

#实时语音转写API文档

#接口说明

#接口Demo

#接口要求

#接口调用流程

#握手阶段

#signa生成

#请求示例

#返回值

#实时通信阶段

#上传数据

#上传结束标志

#接收转写结果

#接收错误信息

#白名单

#错误码

#常见问题

#实时语音转写支持什么平台？

#实时语音转写支持什么语言？

#支持的音频是什么格式？

#实时语音转写支持的音频时长有什么限制？

#实时语音转写的分片时长200ms是什么意思？

#如何购买实时语音转写？

#实时语音转写如何添加IP白名单？

#实时语音转写如何免费试用？

#实时语音转写和语音听写的区别有哪些？

#实时语音转写接口返回10105，如何解决？

#调用实时语音转写接口报10110错误码，如何解决？

#实时语音转写支不支持离线？

#实时语音转写如果一次连接使用时长超出了剩余时长怎么办？

#实时语音转写API文档

#接口说明

实时语音转写（Real-time ASR）基于深度全序列卷积神经网络框架，通过 WebSocket 协议，建立应用与语言转写核心引擎的长连接，开发者可实现将连续的音频流内容，实时识别返回对应的文字流内容。
支持的音频格式：采样率为16K，采样深度为16bit的pcm_s16le音频

#接口Demo

目前仅提供部分开发语言的demo，其他语言请参照下方接口文档进行开发。

#接口要求

集成实时语音转写API时，需按照以下要求。

内 容	说 明
请 求 协 议	ws[s] (为提高安全性，强烈推荐wss)
请 求 地 址	ws[s]: //ai.abcpen.com/v1/asr/ws?{请求参数} 注：服务器IP不固定，为保证您的接口稳定，请勿通过指定IP的方式调用接口，使用域名方式调用
接 口 鉴 权	签名机制，详见 数字签名
字 符 编 码	UTF-8
响 应 格 式	统一采用JSON格式
开 发 语 言	任意，只要可以向笔声云服务发起WebSocket请求的均可
音 频 属 性	采样率16k、位长16bit、单声道
音 频 格 式	pcm
数 据 发 送	建议音频流每200ms发送6400字节
语 言 种 类	中文普通话、中英混合识别、英文

#接口调用流程

注：若需配置IP白名单，请发送邮件到support@abcpen.com

实时语音转写接口调用包括两个阶段：握手阶段和实时通信阶段。

#握手阶段

接口地址

```
ws://ai.abcpen.com/v1/asr/ws?{请求参数}
或
wss://ai.abcpen.com/v1/asr/ws?{请求参数}
```

参数格式

```
key1=value1&key2=value2...（key和value都需要进行urlencode）
```

参数说明

参数	类型	必须	说明	示例
appid	string	是	笔声开放平台应用ID	595f23df
ts	string	是	当前时间戳，从1970年1月1日0点0分0秒开始到现在的秒数	1512041814
signa	string	是	加密数字签名（基于HMACSHA1算法）	lrrzsJeOFk1NGfjHW6SkHUoN9CU=
lang	string	否	实时语音转写语种，不传默认为中文	语种类型：中文、中英混合识别：cn；英文：en。传参示例如："lang=en" 若未授权无法使用会报错10110
punc	string	否	标点过滤控制，默认返回标点，punc=0会过滤结果中的标点	0
pd	string	否	垂直领域个性化参数: 法院: court 教育: edu 金融: finance 医疗: medical 科技: tech 运营商: isp 政府: gov 电商: ecom 军事: mil 企业: com 生活: life 汽车: car	设置示例: pd="edu" 参数pd为非必须设置，不设置参数默认为通用

#signa生成

1.获取baseString，baseString由appid和当前时间戳ts拼接而成，假如appid为595f23df，ts为1512041814，则baseString为

```
595f23df1512041814
```

2.对baseString进行MD5，假如baseString为上一步生成的595f23df1512041814，MD5之后则为

```
0829d4012497c14a30e7e72aeebe565e
```

3.以apiKey为key对MD5之后的baseString进行HmacSHA1加密，然后再对加密后的字符串进行base64编码。

假如apiKey为d9f4aa7ea6d94faca62cd88a28fd5234，MD5之后的baseString为上一步生成的0829d4012497c14a30e7e72aeebe565e，则加密之后再进行base64编码得到的signa为

```
IrrzsJeOFk1NGfJHW6SkHUoN9CU=
```

备注：

- apiKey：接口密钥，在应用中添加实时语音转写服务时自动生成，调用方注意保管；
- signa的生成公式：HmacSHA1(MD5(appid + ts), api_key)，具体的生成方法详见【[调用示例](#)】；

#请求示例

```
ws://ai.abcpen.com/v1/asr/ws?
appid=595f23df&ts=1512041814&signa=IrrzsJeOFk1NGfJHW6SkHUoN9CU=&pd=edu
```

#返回值

结果格式为json，字段说明如下：

参数	类型	说明
action	string	结果标识，started:握手，result:结果，error:异常
code	string	结果码(具体见 错误码)
data	string	结果数据
desc	string	描述
sid	string	会话ID

其中sid字段主要用于DEBUG追查问题，如果出现问题，可以提供sid帮助确认问题。

成功

```
{
  "action": "started",
  "code": "0",
  "data": "",
  "desc": "success",
  "sid": "rta000000a@ch312c0e3f63609f0900"
}
```

```
{
  "action": "error",
  "code": "10110",
  "data": "",
  "desc": "invalid authorization|illegal signa",
  "sid": "rta0000000b@ch312c0e3f65f09f0900"
}
```

#实时通信阶段

握手成功后，进入实时通信阶段，此时客户端的主动操作有两种：上传数据和上传结束标识，被动操作有两种：接收转写结果和错误

#上传数据

在实时转写过程中，客户端不断构造binary message发送到服务端，内容是音频的二进制数据。此操作的频率影响到文字结果展现的实时性。

注意：

1.建议音频流每200ms发送6400字节，发送过快可能导致引擎出错； 2.音频发送间隔超时时间为15秒，超时服务端报错并主动断开连接。

#上传结束标志

音频数据上传完成后，客户端需发送一个特殊的binary message到服务端作为结束标识，内容是：

```
{"end" : true}
```

#接收转写结果

交互过程中，服务端不断返回 text message（转写结果）到客户端。当所有结果发送完毕后，服务端断开连接，交互结束。

结果示例：

```
{
  "action": "result",
  "code": "0",
  "data": "{\\\"cn\\\":{\\\"st\\\":{\\\"bg\\\":\\\"820\\\",\\\"ed\\\":\\\"0\\\",\\\"rt\\\":\\\"\\\"\\\"ws\\\":{\\\"cw\\\":{\\\"w\\\":\\\"啊\\\",\\\"wp\\\":\\\"n\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"喂\\\",\\\"wp\\\":\\\"n\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"!\\\",\\\"wp\\\":\\\"p\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"你好\\\",\\\"wp\\\":\\\"n\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"!\\\",\\\"wp\\\":\\\"p\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"我\\\",\\\"wp\\\":\\\"n\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"是\\\",\\\"wp\\\":\\\"n\\\"}},\\\"wb\\\":0,\\\"we\\\":0},{\\\"cw\\\":{\\\"w\\\":\\\"上\\\",\\\"wp\\\":\\\"n\\\"}},\\\"wb\\\":0,\\\"we\\\":0}}},\\\"type\\\":\\\"1\\\"}},\\\"seg_id\\\":5}\\n",
  "desc": "success",
  "sid": "rta0000000e@ch312c0e3f6bcc9f0900"
}
```

其中data为转写结果的json字符串

```
data:
{
  "cn":{
    "st":{
      "bg":"820",
      "ed":"0",
      "rt":[{
        "ws":[{
          "cw":[{
            "w":"啊",
            "wp":"n"
          }],
          "wb":0,
          "we":0
        },{
          "cw":[{
            "w":"喂",
            "wp":"n"
          }],
          "wb":0,
          "we":0
        },{
          "cw":[{
            "w":"!",
            "wp":"p"
          }],
          "wb":0,
          "we":0
        },{
          "cw":[{
            "w":"你好",
            "wp":"n"
          }],
          "wb":0,
          "we":0
        },{
          "cw":[{
            "w":"!",
            "wp":"p"
          }],
          "wb":0,
          "we":0
        },{
          "cw":[{
            "w":"我",
            "wp":"n"
          }],
          "wb":0,
          "we":0
        },{
          "cw":[{
            "w":"是",
            "wp":"n"
          }],
          "wb":0,
```

```
        "we":0
      },{
        "cw":[{
          "w":"上",
          "wp":"n"
        }],
        "wb":0,
        "we":0
      }]
    }],
    "type":"1"
  }
},
"seg_id":5
}
```

转写结果data字段说明如下：

字段	含义	描述
bg	句子在整段语音中的开始时间，单位毫秒(ms)	中间结果的bg为准确值
ed	句子在整段语音中的结束时间，单位毫秒(ms)	中间结果的ed为0
w	词识别结果	
wp	词标识	n-普通词；s-顺滑词（语气词）；p-标点
wb	词在本句中的开始时间，单位是帧，1帧=10ms 即词在整段语音中的开始时间为(bg+wb*10)ms	中间结果的 wb 为 0
we	词在本句中的结束时间，单位是帧，1帧=10ms 即词在整段语音中的结束时间为(bg+we*10)ms	中间结果的 we 为 0
type	结果类型标识	0-最终结果；1-中间结果
seg_id	转写结果序号	从0开始

#接收错误信息

交互过程中，在服务端出现异常而中断服务时（如会话超时），会将异常信息以 text message 形式返回给客户端并关闭连接。

#白名单

在调用该业务接口时

- 若关闭IP白名单，接口认为IP不限，不会校验IP。
- 若打开IP白名单，则服务端会检查调用方IP是否在笔声开放平台配置的IP白名单中，对于没有配置到白名单中的IP发来的请求，服务端会拒绝服务。

IP白名单规则

- IP白名单，在 控制台-我的应用-相应服务的应用管理卡片上 编辑，保存后五分钟左右生效；
- 不同Appid的不同服务都需要分别设置IP白名单；
- IP白名单需设置为外网IP，请勿设置局域网IP。
- 如果服务器返回结果如下所示(illegal client_ip)，则表示由于未配置IP白名单或配置有误，服务端拒绝服务。

```
{
  "action": "error",
  "code": "10105",
  "data": "",
  "desc": "illegal access|illegal client_ip: xx.xx.xx.xx",
  "sid": "rta..."
}
```

#错误码

错误码	描述	说明	处理方式
0	success	成功	
10105	illegal access	没有权限	检查apiKey, ip, ts等授权参数是否正确
10106	invalid parameter	无效参数	上传必要的参数，检查参数格式以及编码
10107	illegal parameter	非法参数值	检查参数值是否超过范围或不符合要求
10110	no license	无授权许可	检查参数值是否超过范围或不符合要求
10700	engine error	引擎错误	提供接口返回值，向服务提供商反馈
10202	websocket connect error	websocket连接错误	检查网络是否正常
10204	websocket write error	服务端websocket写错误	检查网络是否正常，向服务提供商反馈
10205	websocket read error	服务端websocket读错误	检查网络是否正常，向服务提供商反馈
16003	basic component error	基础组件异常	重试或向服务提供商反馈
10800	over max connect limit	超过授权的连接数	确认连接数是否超过授权的连接数

#常见问题

#实时语音转写支持什么平台？

答：实时转写只支持webapi接口，开放平台“实时语音转写”需要WebSocket接入，针对是有编程基础的开发者用户。如果您是个人用户，不想通过编程方式直接实现语音转写功能，可以去笔声听见官网，了解语音转写功能的更多详情。

#实时语音转写支持什么语言？

答：中文普通话、中英混合识别、英文，小语种以及中文方言可以到控制台-实时语音转写-方言/语种处添加试用或购买。

#支持的音频是什么格式？

答：采样率为16K，采样深度为16bit的pcm_s16le音频

#实时语音转写支持的音频时长有什么限制？

答：实时语音转写可以实时识别持续的音频流，结果是实时返回，音频流长度理论上不做限制，典型的应用场景是大会或者直播的实时字幕。

#实时语音转写的分片时长200ms是什么意思？

答：可以理解为上传的间隔为200ms，建议音频流每200ms向服务器发送6400字节，发过快可能导致引擎出错，音频发送间隔超时时间为15s，超时服务端报错并主动断开连接。

#如何购买实时语音转写？

答：登录笔声开放平台，进入实时语音转写页面，点击“申请购买”按钮，在线购买时长与路数即可。

#实时语音转写如何添加IP白名单？

答：登录笔声开放平台，点击右上角的“控制台”，点击“我的应用”，选择到所创建的实时语音转写Web api应用平台，点击IP白名单“管理”按钮，即可添加IP白名单。

#实时语音转写如何免费试用？

答：可在实时语音转写服务的产品页面，直接领取免费使用权限；到期后可直接在控制台点击购买时长和授权（价格可见）

#实时语音转写和语音听写的区别有哪些？

答：支持时长：在线语音听写单次会话支持60s以内的语音转文字；实时语音转写的音频流长度理论上不做限制

支持语种：在线语音听写除中文普通话和英文外，支持12个语种，25种方言；实时语音转写支持中文普通话、中英混合识别、英语、开通的小语种以及中文方言；

应用场景：在线语音听写主要用于短语音的识别，如聊天输入、语音搜索等；实时语音转写可以实时识别持续的音频流，典型的应用场景是大会或者直播的实时字幕

#实时语音转写接口返回10105，如何解决？

答：未通过服务端校验，请检查appid，apiKey，ip白名单，checksum等授权参数是否正确。

#调用实时语音转写接口报10110错误码，如何解决？

答：没有授权许可或授权数已满，请至控制台查看时长和路数情况，并查看有效期；如果未领取免费包，请至产品页面领取。

#实时语音转写支不支持离线？

答：不支持

#实时语音转写如果一次连接使用时长超出了剩余时长怎么办？

答：首先为了使业务使用不受影响，如果在连接期间使用时长超出，转写功能并不会立刻停止。本次连接断开后时长可能会出现为负数的情况，请在使用过程中关注时长剩余情况并及时购买时长。