语音识别

常见问题

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

常见问题
- 识别效果问题排查
- 服务与计费相关
- 功能相关
- API 与 SDK 相关
- 其他相关

常见问题
识别效果问题排查

最近更新时间：2024-06-25 16:45:11

如您在使用语音识别时，发现转写的结果与您预期有一些差距，可根据本篇文档进行问题排查。

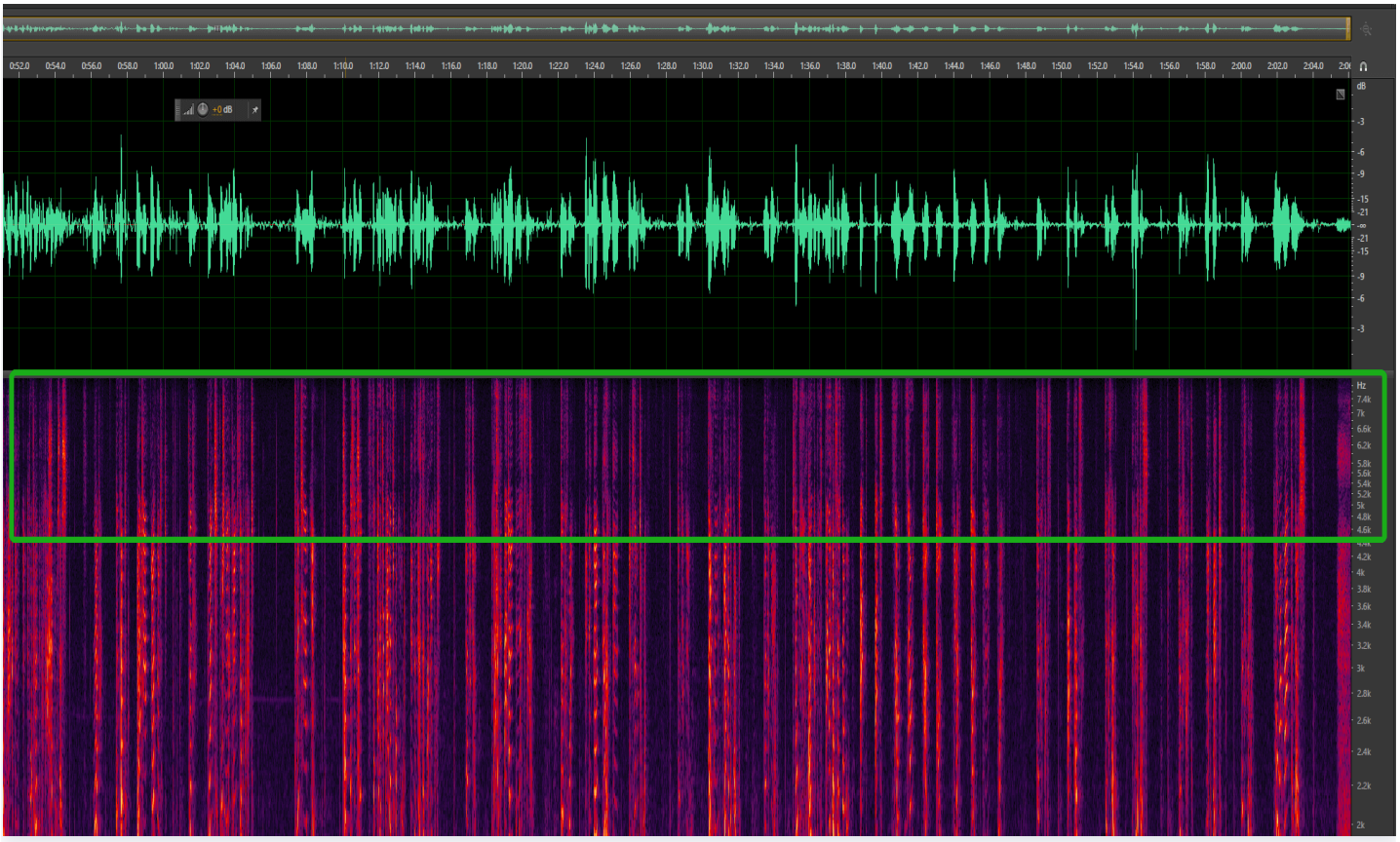
问题排查步骤

分为如下几种常见情况：

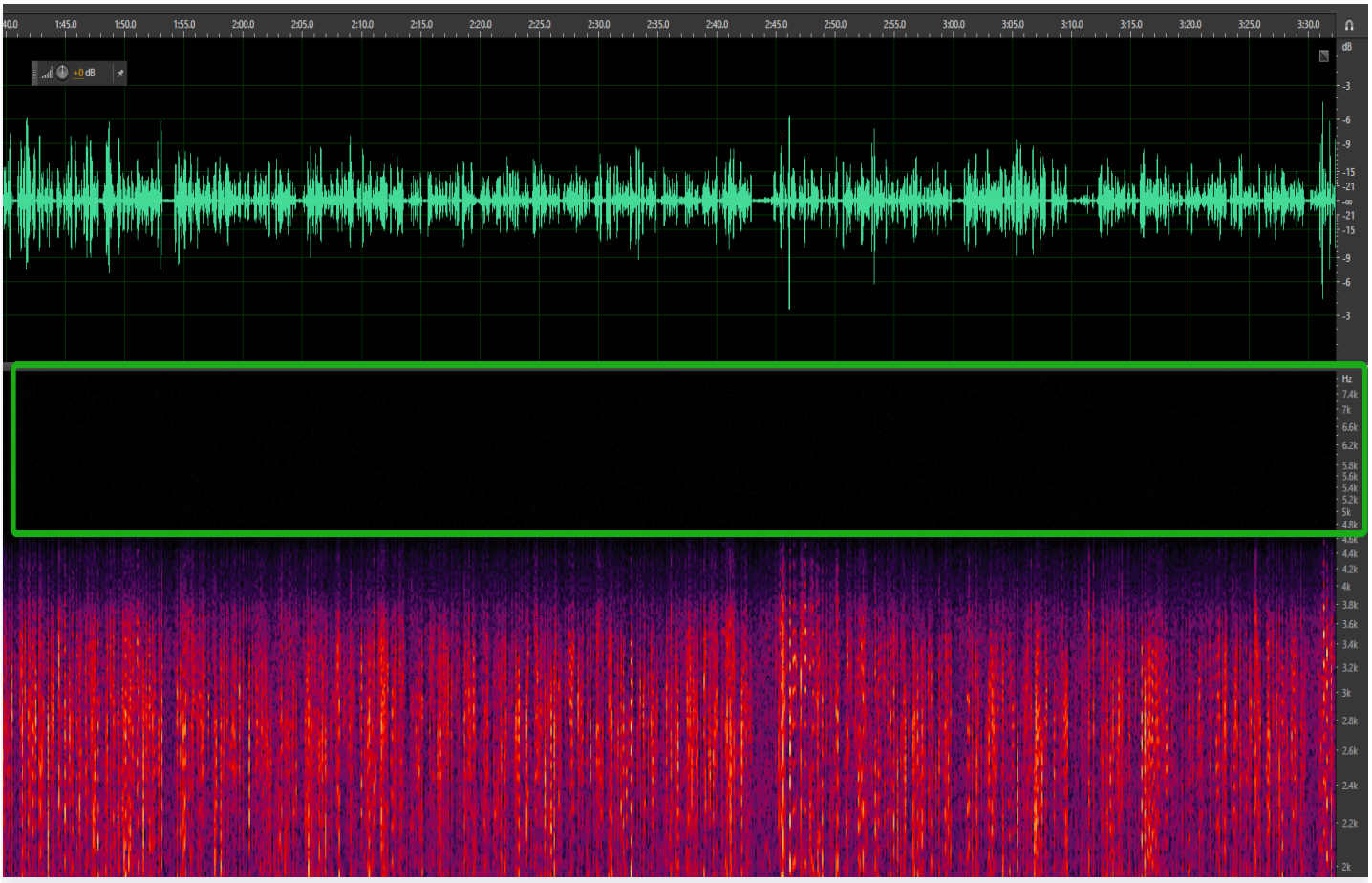
1. 本身音频内容靠正常的人耳听不清楚或者听不懂，这类情况建议对于前端的采音环境进行改造，如采音距离由远场改为近场，采音环境的噪音尽量控制和减少，口音、方言尽量控制在带口音的普通话（即非本地人可听懂）以及改善语速过快造成的吞音。
2. 本身音频内容可以听懂，但识别结果与听到的内容相差很大。这类情况一般是音频信息不满足语音识别服务要求导致的。

- 通过 cooledit、Adobe Audition 或者 FFmpeg 查看音频的详细信息，包括采样率、声道数和位深。语音识别服务目前仅支持8000Hz或者16000Hz采样率、16bits位深，录音文件识别支持单/双声道，实时语音识别和一句话识别仅支持单声道。注：如果您使用的是实时语音识别或一句话识别，音频属性必须严格符合上述要求。
- 查看音频的波形和频谱（Adobe Audition 在视图选项中）来判断实际音频真实的采样率，建议真实采样率要满足语音识别的要求（8k电话引擎模型对应8000kHz采样率，16k非电话引擎模型对应16000kHz采样率）。

真实16000Hz（真实采样率 = 框出来的右侧数值的最高值 × 2，即8kHz × 2=16kHz）音频的波形和频谱如下：



非真正16000Hz（实际为4.6kHz × 2 = 9.2kHz）音频的波形和频谱如下，可以看到在音频在4.6k频段到8k频段信息完全缺失。



3. 本身音频内容可以听懂，并且识别结果与听到的内容相差不大，但有一些特有的名词或者句子识别不好，可通过如下步骤提升识别效果：
- 将识别不好的名词参考 热词使用文档 指引进行热词的添加和使用。
- 将识别不好的名词所在句子或者识别不好的特殊句子参考 自学习模型使用文档 指引进行自学习模型的添加和使用。
4. 本身音频内容可以听懂，并且识别结果与听到的内容相差不大，但会多出来一些识别结果。这类一般是因为噪声造成的，噪声分为两种，一种是非人声的噪音，一种是人声的噪音。本身 ASR 算法对于第一类噪音做过一定的适配和优化，这一类可以提供具体的 badcase 给到腾讯来分析优化，但对于第二类人声噪音较难解决，因为会对于真正需要识别出来的人声的造成误伤。

版权所有：腾讯云计算（北京）有限责任公司

服务与计费相关

最近更新时间：2024-06-25 16:45:11

语音识别常见问题请观看视频：
观看视频

语音识别服务如何开通？

需要用户在 语音识别控制台 开通服务，默认开通的是后付费购买方式。

语音识别服务每月有免费额度吗？

- 录音文件识别免费额度为每月10小时。
- 一句话识别免费额度为每月5000次。
- 实时语音识别免费额度为每月5小时。
- 语音流异步识别免费额度为每月5小时。
- 录音文件识别极速版免费额度为每月5小时。
- 免费额度中的小时数指识别成功的音频时长； 免费额度中的次数指识别成功的次数。

2020年3月25日前，开通免费试用版用户在免费额度用完后将会暂停服务，如用户想继续使用该服务，请在语音识别控制台升级成付费版。

语音识别服务如何计费？

目前有预付费和后付费两种计费模式，预付费支持购买付费资源包，调用量的扣费顺序为：免费资源包 > 付费资源包 > 后付费，具体参见 计费概述 。

> **注意**
> 预付费资源包只能抵扣购买之后产生的调用量，购买资源包之前产生的调用量，会按照后付费方式结算。

如何查询语音识别账单？

登录 腾讯云控制台-费用中心 ，在左侧菜单栏选择**费用账单 > 账单详情**，进入账单详情页根据时间、产品以及计费模式等查看账单详情。

语音识别预付费资源包是否支持余量预警通知？

支持，免费与预付费资源包均支持余量预警，请关注站内信与短信等渠道通知。另外，您可前往 控制台 查看资源包的使用情况，详情请参见 资源包管理 。

功能相关

最近更新时间：2025-05-28 17:49:22

不同使用场景对应的是语音识别哪种服务？

- 实时语音识别适用于有实时性要求的场景，例如语音输入法、语音机器人、会议现场记录等场景。
- 录音文件识别适用于语音时间较长、实时性要求低的场景，例如客服质检、视频字幕生成等场景。
- 录音文件识别极速版适用于语音时间较长、实时性有极高要求的场景，例如视频添加字幕、准实时质检等场景。
- 一句话识别适用于对60秒之内的短音频文件进行识别的场景，例如语音短信、语音搜索等场景。
- 语音流异步识别适用于语音流准实时识别，异步返回文字结果的场景，例如直播审核、音视频审核等场景。

腾讯云语音识别针对于用户传输的音频是否仅供当次识别使用？

是的，腾讯云语音识别针对于用户传入的语音内容仅供当次识别使用，不会进行保存。

录音两个人说话，且录音存储成单声道，识别结果是否把两人的对话分离开？

8K、16K采样率普通话录音文件识别支持单通道双人对话的说话人分离。

支持远场和离在线的语音识别吗？

支持离在线语音识别，详情可参见 离在线 SDK 文档 。

语音识别支持中英文混合场景和地方方言吗？

- 实时语音识别 、一句话识别 、录音文件识别 、录音文件识别极速版 和 语音流异步识别 支持中英文混合识别（使用中文引擎的时候，在少量英文的情况下可以支持中英文混合识别，大量英文可能出现识别率降低的情况），且支持带口音的中文普通话语音识别。
- 实时语音识别 、一句话识别 、录音文件识别 和 录音文件识别极速版 支持上海话、四川话、武汉话、贵阳话、昆明话、西安话、郑州话、太原话、兰州话、银川话等23种方言的识别。

语音识别支持的输入音频时长是多少？

- 一句话识别每次调用支持60秒之内的音频。
- 录音文件识别每次调用支持五小时之内的音频。
- 实时语音音频流中每个数据包的音频分片为200ms。

语音识别支持什么样的音频属性？

接口	音频属性
录音文件识别	采样率（sample rate）：16kHz、8kHz 位深（bit depth）：16bit

	声道（channels）单声道、双声道
一句话识别	采样率（sample rate）：16kHz、8kHz 位深（bit depth）：16bit 声道（channels）：单声道
录音文件极速版	采样率（sample rate）：16kHz、8kHz 位深（bit depth）：16bit 声道（channels）：单声道、双声道
实时语音识别	采样率（sample rate）：16kHz、8kHz 位深（bit depth）：16bit 声道（channels）：单声道
说话人验证	采样率（sample rate）：16kHz 位深（bit depth）：16bit 声道（channels）：单声道
虚拟号真人判定	采样率（sample rate）：8kHz 位深（bit depth）：16bit 声道（channels）：单声道

一句话识别和录音文件识别的音频数据支持哪种传输方法和格式？

采用 HTTP 协议传输，POST 方法，音频数据传输如下两种方式：

1. 音频数据使用 base64 编码后，随 HTTP body 传输。
2. 若采用 url 下载，body 中的数据可不填，在请求参数内填入音频 url。

实时语音识别中，如果一段音频包含多句话该如何提高准确率？

建议开启 VAD 功能，VAD（Voice Activity Detection）是指对音频进行分段的技术，如果一段音频中包含了多句话，VAD 可以检测多句话之间的停顿进行自动断句，从而达到更高的识别准确度。

语音识别是否支持同步的结果调用形式？

- 实时语音识别支持同步返回识别结果。
- 一句话识别支持快速返回识别结果。
- 录音文件识别支持回调和轮询两种异步调用形式。

语音识别可以将普通话的录音文件转为英文文本吗？

不能。语音识别目前无法将普通话的录音文件转为英文文本。

语音识别是否支持评测？

不支持。

语音识别识别出来的文字能否复制?

语音识别识别出来的文字无法复制，复制功能需要接入之后自己前端开发实现。

购买了录音文件识别资源包，怎么导入文件进行识别?

可进入 语音识别控制台 功能体验页面导入文件，也可调用 API 和接入 SDK 使用。

录音转写功能都支持哪些上传文件格式?

录音转写功能支持 wav、mp3、m4a、flv、mp4、wma、3gp、amr、aac、ogg-opus、flac 格式。

实时语音识别能否设置识别的最长时间?

不能设置识别最长时间。如果不需要，直接断开即可。

语音识别是否支持 MRCP 协议?

MRCP 暂未对外开放，如有需要，可通过 售前咨询 。

有没有 SaaS 化的可以直接提供给客户用?

语音识别支持私有化部署，需要商务对接跟进，可通过 售前咨询 。

音频大于5小时或者文件超过1GB如何切割?

可以使用 ffmpeg 命令来切割音频/视频，例如音频时长为3小时，切割为三个1小时的音频，可使用如下命令：

```
ffmpeg -ss 00:00:00 -i input.wav -c copy -t 3600  output_1.wav

ffmpeg -ss 01:00:00 -i input.wav -c copy -t 3600  output_2.wav

ffmpeg -ss 02:00:00 -i input.wav -c copy -t 3600  output_3.wav
```

其中-ss 的参数为切割开始时间，-i 为切割文件名，-t 为切割音频秒数。

语音识别英文录音文件如何转换成中文?

语音识别功能是把音频内容转换成文本，不支持中英文翻译。

实时语音识别后的文字如何保存?

实时语音识别实时返回文字，您可以自行在本地进行保存。

语音识别支持什么语言?

实时语音识别支持中文普通话、英文、韩语、粤语、日语、泰语和上海话等方言，详情请参见 实时语音识别（websocket）。

一句话识别和录音文件识别支持中文普通话、英文、粤语、日语和上海话等方言的识别，详情请参见 录音文件识别 、一句话识别 。

语音识别能否保存语音文件?

语音识别上传的音视频文件是不会保存的，识别成功后存储的是识别后的文本文件，文本文件会在服务器上保存7天。保存语音文件会影响识别结果，目前结果是直接返回的，您可以在业务侧自行实现语音文件的保存，您可将语音文件保存在本地服务器或者数据库中。

语音识别的录音文件识别接口是否支持过滤语气词?

录音文件识别支持过滤语气词。具体使用方法请参见 录音文件识别请求 。

语音识别中录音文件识别接口是否支持过滤标点符号?

录音文件识别接口是支持过滤标点符号的。具体使用方法请参见 录音文件识别请求 。

语音识别准确率是多少?

腾讯云语音机器人系统在国家电子计算机质量监督检验中心出具的测试报告中，对于数据采样率16k、16bit、原始不压缩的 wav 或 pcm 格式的音频数据，中文语音识别的字准率为97.40%（结果保留两位小数），英文（美式）语音识别的字准率不低于88.00%（结果保留两位小数）。但请您知悉，前述字准率仅为第三方实验测试数据，仅供您参考，不作为腾讯云语音识别服务准确性的承诺保证。

语音识别的录音文件识别接口是否支持阿拉伯数字智能转换?

录音文件识别是支持阿拉伯数字智能转换的。具体使用方法请参见 录音文件识别请求 。

版权所有：腾讯云计算（北京）有限责任公司

API 与 SDK 相关

最近更新时间：2025-04-09 10:47:52

语音识别接口的 HTTP 请求返回鉴权失败?

可对照参数表检查自己的参数是否正确上传。若想快速接入，推荐使用官网提供的 SDK 。

语音识别服务识别结果报错无效的 URL 地址?

用户提供的 URL 地址需要是公网的 URL，能被腾讯云访问。可使用腾讯云提供的 COS 服务存放音频并使用相关的 URL。也要请用户排查防火墙是否拦截，是否内网 IP，是否存放于其他服务提供商无法被腾讯云下载等问题。

语音识别调用接口服务时报错"未注册的 AppId"?

用户未注册，用户需要按照语音识别入门开通语音识别服务方可使用服务。

语音识别接口会限制音频文件的采样率吗?

接口不会限制，但是采样率不符合标准，会影响到识别效果。

录音文件识别报错：Audio channel num not match!是什么原因?

音频的声道数需要跟差数 channel 相匹配，如音频声道数是2，那么channel = 2。

录音文件识别过了很久都没有收到回调怎么办?

请确认回调服务正常，且回调地址是否在公网下可访问。
如果回调地址是 https 的，请确认证书经过合法 CA 机构认证。

一句话识别报错：Failed to recognize audio! 是什么原因?

确认下上传的音频格式是否与参数 VoiceFormat 设置的的一样，如果不一样是可能报这个错误的。

使用语音识别 SDK 的情况下报4002鉴权失败的错误怎么排查?

使用 SDK 的情况下报这个错误是常见的原因是密钥信息不正确如：appid、secret_id、secret_key；另外如果使用的是子账户，主账户需要给子账户开通权限否则也是会报这个错误。

实时语音识别报：4008客户端上传超时?

这个错误是因为服务端超过15s没有收到音频数据，主动断开链接报的错误，建议如果没有发送数据了可以主动发送尾包断开链接。
使用 Android、iOS、JS、 小程序 SDK 内置的麦克风采集音频识别时报这个错误，有可能的原因是麦克风采集不到音频了，可按以下情况排查：

1. 是否存在麦克风占用的情况如：同时使用有 RTC 或者其它有使用麦克风的功能或 SDK（如一定要使用这些功能可使用语音识别 SDK 中提供的自定义音频源接口）。

2. 是否有存在同时开启两个识别连接的情况（SDK 只支持同时一路识别）。
3. iOS SDK 报这个错误也可能是 AVAudioSession 的状态被改变了不再是可录音状态，可在识别前修改为可录音状态。

其他相关

最近更新时间：2024-06-25 16:45:11

语音识别如何接入?

语音识别目前支持 API 和 SDK 接入，推荐 SDK 接入，详情可参见 语音识别入门 。

语音识别怎么进行功能体验?

可以通过微信搜索“腾讯云 AI 语音”小程序，选择语音识别进行体验；在 语音识别控制台 的功能体验模块，用户可以通过上传文件或者 URL 进行体验。

影响语音识别结果准确率的因素有哪些?

远离拾音器、明显噪声、严重口音等因素会影响语音识别准确率。

如何查看音频格式和属性?

Windows 系统下：
可以下载相关软件查看和修改音频格式：Adobe Audition CS6。
Linux 或者 macOS 系统下：
用 **file** 命令查看，例如：**file test.wav**
结果：

```
test.wav: RIFF (little-endian) data, WAVE audio, Microsoft PCM, 16 bit, mono 8000 Hz
```

此音频的采样率为8k，采样精度为16bit，声道为 mono，即单声道（双声道为 stereo）。

语音识别控制台功能体验怎样上传大于 5M 的文件?

语音识别控制台上作为功能体验提供给您进行测试，如果您的测试文件比较大，建议您采用上传音频 URL 方式，建议音频时长不能大于五个小时。

录音文件转文字，返回文字用时多久?

录音文件转文字返回结果受网络和音频长度、录音环境、语言标准等因素影响，具体时间需要根据参数来决定。

版权所有：腾讯云计算（北京）有限责任公司