ALIBABA CLOUD

阿里云

智能语音交互

常见问题

文档版本：20220509

阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。

2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。

3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。

4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。

5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。

6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	危险 重置操作将丢失用户配置数据。
警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	警告 重启操作将导致业务中断，恢复业务时间约十分钟。
注意	用于警示信息、补充说明等，是用户必须了解的内容。	注意 权重设置为0，该服务器不会再接受新请求。
说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击**设置**> **网络**> **设置网络类型**。
粗体	表示按键、菜单、页面名称等UI元素。	在**结果确认**页面，单击**确定**。
`Courier字体`	命令或代码。	执行 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *Instance_ID*
[] 或者 [a|b]	表示可选项，至多选择一个。	`ipconfig [-all|-t]`
{} 或者 {a|b}	表示必选项，至多选择一个。	`switch {active|stand}`

目录

1.产品公共FAQ 05
2.语音识别FAQ 10
3.语音合成FAQ 23
4.SDK FAQ 26
5.计费定价FAQ 30
6.自学习平台FAQ 31

1.产品公共FAQ

本文汇总了您在使用智能语音交互产品过程中的常见问题。

产品公共常见问题主要分为以下几类：

- 功能类
  - 使用阿里云音视频通信RTC如何调用语音识别服务？
  - 智能语音交互服务中语音识别和语音合成用到的端口是哪些？
  - 开通商用或者扩容并发，多久才会在控制台上显示？
  - 智能语音交互创建的项目数量有限制吗？
  - 控制台一个项目下支持多个基础模型吗？
  - 有没有H5语音唤醒相应的技术和插件？
  - 如何在阿里云查询账号下ASR语音识别服务的使用情况？
  - 智能语音交互服务有哪些域名？这些域名对应IP地址有哪些？
  - 找不到AccessKey ID和AccessKey Secret在哪里？
  - 如何查看UID？
  - 如何提工单？
  - 智能语音交互是否有供参考的教程？
- 性能类
  - ASR语音识别和TTS语音合成超并发会有什么现象？
  - 免费用户有什么限制？
  - 并发数是如何计算的？
- SDK使用类
  - 智能语音交互是否提供相关的源码呢？
  - 智能语音交互的Android SDK支持在多进程中初始化多个实例吗？
  - 语音合成SDK播放声音有播放完成的监听吗，确认下有没有播放语音模块？
  - 音频数据必须连续发送吗？
  - 音频数据发送中断后，为什么还会收到服务器发回的数据？
  - 初始化失败的可能原因？
  - 开始识别失败的可能原因？
  - 为何开始识别后没有识别结果？
- 计费类
  - 实时语音识别最大并发可以买多少路？
  - 英文的识别，需要购买扩展语言包还是要买路数？

功能类

使用阿里云音视频通信RTC如何调用语音识别服务？

使用阿里云音视频通信RTC服务可以集成智能语音交互服务，具体请参见语音数据处理。

智能语音交互服务中语音识别和语音合成用到的端口是哪些？

对于HTTP和HTTPS协议，端口号为80；对于Websocket，端口号为443。

开通商用或者扩容并发，多久才会在控制台上显示？

可用并发路数为实时更新显示，昨日并发路数为T+1更新。另外，控制台显示的计量数据是T+1规则，第二天更新数据。

智能语音交互创建的项目数量有限制吗？

目前没有限制。

控制台一个项目下支持多个基础模型吗？

不支持。一个Appkey对应一个模型。

有没有H5语音唤醒相应的技术和插件？

设备端暂不支持此功能，云端支持。如果是设备端，一般会采用端云结合的方案。设备端侧有一个唤醒模型，设备端唤醒后结合云端模型二次确认降低误唤醒的概率。

如何在阿里云查询账号下ASR语音识别服务的使用情况？

您可以在控制台直观地查看智能语音交互服务的调用情况，包括时长、次数、并发路数等，根据运营数据判断当前使用是否合理，并决策是否需要增加或减少相关服务调用量。具体请参见服务用量。

智能语音交互服务有哪些域名？这些域名对应IP地址有哪些？

智能语音交互服务的域名为 `nls-meta.cn-shanghai.aliyuncs.com` 和 `nls-gateway.cn-shanghai.aliyuncs.com` ，端口为80、443，支持HTTPS和Websocket协议。 可通过命令 `dig nls-gateway.cn-shanghai.aliyuncs.com` 或 `dig nls-meta.cn-shanghai.aliyuncs.com` 查看域名对应的IP地址。这些IP地址是动态IP，可能会不定期有更新，需要调用方自行监控。

找不到AccessKey ID和AccessKey Secret在哪里？

请参见开通服务，获取AccessKey ID和AccessKey Secret。

如何查看UID？

1. 登录阿里云控制台。
2. 将鼠标悬浮在右上角用户头像上，弹出的面板信息中的**账号ID**即为UID。

如何提工单？

1. 全部的log信息，包含报错部分。
2. 相关taskid、Appkey等信息（不含账号密码信息），处理工单的同学将发送授权同意请求给您，以便您勾选同意。

智能语音交互是否有供参考的教程？

以下是相关的功能的视频介绍：

- 音频基础知识+智能语音控制台介绍
- ASR产品使用介绍
- 自学习平台
- 语音合成

性能类

ASR语音识别和TTS语音合成超并发会有什么现象？

超并发可能会出现一下情况：

1. 查看日志会有大量超时现象，具体服务状态码为40000005，表示请求数量过多。
2. 如果并发量较大时会导致ASR语音识别和TTS语音合成断开连接。

免费用户有什么限制？

- 一句话识别、实时语音识别：最多同时发送2路语音识别。
- 录音文件识别：每自然日最多识别2小时时长的录音文件。

并发数是如何计算的？

并发数指同一个账号（阿里云UID）同时在处理的请求数。

一般语音请求的处理都会延续一段时间，例如用户新建一个语音识别请求，持续发送语音数据给服务端，这时并发数就是1；在这个请求处理的同时，同一用户又新建了另一个请求，开始发送语音数据，这时服务端同时在处理这个账号的两个请求，并发数就变成2。

SDK使用类

智能语音交互是否提供相关的源码呢？

SDK协议及源码已经开源，具体请参见Github。目前开源的有C++、Java和Python语言版本。

如有需要研究智能语音交互的协议架构，可以参考Github中的源码实现。建议您直接使用智能语音交互产品提供的SDK版本进行集成，具体请参见从这里开始。对自行研究协议实现的API很难提供更多支持。

智能语音交互的Android SDK支持在多进程中初始化多个实例吗？

目前不支持。

语音合成SDK播放声音有播放完成的监听吗，确认下有没有播放语音模块？

播放不属于SDK，SDK只有合成事件。

音频数据必须连续发送吗？

音频数据必须连续发送。

服务端在超过一定时间未接到语音数据（10秒）会超时断开连接，返回40000004错误信息。如果需要再次发送数据，客户端需要重新发起请求。

音频数据发送中断后，为什么还会收到服务器发回的数据？

音频数据未连续发送超时中断后，服务器上如果还有之前未处理完的数据，就会继续返回这些数据的识别结果，但是整个句子的识别已经是错误的了。

初始化失败的可能原因？

请检查是否使用正确的AccessKey ID、AccessKey Secret生成Access Token，并填入正确Appkey、Access Token、Workspace等必选参数。

开始识别失败的可能原因？

SDK为单例模式，请确认上一个识别已经结束后再开始新的识别。

为何开始识别后没有识别结果？

确认如下信息：

- 初始化成功。
- 开始识别接口调用成功，且正确使用参数vad_mode。
- 有音频状态回调返回且已正确开启录音。

无问题的情况下，若仍然没有识别结果，则一般会有EVENT_ASR_ERROR事件发生，根据事件携带的错误码进行定位。

计费类

实时语音识别最大并发可以买多少路？

试用版免费2路并发，3个月有效期；开通商用后默认200路并发，可单独扩充并发资源包。

英文的识别，需要购买扩展语言包还是要买路数？

英文识别需要购买扩展包和购买路数。

2.语音识别FAQ

本文汇总了您在使用语音识别服务时的常见问题。

语音识别类常见问题主要分为以下几类：

- 功能类
  - 实时转写说话有停顿，但是语音识别不断句怎么办？
  - 语音识别能自动断开多句话吗？
  - 语音识别服务支持离线功能吗？
  - 语音识别支持哪些模型？
  - 语音识别是否可以混合识别极少量英文单词和字母？
  - 开启ITN（逆文本规整）后，中文数字混合时为什么并不是全部转为阿拉伯数字？
  - 录音文件识别的enable_sample_rate_adaptive和极速版本里的sample_rate，这两个接口是一样的吗？
  - 录音转文本能区分坐席和客户吗？
  - 智能语音交互的一句话识别，标点符号是根据什么来判断逗号和句号的？
  - 离线文件转写如何区分左右声道？
  - 语音识别可以支持多个词表吗？
  - 设置录音文件识别服务的版本，"4.0"和"2.0"两个版本有什么区别？
  - 在电话端支持哪些国家的语音识别？
  - 在语音识别的服务中，有没有请求参数是音频文件地址，返回参数是转写文本？
  - 实时语音转写能和录音文件识别一样加入音轨ID吗？
  - 录音文件识别可以生成SRT字幕文件吗？
  - 语音识别服务支持哪些编码格式的音频？
  - 语音识别服务支持哪些采样率？
  - 怎么查看音频文件的采样率？
  - 语音识别服务支持的方言模型和语种都有哪些？
  - 语音识别能否自动断开多句话？
  - 实时识别和录音文件转写分别支持哪些语音格式？
- 性能类
  - 语音识别的识别准确率怎么计算？
  - 语音识别模型的字准率能达到多少？
  - 录音文件识别极速版延迟是多少？
  - 8k模型可以识别16k的音频吗？
  - 录音文件识别极速版调用频率有限制吗？
  - 粤语的识别准确率是多少？
  - 15秒左右的录音文件识别大概需要多久能转换成文本呢？

- 语音转文本有没有优先级？比如现在正在转写任务，突然有紧急的转写任务，能调整处理优先级吗？
- 针对两个用户打电话场景，哪个模型效果比较好？
- 服务请求时长限制？
- “流式”模式和“非流式”模式识别的区别？
- 什么是ASR尾点延迟？

- 效果类
  - 对于识别不准的词该如何进行优化？
  - 单字识别不出来是什么原因？
  - 热词效果如果不佳是否可以自主调节权重？
  - 录音文件识别时间戳不准，如何解决？
  - 语音识别太灵敏、无效声音（噪音等）被识别出了文字怎么办？
  - 如何提高标点断句的效果？
  - 实时场景中，已经开启了标点断句，为什么效果还是不理想？
  - 录音文件识别存在一次请求后返回两次相同的结果的情况吗？
  - 实时语音识别遇到识别慢、超时问题，如何排查？
  - 为什么语音识别准确率很低，有时只识别出几个字？
  - 确认调用方式和采样率都没问题，识别还是不准确怎么办？
- SDK使用类
  - 一句话识别录入的demo是使用Websocket进行识别展示的吗？
  - 实时语音识别服务有Python SDK吗？
  - 语音识别的返回结果JSON中endtime=-1是什么意思？
- 计费类
  - 录音文件识别极速版不支持试用吗？

功能类

实时转写说话有停顿，但是语音识别不断句怎么办？

1. 如果是vad断句情况下，实时转写的vad断句依赖对音频中静音数据的判断，如果上游不发送静音音频，服务端则无法识别用户说话是否有停顿。如果确认是上游没有发送静音音频，则系统通过对实时转写服务的时间戳和实际音频的时间戳对比。如果发现服务端的判断音频时长比实际音频时长短，说明静音时服务端没有收到用户发的静音数据。
2. 在开启语义断句情况下，有可能是后处理模型的效果问题。

解决方案：在用户停顿时持续地向服务端发送静音数据。

语音识别能自动断开多句话吗？

实时语音识别服务可以断开多句话。一句话识别服务的每个请求只对应一句话，无法断开。

语音识别服务支持离线功能吗？

目前不支持本地离线的语音识别，必须把音频数据发送到服务端做识别。

语音识别支持哪些模型？

可以在智能语音交互控制台中项目功能配置里查看具体的模型种类，目前有8k和16k两种采样率的模型，每个采样率下面又有多个领域模型，可以按需选择。

语音识别是否可以混合识别极少量英文单词和字母？

可以的，中文普通话模型支持对中英文混杂的音频进行识别。

开启ITN（逆文本规整）后，中文数字混合时为什么并不是全部转为阿拉伯数字？

是否要转成阿拉伯数字，系统是用模型来判断的，并不是所有数字都需要转成阿拉伯数字，模型的判断主要准则是一般书面文本中常用的形态。

录音文件识别的enable_sample_rate_adaptive和极速版本里的sample_rate，这两个接口是一样的吗？

不是。极速版的sample_rate就是采样率，enable_sample_rate_adaptive是个开关，极速版默认采样率16k，不需要这个开关。

录音转文本能区分坐席和客户吗？

语音识别引擎只能区分出说话的不同角色，角色对应的身份引擎是无法识别的，需要用户从业务的角度自行判断。建议您在存储录音时按照角色分类存储。

智能语音交互的一句话识别，标点符号是根据什么来判断逗号和句号的？

结合音频的声学特征和对识别结果文本做语音分析后做标点处理。

离线文件转写如何区分左右声道？

语音识别引擎无法区分左右声道，当多声道音频送入语音识别服务进行识别时，返回结果会用channel_id字段来标记多个音轨。如果采集顺序固定，可以根据channel_id区分对应声道。具体可参见接口说明。

语音识别可以支持多个词表吗？

一次可使用一个词表，每次只能传一个vocabulary_id，如果觉得不够可以使用定制模型。具体可参见使用POP API创建业务专属热词。

设置录音文件识别服务的版本，"4.0"和"2.0"两个版本有什么区别？

由于历史原因，早期发布的录音文件识别服务（默认为2.0版本）的回调方式和轮询方式的识别结果在JSON字符串的风格和字段上均有不同。录音文件识别服务在4.0版本对回调方式做了优化，使得回调方式的识别结果与轮询方式的识别结果保持一致，均为驼峰风格的JSON格式字符串。具体可参见接口说明。

在电话端支持哪些国家的语音识别？

电话8k语音目前支持的外语语种为英语，非电话16k语音支持更多外语语种。语种模型支持列表详见功能特性。

在语音识别的服务中，有没有请求参数是音频文件地址，返回参数是转写文本？

可使用录音文件识别功能，具体请参见接口说明。

实时语音转写能和录音文件识别一样加入音轨ID吗？

不能，音轨ID是录音文件专用的。实时转写只有单通道语音，不需要channel区分。

录音文件识别可以生成SRT字幕文件吗？

目前没有。需要根据返回结果中的时间信息自己进行拼接。

语音识别服务支持哪些编码格式的音频？

每种服务支持的格式不尽相同，请参见各服务中的说明。您可以使用常见音频编辑软件如Audacity，查看音频文件的编码格式。

语音识别服务支持哪些采样率？

一般支持8000 Hz、16000 Hz的采样率。 电话客服场景通常是8000采样率，如果是手机App、PC端工具、网页H5类场景，通常是16000 Hz采样率（可能会有32、44k采样率，但开发时需要调用方将采样率调整为16k）。其他采样率的录音数据需要预先自行转码。

录音文件转写可以支持其他采样率，例如32k、44k等等。

怎么查看音频文件的采样率？

可以使用常见音频编辑软件如Audacity查看音频文件的采样率，也可以使用开源命令行工具FFmpeg查看。

语音识别服务支持的方言模型和语种都有哪些？

语音识别目前支持的语种和方言模型如下：

- 语种

语言	模型名称	采样率	标点	ITN	顺滑	语义断句	声音和文本对齐
英语	通用-英文，教育直播-英文，教育内容分析-英文，	16k	支持	支持	支持	不支持	支持
	电话客服（通用）	8k	支持	支持	支持	不支持	不支持
日语	通用-日语	16k	支持	不支持	不支持	不支持	不支持
西班牙语	通用-西班牙语	16k	支持	不支持	不支持	不支持	不支持
阿拉伯语	通用-阿拉伯语	16k	支持	不支持	不支持	不支持	不支持
哈萨克语	通用-哈萨克语	16k	支持	不支持	不支持	不支持	不支持
韩语	通用-韩语	16k	支持	不支持	不支持	不支持	不支持
泰语	通用-泰语	16k	支持	不支持	不支持	不支持	不支持

语言	模型名称	采样率	标点	ITN	顺滑	语义断句	声音和文本对齐
印尼语	通用-印尼语	16k	支持	不支持	不支持	不支持	不支持
	电话客服（通用）	8k	支持	不支持	不支持	不支持	不支持
俄语	通用-俄语	16k	支持	不支持	不支持	不支持	不支持
越南语	通用-越南语	16k	支持	不支持	不支持	不支持	不支持
法语	通用-法语	16k	支持	不支持	不支持	不支持	不支持
德语	通用-德语	16k	支持	不支持	不支持	不支持	不支持
意大利语	通用-意大利语	16k	支持	不支持	不支持	不支持	不支持
印地语	通用-印地语	16k	支持	不支持	不支持	不支持	不支持
马来语	通用-马来语	16k	支持	不支持	不支持	不支持	不支持
菲律宾语	通用-菲律宾语	16k	支持	不支持	不支持	不支持	不支持
	电话客服（通用）	8k	支持	不支持	不支持	不支持	不支持
泰米尔语	通用-泰米尔语	16k	支持	不支持	不支持	不支持	不支持
葡萄牙语	通用-葡萄牙语	16k	支持	不支持	不支持	不支持	不支持
土耳其语	通用-土耳其语	16k	支持	不支持	不支持	不支持	不支持
波兰语	通用-波兰语	16k	支持	不支持	不支持	不支持	不支持
乌克兰语	通用-乌克兰语	16k	支持	不支持	不支持	不支持	不支持

语言	模型名称	采样率	标点	ITN	顺滑	语义断句	声音和文本对齐
罗马尼亚语	通用-罗马尼亚语	16k	支持	不支持	不支持	不支持	不支持
荷兰语	通用-荷兰语	16k	支持	不支持	不支持	不支持	不支持

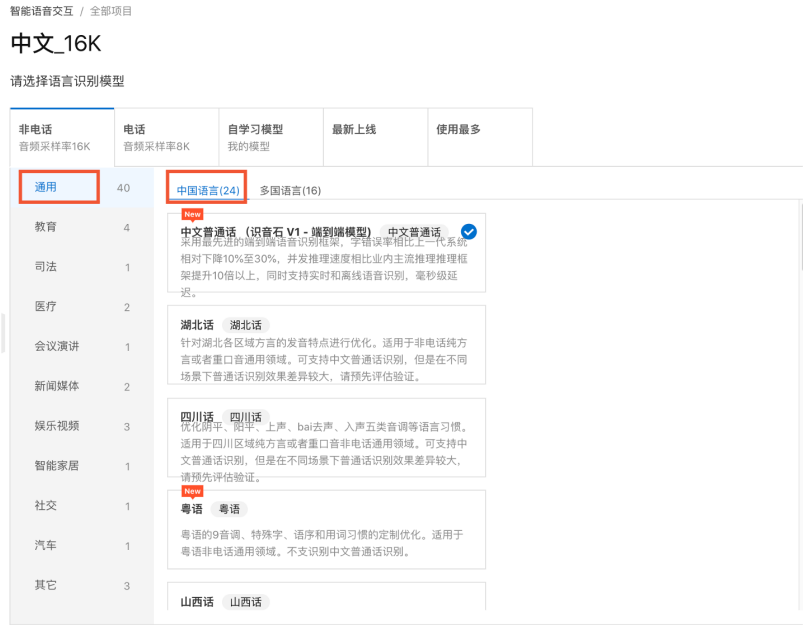
- 方言

语言	模型名称	采样率	标点	ITN	顺滑	语义断句	声音和文本对齐
粤语	通用-粤语	16k	支持	支持	支持	不支持	支持
	电话客服（通用）	8k	支持	支持	支持	不支持	支持
四川话	通用-四川话	16k	支持	支持	支持	支持	支持
	电话客服（通用）	8k	支持	支持	支持	支持	支持
湖北话	通用-湖北话	16k	支持	支持	支持	支持	支持
	通用-湖北话	8k	支持	支持	支持	支持	支持
上海话	通用-上海话	16k	支持	支持	支持	支持	不支持
湖南话	通用-湖南话	16k	支持	支持	支持	支持	支持
河南话	通用-河南话	16k	支持	支持	支持	支持	支持
	通用-河南话	8k	支持	支持	支持	支持	支持
浙江话	通用-浙江话	16k	支持	支持	支持	支持	不支持
东北话	通用-东北话	16k	支持	支持	支持	支持	支持

语言	模型名称	采样率	标点	ITN	顺滑	语义断句	声音和文本对齐
山东话	通用-山东话	16k	支持	支持	支持	支持	支持
天津话	通用-天津话	16k	支持	支持	支持	支持	支持
陕西话	通用-陕西话	16k	支持	支持	支持	支持	支持
山西话	通用-山西话	16k	支持	支持	支持	支持	支持
贵州话	通用-贵州话	16k	支持	支持	支持	支持	支持
云南话	通用-云南话	16k	支持	支持	支持	支持	支持
甘肃话	通用-甘肃话	16k	支持	支持	支持	支持	支持
维吾尔语	通用-维吾尔语	16k	不支持	不支持	不支持	不支持	不支持
	通用-维吾尔语	8k	不支持	不支持	不支持	不支持	不支持
苏州话	通用-苏州话	16k	支持	支持	支持	支持	不支持
闽南语	通用-闽南语	16k	支持	支持	支持	支持	不支持
江西话	通用-江西话	16k	支持	支持	支持	支持	支持
宁夏话	通用-宁夏话	16k	支持	支持	支持	支持	支持
广西话	通用-广西话	16k	支持	支持	支持	支持	支持
	通用-广西话	8k	支持	支持	支持	支持	支持

语言	模型名称	采样率	标点	ITN	顺滑	语义断句	声音和文本对齐
中文普通话	识音石 V1 - 端到端模型，教育内容分析，医疗内容分析，新闻媒体内容分析，娱乐视频内容分析，音视频离线转写（升级版），新零售领域识别模型，出行领域识别模型，电话客服（升级版）	16k	支持	支持	支持	支持	支持

最新的模型支持情况可以登录智能语音交互控制台，在项目配置中查看。



语音识别能否自动断开多句话？

实时语音识别服务可以断开多句话；一句话识别服务的每个请求只对应一句话，无法断开。

实时识别和录音文件转写分别支持哪些语音格式？

	8k 8bit	8k 16bit	16k 8bit	16k 16bit	单声道	双声道	支持格式
录音文件识别	√	√	√	√	√	√	PCM（无压缩的PCM或WAV文件）
实时语音识别	×	√	×	√	√	×	WAV格式、MP3格式

性能类

语音识别的识别准确率怎么计算？

行业通常使用错误率来统计识别效果，中文常用CER（字错误率），英文常用WER（词错误率）。计算方式：(插入错误字数ins+删除错误字数del+替换错误字数sub)/总字数。以下图为例：

```
%WER  15.07  [  2165  /  14365,  74  ins,  385  del,  1706  sub  ]
%SER  67.75  [  645  /  952  ]
Scored  952  sentences,  0  not  present  in  hyp.
```

这批数据的准确率=（14365-74-385-1706）/14365=84.93%。快速计算方式为：100-15.07=84.93。

语音识别模型的字准率能达到多少？

关于达摩院智能语音交互语音识别准确度的数字，我们通过了CNAS（国家软件测试中心）的评测，国家软件中心对语音识别算法准确度测试中，在60分呗以下的降噪环境中，用普通话在距离耳麦1厘米的位置，以240字/小时的匀速朗读样本量1207字的测试下，我们经过5轮测试的结果，识别准确率均大于98%。该准确度经过国家软件测试中心的标准认证。

而在现实的使用过程当中，可能会受到耳麦质量、背景杂音、口音差异等原因导致准确度有一定的偏差，对于数据格式为8k、16bit、双通道分轨（用户或客服双轨）的pcm或者wav格式，信噪比在20dB以上的语音，绝大部分商用场景下我们能保障85%的准确度，确保您有效使用。

录音文件识别极速版延迟是多少？

录音文件识别极速版服务承诺10秒内完成30分钟的音频识别，指的是从收到全部音频到完成识别的时间，音频上传的速度和客户端带宽等因素相关，时长可能会有不同。在服务端返回的识别结果中包含latency字段，记录了服务端处理时长。

8k模型可以识别16k的音频吗？

不可以。8K模型和16K模型只支持识别对应采样率的音频。

录音文件识别极速版调用频率有限制吗？

没有。但对并发有限制，并发数在控制台上查看。

粤语的识别准确率是多少？

一句话识别参考准确率为85%左右，实际结果受语料数据和粤语标准度影响。

15秒左右的录音文件识别大概需要多久能转换成文本呢？

录音文件识别是离线API。对于免费用户的识别任务在24小时内完成并返回识别文本；付费用户的识别任务在3小时内完成并返回识别文本。60秒以内的短音频建议客户使用一句话识别，时效更好。

语音转文本有没有优先级？比如现在正在转写任务，突然有紧急的转写任务，能调整处理优先级吗？

暂不支持这个操作，文件转写目前的时效性还是比较快的。

针对两个用户打电话场景，哪个模型效果比较好？

目前建议都使用新一代端到端“识音石”识别模型，综合效果性能比较好。

服务请求时长限制？

- 一句话识别支持60秒以内的实时语音。
- 实时语音识别不限制时长。

“流式”模式和“非流式”模式识别的区别？

“非流式”模式也称为普通模式，普通模式下，服务判断用户整句话说完后才返回一次识别结果；而“流式”模式下用户一边说话一边返回识别结果，在句子结束的识别结果前会有很多中间结果。

什么是ASR尾点延迟？

尾点延迟的定义是调用端发送音频结束到完成识别的时间。

目前语音实时识别，如一句话识别、实时转写接口的延迟在300毫秒左右，视模型、音频差异而略有不同。

一句话识别RESTful接口因为批量接收音频，识别时长和音频时长相关。不考虑网络开销，一句话识别RESTful接口处理时长和音频时长近似线性关系，简单计算可以认为：接口处理时长=音频时长*0.2。例如，1分钟音频处理时长约为12秒。实际线上性能会随模型的不同和服务器负载略有差异。

效果类

对于识别不准的词该如何进行优化？

1. 首先考虑无标注优化：
    i. 使用业务相关语料进行定制语言模型优化。业务语料包括业务关键词、业务相关的句子和篇章等。训练语料中要尽可能的对词进行泛化。(比如把“银税e贷是什么”、“如何办理银税e贷”等等相关话术加入到训练语料中)

ii. 针对依然识别不好的业务关键词，再以复制多行或者提高模型权重的方式进行定制语言模型加强。

iii. 个别解决不好的业务关键词，使用泛热词进行优化。

2. 其次考虑有标注优化：

i. 如果主要是因为口音等问题导致的整体识别效果不好，并且无标注优化方式无法解决到满意程度，可以开始声学模型优化。

ii. 声学模型优化需要标注数据，标注本身也可以加入业务相关语料中进行语言模型优化。

单字识别不出来是什么原因？

因为单字发音单元短，且没有上下文语义，所以单字的识别是难点，添加热词效果不一定好，建议采用定制化模型优化。您可以提交工单，将识别差的样例及语音（badcase）提供给技术人员，我们分析后会给您提供针对性建议。

热词效果如果不佳是否可以自主调节权重？

人名和地名类热词不支持设置权重，业务专属热词支持权重。目前在控制台上创建业务专属热词时不支持设置权重，如果需要调整权重您可以通过API进行维护。

录音文件识别时间戳不准，如何解决？

可通过设置时间戳校准功能参数enable_timestamp_alignment为true来调准。详情请参见接口说明。

语音识别太灵敏、无效声音（噪音等）被识别出了文字怎么办？

可以通过设定参数 `speech_noise_threshold` 的值来修改VAD噪声阈值。

`speech_noise_threshold` 参数区间是[-1，1]，取值越小越灵敏，可能会有更多噪音被当成语音被误识别；取值越大，可能更多语音段会被当成噪音而没有被识别。例如设为0.6，如果仍觉得太灵敏，可以继续尝试设置为0.7。如果发现有丢字、漏识别，需要将该值调小，例如0.5、0.2甚至是-0.2等。

代码示例：

- Java

```
transcriber.addCustomedParam("speech_noise_threshold", -0.1);
```

- C++

```
request->setPayloadParam("speech_noise_threshold",-0.1);
```

远场识别为什么会经常丢字？如何提高远场识别效果？

这是因为远、近场的VAD阈值不一样，建议调节参数 `speech_noise_threshold` 。

`speech_noise_threshold` 参数区间是[-1，1]，取值越小越灵敏，可能会有更多噪音被当成语音被误识别；取值越大，可能更多语音段会被当成噪音而没有被识别。例如设为-0.2，如果丢字现象仍然比较严重，可以继续调小至如-0.3、-0.4；如果发现较多噪声被误识别了，也可以适当调大，例如-0.1、0等。

ASR识别很离谱，不准确怎么办？

一般来说，ASR模型识别率均有一定的保证。如果在所有情况下语音识别都不准确，或者识别率很低，往往需要整体考虑是否有什么地方配置错误，例如实际语音的采样率（在线识别场景ASR只支持8k 16bit或者16k 16bit）、调用时设置的采样率参数（8000或者16000）、ASR服务端模型（8k或者16k），这三者需要保持一致。

> **注意**
>
> 如果是公共云ASR调用，需要确认阿里云控制台上该appkey所选择的模型采样率；如果是专有云ASR环境，则需要确认service/resource/asr/default/models/readme.txt 文件中所定义的采样率。

有些词汇总是识别不准怎么办？

在某些情况下，确实存在某些词汇识别不准的情况，如特定名词。针对此现象（下面以词汇“银税e贷”为例），我们建议有：

- 通过自学习平台的定制语言模型训练优化，例如把“银税e贷是什么”、“如何办理银税e贷”等等相关话术加入到文本语料中进行训练。
- 通过自学习平台的热词优化，例如把“银税e贷”作为热词进行训练，设置相应权重。
- 专有云可通过ASR的白名单优化，例如“银税e贷”更容易为误识别为“银税一袋”，则可以在service/resource/asr/default/models/nn_itn/correct.list中按预定格式进行设置，第一列是误识别的文本，第二列是正确文本，注意该操作要求重启ASR生效。

如何提高标点断句的效果？

默认VAD断句，可以添加参数 `enable_semantic_sentence_detection` 使用语义断句。如果是实时识别，请确认中间结果（参数 `enable_intermediate_result` ）也是开启的。

实时场景中，已经开启了标点断句，为什么效果还是不理想？

可以确认下是否开启中间结果，实时场景的语义断句需要配合中间结果使用。

录音文件识别存在一次请求后返回两次相同的结果的情况吗？

此类现象大部分是由于用户提交的语音文件是双声道，且两个声道语音内容相同造成的。如果是这种情况，属正常现象，设置参数first_channel_only为true，只识别首个声道即可解决。

实时语音识别遇到识别慢、超时问题，如何排查？

排查方式：

- 运行阿里云提供的示例，和您的服务对比运行状态，记录并提供日志信息。
- 记录请求对应的taskid，方便排查问题。
- 客户端使用TCPDump（Linux）/Wireshark（Windows）等抓包工具，确定网络状况。

为什么语音识别准确率很低，有时只识别出几个字？

请检查音频数据的采样率与管控台应用的模型是否一致，以及音频是否是单通道录音。

> **说明**
>
> 只有录音文件识别支持双通道的录音。

确认调用方式和采样率都没问题，识别还是不准确怎么办？

您可以通过如下两种方式提高识别准确率：

- 使用自定义热词功能，快速、实时提高准确率，详情请参见热词概述。
- 开通自学习模型训练，通过模型定制的方式提高大量文本的识别率，详情请参见语言模型定制概述。

SDK使用类

一句话识别录入的demo是使用Websocket进行识别展示的吗？

是Websocket。SDK就是采用的Websocket协议，RESTful API接口是HTTP协议。SDK接口详情请参见接口说明。

实时语音识别服务有Python SDK吗？

暂不支持。

语音识别的返回结果JSON中endtime=-1是什么意思？

表示当前句子未结束。当语音识别模式为“流式”时，才会存在中间结果。

计费类

录音文件识别极速版不支持试用吗？

对，录音文件极速版暂不支持试用，需要付费才可以。

3.语音合成FAQ

本文汇总了您在使用语音合成服务时的常见问题。

语音合成类常见问题主要分为以下几类：

- 功能类
  - 为什么TTS语音合成的语音和wav文件显示的时间长度不一致？例如语音文件显示长度是7秒钟，但实际语音只有不到5秒？
  - 语音合成时间戳功能是什么？
  - 语音合成时，能否控制一串数字是按数字来整体播报还是按字符来单独播报，有参数可以控制吗？
  - TTS语音合成服务有克隆技术吗？
  - 对于多音字，TTS语音合成服务发音的策略是怎么样的？
  - 个性化人声定制支持时间戳吗？
  - 长文本语音合成有调用限制吗？
- 性能类
  - 为什么TTS语音合成服务的调用有字数限制？
  - 为什么语音合成速度慢，延迟非常大？
  - 语音合成的读音正确率怎么样？
  - 语音合成的发音读错怎么办？多音字如何控制发音？
  - 为什么不同声色的语音合成音产生的延迟不一样？
  - 语音合成的时候可以识别哪些标点符号？
  - 语音合成支持部分文本调速吗？

功能类

为什么TTS语音合成的语音和WAV文件显示的时间长度不一致？例如语音文件显示长度是7秒钟，但实际语音只有不到5秒？

TTS是流式合成机制，也就是边合成边返回数据，因此保存下来的WAV文件头是一个预估的值，有一定的误差。如果对于时长要求较为严格，您可以设置format为pcm，在获取的完整的一句合成结果文件中自行添加WAV头信息，这样就会得到更为精确的时长。具体请参见接口说明。

语音合成时间戳功能是什么？

语音实时合成服务在输出音频流的同时，可输出每个汉字/英文单词在音频中的时间位置，即时间戳，时间戳功能又叫字级别音素边界接口。该时间信息可用于驱动虚拟人口型、做视频配音字幕等。具体请参见语音合成时间戳功能介绍。

语音合成时，能否控制一串数字是按数字来整体播报还是按字符来单独播报，有参数可以控制吗？

您可以尝试使用SSML功能。SSML是一种基于XML的语音合成标记语言，SSML不仅可以控制语音合成能读什么，更可以控制语音合成怎么读，包括控制断句分词方式、发音、速度、停顿、声调、音量等特征，甚至加入背景音乐。具体请参见SSML标记语言介绍。

TTS语音合成服务有克隆技术吗？

你可以使用个性化人声定制服务。个性化人声定制服务可以帮助您通过少量的录音，训练得到自己的声音模型，快速“克隆”个性化的声音，进而使用该声音讲故事、播天气、读小说、导航播报等。具体请参见界面体验。

对于多音字，TTS语音合成服务发音的策略是怎么样的？

当遇到不是词组的多音字，TTS语音合成转换的时候会根据上下文进行多音字的预测，并给出一个发音。

个性化人声定制支持时间戳吗？

目前定制的人声还不支持时间戳功能。

长文本语音合成有调用限制吗？

长文本语音合成功能提供了将超长文本（如千字或者万字）合成为语音二进制数据的功能。支持输出PCM、WAV和MP3编码格式数据；支持设置语速、语调和音量；支持设置男声、女声。您可以通过实时和异步方式获取合成结果。

长文本语音合成服务和语音合成服务的差异在于：语音合成服务只能支持300字符以下的文本，而长文本语音合成是为了满足更多用户对千字或者万字文本合成需求，最多支持10万字的一次性快速合成调用。具体请参见接口说明。

性能类

为什么TTS语音合成服务的调用有字数限制？

TTS语音合成服务调用有字数限制，是为了避免服务端资源浪费，一次性合成太多字最终未必会使用上。如果通过用API或SDK调用，可以分段调用后拼接；如果是MRCP协议调用，多用于客服或者呼叫中心场景，太多字数的TTS语音合成播放效果会持续播放较长时间，不符合人机交互逻辑，通常会被打断或提前结束。如果是超长文本，如果是千字或万字的新闻播放，可使用长文本语音合成接口，支持10万字的一次性快速合成调用。具体请参见接口说明。

为什么语音合成速度慢，延迟非常大？

随着语音合成效果的不断提升，算法的复杂度也越来越高，对用户而言，可能会遇到合成耗时变长的可能，在计算量较大的高级音色上相对更明显。因此我们建议使用流式合成机制，也就是边接收服务端返回的合成数据，边保存或者播放，可以显著改善语音合成延迟问题。

首先确认统计的是否是文本全部合成的耗时，一般只需要关注首包延迟，即客户端发送完合成请求后到第一次收到服务端返回的二进制流的时间差，即为首包延迟。

语音合成的读音正确率怎么样？

语音合成（TTS）是概率模型，目前业界能做到的读音正确率在96%~98%之间，阿里云智能语音交互产品在通用场景下测试准确率在97%左右。这意味着不是所有读音错误都能被修复掉，建议您可以通过换字或使用SSML功能。

语音合成的发音读错怎么办？多音字如何控制发音？

您可以通过以下几种方式处理：

- 可以尝试将多音字替换成同音的其他汉字快速解决发音问题。
- 您可以尝试使用SSML功能。SSML是一种基于XML的语音合成标记语言，SSML不仅可以控制语音合成能读什么，更可以控制语音合成怎么读，包括控制断句分词方式、发音、速度、停顿、声调、音量等特征，甚至加入背景音乐。具体请参见SSML标记语言介绍。
- 您可以提交工单或者添加钉钉群（35829031）进行反馈，此类方式修复周期稍长。

为什么不同声色的语音合成音产生的延迟不一样？

语音合成的实时率与模型算法的复杂度有关。最快的模型1秒内可合成33秒音频，最慢的模型1秒内可合成0.7秒的音频。普通音色和精品音色的时延不同，算法效果越好的音色相对来说耗时更长。

语音合成的时候可以识别哪些标点符号？

特殊符号也会读出相应的发音。例如：α、β、γ、ρ、sin、cos、tan；“百分号”会读成百分之几，“冒号”和“括号”会做停顿处理，“书名号”和“破折号”目前不支持识。 对于特殊符号的处理，TTS语音合成服务和正常人说话效果是相同的，该停顿的时候会停顿。

语音合成支持部分文本调速吗？

支持，您可以尝试使用SSML功能。具体请参见SSML标记语言介绍。

4.SDK FAQ

本文汇总了您在使用SDK的常见问题。

如何使用SDK设置泛热词？

SDK中使用POP API训练的泛热词，是通过控制台配置的业务专属热词表与项目Appkey绑定的，您无需自行设置；而通过POP API训练获取的业务专属热词表，需要在SDK中设置其词表ID，且SDK设置热词的优先级更高，若与控制台一起使用，将覆盖控制台设置结果。请参考使用SDK设置业务专属热词，将为您介绍在一句话识别、实时语音识别、录音文件识别中如何设置泛热词。

SDK报错“DNS resolved timeout”是什么问题？

查看 `/etc/resolv.conf` 文件中nameserver的设置，建议增加并优先使用以下配置：

`nameserver 114.114.114.114` 。

如何使用SDK设置自学习模型？

如果是通过控制台创建的自学习模型，可在项目切换模型时选择该模型，发布上线后将与Appkey绑定，您无需在代码中自行设置；而通过POP API训练获取的自学习模型，需要在SDK中设置其模型ID才可以使用。请参考使用SDK 2.0设置自学习模型，将为您介绍在一句话识别、实时语音识别、录音文件识别中如何设置自学习模型。

是否有Android和iOS的SDK，能否用在专有云下？

有SDK，在专有云安装包里默认不提供，可以通过阿里云帮助中心对应的服务文档中下载，如实时语音识别的Android SDK和iOS SDK。移动端SDK可以调用公共云ASR、TTS服务，也可以用在专有云环境下。

一句话识别、实时语音识别SDK中，send接口参数含义及使用方式？

以Java为例。java SDK中，一句话识别和实时语音识别分别提供了三个重载的 `send()` 接口。如下：

```
public void send(InputStream ins);
public void send(InputStream ins, int batchSize, int sleepInterval);
public void send(byte[] data);
```

三个接口使用时要保证持续、实时地向服务端发送语音数据。

demo是用语音文件模拟实时语音流的速度发送语音，通常一次发送间隔时间为100ms或200ms（sleepInterval）的语音数据，数据量（batchSize）和采样率有关：发送间隔过大，会导致延迟较大，容易断连；发送间隔过小，会消耗服务端和网络资源。可以适当实验调整到合适数值。

第2个接口中 `ins` 为模拟语音流，需要控制发送速率。ins中的数据每间隔100ms，发送3200字节（16000采样率）。调用示例：

```
public void send(ins, 3200, 100); // 16khz语音
```

第3个接口中 `data` 为一次性发送的数据，控制循环调用的间隔，调用示例：

```
recognizer.send(data); // 100ms语音数据
try {
 Thread.sleep(100);
} catch (InterruptedException e) {
 e.printStackTrace();
}
```

如何结合SDK日志，分析延迟问题？

以Java SDK日志为例。

- 一句话识别的延迟为一句话说完开始，到收到最终识别结果为止，消耗的时间。

  在日志中搜索关键字 `StopRecognition` 以及 `RecognitionCompleted` ，分别找到语音发送完毕时的日志，以及一句话识别结束的日志。记录的时间差即为SDK端记录的一句话延时，如下日志延迟为：984-844=140（ms）。

```
14:24:44.844 DEBUG [           main] [c.a.n.c.transport.netty4.NettyConnection] thread:1,
send:{"header":{"namespace":"SpeechRecognizer","name":"StopRecognition","message_id":"bcc
ac69b505f4e2897d12940e5b38953","appkey":"FWpPCaVYDRp6J1rO","task_id":"8c5c28d9a40c4a229a5
345c09bc9c968"}}
14:24:44.984 DEBUG [ntLoopGroup-2-1] [  c.a.n.c.p.asr.SpeechRecognizerListener] on messag
e:{"header":{"namespace":"SpeechRecognizer","name":"RecognitionCompleted","status":200000
00,"message_id":"2869e93427b9429190206123b7a3d397","task_id":"8c5c28d9a40c4a229a5345c09bc
9c968","status_text":"Gateway:SUCCESS:Success."},"payload":{"result":"北京的天气。","durati
on":2959}}
```

- 语音合成关注首包延迟，即从发送合成请求开始，到收到第一个语音包为止，消耗的时间。

  日志中搜索关键字 `send` ，找到这条日志和紧随其后的一条收到语音包的日志。记录的时间差即为SDK端记录的首包延时。如下日志延时为1035-813=222（ms）。

```
14:32:13.813 DEBUG [           main] [c.a.n.c.transport.netty4.NettyConnection] thread:1,
send:{"payload":{"volume":50,"voice":"Ruoxi","sample_rate":8000,"format":"wav","text":"国
家是由领土、人民、文化和政府四个要素组成的，国家也是政治地理学名词。从广义的角度，国家是指拥有共同的语
言、文化、种族、血统、领土、政府或者历史的社会群体。从狭义的角度，国家是一定范围内的人群所形成的共同体
形式。"},"context":{"sdk":{"name":"nls-sdk-java","version":"2.1.0"},"network":{"upgrade_co
st":160,"connect_cost":212}},"header":{"namespace":"SpeechSynthesizer","name":"StartSynth
esis","message_id":"6bf2a84444434c0299974d8242380d6c","appkey":"FWpPCaVYDRp6J1rO","task_i
d":"affa5c90986e4378907fbf49eddd283a"}}
14:32:14.035  INFO [ntLoopGroup-2-1] [  c.a.n.c.protocol.tts.SpeechSynthesizer] write arr
ay:6896
```

- 实时语音识别SDK日志类似一句话识别，可以从日志中计算语音末尾处延迟（关键字：`StopTranscription` 和 `TranscriptionCompleted` ）。
- RESTful形式访问，客户端自带日志中没有体现延迟。需要用户自己编写代码，或者查看服务端日志。

Token如何使用？

公共云Token在不同项目间、不同进程间、不同线程间都可以共用，同时一定要注意在Token过期前提前重新获取Token；出于安全考虑，建议您在服务端集成Token SDK，客户端在服务端获取Token。详情请参见获取Token。

专有云Token目前默认都是default，暂时无需修改。

Token重新获取会不会导致已获取的Token失效？

Token重新获取不会影响已获取Token的有效性，有效性只与时间有关。Token的有效期通过服务端响应消息的 `ExpireTime` 参数获取。更多信息，请参见获取Token协议说明。

是否可以提供服务的IP白名单？

服务器的IP比较多，且会随着扩容、机器故障置换等原因动态变化，当前的IP列表不代表后续都可以使用。建议增加以下两个域名的访问规则：nls-meta.cn-shanghai.aliyuncs.com、nls-gateway.cn-shanghai.aliyuncs.com，开通80、443端口，HTTP和websocket协议。

C++ SDK语音合成时传入的文本没有采用UTF-8编码会有什么错误信息？

如果传入的文本没有采用UTF-8编码，在文本中含有中文字符时，语音合成SDK调用start函数会失败，返回错误信息： `Socket recv failed, errorCode: 0` 。错误码为0表示服务端已经关闭了连接，此时应检查传入的文本是否采用UTF-8编码。

如何进行log控制与音频保存？

SDK提供多级log控制，在初始化接口中配置。

同时提供音频数据的保存方便问题定位，需要设置save_wav和debug_path初始化参数，详情请参见接口说明。

> **说明**
>
> 实时语音识别的save_wav和debug_path参数含义与一句话识别相同。

调用上有什么限制？

SDK已经对语音服务的访问做了封装，对您而言只要调用开始接口，在回调中进行适当事件处理。一般需要处理错误事件和识别结果事件。注意不能在回调中直接调用SDK的接口，可能导致死锁发生。

为什么链接不到framework？

framework中代码采用Objective-C和C++混合编写而成，所以需要使用.mm后缀文件进行调用，同时请确保工程的头文件路径与库文件路径设置正确。

SDK错误事件如何定位原因？

在SDK上报错误事件时，首先根据错误码查看原因，一般能够根据错误码获得问题的初步原因。

如果错误码为DEFAULT_NLS_ERROR，需要查看事件返回的结果字段，其中包含服务端的错误码、错误信息和task_id，然后根据这些信息进行分析或提交工单获得支持。

是否支持后台处理？

SDK本身不限制前后台，iOS SDK的样例工程默认仅支持前台处理，如果您需要支持后台处理，可以做如下修改：

1. 在工程Info.list中添加Required background modes配置，并在该配置下添加item，Value设置为 `App plays audio or streams audio/video using AirPlay` 。

附件预览

Key	Type	Value
▼ Information Property List	Dictionary	(18 items)
Localization native development re...	String	$(DEVELOPMENT_LANGUAGE)
Bundle display name	String	AliyunNlsDemo
Executable file	String	$(EXECUTABLE_NAME)
Bundle identifier	String	$(PRODUCT_BUNDLE_IDENTIFIER)
InfoDictionary version	String	6.0
Bundle name	String	$(PRODUCT_NAME)
Bundle OS Type code	String	APPL
Bundle version string (short)	String	1.0
Bundle version	String	1
Application requires iPhone enviro...	Boolean	YES
Privacy - Microphone Usage Desc...	String	应用需要访问您的录音权限
▼ Required background modes	Array	(1 item)
Item 0	String	App plays audio or streams audio/video using AirPlay
Application supports iTunes file sh...	Boolean	YES
Launch screen interface file base...	String	LaunchScreen
Main storyboard file base name	String	Main
▶ Required device capabilities	Array	(1 item)
▶ Supported interface orientations	Array	(3 items)
▶ Supported interface orientati...	Array	(4 items)

关闭

2. 在录音模块中进入后台时，不停止录音。亦即NLSVoiceRecorder.m中_appResignActive接口中不做停止录音调用。

附件预览

NUIdemo 〉 NUIdemo 〉 audio 〉 NLSVoiceRecorder.m 〉 -_appResignActive

```
347 - (void)_unregisterForBackgroundNotifications {
348     [[NSNotificationCenter defaultCenter] removeObserver:self];
349 }
350
351
352 - (void)_appResignActive {
353     _inBackground = true;
354 //    AudioSessionSetActive(NO);
355 }
356
357 - (void)_appEnterForeground {
358     _inBackground = false;
359 }
360
361 @end
362
```

关闭

5.计费定价FAQ

本文汇总了您在使用智能语音交互产品中关于计费相关的常见问题。

录音文件识别所有的调用都会计费，还是只有识别成功的才会计费？

只有服务调用成功才会计费。

智能语音如何区分不同项目的计费？

不同项目可以对应不同的Appkey，可以根据Appkey查看用量，计算计费情况。

刚刚购买了资源包，为什么控制台上显示不出来？

如果当日购买资源包，在次日（T+1）才会在控制台上显示余量。

语音识别和语音合成调用如果出错是否计费？

请求的状态码如果是2xx，则正常计费，其他状态码如4xx或者5xx不会计费。

6.自学习平台FAQ

本文汇总了您在自学习平台时的常见问题。

语音识别类常见问题主要分为以下几类：

- 功能类
  - 自学习模型的调用限制是什么？
  - 控制台与POP API设置自学习模型的区别是什么？
  - 如何通过控制台添加热词？
  - 如何通过控制台设置泛热词的权重？
  - 如何通过POP API创建热词词表？
  - 热词有数量限制吗？
  - 除了热词之外，是否还有手段可以优化识别效果？
- 效果类
  - 自学习模型如何优化训练语料？

功能类

自学习模型的调用限制是什么？

自学习模型的调用限制具体请参见调用限制。

控制台与POP API设置自学习模型的区别是什么？

具体请参见控制台与POP API设置自学习模型的区别。

如何通过控制台添加热词？

具体请参见创建热词。

如何通过控制台设置泛热词的权重？

控制台不支持调整泛热词的权重，您可以通过POP API进行调整。具体请参见使用POP API创建自学习模型。

如何通过POP API创建热词词表？

您可以使用POP API，在客户端自行操作训练业务专属热词，而无需依赖控制台的设置。可执行创建词表（`CreateAsrVocab`）、获取词表（`GetAsrVocab`）、更新词表（`UpdateAsrVocab`）、删除词表（`DeleteAsrVocab`）和列举所有词表（`ListAsrVocab`）等相关操作。具体请参见使用POP API创建自学习模型。

热词有数量限制吗？

单个词表中热词数量最多为128个，每个账号可创建10个词表。如果您的热词数量超过限制，建议使用自学习平台训练定制语言模型。

除了热词之外，是否还有手段可以优化识别效果？

您还可以使用自学习平台中的语言模型定制功能，进行文本优化。

效果类

自学习模型如何优化训练语料？

具体请参见优化建议。