Automatic Speech Recognition

FAQs

Product Documentation

Contents

FAQs
- Recognition Effect
- Service and Billing
- Features
- API and SDK
- Others

FAQs

Recognition Effect

Last updated：2022-07-06 09:52:58

If you find that there is some gap between the transcription result and your expectation when using ASR, you can troubleshoot the problem according to this document.

Troubleshooting Steps

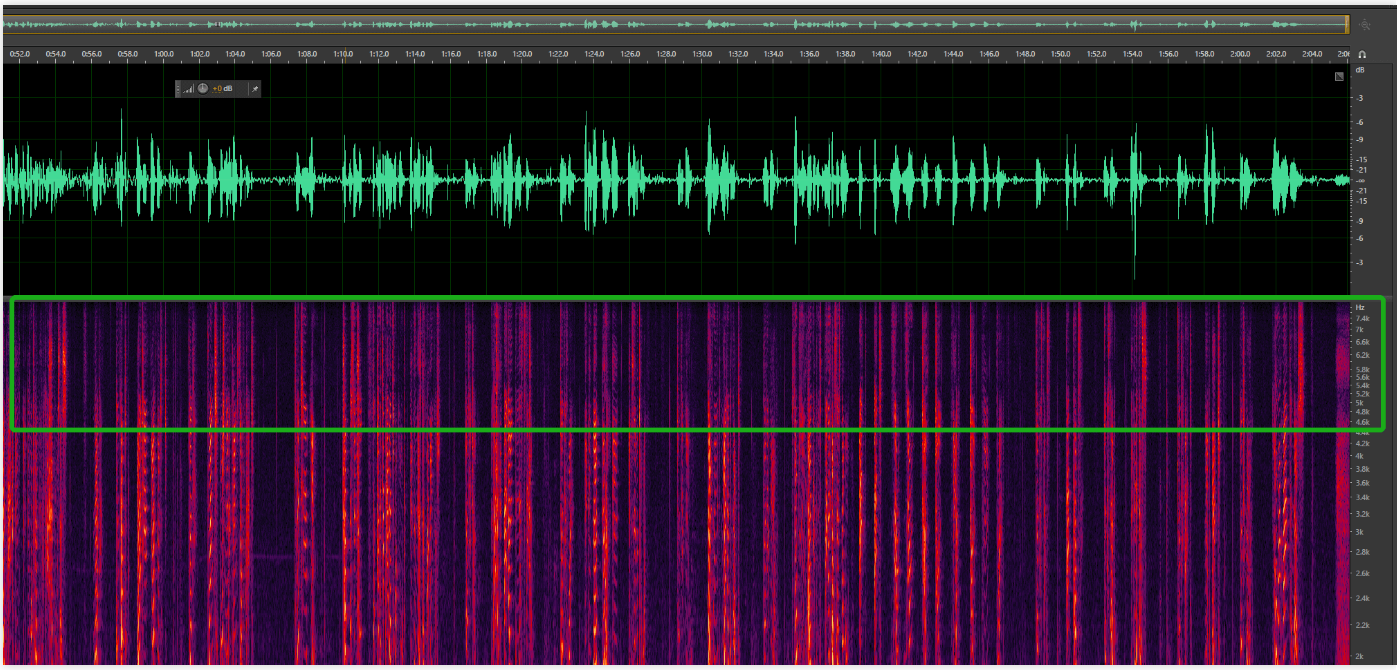
Common problems include the following:

1. The audio content is not clear or comprehensible by ordinary people. In this case, we recommend you transform the audio capture environment on the frontend, for example, changing from far field to near field for audio capture, controlling and reducing noises in the environment, using standard universal language without accent or dialect (i.e., language comprehensible by non locals), and reducing slurs caused by fast speech.
2. The audio content is comprehensible, but the recognition result is very different from what is heard. This problem is generally caused by the failure of the audio information to meet the requirements of ASR.

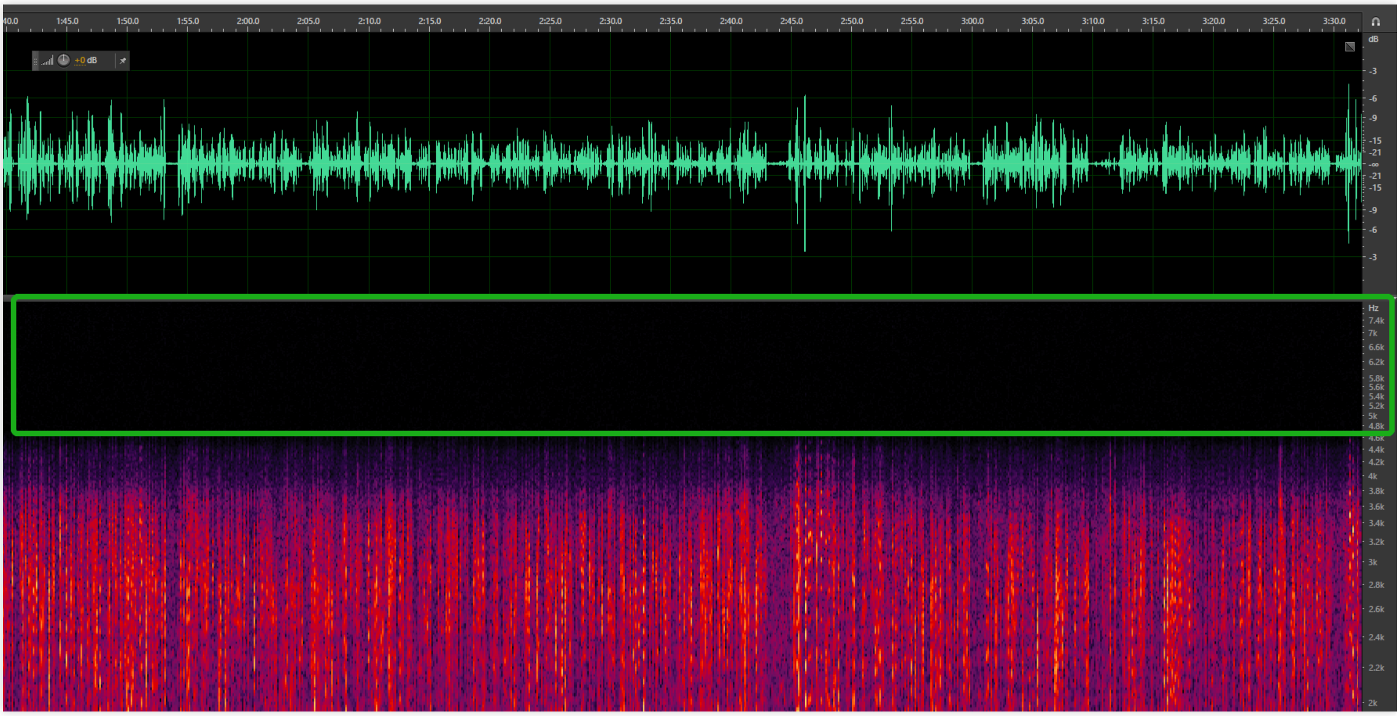
- View detailed audio information in Cool Edit, Adobe Audition, or FFmpeg, including sample rate, number of channels, and bit depth. ASR currently only supports audios with a sample rate of 8,000 Hz or 16,000 Hz and a bit depth of 16-bit (specifically, real-time speech recognition supports only mono-channel audios). Note that if you use real-time speech recognition, the audio attributes must strictly meet the above requirements.

- View the audio waveform and spectrum (in the **View** option in Adobe Audition) to determine the real sample rate of the audio, which should meet the requirements of ASR (8 kHz sample rate for 8 kHz phone call engine model or 16 kHz for 16 kHz non-phone call engine model).

  The waveform and spectrum of a true 16,000 Hz audio is as follows (true sample rate = the highest value on the right in the box * 2, i.e., 8 kHz x 2 = 16 kHz):



The waveform and spectrum of a fake 16,000 Hz audio is as follows (which is actually 4.6 kHz * 2 = 9.2 kHz). It can be seen that the audio information is completely missing in the 4.6k–8k frequency band.



3. The audio content is comprehensible, and the recognition result is not much different from what is heard, but some unique nouns or sentences are poorly recognized. The recognition effect can be improved as follows:
4. The audio content is comprehensible, and the recognition result is not much different from what is heard, but there are some extra words recognized. This problem is generally caused by noise. There are two types of noise: non-

human noise and human noise. ASR's algorithms are optimized and adapted for non-human noise, and you can submit bad cases caused by such noise to Tencent Cloud for further analysis and optimization. If the problem is caused by human noise, it is hard to be solved, because the human speech to be recognized may be hindered by noise reduction.

Service and Billing

Last updated：2022-07-06 09:53:11

How do I activate ASR?

You can activate ASR in the ASR console, which is pay-as-you-go by default.

How is ASR billed?

Currently, ASR is pay-as-you-go only. For more information, see Billing Overview.

Features

Last updated：2023-04-24 17:22:57

Which ASR service should I choose in different scenarios?

- Real-time speech recognition is applicable to scenarios with requirements for real-timeness, such as voice input method, voice robot, and meeting recording.

Are far-field and offline speech recognition features supported?

No. Currently, only real-time speech recognition supports offline speech recognition on mobile devices. If you have such needs, submit a ticket for assistance.

Does ASR support recognizing speeches in Chinese-English mix and dialects?

- The Mandarin engine can recognize speeches in Chinese-English mix (at the word level) and accented Mandarin.
- ASR supports recognizing Mandarin and English.

> Note：
> For recognition of Malaysian, Vietnamese, Hindi, Turkish, Arabic, and other languages, submit a ticket.

How long can an ASR input audio be?

- In real-time speech recognition, each audio segment of a data packet in the audio stream is 200 ms in length.

What audio attributes does ASR support?

For the detailed specifications of ASR on audio attributes, see ASR.

In real-time speech recognition, if the audio contains multiple sentences, how do I increase the recognition accuracy?

We recommend you enable the voice activity detection (VAD) feature for audio segmentation. If the audio contains multiple sentences, VAD can detect the pauses between them and automatically divide the audio into different sentences, achieving a higher recognition accuracy.

Does ASR support sync result call?

- Real-time speech recognition supports sync recognition result return.

Does ASR support evaluation?

No.

Can I copy the text returned by ASR?

No. You need to develop the text copy feature on your frontend by yourself after connecting to ASR.

Can I set the longest recognition time for real-time speech recognition?

No, but you can directly stop recognition at any time.

Does ASR support the MRCP protocol?

Currently, MRCP is not open to external businesses. If you want to use it, submit a ticket for assistance.

Is ASR available as an SaaS service?

ASR supports on-premises deployment. If you need this, submit a ticket for assistance.

API and SDK

Last updated：2022-07-06 09:53:38

What should I do if HTTP requests to ASR APIs return an authentication failure?

Check whether your parameters are uploaded correctly against the parameter table. For quick connection, we recommend you use the SDK provided at our official website.

What should I do if a recognition result error is reported due to invalid URL in ASR?

The URL you provide must be a public network URL accessible by Tencent Cloud. You can use Tencent Cloud COS to store audio files and use relevant URLs. You also need to check whether the firewall is blocking access, whether the URL is at a private IP, and whether the audio files stored at other service providers can be downloaded by Tencent Cloud properly.

What should I do if "Unregistered AppId" is reported during ASR API call?

You are not registered yet. To use ASR, you must activate it first as instructed in Getting Started.

Do ASR APIs have restrictions on the sample rate of audio files?

APIs don't restrict the sample rate of audio files, but if the sample rate is non-compliant, the recognition effect will be compromised.

Others

Last updated：2022-08-29 09:35:52

How do I connect to ASR?

ASR currently supports connection via API and SDK (recommended). For more information, see Getting Started.

What factors affect the accuracy of speech recognition?

Factors such as being far away from the mic, obvious noise, and heavy accent affect the accuracy of speech recognition.

How do I view audio format and attributes?

Windows:
You can download software tools such as Adobe Audition CS6 to view and modify the audio format.
Linux or macOS:
Use the **file** command, such as **file test.wav**.
Result:

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

The sample rate of this audio is 8 kHz, the bit depth is 16-bit, and the channel is mono (as compared to stereo).