

ALIBABA CLOUD

阿里云

智能语音交互
常见问题

文档版本：20200820

 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。
<code>Courier</code> 字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
<i>斜体</i>	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1. 服务开通问题	05
2. 服务使用问题	06
3. 语音识别问题排查	08

1. 服务开通问题

本文为您解答服务开通的相关问题。

1. 如何开通服务？

关于开通智能语音服务，请参见[开通服务](#)。

2. 如何确认我开通了服务？

登录[智能语音交互服务](#)，查看是否有管理控制台按钮。如果没有说明未开通成功，需要单击 **立即开通**。

3. 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。

4. 智能语音服务可以使用RAM子账号吗？

建议关注安全性的用户使用RAM子账号。在管控台的账户头像的下拉列表，单击**访问控制**，创建并管理RAM用户。关于RAM请参见[访问控制基本概念](#)。

2. 服务使用问题

本文为您解答服务使用的相关问题。

appkey是什么？

您可能有多个业务需要智能语音服务，如客服场景和司法场景，每个场景需要的语音能力是不同的。appkey就是用来区别不同业务场景的标识。只有在appkey项目中设置了合适的场景，才能获得理想的效果。

“流式”模式和“非流式”模式识别的区别？

“非流式”模式也叫“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。

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？

目前语音识别服务仅支持16KHz和8KHz两种采样率，其他采样率如48KHz建议重采样到16KHz，再调用语音识别服务。请注意选择和语音文件采样率对应的appkey。

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。

语音识别服务支持离线功能吗？

目前不支持本地离线的语音识别，必须将音频数据发送至服务端做识别。

智能语音交互能的域名是什么？

智能语音交互服务的域名如下：<wss://nls-gateway.cn-shanghai.aliyuncs.com/ws/v1>。

现在有对识别结果进行敏感词屏蔽吗？

目前未提供此功能。您获取识别结果后可以按需处理。

语音识别服务支持英文识别吗？

支持英文识别。请在管控台的配置项目语音识别模型中选择英文模型。要求语音采样率为16KHz。该服务支持但不限于英式、美式及中式等口音的英语。

语音识别服务支持方言识别吗？

语音识别服务支持方言识别。具体方言模型请在管控台中进行设置，请参见[管理项目](#)。

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。

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。

服务请求时长限制？

- 一句话识别支持60s以内的实时语音。
- 实时语音识别不限制时长。

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，token的有效性只与时间有关。

是否可以提供服务的IP白名单？

由于服务器的IP范围很广，不能提供IP白名单。

实时语音识别，识别慢、超时问题？

排查方式：

1. 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。
2. 记录请求对应的taskid，方便排查问题。
3. 客户端使用TCPDump (Linux) /Wireshark (Windows) 等抓包工具，确定网络状况。

3. 语音识别问题排查

本文为您介绍语音识别中出现问题的排查步骤及解决方案。

排查步骤

1. 使用cooledit或者Adobe Audition软件查看语音格式，播放试听并查看分轨情况、波形、能量和频谱图。
ASR识别标准格式：8KHz或16KHz采样率、16bit采样位数、单声道的语音数据（录音文件识别服务可支持双声道语音数据识别）。
2. 检查管控台项目中使用的模型是否支持音频采样率和场景。
3. 播放声音进行试听，重点关注如下两点：
 - i. 是否存在噪音。若存在，是人噪（人发出的声音或者远场非主说话人的声音）还是非人噪（如敲桌子、开门、汽车鸣笛）。
 - ii. 发音清晰度、辨识度如何。是否存在吞音、语速过快或者重口音、方言等情形。
4. 查看波形、能量和频谱图，重点关注波形幅度和频段信息。
 - i. 波形幅度是否过小或过大。以8KHz采样率语音为例，图1为正常语音波形；图2波形幅度过小，语音能量过低；图3波形幅度过大，有可能造成超出系统范围被截断的情况。

图1

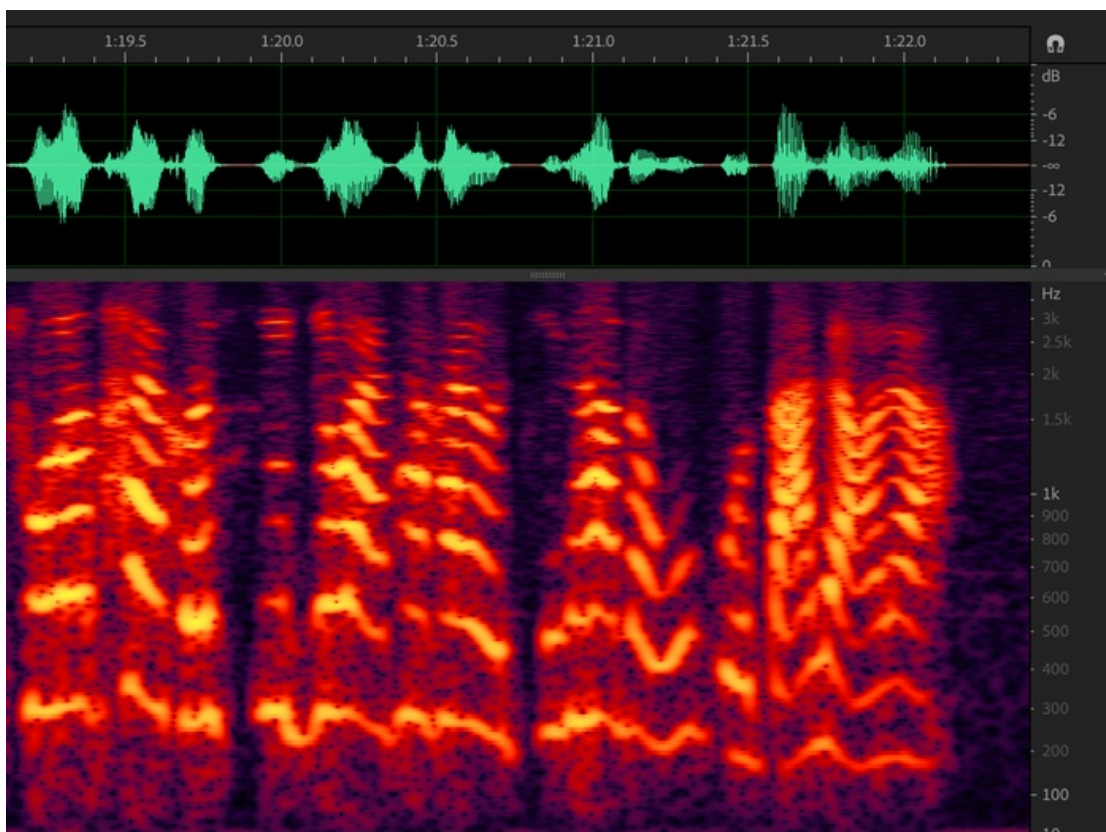


图2

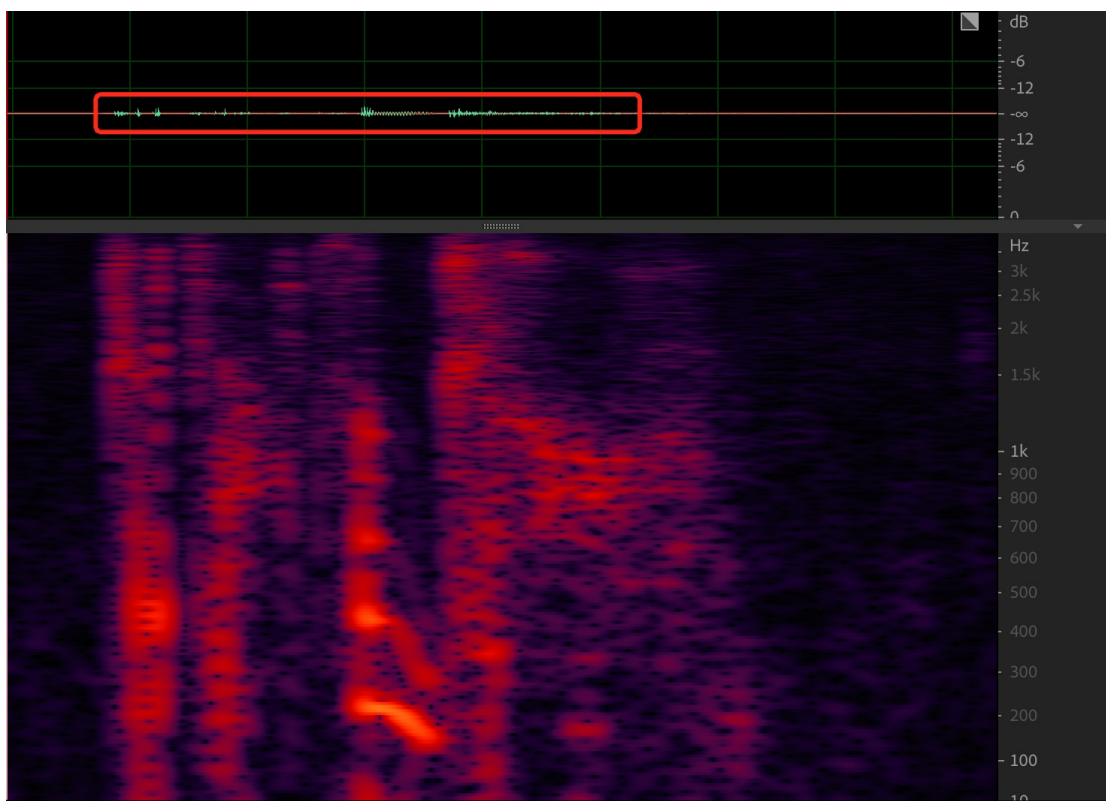
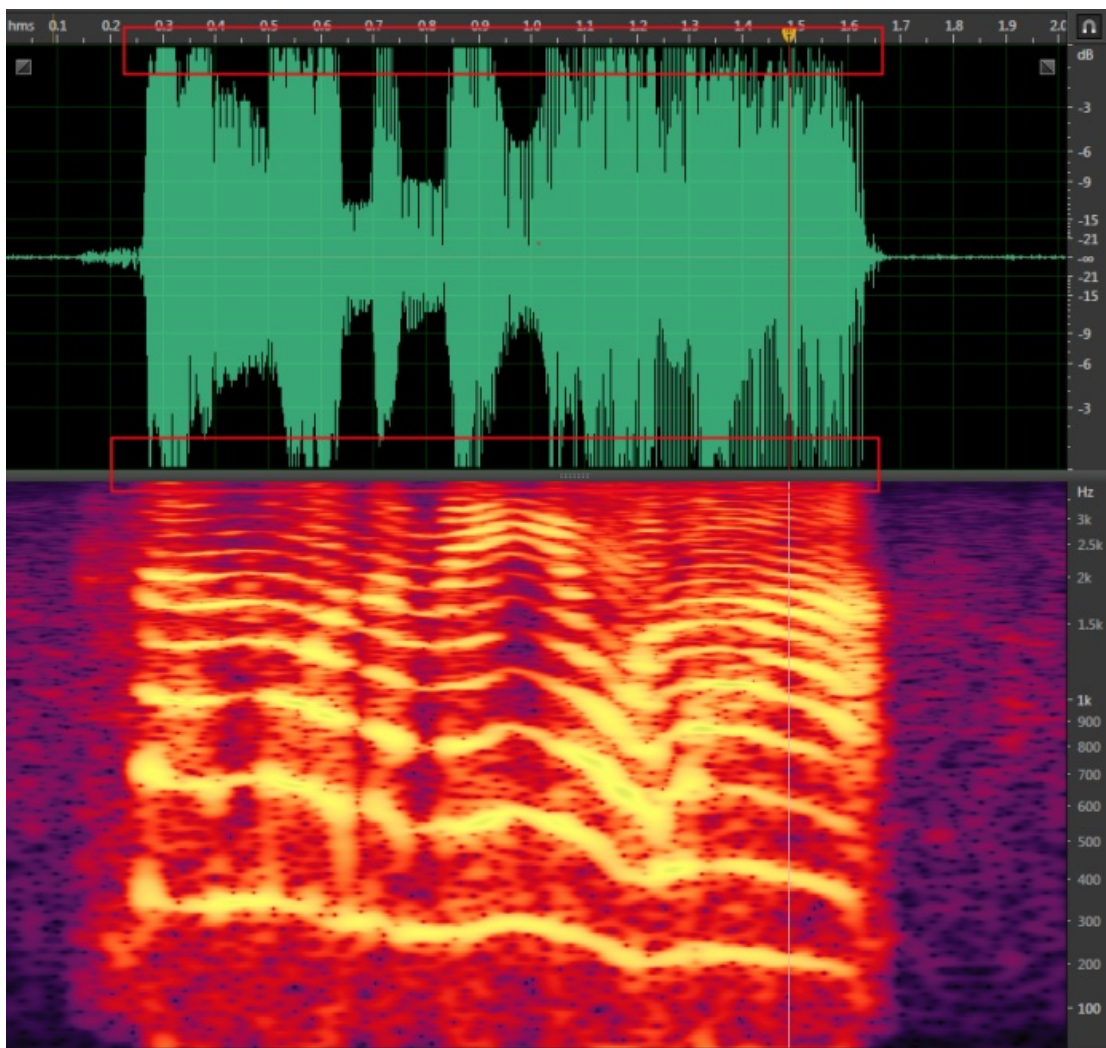


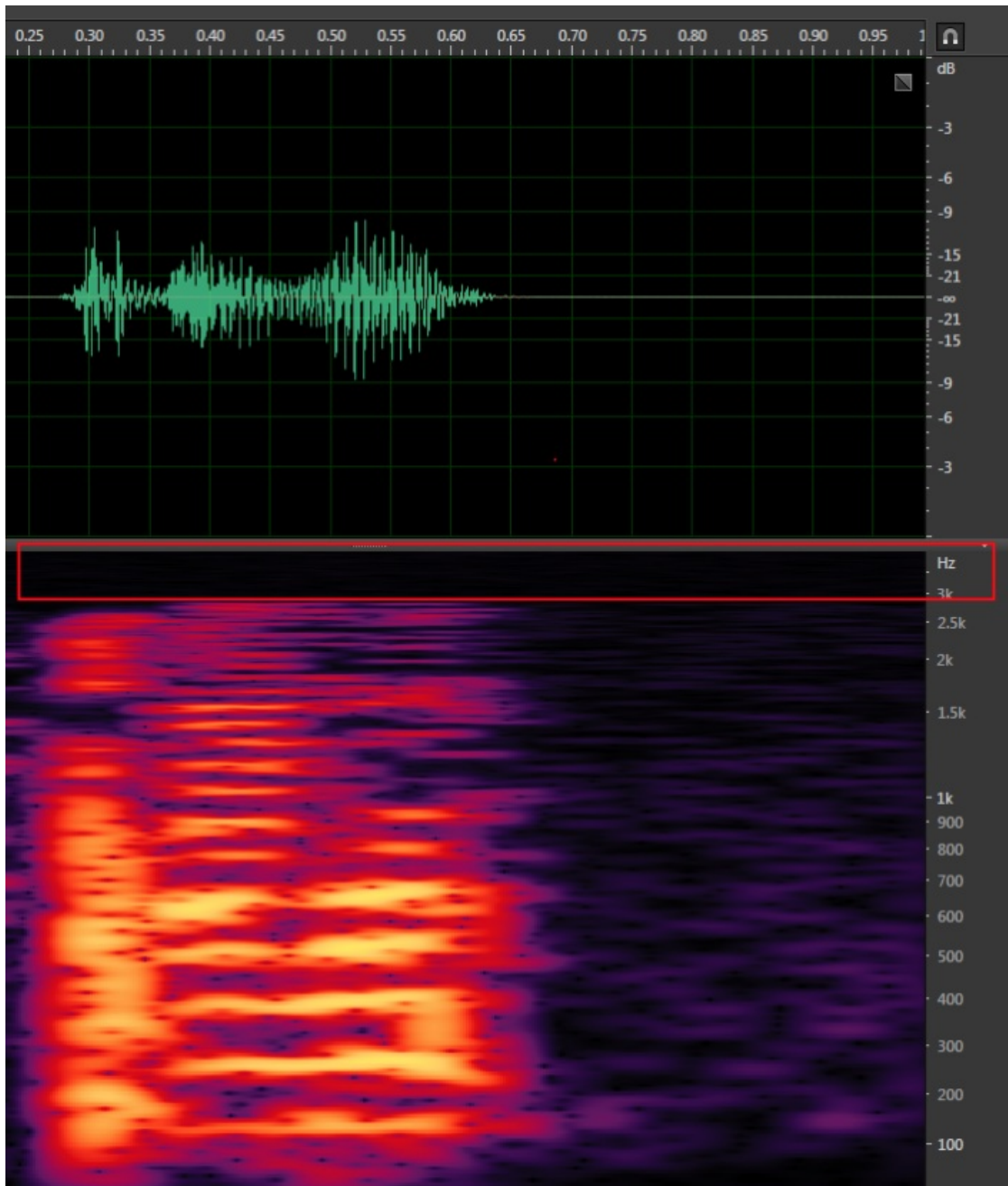
图3



ii. 频段信息是否符合要求。

以8KHz采样率语音为例，实际频段信息只有 $3 \times 2 = 6\text{K}$ （最高频段信息2倍即为实际采样率）。

图4



② 说明

对于录音文件识别服务，还需查看是合轨还是分轨数据：

以客服场景为例，合轨指客户和客服的声音存在一个声道，不免会有语音重叠的时刻；分轨指客户和客服的声音分开存储在两个声道。

5. 是否使用了热词或者自学习模型。

解决方案

② 说明

语音识别不可能达到100%识别率。

- 在项目中选择支持的音频采样率和场景的模型。
- 出现“吞音、辨识度不高、听不懂”等情况无法解决时：
 - 如果存在方言和重口音，可能由于ASR的训练数据覆盖不全造成识别错误，请联系阿里智能语音交互工程师进一步评估。
 - 如果有大量的重口音（非方言）识别需求，请联系阿里智能语音交互工程师进一步评估。
- 如果人噪被误识别，此问题很难解决。
噪声模型优先考虑只要是人发出的声音就会被送进ASR识别。
- 如果存在非人噪被误识别，您可以多收集一些噪声数据，提供给阿里云进行噪声模型优化。
- 如果波形幅度不大能量过低造成识别数据丢失，可能是由于音量太小被噪声模型当成噪声处理。
建议调整收音设备，或减小说话人与收音设备的距离。
- 如果波形幅度过大能量过高造成识别错误，可能是由于音量太大被截幅而语音失真，造成识别错误。
建议调整收音设备，或者说话人离收音设备远些。
- 如果频段信息不完整，可能会造成识别不准确，ASR模型的标准训练数据要求为频段完整的8K/16K采样率数据。
建议确保频段信息完整的基础上，对识别不准确的地方使用自学习模型进行优化。
- 如果使用了热词，业务专属热词的权重不宜过高，权重太高可能会引起语句被截断，导致后续语音无法进行识别。
- 针对一般的识别错误，可以使用语音模型优化，将识别不佳的句子（非单个词）进行多复制几遍的操作。
- 对于使用录音文件识别服务，如果是合轨数据造成识别不准确。
建议采用分轨进行存储。

如果以上条件都不符合，或者尚未能解决您的问题，请提供识别有问题的语音数据以及该数据的正确识别结果和错误识别结果，简单描述该识别问题并提交工单，并补充以下完整信息：

信息	说明
部署方式	公共云1.0/公共云2.0
调用服务	一句话识别/实时语音识别/录音文件识别
业务使用场景	无。
数据采样率	8K/16K

信息	说明
热词功能	是否使用热词功能。
语言模型定制	是否使用语言模型定制且按照优化建议进行操作，详情请参见 训练语料优化建议 。