接口说明

语音识别（ASR）可以将语音信息转换为文字信息。该能力通过WebSocket API的方式提供给开发者，相较于SDK，该方式具有轻量、跨平台、跨开发语言的特点。

接口要求

项目	说明
请求地址	ws://api.baller-tech.com/v1/service/ws/v1/asr
字符编码	UTF-8
WebSocket版本	13 (RFC 6455)
响应格式	统一采用JSON格式

调用流程

1. 通过hmac-sha256计算签名，向服务器端发送WebSocket协议握手请求。
2. 握手成功之后，通过WebSocket连接上传和接收数据。
3. 请求方接收到服务器端推送的结果返回结束标记后断开WebSocket连接

音频数据发送模式

向服务器发发送音频数据时，可以一次性的将音频数据发送到的服务器，也可以将音频数据分多次发送到服务器，不论使用那种模式向服务器发送音频数据，识别结果的推送方式是一样的，服务器会分多次推送识别结果。两种模式的适用场景如下：

1. 已经录制好的音频且时长较短（一般60秒内）时，即可以一次性将音频数据发送到服务器，也可以分多次发送到服务器；
2. 已经录制好的音频且时长较长时，分多次将音频数据发送到服务器；
3. 音频数据实时录取，实时识别时，分多次将音频数据发送到服务器。

握手和接口鉴权

在WebSocket的握手阶段，请求方需要对请求进行签名，服务端会根据签名检查请求的合法性。握手时请求方将签名相关的参数经过url编码后加到请求地址的后面，具体的参数和示例如下：

```
ws://api.baller-tech.com/v1/service/ws/v1/asr?
authorization=xxxx&host=xxxx&date=xxx
```

参数	类型	说明	示例
host	string	请求的主机	api.baller-tech.com
date	string	当前GMT格式的时间	Fri, 10 Jan 2020 07:31:50 GMT
authorization	string	鉴权信息Base64编码后的数据	-

握手和鉴权参数详细介绍

date介绍

1. date必须是GMT+0时区的符合RFC1123格式的日期和时间，星期和月份只能使用英文表示
2. 服务端允许date的最大偏差为300秒，超出此偏差请求会被拒绝

authorization介绍

authorization使用base64编码前的格式如下json格式

```
{
    "app_id": "1172448516240310275",
    "signature": "qaIpgE3Ecs78g6GRFxQBJKgdna28b7ronAcsDCsO+Zw="
}
```

app_id介绍

1. 由北京大牛儿科技发展有限公司统一分配。

signature介绍

1. signautre 是使用hmac-sha256对参数进行签名后并base64编码的字符串。
2. signautre 使用hmac-sha256签名前的原始字段由三部分构成，分别为app_id、date、host。每一部分使用换行符(\n)进行分割，“:”号前后无空格。

```
app_id:1172448516240310275
date:Fri, 10 Jan 2020 07:31:50 GMT
host:api.baller-tech.com
```

3. 使用hmac-sha256算法，结合app_key（由北京大牛儿科技发展有限公司统一分配）对signautre的原始字段进行签名。
4. 对签名数据进行base64编码，生成signature的字段值。

握手和鉴权消息响应

1. 接口鉴权成功时，WebSocket握手回复报文的状态码为101。
2. 接口鉴权失败时，WebSocket握手回复报文的状态码为403，可以通过响应行的原因短语查看接口鉴权失败原因。
3. 接口鉴权失败时，响应报文的主体中会返回json格式的数据，包含了以下信息

参数	类型	说明
task_id	string	本次任务的标识，如果对请求有疑问，可以将task_id提供给我公司进行排查
message	string	接口鉴权失败的原因，与响应行中的原因短语相同

数据的发送和接收

握手成功之后，请求方和服务器会建立WebSocket的连接，请求方将数据通过WebSocket发送给服务器，服务器有识别结果的时候，会通过WebSocket连接推送识别结果到请求方。请求方和服务器通过json的格式交换数据。

请求方发送数据时使用的参数

参数名	类型	是否每帧必须	描述
business	obj	否	业务参数，仅在握手成功后首帧中上传
data	obj	是	数据流参数，握手成功后所有帧中都需要上传

业务参数(business)

参数名	类型	是否必须	描述
language	string	是	音频的语种；参见支持的语种和采样格式
sample_format	string	是	音频采样格式；参见支持的语种和采样格式
audio_format	string	是	音频格式；参见 参见支持的音频格式
service_type	string	否	服务类型: sentence: 句子识别（默认值，任务有时长限制） realtime: 实时识别（任务无时长限制）
vad	string	否	是否启用端点检测: on : 启用（默认值） off: 不启用
dynamic_correction	string	否	是否启用动态纠正: on : 启用（暂不支持） off: 不启用（默认值）

sample_format 介绍

根据RFC对MIME格式的定义，使用audio/Lxx;rate=xxxxx 表明采样格式，audio/L后面的数字表示音频的采样点大小（单位bit），rate=后面的数字表示音频 的采样率（单位hz）。

比如audio/L16;rate=16000表示音频数据为16000hz，16bit的pcm音频数据

数据流参数（data)

参数名	类型	是否必须	描述
input_mode	int	是	可选值为: once continue end
audio	string	是	经过base64编码后的音频数据

input_mode 介绍

一次性将音频数据发送到服务器时，input_mode应设置为once。当分多次将音频数据发送到服务器时，如果不是本次识别事务的最后一次，input_mode应设置为continue；如果是本次识别事务的最后一次应设置为end。

```
{
    "data": {
        "input_mode": "once",
        "audio":
"AAAFAAoADwAXAB0AJgA0AEIATABPAE8AUQBRAEgAOwA0AC8AJwAcABUAEQAJAAIAAgADAAAA+P="
    },
    "business": {
        "language": "mon",
        "service_type": "sentence",
        "sample_format": "audio/L16;rate=16000",
        "audio_format": "raw",
    }
}
```

服务器推送结果的参数

语音识别时，会将传入的音频分为不同的子句，每次推送的结果是一个子句的结果。子句的识别结果分为最终结果和非最终结果两种状态；最终状态表示结果为当前子句的最终结果，之后再推送的结果为新子句的结果；非最终状态表示结果为当前子句的中间状态的结果，之后再推送的结果还是该子句的识别结果。

一般我们只需关注最终状态的识别结果即可，如果需要更快速的让用户的看到部分识别结果，并动态的调整用户看到的识别结果时，才需要考虑非最终状态的结果。

参数名	类型	描述
task_id	string	本次任务的id，仅在第一帧中返回，如果对请求有疑问，可以将task_id提供给我公司进行排查
code	int	请求处理的结果码
message	string	错误提示
is_end	int	结果返回是否结束（0-未结束; 1-结束），当为1时，请求方需关闭WebSocket
data	string	子句的识别结果
is_complete	int	子句结果是否是最终的（0：非最终结果；1：最终结果）
begin	int	子句的起始位移，单位毫秒
end	int	子句的结束位移，单位毫秒

子句位移的介绍

需在以下条件都满足时begin、end字段的值有效：

1. 业务参数中启用了vad。
2. 推送结果中is_complete字段的值为1。
3. 推送结果data字段包含识别的结构。

特殊请情况说明：

当启用vad后，每个任务最后一次推送的识别结果只有一个标点符号，此时推送结果的is_complete字段为1，但begin和end字段为0。

```
{
    "code": 0,
    "message": "success",
    "is_end": 0,
    "data": "xxxxx",
    "is_complete": 1,
    "begin": 245,
    "end": 5600,
    "task_id": "1172448516240310275-2903dc7e3ab65879b4fc66055720ec09"

}
```

支持的语种以及采样格式

语种	对应的language字段	支持的采样格式	对应的sample_format
哈语	kaz	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
蒙语	mon	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
藏语（安多）	tib_ad	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
藏语（康巴）	tib_kb	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
藏语（卫藏）	tib_wz	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
维语	uig	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
汉语	zho	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
壮语	zha	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
彝语	iii	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000
朝鲜语	kor	采样率：16000hz 采样点大小：16bits	audio/L16;rate=16000

支持的音频格式

音频格式	对应的audio_format字段
raw	未压缩的pcm
mp3	mp3格式
wav	wav格式
m4a	m4a格式
ogg_opus	ogg封装后的opus音频编码
ogg_speex	ogg封装后的speex音频编码

m4a格式说明

部分m4a文件的moov atom位于文件的尾部，无法做的实时解码。本 接口处理的m4a文件，需要moov atom位于文件的头部，可以使用ffmpeg将moov atom移动到文件头部

```
ffmpeg -i input.m4a -movflags faststart -acodec copy output.m4a
```