**通用TTS-HTTP接口文档**

**百度智能云团队提供**

**(版权所有,翻版必究)**

**目 录**

[1 背景 3](#_Toc54108776)

[2 接口定义 3](#_Toc54108777)

[3 错误码 6](#_Toc54108778)

[4 SSML（语音合成标记语言）使用帮助 7](#_Toc54108779)

[4.1概述 7](#_Toc54108780)

[4.2定制化TTS标签说明 8](#_Toc54108781)

## 1 背景

该文档旨在为使用语音合成的客户提供接口调用说明，帮助客户理解API的功能以及如何使用语音合成服务。

## 2 接口定义

语音合成支持接口支持GET和POST两种方式。相关参数见下表

不论是将参数放到URL地址中，还是利用表单的方式放入到http body中，参数都需要经过URLENCODE操作。如果是直接在浏览器地址栏输入则不需要。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 请求 | | GET/POST http://127.0.0.1:8088/text2audio | | | |
| 描述 | | 合成 | | | |
| 请求参数说明 | | | | | |
| 参数名称 | 必选 | 类型 | 样例 | 说明 | |
| tex | 是 | string | 百度语音合成  百度人民很行(xing2) | 合成的文本，使用UTF-8编码，请注意文本长度必须小于1024字节；如果是多音字，可以在字后面加括号来标注发音。 | |
| lan | 是 | string | zh | 语言选择，填写zh | |
| pdt | 是 | uint | 993 | 填写993 | |
| ctp | 是 | uint | 1 | 填写1 | |
| cuid | 是 | string | fa:16:3c:40:38:4a | 用户唯一标识，用来区分用户，可以填写机器 MAC 地址，长度为60以内 | |
| spd | 否 | uint | 5 | 语速，取值0-15，默认为5中语速 | |
| pit | 否 | uint | 5 | 音调，取值0-15，默认为5中语调 | |
| vol | 否 | uint | 5 | 音量，取值0-15，默认为5中音量 | |
| aue | 否 | Uint | 3 | 输出格式，默认为3，MP3; 4为pcm-16k；5为pcm-8k；6为wav（内容同pcm-16k）; 注意aue=4，5，6是语音识别要求的格式，但是音频内容不是语音识别要求的自然人发音，所以识别效果会受影响。 | |
| 响应说明 | | | | | |
| 如果合成成功，下行数据为二进制语音文件   * aue=3 ，返回为二进制mp3文件，具体header信息 Content-Type: audio/mp3 * aue=4，返回为二进制pcm文件，具体header信息Content-Type:audio/basic;codec=pcm;rate=16000;channel=1 * aue=5，返回为二进制pcm文件，具体header信息Content-Type:audio/basic;codec=pcm;rate=8000;channel=1 * aue=6 ，返回为二进制wav文件，具体header信息 Content-Type: audio/wav   如果合成出现错误，则会返回json结果，具体header信息为：Content-Type: application/json。 | | | | | |
| 错误返回参数说明 | | | | | |
| 字段名称 | | 类型 | 样例 | | 说明 |
| err\_no | | int | 500 | | 参考错误码 |
| err\_msg | | string | notsupport. | | 服务调用提示信息 |
| 合成接口示例 | | | | | |
| # Get请求方式示例  # 将所有的参数都填写到URL地址中，可以通过浏览器播放合成的语音结果。  http://<tts-endpoint>/text2audio?tex=你好&lan=zh&pdt=993&ctp=1&cuid=test&per=100  # Post请求方式示例  # 将文本以及其他参数写入到body里面，利用表单的方式将参数传递到服务端。所有的参数都在body中。body里面的数据为：  tex=你好&lan=zh&pdt=993&ctp=1&cuid=test  # 错误放回示例  HTTP/1.1 200 OK  Content-Length: 51  Content-Type: application/json    {  "err\_no":500,  "err\_msg": "notsupport."  } | | | | | |

##### 2.1 iOS接入

##### 版本及依赖

Object-C版本SDK支持文字转音频后直接播放，或者返回音频文件路径，由开发者自行处理音频文件。代码可见：<https://github.com/baidubce/pie/blob/master/ios-demo/BaiduARS>

版本要求:

支持iOS 8.0+

依赖包：详见工程代码中的Podfile文件

#### demo code

1、执行pod install 安装sdk运行所需要的依赖库文件

2、通过BaiduASR.xcworkspace文件打开工程，编译运行

3、Instance初始化及参数配置。

| **参数** | **类型** | **默认值** | **描述** |
| --- | --- | --- | --- |
| per | String | / | 说话人ID |
| spd | String | 5 | 语速，取值0-15，默认中语速5 |
| pit | String | 5 | 语调，取值0-15，默认中语调5 |
| vol | Sring | 5 | 音量，取值0-15，默认中音量5 |

示例代码：

TTSConfig \*config = [TTSConfig config];

[[BDTTSInstance sharedInstance] setPer:config.per];

[[BDTTSInstance sharedInstance] setPit:config.pit];

[[BDTTSInstance sharedInstance] setSpd:config.spd];

[[BDTTSInstance sharedInstance] setVol:config.vol];

4、方法调用

直接调用BDTTSInstance类的- (**void**)ttsWithText:(NSString \*)text isPlay:(**BOOL**)isPlay complete:(**void** (^ **\_Nullable** )(NSURL \*filePath))complete failure:(**void** (^ **\_Nullable** )(NSError \*error))failure;方法。 参数释义见SDK内部.h文件释义部分，按照需求传入对应的参数即可。

示例代码：

[[BDTTSInstance sharedInstance] ttsWithText:text isPlay:**YES** complete:^(NSURL \* **\_Nonnull** filePath) {

// 调用结束，调用方自行处理逻辑

} failure:^(NSError \* **\_Nonnull** error) {

// 调用失败，调用方自行处理逻辑

}];

## 3 错误码

合成接口错误码如下表所示：

|  |  |  |
| --- | --- | --- |
| 错误码 | 说明 | 备注 |
| 500 | 不支持输入 |  |
| 501 | 输入参数不正确 |  |
| 503 | 合成后端错误 |  |

**注意：tex 字段 2 次 urlencode**

由于urlencode有两个标准 RFC 1738和RFC 3986. 百度为了更好地兼容，支持1次及2次urlencode， 其中2次urlencode可以覆盖全部的特殊字符。因而推荐传递tex 参数时做2次urlencode编码。测试用例：“1+1=2”。 一次urlencode时，“+”可能会没有合成。

## 4 SSML（语音合成标记语言）使用帮助

### 4.1概述

如需使用SSML功能，请求TTS服务时，需设置参数xml=1

#### 4.1.1 测试环境

*内网试用页面：<http://insight.speech.baidu.com/speechinsight/tts>  
地址栏填写测试地址：<http://tts.baidu.com/text2audio?xml=1>  
文本框中填写需要验证SSML文本。*

#### 4.1.2 定制化TTS

订制化TTS标签：通过对合成文本加入标签的方式定制要实现的合成效果；

#### 4.1.3 SSML支持计划

* [say-as](http://wiki.baidu.com/pages/viewpage.action?pageId=329463001#say-as)标签，并扩展加入英文单词是按词读还是按字母读；
* 增加背景音标签：[background](http://wiki.baidu.com/pages/viewpage.action?pageId=329463001" \l "background)；
* [audio](http://wiki.baidu.com/pages/viewpage.action?pageId=329463001#audio)标签
* 使用限制
* audio标签不支持嵌套audio/background（内层标签不生效）；
* background/标签不支持嵌套自身，内层标签不生效；
* sub/say-as标签不支持嵌套其他任何标签，会引发解析错误，导致标签按字母朗读；
* 非汉字文本内部出现标签影响语义转换，建议使用汉字形式请求；
* &和<符号在XML中为非法字符，使用前需进行转义操作，分别为&amp;和&lt;，另外，>(&gt;)、”(&quot;)、’(&apos;)也建议使用前进行转义
* 请求文本总长度（包括SSML标签）应少于等于1024字节

### 4.2定制化TTS标签说明

#### 4.2.1 基础标签

1. **speak**

* 标签: speak
* 解释:

*设置文本层级， speak为ssml标注文本的边界；*

* **speak为强制标签，ssml标注文本必须要加上speak边界；**

1. **audio**

* 标签: audio
* 解释:

*在合成中引入外部的音频，音频以服务器可以访问的的地址给出，目前支持16K采样和24K采样,16bit,单声道，44字节头的wave格式文件；  
目前使用audio标签，要求用户必须将相应的音频文件上传到百度云bos平台，在使用过程中提供bos地址；  
单个请求最大限制3个并列的audio资源，单个audio资源大小限制为3M；  
使用前需要前把音频转换为支持的格式，推荐使用ffmpeg。命令参考如下（background标签等同）：****使用示例：ffmpeg -i <input-file> -acodec pcm\_s16le -b:a 16k -ar 16000 -ac 1 -flags bitexact <output-file.wav>***

***audio标签不支持嵌套audio/background****；  
audio标签支持单标签和双标签；如果是双标签的话，当音频无法访问时，将合成嵌套文本；*

| **属性** | **值** |
| --- | --- |
| src | * 音频地址 |

1. <speak>
2. 播放周杰伦的听妈妈的话。<audio src="http://tis-audio.bj.bcebos.com/tingmamadehua.wav?authorization=bce-auth-v1%2F96d3a871ef2a41ae92abc41ae6acf932%2F2018-03-05T02%3A31%3A13Z%2F-1%2Fhost%2F792bff900a458fd52c07893f27957263a3b2719ee38e1a79212bd980ddd80156"/> 播放完了，好听吗？
3. </speak>
4. **say as**

* 标签: say-as
* 解释:

*设置数字、符号等的读法， 属性type；*

| **属性值** | **功能支持** | **使用对象** |
| --- | --- | --- |
| number:ordinal | 整数读法 | 纯阿拉伯数字串、带有英文逗号的纯阿拉伯数字串，数字个数（英文逗号不计数）小于13 |
| number:digits | 数字串读法 | 纯阿拉伯数字串 |
| number:score | 比分读法 | 带有一个冒号或中划线-表示的比分读法，冒号或中划线-表前是纯阿拉伯数字串、带有英文逗号的纯阿拉伯数字串，冒号或中划线-表后也是 |
| number:fraction | 分数读法 |  |
| telephone | 电话号码读法 |  |
| number | * 整数读法 * 负整数读法 * 浮点数读法 * 负浮点数读法 * 比分读法 * 分数读法 * 负分数读法 * 科学计数读法 |  |

* 使用示例：

| **加标签示例** | **处理后文本** |
| --- | --- |
| <say-as type="number:ordinal">123</say-as> | 一百二十三 |
| <say-as type="number:digits">123</say-as> | 一二三 |
| <say-as type="number:score">12:123</say-as> | 十二比一百二十三 |
| <say-as type="number:fraction">12/123</say-as> | 一百二十三分之十二 |
| <say-as type="telephone">110</say-as> | 幺幺零 |
| <say-as type="number">123</say-as> | 一百二十三 |
| <say-as type="number">-123</say-as> | 负一百二十三 |
| <say-as type="number">123.45</say-as> | 一百二十三点四五 |
| <say-as type="number">-123.45</say-as> | 负一百二十三点四五 |
| <say-as type="number">123:45</say-as> | 一百二十三比四十五 |
| <say-as type="number">123/45</say-as> | 四十五分之一百二十三 |
| <say-as type="number">-123/45</say-as> | 负四十五分之一百二十三 |
| <say-as type="number">3e+11</say-as> | 三乘以十的十一次幂 |
| <say-as type="number">3e-11</say-as> | 三乘以十的负十一次幂 |
| <say-as type="number">-3e+11</say-as> | 负三乘以十的十一次幂 |
| <say-as type="number">-3e-11</say-as> | 负三乘以十的负十一次幂 |
| <say-as type="number">1.23e+11</say-as> | 一点二三乘以十的十一次幂 |
| <say-as type="number">1.23e-11</say-as> | 一点二三乘以十的负十一次幂 |
| <say-as type="number">-1.23e+11</say-as> | 负一点二三乘以十的十一次幂 |
| <say-as type="number">-1.23e-11</say-as> | 负一点二三乘以十的负十一次幂 |
| <say-as type="number">3E+07</say-as> | 三乘以十的七次幂 |
| <say-as type="number">3E-07</say-as> | 三乘以十的负七次幂 |
| <say-as type="number">-3E+07</say-as> | 负三乘以十的七次幂 |
| <say-as type="number">-3E-07</say-as> | 负三乘以十的负七次幂 |
| <say-as type="number">1.23E+07</say-as> | 一点二三乘以十的七次幂 |
| <say-as type="number">1.23E-07</say-as> | 一点二三乘以十的负七次幂 |
| <say-as type="number">-1.23E+07</say-as> | 负一点二三乘以十的七次幂 |
| <say-as type="number">-1.23E-07</say-as> | 负一点二三乘以十的负七次幂 |

1. **sub**

* 标签: sub
* 解释:

*替换，属性alias；*

* 使用限制:

*alias值可包含中文或者字母，不支持包含数字；*

* 使用示例：
  1. <speak>
  2. 百度的下一个使命聚焦在<sub alias="人工智能">AI</sub>上。
  3. </speak>

1. **silence**

* 标签：silence
* 解释：

*设置静音，在文本播报的开头或者结尾增加静音片段;单位支持ms、s两个级别，最大值为10s，如果设置超过10s的统一按照10s处理；*

* 使用实例：
  1. <speak>
  2. <silence time="5s"></silence> 测试一下
  3. </speak>

1. **phoneme**

* 标签: phoneme
* 解释:

*多音字注音，属性py；*

* 使用示例：
  1. <speak>
  2. 迎着<phoneme py="zhao1">朝</phoneme>阳的<phoneme py="chao2">朝</phoneme>阳群众
  3. </speak>

#### 4.2.2 扩展标签

1. **background**

* 标签: background
* 解释:

*设置背景声，和audio标签的区别是background是与合成声音同时播报，而audio是插入在合成声音中间的，目前仅支持16K采样,16bit,单声道，44字节头的wave格式文件；  
目前使用background标签，要求用户必须将相应的音频文件上传到百度云bos平台，在使用过程中提供bos地址；****background标签不支持嵌套background****；*

| **属性** | **值** |
| --- | --- |
| src | * 音频地址 |

* 使用示例：
  1. <speak>
  2. <background src="http://tis-audio.bj.bcebos.com/lantern\_festival.wav?authorization=bce-auth-v1%2F96d3a871ef2a41ae92abc41ae6acf932%2F2018-03-05T02%3A31%3A01Z%2F-1%2Fhost%2F40a27013eb5244b349301dc1775747ccf3254a1e65b29a6109f38a4479f4cc29" repeat="yes">欢迎使用百度语音合成服务</background>。
  3. </speak>

1. **say-as**

* 标签: say-as
* 解释:

*say-as属性interpret-as加入两个新值，仅对英文有效；*

| **属性** | **值** |
| --- | --- |
| interpret-as | * as-character : 按字母读 * as-word： 按词读（暂不生效） |

* 使用示例：
  1. <speak>
  2. 美国宇航局简称读作<say-as interpret-as="as-word">NASA</say-as>还是<say-as interpret-as="as-character">NASA</say-as>。
  3. </speak>

1. **poem**

* 标签: poem
* 解释:

*设置诗词，属性name；*

| **属性** | **值** |
| --- | --- |
| poem | * wuyan：五言诗 * qiyan：七言诗 * songci：宋词 |

* 使用示例：
  1. <speak>
  2. 背一首唐诗吧，<poem name="wuyan">床前明月光，疑是地上霜。举头望明月，低头思故乡。</poem>
  3. </speak>

1. **space**

* 标签: space
* 解释:

*通过设置属性type，在所包含文本的空格处生成停顿；*

| **属性** | **值** |
| --- | --- |
| type | * title：在标题空格处生成停顿 |

* 使用示例：
  1. <speak>
  2. <space type="title">20余人花百万买来的北京户口被注销 涉事民警被判16年</space>
  3. </speak>