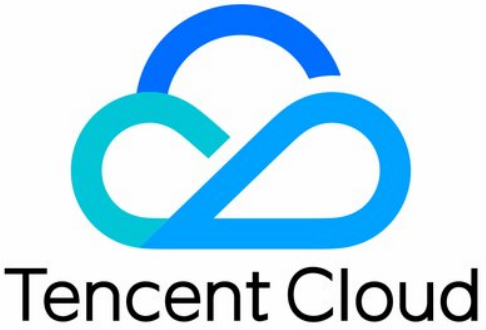
Automatic Speech Recognition FAQs



Copyright Notice

Trademark Notice

Service Notice

This document provides an overview of the as-is details of Tencent Cloud's products and services in their entirety or part. The descriptions of certain products and services may be subject to adjustments from time to time.

The commercial contract concluded by you and Tencent Cloud will provide the specific types of Tencent Cloud products and services you purchase and the service standards. Unless otherwise agreed upon by both parties, Tencent Cloud does not make any explicit or implied commitments or warranties regarding the content of this document.

Contact Us

We are committed to providing personalized pre-sales consultation and technical after-sale support. Don't hesitate to contact us at 4009100100 or 95716 for any inquiries or concerns.

Contents

FAQs
Recognition Effect Troubleshooting

Last updated: 2025-04-03 15:49:05

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.

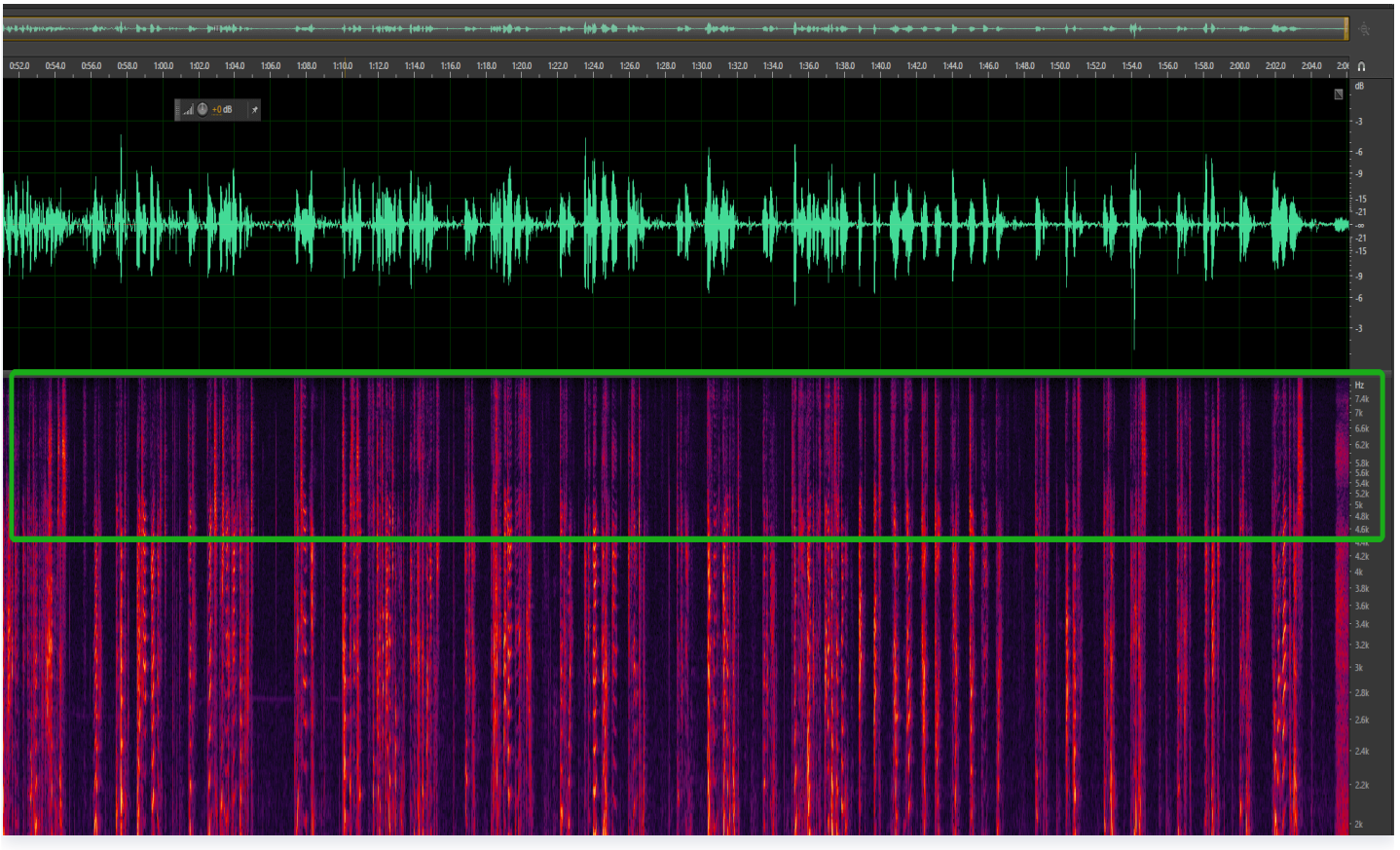
Troubleshooting Steps

Common problems include the following:

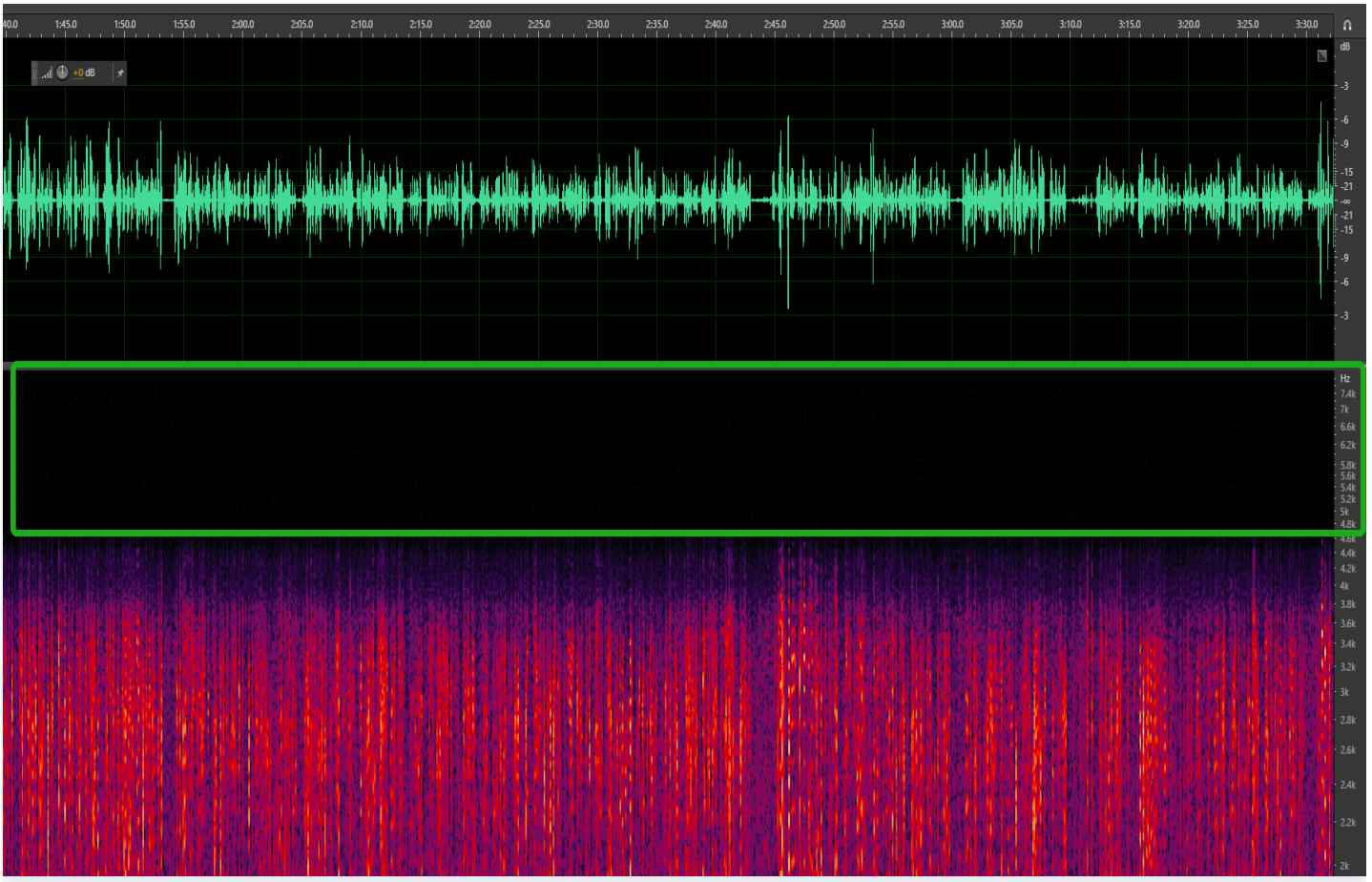
1. The audio content is not clear or comprehensible by ordinary people. In this case, we recommend you transform the audio capture environment on the frontend, for example, changing from far field to near field for audio capture, controlling and reducing noises in the environment, using standard universal 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 sampling rate, number of sound channels, and bit depth. ASR currently only supports audios with a sampling rate of 8,000 Hz or 16,000 Hz and a bit depth of 16-bit. Recording file recognition supports mono and stereo channels, while real-time speech recognition and single-sentence recognition support only mono-channel. Note that if you use real-time speech recognition or single-sentence recognition, the audio attributes must strictly meet the above requirements.
- View the audio waveform and spectrum (in the view options of Adobe Audition) to determine the true sampling rate of the audio. It is recommended that the actual sampling rate meets ASR requirements (8k telephone engine model corresponds to 8,000 kHz sampling rate, 16k non-telephone engine model corresponds to 16,000 kHz sampling rate).

The waveform and spectrum of true 16000Hz (true sampling rate = highest value on the right of the boxed values × 2, i.e., 8kHz × 2=16kHz) audio are as follows:



The waveform and spectrum of non-true 16000Hz (actual value is 4.6kHz × 2 = 9.2kHz) audio are as follows. You can see that audio information is completely missing in the 4.6kHz to 8kHz 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:
   - Refer to the Hotword Operation Instructions for the addition and usage of hotwords for poorly recognized nouns.
   - Refer to the Self-learning Model Operation Instructions for the addition and usage of self-learning models for sentences with poorly recognized nouns or sentences with poorly recognized special cases.
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 specific bad cases caused by such noise to Tencent for further analysis and optimization. However, it is difficult to solve the problem caused by human noise because it may cause false positives for the human speech that needs to be recognized.

Service and Billing

Last updated: 2025-04-03 15:49:16

For common issues with Automatic Speech Recognition, please watch the video:
Watch video

How To Activate the ASR Service?

Users need to activate the service in the ASR Console. The default purchase method is pay-as-you-go.

Is There a Free Quota For the ASR Service Each Month?

- The free quota for recording file recognition is 10 hours per month.
- The free quota for single-sentence recognition is 5000 times per month.
- The free quota for real-time ASR is 5 hours per month.
- The free quota for async recognition of voice streams is 5 hours per month.
- The free quota for the ultra-fast version of recording file recognition is 5 hours per month.
- The hours in the free quota refer to the duration of recognition successful audio; the count in the free quota refers to the number of recognition successful times.

Before March 25, 2020, users who activate the free trial version will have their service suspended after the free tier is exhausted. If users want to continue using the service, please upgrade to the paid edition in the ASR Console.

How Is the ASR Service Billed?

Currently, there are two billing modes: prepayment and pay-as-you-go. Prepayment supports the purchase of paid resource packages. The deduction order for call volume is: free resource package > paid resource package > pay-as-you-go. For details, see Billing Overview.

> **Note**
> Prepaid resource packages can only be used to offset the calling volume generated after purchase. The calling volume generated before purchasing the resource package will be settled on a pay-as-you-go basis.

How To Query the ASR Bill?

Log in to the Tencent Cloud Console - Billing Center, select **Bill > Bill Details** from the left menu, and enter the bill details page to view bill details by time, product, and billing mode.

Does the ASR Prepaid Resource Package Support Remaining Quota Alert Notifications?

Support. Both free and prepaid resource packages support balance alerts. Please pay attention to notifications via Message Center and SMS. Additionally, you can go to the Console to check the usage of the resource package. For details, see Resource Package Management.

Functionality

Last updated: 2025-04-03 15:49:28

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.
- Recording file recognition is suitable for scenarios with longer speech duration and low real-time requirements, such as customer service quality inspection and video subtitles generation.
- Ultrafast recording file recognition is suitable for scenarios with longer speech duration and extremely high real-time requirements, such as adding subtitles to videos and quasi-real-time quality inspection.
- One-sentence recognition is suitable for recognizing short audio files within 60 seconds, such as voice messages and voice search.
- Asynchronous stream recognition is suitable for quasi-real-time recognition of voice streams, returning text results asynchronously, such as live streaming review and audio/video review.

Is the Audio Transmitted By Users To Tencent Cloud ASR Used Only For the Current Recognition?

Yes, the audio content transmitted by users to Tencent Cloud ASR is used only for the current recognition and will not be saved.

If Two People Are Talking In a Recording Stored As Mono, Will the Recognition Result Separate Their Dialogue?

8K and 16K sampling rate Mandarin recording file recognition supports speaker separation for single channel dual-person dialogue.

Are Far-Field and Online/Offline Speech Recognition Supported?

Supports both online and offline speech recognition. For details, refer to the offline section in the SDK Documentation.

Does ASR Support Recognizing Speeches In Chinese-English Mix and Dialects?

- Real-Time Speech Recognition, One-Sentence Recognition, Recording File Recognition, Ultrafast Recording File Recognition, and Asynchronous Stream Recognition support

mixed recognition of Chinese and English (when using the Chinese engine, it can support mixed recognition of Chinese and English in the case of a small amount of English, but the recognition rate may decrease with a large amount of English) and support Mandarin with an accent.

- Real-Time Speech Recognition, One-Sentence Recognition, Recording File Recognition, and Ultrafast Recording File Recognition support the recognition of 23 dialects, including Shanghai dialect, Sichuan dialect, Wuhan dialect, Guiyang dialect, Kunming dialect, Xi'an dialect, Zhengzhou dialect, Taiyuan dialect, Lanzhou dialect, and Yinchuan dialect.

What Is the Supported Input Audio Duration For ASR?

- One-sentence recognition supports audio within 60 seconds per call.
- Recording file recognition supports audio within five hours per call.
- 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?

API	Audio Properties
Recording file recognition	Sampling rate: 16kHz, 8kHz Bit depth: 16bit Channels: mono, stereo
One-Sentence recognition	Sampling rate: 16kHz, 8kHz Bit depth: 16bit Channels: mono
Ultrafast recording file	Sampling rate: 16kHz, 8kHz Bit depth: 16bit Channels: mono, stereo
Real-Time speech recognition	Sampling rate: 16kHz, 8kHz Bit depth: 16bit Channels: mono
Speaker verification	Sampling rate: 16 kHz Bit depth: 16 bit Channels: Mono
Virtual number human detection	Sampling rate: 8 kHz Bit depth: 16 bit Channels: Mono

What Transmission Methods and Formats Are Supported For Audio Data In Single-Sentence Recognition and Recording File Recognition?

Transmit using HTTP Protocol, POST method. The audio data can be transmitted in the following two ways:

1. Audio data is encoded using base64 and transmitted with the HTTP body.
2. If using URL download, the data in the body can be left empty, and the audio URL should be included in the request parameter.

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 Synchronous Result Invocation?

- Real-time speech recognition supports sync recognition result return.
- One-Sentence Recognition supports quick return of recognition results.
- Recording file recognition supports two forms of asynchronous invocation: callback and polling.

Can ASR Convert a Mandarin Recording File To English Text?

No. ASR currently cannot convert Mandarin recording files to English text.

Does ASR Support Evaluation?

It is not supported.

Can the Text Recognized By ASR Be Copied?

The text recognized by ASR cannot be copied. The copy feature requires frontend development after integration.

After Purchasing a Recording File Recognition Resource Package, How Do I Import Files For Recognition?

You can import files on the ASR console feature experience page, or use the API and SDK.

What File Upload Formats Are Supported By the Recording Transcription Feature?

The transcription feature supports WAV, MP3, M4A, FLV, MP4, WMA, 3GP, AMR, AAC, OGG-OPUS, and FLAC formats.

Can I Set the Longest Recognition Time For Real-Time Speech Recognition?

The maximum recognition time cannot be set. If not needed, just disconnect.

Does ASR Support MRCP Protocol?

MRCP is not yet available to the public. If needed, please contact Presales Inquiry.

Is There a SaaS Solution That Can Be Provided Directly To Customers?

ASR supports private deployment, which requires business coordination and follow-up. You can contact Presales Inquiry.

How To Cut Audio Longer Than 5 Hours or Files Larger Than 1GB?

You can use the ffmpeg command to cut audio/video. For example, if the audio duration is 3 hours and you want to cut it into three 1-hour audios, you can use the following command:

```
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
```

The -ss parameter is the start time of the cut, -i is the file name of the cut, and -t is the duration of the cut audio in seconds.

How To Convert an English Recording File To Chinese Using ASR?

The ASR feature converts audio content into text and does not support Chinese-English translation.

How To Save the Text After Real-Time Speech Recognition?

Real-time speech recognition returns text in real-time, which you can save locally.

What Languages Does ASR Support?

Real-time speech recognition supports Mandarin, English, Korean, Cantonese, Japanese, Thai, and Shanghai dialect. For details, refer to Real-Time Speech Recognition (WebSocket). One-sentence recognition and recording file recognition support Mandarin, English,

Cantonese, Japanese, and Shanghai dialect. For details, refer to Recording File Recognition and One-Sentence Recognition.

Can ASR Save Voice Files?

The audio and video files uploaded for ASR are not saved. After successful recognition, the recognized text file is stored on the server for 7 days. Saving audio files affects the recognition result, which is currently returned directly. You can implement audio file saving on the business side, storing the audio files on a local server or in a database.

Does the Recording File Recognition API In ASR Support Filtering Modal Particles?

Recording file recognition supports filtering modal particles. For specific usage, refer to Recording File Recognition Request.

Does the Recording File Recognition API In ASR Support Filtering Punctuation?

The recording file recognition API supports filtering punctuation. For specific usage, refer to Recording File Recognition Request.

What Is the Accuracy Of ASR?

In the test report issued by the National Quality Supervision and Inspection Center for Electronic Computers, Tencent Cloud's voice robot system achieved a character accuracy rate of 97.40% for Chinese speech recognition (results rounded to two decimal places) and no less than 88.00% for American English speech recognition (results rounded to two decimal places) for audio data with a sampling rate of 16k, 16bit, and in raw uncompressed wav or pcm format. However, please be aware that the aforementioned character accuracy rates are only third-party experimental test data for your reference and do not constitute a guarantee of the accuracy of Tencent Cloud's speech recognition service.

Does the Recording File Recognition API In ASR Support Intelligent Conversion Of Arabic Numerals?

Recording file recognition supports intelligent conversion of Arabic numerals. For specific usage, refer to Recording File Recognition Request.

API and SDK

Last updated: 2025-04-03 15:49:41

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 integration, we recommend you use the official SDK.

ASR Service Recognition Result Reports an Invalid URL Address?

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?

User not registered. The user needs to activate the ASR service according to the ASR Getting Started guide before using the service.

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.

Recording File Recognition Error: Audio Channel Num Not Match! What'S the Reason?

The number of sound channels in the audio must match the channel parameter. For example, if the audio has 2 sound channels, then channel = 2.

What Should I Do If No Callback Is Received For a Long Time After Recording File Recognition?

Ensure the callback service is functioning properly and the callback address is accessible over the public network.
If the callback address uses https, ensure the certificate is authenticated by a legitimate CA.

Single-Sentence Recognition Error: Failed To Recognize Audio! What'S the Reason?

Confirm whether the uploaded audio format matches the VoiceFormat parameter setting. If not, this error may occur.

How To Troubleshoot the 4002 Authentication Failure Error When Using the ASR SDK?

Common causes of this error when using the SDK include incorrect key information such as appid, secret_id, and secret_key. Additionally, if a subaccount is used, the main account needs to grant the subaccount permission, otherwise, this error will occur.

Real-Time ASR Error: 4008 Client Upload Timeout?

This error occurs because the server-side did not receive audio data for over 15 seconds and actively disconnected the link. It is recommended to actively send a tail packet to disconnect the link if no data is being sent.

Other Related

Last updated: 2025-04-03 15:49:58

How To Integrate ASR?

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

How To Experience ASR Features?

You can search for "Tencent Cloud AI Voice" mini program on WeChat and select Speech Recognition to experience it. In the Speech Recognition Console feature experience module, users can experience it by uploading files or URLs.

What Factors Affect the Accuracy Of Speech Recognition Results?

Factors such as being far away from the microphone, obvious noise, and heavy accent can 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).

How To Experience the Functionality Of Uploading Files Larger Than 5M In the ASR Console?

The Speech Recognition Console provides a feature experience for you to test. If your test file is large, it is recommended to use the audio URL upload method. The audio duration should not exceed five hours.

How Long Does It Take To Convert a Recording File To Text?

The result of converting recording files to text is affected by network, audio length, recording environment, and language standards. The specific time depends on the parameters.