语音识别

常见问题

产品文档

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

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。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。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。

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

版权所有：腾讯云计算（北京）有限责任公司

常见问题
识别效果问题排查

最近更新时间：2022-07-06 09:53:04

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。

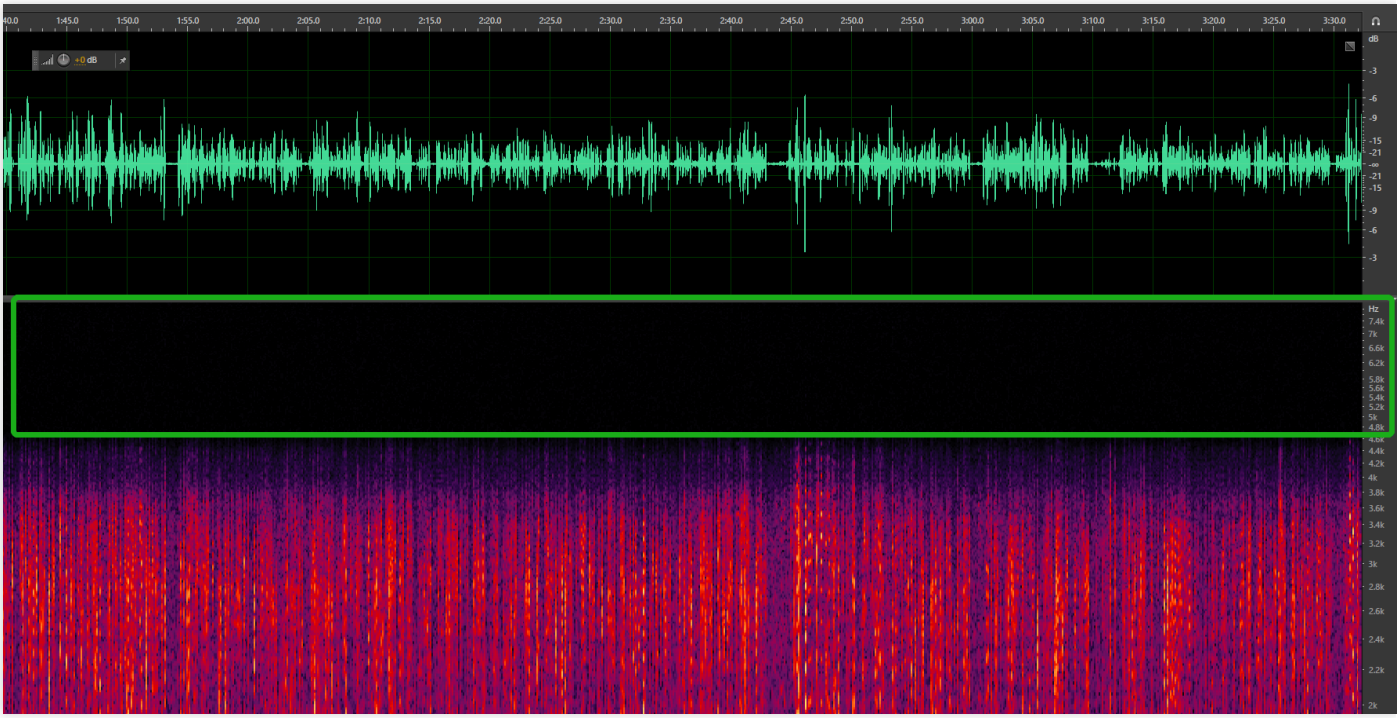
- 通过 cooledit、Adobe Audition 或者 FFmpeg 查看音频的详细信息，包括采样率、声道数和位深。语音识别服务目前仅支持8000Hz或者16000Hz采样率、16bits位深，实时语音识别仅支持单声道。注：如果您使用的是实时语音识别，音频属性必须严格符合上述要求。

- 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）。

版权所有：腾讯云计算（北京）有限责任公司

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：

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：2022-07-06 09:53:17

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。

语音识别服务如何计费？

目前仅有后付费计费模式，具体参见 计费概述。

版权所有：腾讯云计算（北京）有限责任公司

功能相关

最近更新时间：2022-08-29 09:34:07

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。

支持远场和离线的语音识别吗？

不支持。目前仅实时语音识别支持移动端离线语音识别，如有需要请 提交工单 咨询。

语音识别支持中英文混合场景和地方方言吗？

- 普通话引擎支持单词级别的中英文混合识别，且支持带口音的中文普通话语音识别。
- 实时语音识别支持中文普通话和英文。

> 说明：
> 若有马来语 越南语 印地语 土耳其和阿拉伯语等更多语种需求，可提交工单。

语音识别支持的输入音频时长是多少？

- 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？

语音识别服务的对于音频属性的详细规定请参见 语音识别服务详细信息页。

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。

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。

版权所有：腾讯云计算（北京）有限责任公司

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，可提交工单。

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，可提交工单。

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：2022-07-06 09:53:43

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的时报错"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。

版权所有：腾讯云计算（北京）有限责任公司

其他相关

最近更新时间：2022-08-29 09:34:54

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。

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）。