

口语评测（新版）

API 文档



腾讯云

【 版权声明 】

©2013–2025 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

API 文档

智聆口语评测（新版）相关接口

API 文档

智聆口语评测（新版）相关接口

最近更新时间：2025-07-18 15:56:31

接口描述

本接口服务采用 websocket 协议，对实时音频流进行评测，同步返回识别结果，达到“边说边评测发音”的效果。

在使用该接口前，需要前往智聆口语评测控制台 [开通服务](#)，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

集成实时语音识别 API 时，需按照以下要求。

内容	说明
语言种类	支持中文普通话和英语。可通过接口参数 server_engine_type 设置对应语言类型
音频属性	● 采样率：16000Hz ● 采样精度：16bits ● 声道：单声道（mono）
音频格式	pcm、wav、mp3、speex
请求协议	wss 协议
请求地址	wss://soe.cloud.tencent.com/soe/api/{请求参数}
接口鉴权	签名鉴权机制
响应格式	统一采用 JSON 格式
数据发送	建议每40ms发送40ms时长（即1:1实时率）的数据包，对应16k 采样率的 pcm 大小为1280 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数为50路，如您有提高并发限制的需求，请前往 购买页 购买额外并发路数。
------	---

接口调用流程

接口调用流程分为两个阶段：握手阶段和识别阶段。两阶段后台均返回 text message，内容为 json 序列化字符串，以下是格式说明：

字段名	类型	描述
code	Integer	状态码，0代表正常，非0值表示发生错误
message	String	错误说明，发生错误时显示这个错误发生的具体原因，随着业务发展或体验优化，此文本可能会经常保持变更或更新
voice_id	String	音频流唯一 id，由客户端在握手阶段生成并赋值在调用参数中
message_id	String	本 message 唯一 id
result	SentenceInfo	最新评测结果
final	Integer	该字段返回1时表示音频流全部识别结束

握手阶段

请求格式

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

```
wss://soe.cloud.tencent.com/soe/api/<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...(生成签名的时候不需要对参数urlencode，实际请求时key和 value 都需要进行 urlencode)
```

参数说明：

参数名称	必填	类型	描述
------	----	----	----

secretid	是	String	腾讯云注册账号的密钥 SecretId，可通过 API 密钥管理页面 获取
timestamp	是	Integer	当前 UNIX 时间戳，单位为秒。如果与当前时间相差过大，会引起签名过期错误
expired	是	Integer	签名的有效期截止时间 UNIX 时间戳，单位为秒。expired 必须大于 timestamp 且 expired - timestamp 小于90天
nonce	是	Integer	随机正整数。用户需自行生成，最长10位
server_engine_type	是	String	模型引擎类型 16k_zh：中文标准版； 16k_en：英文标准版；
voice_id	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。
voice_format	否	Int	语音编码方式，可选，默认值为0 0: pcm； 1: wav； 2: mp3； 4: speex(sp)
signature	是	String	接口签名参数
text_mode	否	Integer	输入文本模式 0: 普通文本（默认） 1: 音素结构 文本
ref_text	否	String	被评估语音对应的文本，仅支持中文和英文。 - 句子模式下不超过 30个 单词或者中文文字。 - 段落模式下不超过 120 个单词或者中文文字。 - 中文评估使用 utf-8 编码。 - 自由说模式 RefText 可以不填。关于 RefText 的文本键入要求，请参考 评测模式介绍。 如需要在评测模式下使用自定义注音（支持中英文），可以通过设置 TextMode 参数实现，设置方式请参考 音素标注 。 示例值：apple
keyword	否	String	主题词和关键词示例值：keyword

eval_mode	是	Integer	评测模式 0: 单词/单字模式（中文评测模式下为单字模式） 1: 句子模式 2: 段落模式 3: 自由说模式 4: 单词音素纠错模式 5: 情景评测模式 6: 句子多分支评测模式 7: 单词实时评测模式 8: 拼音评测模式 关于每种评测模式的详细介绍，以及适用场景，请参考 评测模式介绍。 示例值：1
score_coeff	是	Float	评价苛刻指数。取值为[1.0 – 4.0]范围内的浮点数，用于平滑不同年龄段的分数。 1.0: 适用于最小年龄段用户，一般对应儿童应用场景； 4.0: 适用于最高年龄段用户，一般对应成人严格打分场景。 苛刻度影响范围请参考 苛刻指数介绍 示例值：1.0
sentence_info_enabled	否	Integer	输出断句中间结果标识 0: 不输出（默认） 1: 输出，通过设置该参数可以在评估过程中的分片传输请求中，返回已经评估断句的中间结果，中间结果可用于客户端 UI 更新
rec_mode	否	Integer	是否为录音识别模式标识 0: 实时识别（默认） 1: 录音识别 注意：录音模式不支持分片传输音频，需要一次性发送音频（上限 60s）。 单次连接只能发一个分片，得到识别结果后需要关闭此条 websocket 连接，再次识别需要重新建立连接。

signature 签名生成

- 对除 signature 之外的所有参数按字典序进行排序，拼接请求 URL（不包含协议部分：wss://）作为签名原文，这里以 Appid=125922***，SecretId=*****Qq1zhZMN8dv0***** 为例拼接签名原文，则拼接的签名原文为：

```
soe.cloud.tencent.com/soe/api/1306***?
eval_mode=0&expired=1722408159&nonce=42261112&ref_text=hello&score_coeff
=1.5&secretid=*****
```

```
**&sentence_info_enabled=1&server_engine_type=16k_zh&text_mode=0&timesta  
mp=1722321759&voice_format=1&voice_id=4943511b-192c-40f8-b6c9-  
c3df2a827b75
```

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=*****SkqpeHgqmSz*****`，使用 **HmacSha1** 算法进行加密并做 **base64** 编码处理：

```
Base64Encode(HmacSha1("soe.cloud.tencent.com/soe/api/1306***?  
eval_mode=0&expired=1722408159&nonce=42261112&ref_text=hello&score_coeff  
=1.5&secretid=*****  
**&sentence_info_enabled=1&server_engine_type=16k_zh&text_mode=0&timesta  
mp=1722321759&voice_format=1&voice_id=4943511b-192c-40f8-b6c9-  
c3df2a827b75))
```

得到 **signature** 签名值为：

```
G8jDQBRg1JfeBi/YnTjyjekxfDA=
```

3. 将 **signature** 值进行 **urlencode**（必须进行 **URL** 编码，编码函数必须要支持对 **+**、**=** 等特殊字符的编码，否则将导致鉴权失败偶现）后拼接得到最终请求 URL 为：

```
soe.cloud.tencent.com/soe/api/1306***?  
eval_mode=0&expired=1722408159&nonce=42261112&ref_text=hello&score_coeff  
=1.5&secretid=*****  
**&sentence_info_enabled=1&server_engine_type=16k_zh&text_mode=0&timesta  
mp=1722321759&voice_format=1&voice_id=4943511b-192c-40f8-b6c9-  
c3df2a827b75&signature=oUscTnUagB5HpfUxQYIS0Py7GZM%3D
```

请求响应

客户端发起连接请求后，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。

```
{"code":0,"message":"success","voice_id":"RnKu9FODFHK5FPpsrN"}
```

识别阶段

握手成功之后，进入识别阶段，客户端上传语音数据并接收识别结果消息。

上传数据

在识别过程中，客户端持续上传 binary message 到后台，内容为音频流二进制数据。建议每40ms 发送40ms 时长（即1:1实时率）的数据包，对应 pcm 大小为：16k采样率1280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。

音频流上传完成之后，客户端需发送以下内容的 text message，通知后台结束识别。

```
{"type": "end"}
```

接收消息

客户端上传数据的过程中，如果 sentence_info_enabled 打开，则需要同步接收后台返回的实时识别结果，结果示例：

```
{"code":0,"message":"success","voice_id":"RnKu9FODFHK5FPpsrN","message_id":"RnKu9FODFHK5FPpsrN_11_0","result":{"SuggestedScore:-0.36000001430511475 PronAccuracy:-1 PronFluency:-1 PronCompletion:0.20000000298023224 Words:[{Mbtm:760 Metm:230 PronAccuracy:91.225341796875 PronFluency:0.9780682325363159 ReferenceWord: Word:窗 Tag:0 KeywordTag:0 PhoneInfo:[] Tone:{Valid:false RefTone:-1 HypTone:-1}}] SentenceId:0 RefTextId:-1 KeyWordHits:[] UnKeyworddits:[]}}
```

后台识别完所有上传的语音数据之后，最终返回 final 值为1的消息并断开连接。

```
{"code":0,"message":"success","voice_id":"CzhjngBkv8lk5pRUxhpX","message_id":"CzhjngBkv8lk5pRUxhpX_241","final":1}
```

识别过程中如果出现错误，后台返回 code 为非0值的消息并断开连接。

```
{"code":4008,"message":"后台识别服务器音频分片等待超时","voice_id":"CzhjngBkv8lk5pRUxhpX","message_id":"CzhjngBkv8lk5pRUxhpX_241"}
```

开发者资源

SDK

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

SDK 调用示例

- [Golang 示例](#)
- [Java 示例](#)
- [Python 示例](#)

错误码

数值	说明
4000	音频数据发送过多，请1秒内最多发送3秒音频数据注：实时识别的效果是“边说边评测”，1秒内发送的音频数据总时长应为1秒
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	资源包耗尽，请开通后付费或者购买资源包
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	客户端超过15秒未发送音频数据
4009	客户端连接断开
4010	客户端上传未知文本消息
4011	分片音频数据太大，请按照文档指引调整分片大小
4014	输入音频时长超过限制，请结束本次评测，后续请根据评测模式设置音频时长限制
4102	请求参数 ref_text 无效或参考文本为空，请检查 ref_text 是否为空
4103	请求参数 ref_text 包含 OOV 词汇，请使用指定发音或联系我们处理
4104	请求参数 ref_text 的字数超过最大限制，请根据评测模式调整字数后重新初始化
4105	请求音频无人声
4106	输入音频时长超过限制，请结束本次评测，后续请根据评测模式设置音频时长限制

4107	输入音频异常，音频数据指针或音频 度必须为偶数，请参考 API 文档检查音频数据是否正确后重新传输数据
4108	引擎未识别到有效语 ，检查音频数据是否包含有效发音
4109	该评测功能暂不支持，请参考 API 文档检查评测参数
4110	请求参数 ref_text 语法错误，请参考 API 文档检查文本格式，尤其是指定发音格式是否正确
4111	请求参数 ref_text 语法错误，请参考 API 文档检查文本格式，尤其是指定发音格式是否正确
4112	检查参考文本中是否包含大量多音字，可通过发音描述块指定标准发音解决
4113	ref_text 指定发音无效
4114	ref_text 内容无效
4115	RefText 内容和使用引擎不匹配，请仔细检查是否中英文评测引擎使用错误
5000	因机器负载过高、网络抖动等导致失败，请重新发起新识别注：该问题通常为偶发，少量出现可忽略，发起新识别即可
5001	因机器负载过高、网络抖动等导致失败，请重新发起新识别注：该问题通常为偶发，少量出现可忽略，发起新识别即可
5002	因机器负载过高、网络抖动等导致失败，请重新发起新识别注：该问题通常为偶发，少量出现可忽略，发起新识别即可