

捷通华声
灵云公有云能力平台
HTTP 开发手册-VPR

灵云科技 源自清华 服务全球

北京捷通华声科技股份有限公司

Beijing SinoVoice Technology Co., Ltd.

2015 年 12 月

目 录

1. 概述.....	1
2. HTTP 头.....	1
3. VPR 访问接口.....	2
3.1. VPR 注册请求接口.....	2
3.1.1. 访问.....	2
3.1.2. 服务器应答.....	2
3.1.3. 示例.....	2
3.1.4. x-task-config 参数说明.....	3
3.2. VPR 判别请求接口.....	4
3.2.1. 访问.....	5
3.2.2. 服务器应答.....	5
3.2.3. 示例.....	5
3.2.4. x-task-config 参数说明.....	6
3.3. VPR 识别请求接口.....	8
3.2.5. 访问.....	8
3.2.6. 服务器应答.....	8
3.2.7. 示例.....	9
3.2.8. x-task-config 参数说明.....	9

1. 概述

灵云HTTP-VPR接口，客户端通过HTTP-VPR接口使用POST方式与灵云平台进行通讯，返回数据为XML格式。

注意：传入的userid和groupid中如果有空格会被过滤掉。

2. HTTP 头

客户端POST请求时HTTP头信息，必须带上如下：

顺序	名称	值（例子）	含义	备注
1	x-app-key	0c5d5411	应用标识	必选，由捷通公司提供
2	x-sdk-version	5.1	sdk 版本号	必选，使用 http 接口时，可以指定为 5.1
3	x-request-date	2014-6-18 10:10:11	请求时间	必选，每个任务的请求时间不能重复
4	x-task-config	capkey=vpr.cloud, audioformat=alaw8k8b it	任务参数信息	必选，为 name=value 形式，多个参数以逗号隔开 详情请参考后面的 参数解释
5	x-session-key	18EA9384AA34	请求数据签名	必选 x-session-key 生成算法说明： x-session-key = md5(x-request-date + devkey) devkey 是应用开发者密钥，在注册开发者用户时由开发者网站统一分配，用作请求数据签名；devkey 在私有云能力平台中目前固定为 developer_key
6	x-udid	101:1234567890	设备标识	必选，如使用设备取设备标识号，如不使用设备设置为例子中的默认值

注：头的内容必须为英文，仅支持utf8字符编码。

devkey是应用开发者密钥，在注册开发者用户时由开发者网站统一分配，用作请求数据签名；

3. VPR 访问接口

3.1. VPR 注册请求接口

本接口在提交音频数据时，同步在 http 请求的回应消息中返回模型注册的结果。

3.1.1. 访问

(POST) [ServiceUrl]/vpr/register

说明：包体中传输音频数据，单次传输最大为 8MB（未经特殊说明，音频数据都应为单声道的）。

需要以下 HTTP 头：x-app-key、x-sdk-version、x-request-date、x-task-config、x-session-key、x-udid。

3.1.2. 服务器应答

应答格式：XML 文件

XML 的元素（Element）及含义

顺序	名称	值（例子）	含义	备注
1	ResCode	Success/Failed	完毕/处理失败	必选
2	ResMessage	Success/Failed	完毕/处理失败	必选
3	ErrorNo	0	返回错误码	必选
4	ResultToken	0E01123456	结果令牌	成功时必选
5	UserId	123456	用户 ID	第一次注册成功，必选。

3.1.3. 示例

中间结果：

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<ResponseInfo>
```

```
<ResCode>InProgress</ResCode>
```

```
<ResMessage>expect more chunk</ResMessage>

<ErrorNo>7</ErrorNo>

<ResultToken>303_72_90_43398_20150205204125_1</ResultToken>

</ResponseInfo>

最终结果:

<?xml version="1.0" encoding="utf-8"?>

<ResponseInfo>

  <ResCode>Success</ResCode>

  <ResMessage>Success</ResMessage>

  <ErrorNo>0</ErrorNo>

  <ResultToken>303_72_90_43398_20150205204125_1</ResultToken>

  <Result>

    <UserId>userid_303_72_90_43398_20150205204125_1</UserId>

  </Result>

</ResponseInfo>
```

3.1.4. x-task-config 参数说明

以下是有效参数:

✧ **capkey** (声纹识别能力名称, 必选)

Capkey	类型	支持 audioformat
vpr.cloud.recog	8k16bit 模型, 主要应用于电话信道 16k16bit 模型, 只支持 16k 格式音频	alaw8k8bit, ulaw8k8bit, vox6k4bit ,pcm8k16bit, alaw16k8bit, ulaw16k8bit, pcm16k16bit

✧ **userid (用户ID):** 由数字和字母组成的字符串, 长度不超过64个字节。同一组内唯一。

注册时, 用户可以不设置, 服务会生成userid, 返回给用户; 如果用户设置了, 服务会使用用户设置的userid, 不过, 用户必须保证userid的唯一性。注册成功后, 模型训练, 必须设置。

✧ **groupid (组 ID, 可选):** 组的标示符, 对同一应用 (develop_key) 唯一。

通过用户管理接口添加的组 id, 对应有多个用户 id。传递 groupId (该 group 不存在则创建), 就将 user_id 添加到该组中; 不传递的话, 则只是创建 user_id, 添加模型, 不添加组与用户之间的关系。

注意: 在此处创建的组为独享组, 不能与同一 develop_key 下的其他 app 共享。如果需要和其他 app 共享, 请使用《捷通华声灵云公有云能力平台 HTTP 开发手册-用户管理(通用版).doc》中的添加组接口。

例如: develop A 下面有 app B 和 app C。如果 B 通过该接口创建了组, 那么 C 将不能对这个组进行操作 (删除组、向组内添加用户、删除组内用户), 也不能创建一个同名的组。

✧ **audioformat (音频格式, 必选)**

有效参数	参数说明
alaw 8k8bit	8k 8bit a-law 数据
ulaw8k8bit	8k 8bit u-law 数据
pcm8k16bit	8k 16bit PCM 音频数据
vox6k4bit	6k4bit vox 音频数据
alaw16k8bit	16k 8bit a-law 数据
ulaw16k8bit	16k 8bit u-law 数据
pcm16k16bit	16k 16bit PCM 音频数据

✧ **identify(会话标识, 流式音频必选):**

该字段为会话的标识, 和 x-udid 字段一起作为一次会话的唯一标识。在不同会话之间此唯一标识不可相同 (即不同识别任务的[x-udid+identify]不同)。若采用非流式音频方式, 则此参数可不设置。一次会话即指一次完整的 verify 请求、获得结果的过程。流式音频指一次 verify 过程会分段传入音频。非流式音频是指一次传入完整的音频。

✧ **index(识别序号, 流式音频必选):**

某次会话的请求序号, 从 1 开始, 最后一次的序号以负数作为终结。可能的序列号举例:

- -1 此时请求数据只有一块
- 1, 2, -3 此时请求数据有三块, 此时只有最后一块会返回处理结果, 其余均返回 ResCode 为 InProgress。

若采用非流式音频, 此参数为可选; 但如果设置此值, 则必须是-1。

✧ **property (模型参数, 必选, 默认为16k_common)**

有效参数	参数说明
8k_common	8k16bit 模型, 主要应用于电话信道, 支持

	audioformat 为 8k 的情况
16k_common	16k16bit 模型 (默认), 支持 audioformat 为 16k 的情况

上述 8k16bit 及 16k16bit 指的是传入引擎的 PCM 音频格式, 和 audioformat 无关。设置错误会返回 CommitRequest jetcl_commit_request failed。

3.2. VPR 判别请求接口

本接口在提交音频数据和 **userid** 时, 同步在 http 请求的回应消息中返回音频与 **userid** 对应模型的匹配结果。

3.2.1. 访问

(POST) [ServiceUrl]/vpr/verify

说明: 包体中传输音频数据, 单次传输包体最大 8MB。

需要以下 HTTP 头: x-app-key、x-sdk-version、x-request-date、x-task-config、x-session-key、x-udid。

3.2.2. 服务器应答

应答格式: XML文件

XML的元素 (Element) 及含义

顺序	名称	值 (例子)	含义	备注
1	ResCode	Success/Failed	完毕/处理失败	必选
2	ResMessage	Success/Failed	完毕/处理失败	必选
3	ErrorNo	4	返回错误码	必选
4	ResultToken	0E01123456	结果令牌	成功时必选
5	VerifyResult	YES/NO	声纹与模型匹配结果	成功必选
6	Score	90	声纹与模型相似度	成功必选, VerifyResult 的属性

3.2.3. 示例

中间结果:

```
<?xml version="1.0" encoding="utf-8"?>

<ResponseInfo>

  <ResCode>InProgress</ResCode>

  <ResMessage>expect more chunk</ResMessage>

  <ErrorNo>7</ErrorNo>

  <ResultToken>303_72_90_43398_20150205204125_1</ResultToken>

</ResponseInfo>
```

最终结果:

```
<?xml version="1.0" encoding="utf-8" ?>

<ResponseInfo>

  <ResCode> Success </ResCode>

  <ResMessage> Success </ResMessage>

  <ErrorNo>0</ErrorNo>

  <ResultToken>303_72_90_43398_20150205204125_1</ResultToken>

  <Result>

    <VerifyResult Score="99">YES</VerifyResult>

  </Result>

</ResponseInfo>
```

3.2.4. x-task-config 参数说明

以下是有效参数:

✧ capkey (声纹识别能力名称, 必选)

Capkey	类型	支持 audioformat
vpr.cloud.recog	8k16bit 模型, 主要应用于电话信道 16k16bit 模型, 只支持 16k 格式音频	alaw8k8bit, ulaw8k8bit, vox6k4bit , pcm8k16bit, alaw16k8bit, ulaw16k8bit, pcm16k16bit

✧ userid (用户ID, 必选): 注册成功时, 返回的userid。

✧ audioformat (音频格式, 必选)

有效参数	参数说明
------	------

alaw 8k8bit	8k 8bit a-law 数据
ulaw8k8bit	8k 8bit u-law 数据
pcm8k16bit	8k 16bit PCM 音频数据
vox6k4bit	6k4bit vox 音频数据
alaw16k8bit	16k 8bit a-law 数据
ulaw16k8bit	16k 8bit u-law 数据
pcm16k16bit	16k 16bit PCM 音频数据

✧ **identify(会话标识, 流式音频必选):**

该字段为会话的标识, 和 x-udid 字段一起作为一次会话的唯一标识。在不同会话之间此唯一标识不可相同 (即不同识别任务的[x-udid+identify]不同)。若采用非流式音频方式, 则此参数可不设置。一次会话即指一次完整的 verify 请求、获得结果的过程。流式音频指一次 verify 过程会分段传入音频。非流式音频是指一次传入完整的音频。

✧ **index(识别序号, 流式音频必选):**

某次会话的请求序号, 从 1 开始, 最后一次的序号以负数作为终结。可能的序列号举例:

- -1 此时请求数据只有一块
- 1, 2, -3 此时请求数据有三块, 此时只有最后一块会返回处理结果, 其余均返回 ResCode 为 InProgress。

若采用非流式音频, 此参数为可选; 但如果设置此值, 则必须是-1。

✧ **threshold (阈值, 可选):** 服务将音频与指定模型匹配返回相似百分比, 记为相似度。相似度 \geq threshold/100时, 服务器应答中Result为"YES"; 否则, 为"NO"。不设置threshold时, 服务使用默认值50。Threshold取值范围: 0~100整数。

阈值设置参考

文本无关情况 (注册为任意文本, 测试确认为任意文本, 即注册文本和测试确认文本不相同) 使用73。

文本相关情况 (注册文本和测试确认文本为同一文本, 注册文本读3遍, 语音时长5秒): 16k模型使用61, 8k模型使用74。

✧ **注:** 以上建议阈值是等错误率的阈值。具体使用, 请根据使用场景相应的调高 (降低认假率) 或调低阈值 (降低拒真率)。

✧ **property (模型参数, 必选, 默认为16k_common)**

有效参数	参数说明
------	------

8k_common	8k16bit 模型，主要应用于电话信道，支持 audioformat 为 8k 的情况
16k_common	16k16bit 模型 (默认)，支持 audioformat 为 16k 的情况

上述8k16bit及16k16bit指的是传入引擎的PCM音频格式，和audioformat无关。设置错误会返回CommitRequest jetcl_commit_request failed。

3.3. VPR 识别请求接口

本接口在提交音频数据和 **groupid** 时，同步在 http 请求的回应消息中返回音频与 **groupid** 中所有用户的模型的匹配结果，结果为相似度。

3.2.5. 访问

(POST) [ServiceUrl] /vpr/recognise

说明：包体中传输音频数据，单次传输包体最大 8MB。

需要以下 HTTP 头：x-app-key、x-sdk-version、x-request-date、x-task-config、x-session-key、x-udid。

3.2.6. 服务器应答

应答格式：XML文件

XML的元素（Element）及含义

顺序	名称	值（例子）	含义	备注
1	ResCode	Success/Failed	完毕/处理失败	必选
2	ResMessage	Success/Failed	完毕/处理失败	必选
3	ErrorNo	4	返回错误码	必选
4	ResultToken	0E01123456	结果令牌	成功时必选
5	UserId	userid_303_72_90_433 98_20150205204125_1	用户标识	
6	Score	90	相似度	User_Id 的属性

3.2.7. 示例

中间结果:

```
<?xml version="1.0" encoding="utf-8"?>
<ResponseInfo>
  <ResCode>InProgress</ResCode>
  <ResMessage>expect more chunk</ResMessage>
  <ErrorNo>7</ErrorNo>
  <ResultToken>303_72_90_43398_20150205204125_1</ResultToken>
</ResponseInfo>
```

最终结果:

```
<?xml version="1.0" encoding="utf-8" ?>
<ResponseInfo>
  <ResCode> Success </ResCode>
  <ResMessage> Success </ResMessage>
  <ErrorNo>0</ErrorNo>
  <ResultToken>303_72_90_43398_20150205204125_1</ResultToken>
  <ResultCount>2</ResultCount>
  < Result>
    <UserId Score="90">userid_303_72_90_43398_20150205204125_1<UserId>
    <UserId Score="80">userid_303_72_90_43398_20150205204125_3<UserId>
  </Result>
</ResponseInfo>
```

3.2.8. x-task-config 参数说明

以下是有效参数:

✧ **capkey** (声纹识别能力名称, 必选)

Capkey	类型	支持 audioformat
--------	----	----------------

vpr.cloud.recog	8k16bit 模型，主要应用于电话信道 16k16bit 模型，只支持 16k 格式音频	alaw8k8bit, ulaw8k8bit, vox6k4bit , pcm8k16bit, alaw16k8bit, ulaw16k8bit, pcm16k16bit
-----------------	--	---

✧ **candnum**（候选值个数，可选）

限定返回的候选值的最大个数，不允许为负数。

识别候选结果个数，正整数，范围大于等于1以上，缺省值10。

如果超过当前组内最大个数，则返回组内超过阈值的所有模型的相似度结果；

如果小于组内最大个数，返回组内相似度最高且超过阈值的candnum个。

不设置candnum时，服务使用默认值10。

✧ **audioformat**（音频格式，必选）

有效参数	参数说明
alaw 8k8bit	8k 8bit a-law 数据
ulaw8k8bit	8k 8bit u-law 数据
pcm8k16bit	8k 16bit PCM 音频数据
vox6k4bit	6k4bit vox 音频数据
alaw16k8bit	16k 8bit a-law 数据
ulaw16k8bit	16k 8bit u-law 数据
pcm16k16bit	16k 16bit PCM 音频数据

✧ **identify**(会话标识，流式识别必选)：

该字段为会话的标识，和 x-udid 字段一起作为一次会话的唯一标识。在不同会话之间此唯一标识不可相同（即不同识别任务的[x-udid+identify]不同）。若采用非流式识别方式，则此参数可不设置。一次会话即指一次完整的识别过程。

✧ **index**(识别序号，流式识别必选)：

某次会话的请求序号，从 1 开始，最后一次的序号以负数作为终结。可能的序列号举例：

- -1 此时请求数据只有一块
- 1, 2, -3 此时请求数据有三块，此时只有最后一块会返回处理结果，其余均返回 ResCode 为 InProgress。

若采用非流式识别，此参数为可选；但如果设置此值，则必须是-1。

✧ **groupid**（组ID，必选）：通过用户管理接口添加的组id，对应有多个用户id。同一应用（develop_key）下唯一。

- ✧ **threshold（阈值，可选）**：服务将音频与指定模型匹配返回相似百分比，记为相似度。服务会在返回结果中相似度 \geq threshold/100的前candnum个结果。不设置threshold时，服务使用默认值50。Threshold取值范围：0~100整数。

阈值设置参考

文本无关情况（注册为任意文本，测试确认为任意文本，即注册文本和测试确认文本不相同）使用73。

文本相关情况（注册文本和测试确认文本为同一文本，注册文本读3遍，语音时长5秒）：16k模型使用61，8k模型使用74。

注：以上建议阈值是等错误率的阈值。具体使用，请根据使用场景相应的调高（降低认假率）或调低阈值（降低拒真率）。

- ✧ **property（模型参数，必选，默认为16k_common）**

有效参数	参数说明
8k_common	8k16bit 模型，主要应用于电话信道，支持 audioformat 为 8k 的情况
16k_common	16k16bit 模型 (默认)，支持 audioformat 为 16k 的情况

上述8k16bit及16k16bit指的是传入引擎的PCM音频格式，和audioformat无关。设置错误会返回CommitRequest jetcl_commit_request failed。

附录 1：通用模块返回错误码

错误码	描述
0	成功
-1	初始化错误，目前不会返回(避免影响 apache 启动)
-2	加载初始化配置失败
-3	初始化配置出错
-4	参数错误
-5	私有云返回，部分接口不支持
-6	Appkey 错误，转换为 app_serial_id 失败
-7	无效的 udid，格式不正确
-8	sessionkey 验证失败
-9	授权版本信息不对，sdk 版本和授权版本映射
-10	添加缓存失败
-11	获取缓存失败
-12	缓存数据格式不正确
-13	内部访问超时（授权，负载，cu）
-14	内部访问失败（授权，负载，cu）
-15	查询 cu 负载返回失败
-16	udid 注册失败
-17	不能继续访问灵云能力
-18	查询应用信息失败
-19	查询应用授权失败
-20	添加确认数据失败
-21	确认数据格式不正确
-22	DDRinit 失败（数据存放路径错误）
-23	SDRinit 失败（ActiveMQ 无法连接）
-24	应用 domain 不匹配
-25	过载（一般为访问时间太长的请求在某一台 cu 请求过多导致，一般不会返回）
-26	发送和接收的 token 不一致
-27	接收数据与实际发送的不一致，接收到的不完整
-28	私有云获取授权出错
-29	数据库连接失败

附录 2：VPR 模块错误码

错误码	描述
0	成功
14001	错误的音频数据(一般为长度超出限制)

14002	错误的编码（编码类型不支持）
14003	请求参数错误（一般为 audioformat 错误）
14004	请求 token 为空
14005	请求配置串错误（一般为某些配置项没找到）
14006	音频长度小于 0（为兼容老版本而设，此版本返回 14001）
14007	音频数据错误
14008	用户 ID 不存在
14009	组 ID 不存在
14010	平台协议初始化失败
14011	mongodb 初始化失败
14012	服务处理失败 1、内部小错误(new 失败) 2、其他文件格式转 PCM 失败
14013	添加请求数据块识别失败
14014	处理超时
14015	读取模型文件失败
14016	保存模型文件失败
14017	保存模型配置文件失败
14018	设置引擎参数识别
14019	训练失败
14020	识别失败
14021	引擎相关错误号, 初始化失败
14022	多选一失败
14023	提取特征数据失败
14024	获取模型句柄失败
14025	会话创建失败
14026	声纹模型文件路径不存在
14027	获取模型文件列表失败