ALIBABA CLOUD

阿里云

智能语音交互
产品简介

文档版本：20201012

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

1.什么是智能语音交互

智能语音交互（Intelligent Speech Interaction）是基于语音识别、语音合成、自然语言理解等技术，为企业在多种实际应用场景下，赋予产品“能听、会说、懂你”式的智能人机交互功能。适用于多个应用场景中，包括智能问答、智能质检、法庭庭审实时记录、实时演讲字幕、访谈录音转写等场景，在金融、保险、司法、电商等多个领域均有应用。

> 说明
>
> 全新的智能语言交互2.0版本现已发布。您可以使用自学习平台等工具改善语音识别效果，而且我们为您提供了功能更丰富的管理控制台和更易用的SDK，欢迎开通体验。

一句话识别

对时长较短（一分钟以内）的语音进行识别，适用于较短的语音交互场景，如语音搜索、语音指令、语音短消息等，可集成在各类App、智能家电、智能助手等产品中。详情请参见接口说明。

产品优势

- 识别准确率高

  国内独创的字级LC-BLSTM/DFSMN-CTC建模，相对业界传统CTC方法降低了20%的错误率，大幅提高了语音识别的精度。

- 解码速率高

  国内独创的LFR解码技术，在不损失识别精度的情况下，将解码速率提高3倍以上，大幅缩短反馈时间，提升用户体验。

- 独创的模型优化工具

  结合模型优化工具子产品，针对特定的领域定制专属模型，最大限度地提升识别效果。

- 广泛的领域覆盖

  支持金融、保险、司法、电商、智能家居等多个领域。

适用场景

- 语音搜索

  支持各种场景下的语音搜索，如地图导航、浏览器搜索等。可以集成到任何形式的手机应用中，最大限度地解放双手。

- 语音指令

  通过语音命令控制智能设备，实现快捷便利的操作。如控制空调开关、电视换台等。可以集成到智能家居等设备中。

- 语音短消息

  发送或者接收语音短消息时，利用音频转文字能力，实现音频内容快速预览。

实时语音识别

对不限时长的音频流做实时识别，达到“边说边出文字”的效果，内置智能断句，可提供每句话开始结束时间。可用于视频实时直播字幕、实时会议记录、实时法庭庭审记录、智能语音助手等场景。详情请参见接口说明。

产品优势

- 识别准确率高

  国内独创的字级LC-BLSTM/DFSMN-CTC建模，相对业界传统CTC方法降低20%的错误率，大幅提高了语音识别的精度。

- 解码速率高

  国内独创的LFR解码技术，在不损失识别精度的情况下，将解码速率提高了3倍以上，大幅缩短了反馈时间，提升用户体验。

- 独创的模型优化工具

  可以结合模型优化工具子产品，针对特定的领域定制专属模型，最大限度地提升识别效果。

- 广泛的领域覆盖

  支持金融、保险、司法、电商、智能家居等多个领域。

适用场景

- 视频实时直播字幕

  现场演讲场景、直播场景下，将视频中的音频实时转写为字幕，还可以进一步对内容进行管理。

- 实时会议记录

  将会议中的音频实时转写为文字，特别适用于电视会议等远距离场景。

- 实时法庭庭审记录

  将庭审各方在庭审过程中的语音转写为文字，供各方在庭审页面上查看，减少书记员的工作。

- 实时客服记录

  将呼叫中心的语音实时转写为文字，可以实现实时质检等。

录音文件识别

对用户上传的录音文件进行识别，可用于呼叫中心语音质检、庭审数据库录入、会议记录总结、医院病历录入等场景。详情请参见接口说明。

> **说明**
>
> 针对免费用户，系统可在24小时内完成识别并返回识别文本；针对付费客户，系统可在6小时之内完成识别并返回识别文本，一次性上传大规模数据（半小时内上传超过500小时时长的录音）的除外。有大规模数据转写需求的客户，可与售前专家另行沟通。

产品优势

- 识别准确率高

  国内独创的字级LC-BLSTM/DFSMN-CTC建模，相对业界传统CTC方法降低20%的错误率，大幅提高了语音识别的精度。

- 解码速率高

  国内独创的LFR解码技术，在不损失识别精度的情况下，将解码速率提高了3倍以上，大幅缩短了反馈时间，提升用户体验。

- 独创的模型优化工具

  结合模型优化工具子产品，针对特定的领域定制专属模型，最大限度地提升识别效果。

- 广泛的领域覆盖

支持金融、保险、司法、电商、智能家居等多个领域。

适用场景

- 呼叫中心语音质检

  上传呼叫中心的录音文件，通过录音文件识别得到文本，进一步通过文本检索，检查有无违规话术、敏感词等信息。

- 庭审数据库录入

  上传庭审记录的录音文件，进行识别后，将识别文本录入数据库。

- 会议记录总结

  对会议记录的音频文件进行识别，然后通过人工或者自动方法，对会议记录作出总结。

- 医院病历录入

  手术时通过音频记录医生的操作，通过录音文件识别得到文本，提高病例录入效率。

语音合成

通过先进的深度学习技术，将文本转换成自然流畅的语音。目前有多种音色可供选择，并提供调节语速、语调、音量等功能。适用于智能客服、语音交互、文学有声阅读和无障碍播报等场景。详情请参见接口说明。

产品优势

- 技术领先

  兼顾了多级韵律停顿，达到自然合成韵律的目的，综合利用声学参数和语言学参数，建立基于深度学习的多重自动预测模型。

- 多领域覆盖

  在智能家居、车载、导航、金融、银行、保险、证券、运营商、物流、房地产、教育等众多领域积累了大量的词库，使阿里语音合成技术对各领域、各行业的词汇发音更准确。

- 听感自然

  经海量音频数据训练，使合成音真实饱满、抑扬顿挫、富有表现力，MOS评分达到业内顶级水准。

- 深度定制

  根据用户需求定制音库，满足用户的个性化应用需求，提供标准男女声、温柔甜美女声等多风格选择，支持标记语言（SSML）方式的合成方式，音量、语速、音高等参数也支持动态调整。

适用场景

- 智能客服

  提供多行业多场景的智能客服语音合成能力。提高解答效率，提升客户满意度，降低呼叫中心人工成本。

- 智能设备

  为智能家居、音箱、车载和可穿戴设备等赋予一个最有温度的声音。

- 文学有声阅读

  让富有感染力的声音为您讲故事、读小说、播新闻，满足“懒人”的阅读需求。

- 无障碍播报

  无论是健全人还是残疾人，无论是年轻人还是老年人，将文字转成流畅动听的自然语言声音。

语音合成声音定制

为您提供深度定制的TTS（Text to Speech）声音功能：使用先进的深度学习技术，用更少的数据量，更快速高效地定制个性化的TTS声音。将自然流畅的声音输出到服务或设备上。

如果您想体验定制的声音、了解定制流程，请登录阿里云官网。如有任何需求和疑问，请联系：nls_support@service.aliyun.com。

产品优势

- 技术领先

  使用最新推出的KAN-TTS（Knowledge-Aware Neural TTS）语音合成技术，基于深度神经网络和机器学习，将文本转换成真实饱满、抑扬顿挫、富有表现力的语音。合成效果与真人录音相比，几乎可以以假乱真。

- 数据量门槛低

  在中文普通话场景，2000句起即可合成自然流畅效果的声音，加入英文数据后，还可实现中英混读效果。

- 节省成本

  由于数据量门槛低，录音和标注的时间成本大幅减少，尽显价格优势。

- 深度定制

  支持客户指定自有数据合成TTS声音。同时提供海量候选发音人资源，多种音色和风格源备选，且保证顶级录音棚采集高品质录音数据。

适用场景

- 智能客服

  提供多行业多场景的智能客服语音合成能力。提高解答效率，提升客户满意度，降低呼叫中心人工成本。

- 智能设备

  为智能家居、音箱、车载和可穿戴设备等赋予一个最有温度的声音。

- 文学有声阅读

  让富有感染力的声音为您讲故事、读小说和播新闻，满足“懒人”的阅读需求。

- 无障碍播报

  无论是健全人还是残疾人，无论是年轻人还是老年人，将文字转成流畅动听的自然语言声音。

自学习平台

您可以使用自学习平台提升识别效果。自学习平台提供了训练热词和自学习模型两种方式，帮助您提升语音识别服务的识别效果。

产品优势

- 易用

  自学习平台颠覆性地提供一键式自助语音优化方案，极大地降低进行语音智能优化所需要的门槛，让不懂技术的业务人员也可以显著提高自身业务识别准确率。

- 快速

  自学习平台能够在数分钟之内完成业务专属定制模型的优化测试上线， 更能支持业务相关热词的实时优化，一改传统定制优化长达数周甚至数月的漫长交付弊端。

- 准确

  自学习平台优化效果在很多内外部合作伙伴和项目上得到了充分验证，很多项目最终通过自学习平台不仅解决了效果可用性问题，还在项目中超过了竞争对手使用传统优化方式所取得的优化效果。

适用场景

- 热词

  在语音识别服务中，如果在您的业务领域有一些特有的词，默认识别效果较差的情况下可以使用热词功能，将这些词添加到词表，改善识别结果。

- 语言模型定制

  支持上传业务相关的文本语料训练模型，可以在该业务领域中获得更高的识别准确率。如司法、金融等领域。

学习路线

- 计量计费：了解智能语音交互服务的计费情况。
- 快速开始：快速体验智能语音交互服务。
- 开发指南：掌握相关术语、获取Access Token等内容。
- 管控台用户指南：详细了解管控台提供的各项功能。
- 选择需要的服务：一句话识别、实时语音识别、录音文件识别、语音合成等。
- 自学习平台：通过自学习平台的热词、自学习模型提升识别效果。
- 最佳实践：了解智能语音交互服务的最佳实现方式。
- 常见问题：查询常见问题的解决方案。

2.版本说明

本文介绍了智能语音交互产品发布后的更新情况。

2020年8月23日

新增

- 语音合成的SSML增加资源标签，可解析“多模态交互使用的离线资源”，并可取代时间戳中每个字的位置信息。
- 语音合成的RESTful接口支持在管控台配置说话人、音量、语速和语调参数功能，方便接口调参配置。
- 语音合成新增文学场景发音人：艾楠、艾颜、艾浩、艾茗，为您提供更多选择。

优化

实时语音识别默认最大断句时长由60秒缩短至15秒，方便您进行相关接口调用。

修复

- 语音识别16k中文通用模型，改善语音活动检测（Voice Activity Detectio）效果，解决纯静音数据误检出语音的问题。
- 语音识别8k中文客服质检/8k英文客服质检/16k韩语模型：语言模型常规更新，修复部分识别有误的场景。

2020年7月23日

新增

- 自学习模型全面开放免费使用，为您提供零成本个性化语音定制服务，助力业务创新。
- 自学习平台训练流程
  - 新增推荐最佳基线模型，方便您进行训练。
  - 结合自动化测试，增加模型可量化的测试指标结果。
- 长文本Restful接口集成字幕能力对外正式发布，官网开发文档上线。

优化

上线Android/iOS双端新版SDK：

- Android SDK体积减少34.6%、iOS SDK体积减少17.5%，经历日亿次调用次数考验，稳定性极强。
- 完善SDK的状态管理（开/关音频、数据推送等），您可以专注业务实现而无需进行复杂的状态与线程管理。
- 与全链路解决方案保持接口一致。后续可无缝对接唤醒、声纹、对话理解、离线语音合成等智能语音交互场景。

修复

英文后处理效果优化，解决部分情况下，启用标点后识别结果格式错误的问题。

2020年7月9日

优化

一句话识别/实时语音识别/录音文件识别8K音频采样率的英文识别模型更新，在通用测试集字识别准确率没有下降的情况下，提升模型口音覆盖的广度，同时在语言模型上更加通用。

修复

语音合成模型修复如下内容：

- Abby（发音人名称）：降低漏字率。
- Wendy（发音人名称）：解决较长文本合成不稳定的问题。
- 英文场景：解决英文文本出现非标空格导致单词解析失败的情况，提高单词识别准确率。
- 中文场景：修复多音字和分词问题。