

语音交互服务

产品介绍

文档版本 01
发布日期 2025-09-18



版权所有 © 华为技术有限公司 2025。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址：深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址：<https://www.huawei.com>

客户服务邮箱：support@huawei.com

客户服务电话：4008302118

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目 录

1 什么是语音交互服务.....	1
2 功能介绍.....	2
3 应用场景.....	4
4 计费说明.....	5
5 与其他云服务的关系.....	6
6 约束与限制.....	7
6.1 实时语音识别.....	7
6.2 一句话识别.....	7
6.3 实时语音合成.....	7

1 什么是语音交互服务

语音交互服务（Speech Interaction Service，简称SIS）是一种人机交互方式，用户通过实时访问和调用API获取语音交互结果。例如用户通过语音识别功能，将口述音频或者语音文件识别成可编辑的文本，同时也支持通过语音合成功能将文本转换成逼真的语音等提升用户体验。适用场景如语音客服质检、会议记录、语音短消息、有声读物、电话回访等。

使用前必读

用户需要具备编程能力，熟悉Java、Python、IOS等编程语言。

用户通过调用API接口，将语音文件识别成可编辑的文本，然后返回JSON格式的识别结果，用户需要通过编码将识别结果对接到业务系统或保存为TXT、Excel等格式。

首次使用 SIS

如果您是首次使用SIS的用户，建议您学习并了解如下信息：

- 功能介绍
通过[功能介绍](#)章节的内容，了解SIS不同功能的具体介绍，主要包括实时语音识别（Real-time ASR）、一句话识别（Short Sentence Recognition）、语音合成（Text To Speech）功能。
- 入门使用
SIS以开放API的方式提供给用户，您可以参考[《快速入门》](#)学习并使用SIS服务。
- 使用方式
如果您是一个开发工程师，熟悉代码编写，想要直接调用SIS的API使用服务，您可以参考[《API参考》](#)或[《SDK参考》](#)获取详情。
- 由浅入深学习
您可以参考[成长地图](#)，由浅入深学习使用SIS。

2 功能介绍

实时语音识别

实时语音识别服务，用户通过实时访问和调用API获取实时语音识别结果，支持的语言包含中文普通话。

- **文本时间戳**
为音频转换结果生成特定的时间戳，从而通过搜索文本即可快速找到对应的原始音频。
- **智能断句**
通过提取上下文相关语义特征，并结合语音特征，智能划分断句及添加标点符号，提升输出文本的可阅读性。
- **中英文混合识别**
支持在中文句子识别中夹带英文字母、数字等，从而实现中、英文以及数字的混合识别。
- **即时输出识别结果**
连续识别语音流内容，即时输出结果，并可根据上下文语言模型自动校正。
- **自动静音检测**
对输入语音流进行静音检测，识别效率和准确率更高。

产品优势

- **识别准确率高**
采用最新一代语音识别技术，基于深度神经网络（Deep Neural Networks，简称DNN）技术，大大提高了抗噪性能，使识别准确率显著提升。
- **识别速度快**
把语言模型、词典和声学模型统一集成为一个大的神经网络，同时在工程上进行了大量的优化，大幅提升解码速度，使识别速度在业内处于领先地位。
- **多种识别模式**
支持多种实时语音识别模式，如流式识别、连续识别和实时识别模式，灵活适应不同应用场景。
- **定制化服务**
可定制特定垂直领域的语言层模型，可识别更多专有词汇和行业术语，进一步提高识别准确率。

一句话识别

可以实现以内音频到文字的转换。对于用户上传的二进制音频格式数据，系统经过处理，生成语音对应的文字，支持的语言包含、阿拉伯语。

产品优势

- 高识别率
基于深度学习技术，对特定领域场景的语音识别进行优化，识别率达到业界领先。
- 前沿技术
使用工业界成熟的算法，结合学术界最新研究成果，为企业提供独特竞争力优势。
- 可定制化
针对客户的特定场景需求，定制垂直领域的语音识别模型，识别效果更精确。

3 应用场景

- **语音客服质检**
识别客服、客户的语音，转换为文本。进一步通过文本检索，检查有没有违规、敏感词、电话号码等信息。
- **语音短消息**
通过语音发送或者接收短消息时，将音频短消息转文字，提升阅读效率和交互体验。
- **游戏娱乐**
将游戏娱乐中的语音聊天转成文字消息，提升用户阅读效率，提升用户体验。

4 计费说明

计费项

- 按调用次数计费，返回失败的调用不计入次数。

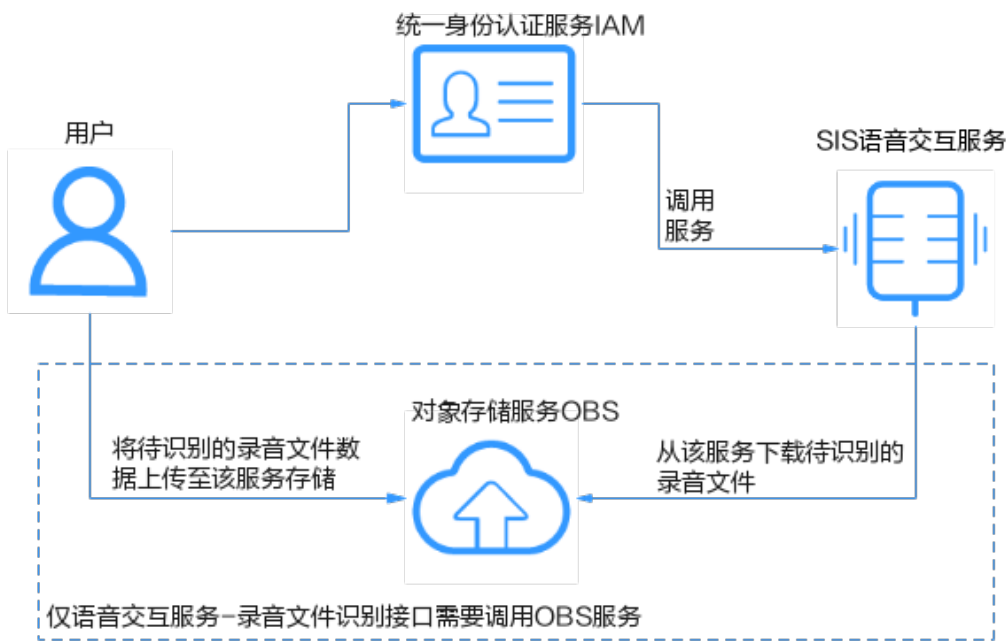
计费模式

- **按需计费**
按需计费是指按照API调用次数阶梯价格计费，详细收费明细请参考[价格计算器](#)。

5 与其他云服务的关系

SIS服务与周边其他服务的关系如图5-1所示。

图 5-1 SIS 与其他服务的关系图



统一身份认证服务

统一身份认证（Identity and Access Management，简称IAM）服务，IAM为语音交互服务提供了用户认证和鉴权功能。

6 约束与限制

6.1 实时语音识别

- 支持“亚太-新加坡”、“中东-利雅得”区域。
- 音频采样率8KHz或者16KHz，采样位数8bit或者16bit。
- 支持中文普通话、英语、阿拉伯语识别。
- 端点检测参数（vad_head、vad_tail、max_seconds）会影响句子的分段结果，存在一定误差。对连续模式和单句模式生效，对流式一句话模式无效。

6.2 一句话识别

- 支持pcm16k16bit、pcm8k16bit、ulaw16k8bit、ulaw8k8bit、alaw16k8bit、wav。
- 音频时长不超过1分钟。
- 支持“亚太-新加坡”、“中东-利雅得”区域。

6.3 实时语音合成

- 支持“中东-利雅得”区域。
- 支持阿语和英语，文本不长于10000字符。
- 支持合成采样率8k、16kHz。