本文档强烈推荐使用Web版式视图阅读

在菜单栏点击“视图”🡪“Web版式视图”完成设置

1.ASR REST API

1.1.简介

百度语音识别通过 REST API 的方式给开发者提供一个通用的 HTTP 接口。 上传需要完整的录音文件，录音文件时长不超过60s。

1.2.语种

普通话、英语、粤语、四川话

1.3.识别模型

百度语音提供2种识别模型

* 搜索模型： 效果同手机百度搜索的语音输入。适合于短语识别，没有逗号。
* 输入法模型：效果同百度输入法的语音输入。适合于长句识别，有逗号。

普通话搜索模型同时能识别简单的常用英语语句，效果同手机百度。

1.4.适用范围

任意操作系统，任意编程语言，只要可以对百度语音服务器发起http请求的，均可以使用本接口。 示例代码见：https://github.com/Baidu-AIP/SPEECH-TTS-CORS 浏览器由于无法跨域请求百度语音服务器的域名，因此无法使用本接口。

1.5.语音格式

格式支持：pcm（不压缩）、wav（不压缩，pcm编码）、amr（压缩格式）。推荐pcm 采样率 ：16000 固定值。 编码：16bit 位深的单声道。

百度服务端会将非pcm格式，转为pcm格式，因此使用wav、amr会有额外的转换耗时。

* [16k 采样率pcm文件样例下载](http://speech-doc.gz.bcebos.com/rest-api-asr/public_audio/16k.pcm)
* [16k 采样率wav文件样例下载](http://speech-doc.gz.bcebos.com/rest-api-asr/public_audio/16k.wav)
* [16k 采样率amr文件样例下载](http://speech-doc.gz.bcebos.com/rest-api-asr/public_audio/16k-23850.amr)

音频文件格式转换请参见文档[【语音识别小工具\音频文件转码】](http://ai.baidu.com/docs#/ASR-Tool-convert/top)

1.6.自定义词库

自定义词库在您网页申请的应用内设置（具体位置参见下图）。 

自定义词库适合短句，保证词库中一模一样的短句可以被识别出，词库中的分词优先级较高。 自定义词库仅对dev\_pid = 1536生效，并且原始音频的采用率为16K。

最好在1万行以内。

副作用：如果用户的测试集中包含大量非自定义词表的query，整体上准确率下降。

1.6.1.举例：

词库定义了1个短句： 1 . 摆渡船来了 百度内部处理的可能的分词结果： 摆渡船 来 了

以下录音的结果

1. 原始音频：摆渡船来了 =>识别结果： 摆渡船来了 【保证结果】
2. 原始音频：摆渡船来了么 =>识别结果： 百度传来了么 【可能结果，不保证】
3. 原始音频：摆渡船来 => 识别结果： 百度传来 【可能结果，不保证】
4. 原始音频：百度传来了喜讯 => 识别结果： 摆渡船传来了喜讯 【不保证，词库内的分词优先级高】

1.7.SDK

目前对识别和合成的REST API，均封装了 Java、Python、PHP、C#、NodeJs、C++ 共6种开发语言的SDK。功能等同于REST API。SDK中 识别使用JSON方式提交本地文件。

2.调用流程

2.1.流程

1. 使用appKey secretKey 访问 https://openapi.baidu.com 换取 token ，详细见“鉴权认证机制”
2. 选择一种HTTP POST 请求格式，参见下一节 ”请求方式“
3. 填写您的参数 ，详细见 ”API请求方式及参数基本说明“

2.2.请求方式

如果您的音频在本地，需要将音频数据放在body中。（推荐方式） 音频在本地，有JSON和raw两种方式提交。这两种提交方式，均不是浏览器表单的提交

2.2.1.json 方式，上传本地文件

* 音频文件，读取二进制内容后， base64 放在speech参数内。
* 音频文件的原始大小, 即二进制内容的字节数，填写“len”字段

由于使用json格式， header为：

Content-Type:application/json

注意 由于base64编码后，数据会增大1/3。

2.2.2.raw方式，上传本地文件

* 音频文件，读取二进制内容后，直接放在body中。
* Content-Length的值即为音频文件的大小。（一般代码会自动生成）。

由于使用raw方式， 采样率和文件格式需要填写在Content-Type中

Content-Type: audio/pcm;rate=16000

2.3.测试demo

您可以下载我们的官方rest api示例查看。 测试demo请在申请应用后，从<http://ai.baidu.com/sdk>上下载

* 语音REST API示例代码 :: 一些简单的demo 有JAVA PHP 和C++
* RestApi SDK 下载 :: rest api 的完整封装，有 JAVA PHP Python Nodejs 和 C# 语言。使用JSON方式上传本地音频文件。**同rest api的功能一致，没有额外功能。**

3.鉴权认证机制

3.1.获取 Access Token

获取AccessToken 需要您在应用管理界面中新建应用，应用列表中即可查看。 开放平台上： 

使用语音识别及合成REST API 需要获取 Access Token。Access Token 是用户身份验证和授权的凭证，语音识别采用的是Client Credentials授权方式，即采用应用公钥（Api Key）、密钥获取Access Token，适用于任何带server类型应用，通过此授权方式获取Access Token仅可访问平台授权类的接口，

使用Client Credentials获取Access Token需要应用在其服务端发送请求（推荐用POST方法）到百度OAuth2.0授权服务的“ https://openapi.baidu.com/oauth/2.0/token ”地址上，并带上以下参数：

* grant\_type：必须参数，固定为“client\_credentials”；
* client\_id：必须参数，应用的 API Key；
* client\_secret：必须参数，应用的 Secret Key;

\*\* 例如：\*\*

https://openapi.baidu.com/oauth/2.0/token?grant\_type=client\_credentials&client\_id=Va5yQRHl\*\*\*\*\*\*\*\*LT0vuXV4&client\_secret= 0rDSjzQ20XUj5i\*\*\*\*\*\*\*\*PQSzr5pVw2&

响应数据包如下所示，其中 “access\_token” 字段即为请求 REST API 所需的令牌, 默认情况下，Access Token 有效期为一个月，开发者需要对 Access Token的有效性进行判断，如果Access Token过期可以重新获取。

**例如：**

HTTP/1.1 200 OK

Content-Type: application/json

Cache-Control: no-store

{

"access\_token": "1.a6b7dbd428f731035f771b8d\*\*\*\*\*\*\*\*.86400.1292922000-2346678-124328",

"expires\_in": 86400,

"refresh\_token": "2.385d55f8615fdfd9edb7c4b\*\*\*\*\*\*\*\*.604800.1293440400-2346678-124328",

"scope": "public",

"session\_key": "ANXxSNjwQDugf8615Onqeik\*\*\*\*\*\*\*\*CdlLxn",

"session\_secret": "248APxvxjCZ0VEC\*\*\*\*\*\*\*\*aK4oZExMB",

}

具体代码示例可以参见：http://ai.baidu.com/docs#/Auth/top。 注意语音服务的调用地址是https://openapi.baidu.com/oauth/2.0/token

4.参数说明

* 格式支持：pcm（不压缩）、wav（不压缩，pcm编码）、amr（压缩格式）；固定16k 采样率；
* 系统支持语言种类 普通话

4.1.识别语言及模型选择

dev\_pid 参数列表

| dev\_pid | 语言 | 模型 | 是否有标点 | 备注 |
| --- | --- | --- | --- | --- |
| 1536 | 普通话(支持简单的英文识别) | 搜索模型 | 无标点 | 支持自定义词库 |
| 1537 | 普通话(纯中文识别) | 输入法模型 | 有标点 | 不支持自定义词库 |
| 1737 | 英语 |  | 有标点 | 不支持自定义词库 |
| 1637 | 粤语 |  | 有标点 | 不支持自定义词库 |
| 1837 | 四川话 |  | 有标点 | 不支持自定义词库 |
| 1936 | 普通话远场 | 远场模型 | 有标点 | 不支持 |

4.2.简介

目前 API 仅支持**整段语音**识别的模式，即需要上传完整语音文件进行识别。文件大小不超过10M，时长不超过60s。 语音数据上传POST方式有2种：

1. JSON格式POST上传本地文件。
2. raw格式POST上传本地文件

正式地址：http://vop.baidu.com/server\_api 或 https://vop.baidu.com/server\_api

4.3.文件样例

推荐16K采样率 pcm文件

* [16k 采样率pcm文件样例下载](http://speech-doc.gz.bcebos.com/rest-api-asr/public_audio/16k.pcm)
* [16k 采样率wav文件样例下载](http://speech-doc.gz.bcebos.com/rest-api-asr/public_audio/16k.wav)
* [16k 采样率amr文件样例下载](http://speech-doc.gz.bcebos.com/rest-api-asr/public_audio/16k-23850.amr)

4.4.JSON方式上传

语音数据和其他参数通过标准 JSON 格式串行化 POST 上传， JSON 里包括的参数：

| 字段名 | 类型 | 可需 | 描述 |
| --- | --- | --- | --- |
| format | string | 必填 | 语音文件的格式，pcm 或者 wav 或者 amr。不区分大小写。推荐pcm文件 |
| rate | string | 必填 | 采样率，16000，固定值 |
| channel | string | 必填 | 声道数，仅支持单声道，请填写固定值 1 |
| cuid | int | 必填 | 用户唯一标识，用来区分用户，计算UV值。建议填写能区分用户的机器 MAC 地址或 IMEI 码，长度为60字符以内。 |
| token | string | 必填 | 开放平台获取到的开发者[[access\_token]](http://ai.baidu.com/docs#/Auth/top)获取 Access Token "access\_token") |
| dev\_pid | int | 选填 | 不填写lan参数生效，都不填写，默认1537（普通话 输入法模型），dev\_pid参数见本节开头的表格 |
| lan | string | 选填，废弃参数 | 历史兼容参数，请使用dev\_pid。如果dev\_pid填写，该参数会被覆盖。语种选择,输入法模型，默认中文（zh）。 中文=zh、粤语=ct、英文=en，不区分大小写。 |
| speech | string | 选填 | 本地语音文件的的二进制语音数据 ，需要进行base64 编码。与len参数连一起使用。 |
| len | int | 选填 | 本地语音文件的的字节数，单位字节 |

* (speech, len)： 开发者可以将语音文件进行 base64编码，放在 “speech”字段中。并将语音数据的原始长度，填写“len”字段；

4.4.1.上传示例(speech, len 参数)

即：JSON格式POST上传本地文件

固定头部header

Content-Type:application/json

请求示例

4K大小的pcm文件（普通话录音）请求： speech 参数填写为 文件内容base64后的结果：

{

"format":"pcm",

"rate":16000,

"dev\_pid":1536,

"channel":1,

"token":xxx,

"cuid":"baidu\_workshop",

"len":4096,

"speech":"xxx", // xxx为 base64（FILE\_CONTENT）

}

返回示例

{"corpus\_no":"6433214037620997779","err\_msg":"success.","err\_no":0,"result":["北京科技馆，"],"sn":"371191073711497849365"}

注意事项

len 字段表示原始语音大小字节数，不是 base64 编码之后的长度。

4.5.Raw 方式上传

即 raw格式POST上传本地文件 语音数据直接放在 HTTP BODY 中，控制参数以及相关统计信息通过 header和url里参数传递。

4.5.1.Header 参数说明

| 字段名 | 数据类型 | 可需 | 描述 |
| --- | --- | --- | --- |
| format | string（格式见下面示例） | 必填 | 语音格式，pcm 或者 wav 或者 amr。不区分大小写，推荐使用pcm文件 |
| rate | int（格式见下面示例） | 必填 | 采样率 16000， 固定值 |

语音数据的采样率和压缩格式在 HTTP-HEADER 里的Content-Type 表明，例：

Content-Type: audio/pcm;rate=16000

4.5.2.url参数说明

| 字段名 | 可需 | 描述 |
| --- | --- | --- |
| cuid | 必填 | 用户唯一标识，用来区分用户，计算UV值。建议填写能区分用户的机器 MAC 地址或 IMEI 码，长度为60字符以内。 |
| token | 必填 | 开放平台获取到的开发者[[access\_token]](http://ai.baidu.com/docs#/Auth/top)获取 Access Token "access\_token") |
| dev\_pid | 选填 | 不填写lan参数生效，都不填写，默认1537（普通话 输入法模型），dev\_pid参数见本节开头的表格 |
| lan | 选填，废弃参数 | 历史兼容参数，请使用dev\_pid。如果dev\_pid填写，该参数会被覆盖。语种选择,输入法模型，默认中文（zh）。 中文=zh、粤语=ct、英文=en，不区分大小写。 |

URL 示例：

http://vop.baidu.com/server*\_api?dev\_*pid=1536&cuid=**\*\*\*\*\****\*&token=1.a6b7dbd428f731035f771b8d\****\*\*\*\*\***\*\*.86400.1292922000-2346678-124328

4.6.接口返回参数

两种上传方式都返回统一的结果，采用 JSON 格式封装，如果识别成功，识别结果放在 JSON的“result”字段中，统一采用 utf-8 方式编码。 （如果使用POST方式的（url,callback）方式，百度服务器会回调用户服务器的callback地址, 返回如下结果）

| 字段名 | 数据类型 | 可需 | 描述 |
| --- | --- | --- | --- |
| err\_no | int | 必填 | 错误码 |
| err\_msg | string | 必填 | 错误码描述 |
| sn | string | 必填 | 语音数据唯一标识，系统内部产生。如果反馈及debug请提供sn。 |
| result | array ( [string,string,…]) | 选填 | 识别结果数组，提供1-5 个候选结果， 优先使用第一个结果。utf-8 编码。 |

4.6.1.识别成功返回 case

{"err\_no":0,"err\_msg":"success.","corpus\_no":"15984125203285346378","sn":"481D633F-73BA-726F-49EF-8659ACCC2F3D","result":["北京天气"]}

4.6.2.识别错误返回 case

{"err\_no":2000,"err\_msg":"data empty.","sn":"481D633F-73BA-726F-49EF-8659ACCC2F3D"}

4.6.3.raw 方式测试说明

curl -i -X POST -H "Content-Type: audio/pcm;rate=16000" "http://vop.baidu.com/server\_api?dev\_pid=1536&cuid=xxxxx&token=1.a6b7dbd428f731035f771b8d\*\*\*\*\*\*\*\*.86400.1292922000-2346678-124328" --data-binary "@/home/test/test.pcm"

4.7.错误码解释

见错误码及常见原因部分

8.错误码及常见原因

8.1.错误码列表

| 错误码 | 用户输入/服务端 | 含义 | 一般解决方法 |
| --- | --- | --- | --- |
| 3300 | 用户输入错误 | 输入参数不正确 | 请仔细核对文档及参照demo，核对输入参数 |
| 3301 | 用户输入错误 | 音频质量过差 | 请上传清晰的音频 |
| 3302 | 用户输入错误 | 鉴权失败 | token字段校验失败。请用appkey 和 app secret生成 |
| 3303 | 原始音频或者服务端问题 | 转pcm失败 | 如下2个原因 1. wav amr音频转码失败，即音频有问题。2. 服务端问题，请将api返回结果反馈至论坛或者QQ群 |
| 3304 | 用户请求超限 | 用户的请求QPS超限 | 请降低识别api请求频率 （qps以appId计算，移动端如果共用则累计） |
| 3305 | 用户请求超限 | 用户的日pv（日请求量）超限 | 请“申请提高配额”，如果暂未通过，请降低日请求量 |
| 3307 | 服务端问题 | 语音服务器后端识别出错问题 | 请将api返回结果反馈至论坛或者QQ群 |
| 3308 | 用户输入错误 | 音频过长 | 音频时长不超过60s，请将音频时长截取为60s以下 |
| 3309 | 用户输入错误 | 音频数据问题 | 服务端无法将音频转为pcm格式，可能是长度问题，音频格式问题等。 请将输入的音频时长截取为60s以下，并核对下音频的编码，是否是8K或者16K， 16bits，单声道。 |
| 3310 | 用户输入错误 | 输入的音频文件过大 | 语音文件共有3种输入方式： json 里的speech 参数（base64后）； 直接post 二进制数据，及callback参数里url。 分别对应三种情况：json超过10M；直接post的语音文件超过10M；callback里回调url的音频文件超过10M |
| 3311 | 用户输入错误 | 采样率rate参数不在选项里 | 目前rate参数仅提供8000,16000两种，填写4000即会有此错误 |
| 3312 | 用户输入错误 | 音频格式format参数不在选项里 | 目前格式仅仅支持pcm，wav或amr，如填写mp3即会有此错误 |

8.2.错误码常见问题及具体分析

8.2.1.3300 错误

语音识别api使用的是HTTP POST方法， BODY里直接放置json， Content-Type头部为 application/json。 并非常见的浏览器表单请求（application/x-www-form-urlencoded或者multipart/x-www-form-urlencoded）。

必填字段：format rate channel 请勿漏填。此外 (speech, len) 及 (url, callback) 这两组参数必须二选一，如果都填，默认处理第一组。并确认 音频时长截取为60s以下。

8.2.2.3309错误

wav和amr的音频，服务端会自动转为pcm，这个过程中导致转码出错。请确认下format及rate参数与音频一致，并确认音频时长截取为60s以下。

8.2.3.3301 错误

识别结果实际为空。可能是音频质量过差，不清晰，或者是空白音频。 有时也可能是pcm填错采样率。如16K采样率的pcm文件，填写的rate参数为8000。

8.2.4.错误反馈

* 结果含有错误码：请提供： 1. 原始音频 2. 返回的完整json 3. 调用的时间点 4. 识别的参数
* 识别结果与期望不符： 请提供 ： 1.原始音频 2. 返回的完整json 3.期望结果 4. 识别的参数 5. 是否偶发
* 调用官方的rest sdk demo报错： 如果是rest sdk的报错，请提供完整报错信息。如果是上述两项，请按上面的说明。