

语音合成同步接口

概述

语音合成（Text-to-Speech, TTS）是将文本转换成语音的技术，可用于有声书、有声新闻、有声广告等应用场景。本文档说明了 HAS 平台目前支持哪些 TTS 服务, 以及业务方如何选择并接入相关接口。

本文档介绍同步合成接口：

📌 同步合成

- 同步合成方式在合成请求发出后会阻塞直到音频合成完毕。主要面向实时性要求高，但对首包时延要求不高的场景，例如新闻、广告、开放平台等。

接口调用

服务地址

- <endpoint>
 - prod: audio-ai.ximalaya.com

鉴权

API接口需要token认证

- 具体token根据每个业务会有不同，申请需联系@喜马拉雅音频算法团队
- token在http header使用示例：

Authorization: Bearer eyJ0eXaiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJleAoiOiJlbnRlZDI6MTYyNzAxMjkzMywic3ViOiJoiaGlksW1vIn0.71Ci19_JDizfGiIYL2Zd70pQXZ

接口状态码说明

code	解释
200000	成功
400001	身份认证失败
400003	该账号无权限访问该资源
400000	默认的客户端错误码
400429	请求太频繁(超过并发限制)
401431	tts字数超过限制
500000	默认的服务端错误
501001	内部调用错误

Token接口(外网访问须要通过该接口动态获取,token有效期24h)

接口描述

- 协议: HTTP
- 方法: POST
- content-type: application/x-www-form-urlencoded

- 地址：<https://<endpoint>/has/v1/user/access-token>

输入参数

- 参数说明:

参数	是否必填	说明	默认值
username	是	username, 由喜马拉雅提供	无
password	是	password, 由喜马拉雅提供	无
timestamp	是	unix time 以秒为单位, 例如 1651507200	无
signature	是	username+password+timestamp的MD5值, password由喜马拉雅提供。	无

- 参考实例:

◦ 请求示例:

```
curl --location --request POST 'https://audio-ai.ximalaya.com/has/v1/user/access-token' \
--header 'accept: application/json' \
--header 'Content-Type: application/x-www-form-urlencoded' \
--data-urlencode 'username=111' \
--data-urlencode 'password=000' \
--data-urlencode 'timestamp=1653460479' \
--data-urlencode 'signature=ed50dd2af84782d0e2293a1e92a0e649'
```

◦ 响应示例:

```
{
  "success": true,
  "code": 200000,
  "message": "ok",
  "data": {
    "token": "Bearer
eyJ0eXaiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJleAoiOiJlbnRlZCI6MTYyNzAxMjkzMywic3ViIjoiaGlksW1vIn0.71Ci19_JDizfGiIYL2Zd7Qj

  }
}
```

音色查询接口

接口描述

查询可用的音色及相关信息

- 类型: GET
- 地址: <https://<endpoint>/has/tts/api/v1/voices>
- 用途: 查询可选的语种对应的音色。一种音色由 3 个维度共同确定(domain, speaker_name, speaker_variant), 即说话人适用的领域, 说话人的名称, 说话人的风格 变种

输入参数

- 参数：无

输出参数

- 返回：每个语种支持的音色列表，每个音色是一个三元组(domain, speaker_name, speaker_variant)
- 返回结果示例：

```
{
  "success": true,
  "code": 200000,
  "message": "ok",
  "data": {
    "zh-cn": [
      {
        "domain": "小说",
        "speaker_name": "喜小道",
        "speaker_variant": "中性(旁白)"
      },
      {
        "domain": "广告",
        "speaker_name": "喜小华",
        "speaker_variant": "中性(旁白)"
      }
    ],
    "en-us": [
      {
        "domain": "小说",
        "speaker_name": "喜小雨",
        "speaker_variant": "中性(旁白)"
      }
    ]
  }
}
```

- 返回值说明：

状态	说明
success	是否成功(code等于2xxxxx为是)
code	参见状态码的说明
message	对应状态码的说明
data	- 语种信息（如zh-cn, en-us） - 每个语种支持的音色列表,每个音色信息包含: - domain - speaker_name - speaker_variant

同步合成接口

接口描述

- 类型: POST
- 地址: <https://<endpoint>/has/tts/api/v1/synth/sync>
- 文本长度限制: 单次调用长度不能超过 2000 字符（其中1个汉字、1个英文字母或1个标点均算作1个字符）

输入参数

- 参数说明:

参数	是否必须	说明	默认值
text	是	需要合成的文本	无
lang	否	需要合成的语种, 目前支持: "zh-cn", "en-us"	"zh-cn"
ssml	否	是否是SSML格式的输入, 默认: false	false

参数	是否必须	说明	默认值
speaker	是	选择哪个音色来合成，由三个维度确定一个音色 (domain, speaker_name, speaker_variant), 可通过`/has/tts/api/v1/voices`查询可用的音色列表	无
sr	否	合成音频的采样率	24000
audio_format	否	合成音频的格式，目前支持：“wav”，“mp3”，“m4a”	“wav”
br	否	当音频格式为“mp3”或“m4a”时生效，设置bitrate，默认“64k”	“64k”
pitch	否	调整音高pitch, 取值范围: [0.0, 100.0] ◦ 0.0 最低沉 ◦ 50.0 表示默认音高 ◦ 100.0 最尖	50.0
rate	否	调整语速，取值范围: [0.0, 100.0] ◦ 0.0 最慢，表示默认语速的 0.5 倍速 ◦ 50.0 表示默认语速 ◦ 100.0 最快，表示默认语速的 2 倍速	50.0
volume	否	调整音量，取值范围: [0.0, 100.0] ◦ 0.0 最轻 ◦ 50.0 表示默认音量 ◦ 100.0 最响	50.0

参数	是否必须	说明	默认值
bgm	否	背景音乐，可设置： ◦ url: ◦ 背景音url ◦ bgm_volume: ◦ 背景音音量(取值与volume无关) ◦ 默认50 ◦ 取值范围：[0,100] ◦ bgm_begin_offset: ◦ bgm相对人声的开始位置 ◦ 默认0s ◦ 取值范围：[-60(seconds), 60(seconds)] ◦ > 0 表示bgm的开头位于 人声开头之后的秒数(先有人声) ◦ < 0 表示bgm的开头位于 人声开头之前的秒数(先有音乐)	null
result_audio_type	否	返回音频的方式，目前支持：“url”，“base64”	“url”
is_free	否	是否免费声音	true
is_audio_transcode	否	声音是否自动转码	false

- 参数示例：

```
{
  "text": "这是一段语音合成的例子，可以带背景音乐，可以调节语速、音量和音高。",
  "lang": "zh-cn",
  "ssml": false,
  "speaker": {"domain": "新闻", "speaker_name": "喜小华", "speaker_variant": "中性(旁白)"},
  "sr": 24000,
  "audio_format": "wav",
  "br": "64k",
  "bgm": {
    "url": "http://nls.alicdn.com/bgm/1.wav",
    "bgm_volume": 50,
    "bgm_begin_offset": 0,
  },
  "rate": 75.0,
  "volume": 25.0,
  "pitch": 60.0
}
```

输出参数

- 返回值说明：

状态	说明
success	是否成功(code等于2xxxxx为是)
code	参见状态码的说明
message	对应状态码的说明

状态	说明
data	◦ audio: ◦ 此处返回的是 base64 编码的音频流，或者是存放音频的 URL，由输入参数`result_audio_type`指定 ◦ 当设置返回音频流时，可通过 b64decode 解码得到音频数据，音频的格式由`audio_format`设置 ◦ 当设置返回 URL 时，可以通过访问返回的 URL 获取音频数据 ◦ subtitle: ◦ 字幕返回的是一个列表，列表中每个元素是一句话的信息。 ◦ 每句话有自己的id、文本、结束时间。时间的单位是毫秒 ◦ duration: ◦ 音频时长 ◦ upload_id: ◦ 喜马拉雅内部上传id(部分业务需要，用不到可忽略)。仅当result_audio_type=url有效。

• 同步合成接口返回结果示例：

```

{
  "success": true,
  "code": 200000,
  "message": "ok",
  "data": {
    "audio": "UklGR...AAA=",
    "subtitle": "[\n      {\n          \"id\": 0,\n          \"text\": \"这是一段语音合成的例子，可以带背景音乐，可以调节语速音量。\", \n          \"end\": 7250\n      }\n    ]",
    "duration": 0.5s,
    "upload_id": ""
  }
}

```

