid",
17         "text": "字节跳动语音合成",
18         "operation": "query",
19     }
20 }
```

返回参数

响应示例

代码块

```
1  {
2      "reqid": "reqid",
3      "code": 3000,
4      "operation": "query",
5      "message": "Success",
6      "sequence": -1,
7      "data": "base64 encoded binary data",
8      "addition": {
9          "duration": "1960",
10     }
11 }
```

注意事项

- websocket 单条链接仅支持单次合成，若需要合成多次，则需要多次建立链接
- 每次合成时 reqid 这个参数需要重新设置，且要保证唯一性（建议使用 uuid.V4 生成）
- operation 需要设置为 submit

返回码说明

常见错误返回说明

1. 错误返回：
"message": "quota exceeded for types: xxxxxxxxx_lifetime"
错误原因：试用版用量用完了，需要开通正式版才能继续使用
2. 错误返回：
"message": "quota exceeded for types: concurrency"
错误原因：并发超过了限定值，需要减少并发调用情况或者增购并发
3. 错误返回：
"message": "Fail to feed text, reason Init Engine Instance failed"
错误原因：voice_type / cluster 传递错误
4. 错误返回：
"message": "illegal input text!"
错误原因：传入的 text 无效，没有可合成的有效文本。比如全部是标点符号或者 emoji 表情，或者使用中文音色时，传递日语，以此类推。多语种音色，也需要使用 language 指定对应的语种
5. 错误返回：
"message": "authenticate request: load grant: requested grant not found"
错误原因：鉴权失败，需要检查 appid&token 的值是否设置正确，同时，鉴权的正确格式为 headers["Authorization"] = "Bearer;\${token}"

6. 错误返回：

"message": "extract request resource id: get resource id: access denied"

错误原因：语音合成已开通正式版且未拥有当前音色授权，需要在控制台购买该音色才能调用。标注免费的音色除 BV001_streaming 及 BV002_streaming 外，需要在控制台进行下单（支付 0 元）