

# 音色变换 开发指南



腾讯云

**【 版权声明 】**

©2013–2024 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

**【 商标声明 】**

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

**【 服务声明 】**

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

**【 联系我们 】**

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

---

## 文档目录

开发指南

声音变换 (websocket)

# 开发指南

## 声音变换（websocket）

最近更新时间：2024-05-15 11:18:11

### 接口描述

本接口服务采用 websocket 协议，允许用户以音频流的形式输入原始声音，同时转换后的声音也以流的形式实时推送回去，可以做到边说话边转换的实时效果。

在使用该接口前，需要 [开通音色变换服务](#)，并进入 [API 密钥管理页面](#) 新建密钥，生成 AppID、SecretID 和 SecretKey，用于 API 调用时生成签名，签名将用来进行接口鉴权。

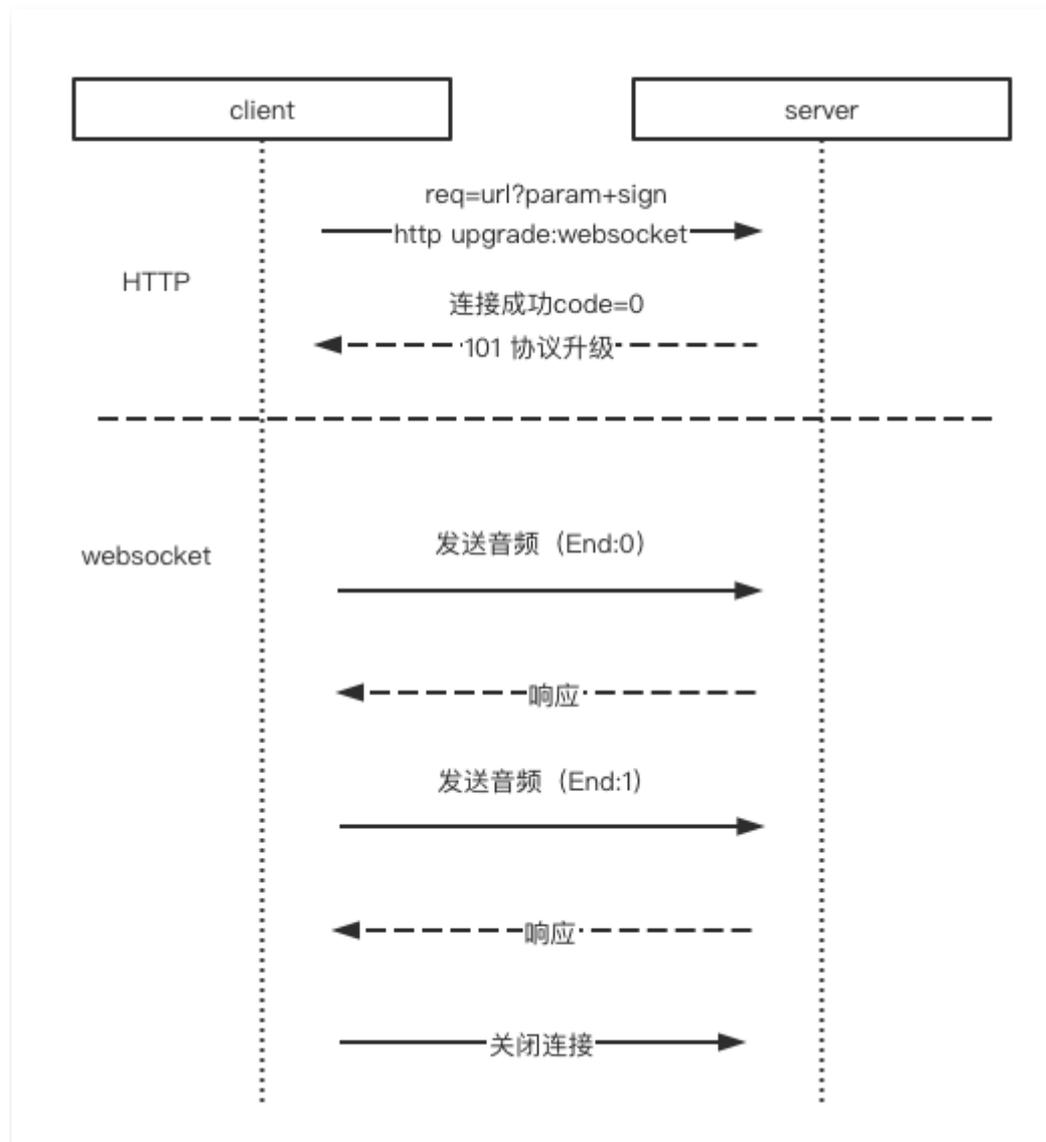
### 接口要求

集成实时语音变换 API 时，需按照以下要求。

类型	说明
语言	支持中文普通话
音频属性	采样率：16000Hz 采样精度：16bits 声道：单声道（mono）
音频格式	pcm
请求协议	wss 协议
请求地址	<code>wss://tts.cloud.tencent.com/vc_stream/&lt;appid&gt;?{请求参数}</code>
接口鉴权	签名鉴权机制，详见 <a href="#">签名生成</a>
数据发送	建议每100ms 发送100ms 时长（即1:1实时率）的数据包，对应 pcm 大小为：16k采样率 3200字节 音频发送速率过快超过1:1实时率或者音频数据包之间发送间隔超过6秒，可能导致引擎出错，后台将返回错误并主动断开连接
并发限制	默认单账号限制并发连接数为10路，如您有提高并发限制的需求，请 <a href="#">提交工单</a> 进行咨询

## 接口调用流程

接口调用流程分为两个阶段：握手阶段和声音变换阶段。



### 握手阶段

握手阶段，客户端主动发起 websocket 连接请求，请求 URL 格式为：

```
wss://tts.cloud.tencent.com/vc_stream/<appid>?{请求参数}
```

其中 <appid> 需替换为腾讯云注册账号的 AppID，可通过 [API 密钥管理页面](#) 获取，{请求参数}格式为：

```
key1=value2&key2=value2...(key 和 value 都需要进行 urlencode)
```

## 请求参数:

参数名称	必填	类型	描述
SecretId	是	String	官网 SecretId
Timestamp	是	Integer	当前 UNIX 时间戳, 可记录发起 API 请求的时间。例如1529223702, 如果与当前时间相差过大, 会引起签名过期错误
AppId	是	Integer	账号 AppId ( 请确保该字段数据类型为整型 int )
Expired	是	Integer	签名的有效期, 是一个符合 UNIX Epoch 时间戳规范的数值, 单位为秒; Expired 必须大于 Timestamp 且 Expired-Timestamp 小于90天
VoiceType	是	Integer	<ul style="list-style-type: none"> <li>301005-小新, 男童声</li> <li>301006-小姐姐, 女声</li> <li>301007-童声, 男童声</li> <li>301008-东北话, 男声</li> <li>301009-影视解说, 男声</li> <li>301010-萝莉, 女声</li> <li>301011-小黄人, 男童声</li> </ul>
SampleRate	是	Integer	音频采样率: 16000: 16k
Codec	是	String	音频格式: pcm: pcm 音频
End	是	Integer	该字段为1时表示音频流全部发送完成
VoiceId	是	String	音频流识别全局唯一标识, 一个 websocket 连接对应一个, 用户自己生成 ( 推荐使用 uuid ), 最长128位
Volume	否	Float	音量大小, 范围[-10, 10], 用于调整变换后音频音量。默认为0, 代表正常音量。音量值大于0, 增加音量。音量值小于0, 降低音量。

## signature 签名生成

1. 对除 signature 之外的所有参数按字典序进行排序，拼接请求 URL 作为签名原文，这里以 `Appld=251139***, SecretId=*****5xz330ynTzIOGvNdggghaG2oYL*****` 为例拼接签名原文，则拼接的签名原文为：

```
tts.cloud.tencent.com/vc_stream/251139***?  
Codec=pcm&End=0&Expired=1670390440&SampleRate=16000&SecretId=*****5xz3  
30ynTzIOGvNdggghaG2oYL*****&Timestamp=1670304040&Voiceld=11111111&Voice  
Type=301002
```

2. 对签名原文使用 SecretKey 进行 HmacSha1 加密，之后再进行 base64 编码。例如对上一步的签名原文 `SecretKey=*****PXQ5WihT5ZsrxYlV6b5n*****`，使用 HmacSha1 算法进行加密并做 base64 编码处理：

```
Base64Encode(HmacSha1("tts.cloud.tencent.com/vc_stream/251139***?  
Codec=pcm&End=0&Expired=1670390440&SampleRate=16000&SecretId=*****5xz3  
30ynTzIOGvNdggghaG2oYL*****&Timestamp=1670304040&Voiceld=11111111&Voice  
Type=301002", "*****PXQ5WihT5ZsrxYlV6b5n*****"))
```

3. 得到 signature 签名值为：

```
H5EZgzPGP5wkjGHw96ijDrrqFVs=
```

4. 将 signature 值进行 `urlencode`（必须进行 URL 编码，否则将导致鉴权失败偶现）后拼接得到最终请求 URL 为：

```
wss://tts.cloud.tencent.com/vc_stream/251139***?  
Codec=pcm&End=0&Expired=1670390440&SampleRate=16000&SecretId=*****5xz3  
30ynTzIOGvNdggghaG2oYL*****&Timestamp=1670304040&Voiceld=11111111&Voice  
Type=301002&Signature=H5EZgzPGP5wkjGHw96ijDrrqFVs%3D
```

## 请求响应

客户端使用最终请求 URL 发起连接请求后，后台建立连接并进行签名校验，校验成功则返回 code 值为 0 的确认消息表示握手成功；如果校验失败，后台返回 code 为非 0 值的消息并断开连接。

## 声音变换阶段

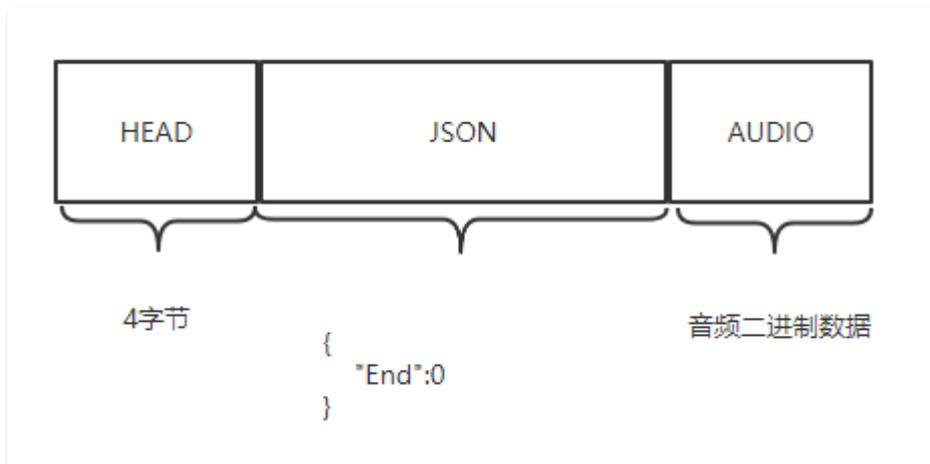
1. 握手成功之后，进入声音变换阶段，客户端上传语音数据并接收变换结果消息。
2. 在变换过程中，客户端持续上传 binary message 到后台，内容为（HEAD + JSON + 音频流）二进制数据。

**说明：**

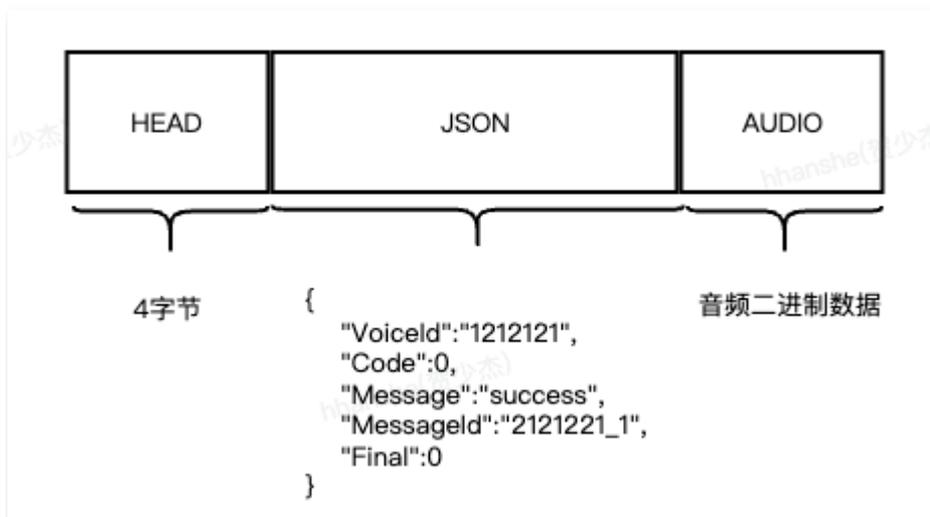
建议音频数据每100ms发送100ms时长（即1:1实时率）的数据包，对应 pcm 大小为：16k采样率3200字节。音频发送速率过快超过1:1实时率或者音频数据包之间发送间隔超过6秒，可能导致引擎出错，后台将返回错误并主动断开连接。

3. 音频流上传完成后，客户端需将 JSON 结构体中 End 参数置为1，通知后台结束声音变换。

**请求参数：**



**响应参数：**



**格式说明：**

- 第一部分 HEAD 四字节的二进制，长度为一个大的32位 int 转换而来，表示第二部分 JSON 数据的长度。

- 第二部分是一个 JSON 的字符串。
- 第三部分是音频的二进制数据，若数据包中不包含音频，则第三部分可以为空。

#### 参数说明：

参数名称	必填	类型	描述
Voiceld	String	是	16位 String 串作为每个音频的唯一标识，用户自己生成
MessageId	String	是	本 Message 唯一 ID
Message	String	是	响应信息
Final	Integer	是	该字段返回1时表示音频流全部变换结束
Code	Integer	是	状态码，0代表正常，非0值表示发生错误

## 开发者资源

### SDK

- [Tencent Cloud SDK for Go](#)
- [Tencent Cloud SDK for Python](#)

### 错误码

错误码	说明
4001	参数不合法，具体详情参考 message
4002	鉴权失败
4003	AppID 服务未开通，请在控制台开通服务
4004	无可使用的免费额度
4005	账户欠费停止服务，请及时充值
4006	账号当前调用并发超限
4007	音频解码失败，请检查上传音频数据格式与调用参数一致
4008	客户端数据上传超时

4100	服务未开通
4102	欠费停服
4103	用户主动停服
4109	资源包余量已用尽
4009	客户端连接断开
5000	后台错误，请重试
5001	后台识别服务器变换失败，请重试
5002	后台识别服务器变换失败，请重试