Speech Interaction Service

Overview

Issue	01
Date	2025-09-18

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process.* For details about this process, visit the following web page:
https://www.huawei.com/en/psirt/vul-response-process
For vulnerability information, enterprise customers can visit the following web page:
https://securitybulletin.huawei.com/enterprise/en/security-advisory

Contents

1 What Is SIS?

Speech Interaction Service (SIS) lets you get instant responses by accessing and calling APIs in real time. For example, through Real-Time Automatic Speech Recognition (RASR), you can transcribe spoken audio or speech files into editable text. Additionally, Text To Speech (TTS) enables text-to-speech conversion, producing lifelike audio for an enhanced user experience. SIS is applicable in various scenarios, including speech-based customer service quality checks, meeting transcriptions, voice messaging, audiobooks, and call-back services.

Prerequisites

You must have programming capabilities and be familiar with the Java, Python, and iOS programming languages.

SIS provides APIs for you to convert speech into editable text and returns the recognition result in JSON format. You need to encode the recognition result and save it to a service system or save it in TXT or Excel format.

Using SIS for the First Time

If you are a first-time user, the following information will help you get familiar with SIS:

- Functions

  Functions describes SIS functions, including Real-time ASR, Short Sentence Recognition, TTS.

- Getting Started

  SIS provides services through open APIs. You can learn how to use SIS by referring to the **Speech Interaction Service Getting Started**.

- Using SIS

  If you are a development engineer familiar with code compilation and want to directly call SIS APIs, see the **Speech Interaction Service API Reference** or **Speech Interaction Service SDK Reference**.

- From Beginners to Experts

  You can learn how to use SIS by referring to **Progressive Knowledge**.

2 Functions

Real-Time ASR

Real-Time ASR allows you to obtain real-time speech recognition results by accessing and invoking the API. Currently, Real-Time ASR supports Mandarin Chinese.

- Text Timestamps

  Generates specific timestamps for the audio conversion result, so that you can quickly find the spot in the original audio clip to confirm the text and adopt if needed.

- Intelligent Text Segmentation

  By extracting semantic features of the context and combining voice features, intelligently segments sentences and adds punctuation marks to improve the readability of the output text.

- Hybrid Recognition

  Supports recognition of English letters/words and digits included in Chinese sentences.

- Instant Result Output

  Continuously recognizes voice streams, outputs results in real time, and automatically corrects the content based on the context language model.

- Automatic VAD

  Performs voice activity detection (VAD) on the input voice streams to improve recognition efficiency and accuracy.

Highlights

- High Recognition Accuracy

  Adopts the latest generation of speech recognition and Deep Neural Network (DNN) technologies to greatly improve the anti-noise performance and recognition accuracy.

- High Speed

  Integrates the language models, dictionaries, and acoustic models into a large neural network featuring impressive optimizations in the engineering to greatly increase the decoding speed, achieving faster recognition.

- Multiple Recognition Modes

  Supports multiple real-time speech recognition modes, including streaming, continuous, and single-sentence, to suit different application scenarios.

- Customization Service

  Allows you to customize the language-layer model in a specific vertical domain to better recognize proprietary words and industry terms, adding a significant boost to accuracy.

Short Sentence Recognition

Short Sentence Recognition converts audio recordings within 30s to text. Specifically, the system processes the binary audio data uploaded by users and generates the corresponding text. The supported language includes English.

Highlights

- High Recognition Rate

  Utilizes the deep learning technology to optimize speech recognition for domain-specific scenarios, enabling an industry-leading recognition rate.

- Cutting-Edge Technologies

  Combines mature speech recognition algorithms currently in active use in the industry with the latest research to empower enterprises with unique competitive advantages.

- Customizable Models

  Increases accuracy by using speech recognition models designed for the specific requirements of the vertical industry you operate in for other specific scenarios.

3 Application Scenarios

- Voice Customer Service Inspection

  Recognizes the speech of the customer service personnel and customer, converts the speech into text, and checks whether it contains any violation, sensitive word, or phone number through text retrieval.

- Voice Message

  Converts voice messages you send or receive into text to deliver higher reading efficiency and interaction experience.

- Gaming and Entertainment

  Converts voice chats into text messages, improving reading efficiency and user experience.

4 Billing

Billing Item

- SIS bills based on the number of API calls, with failed API calls are not counted.

Billing Mode

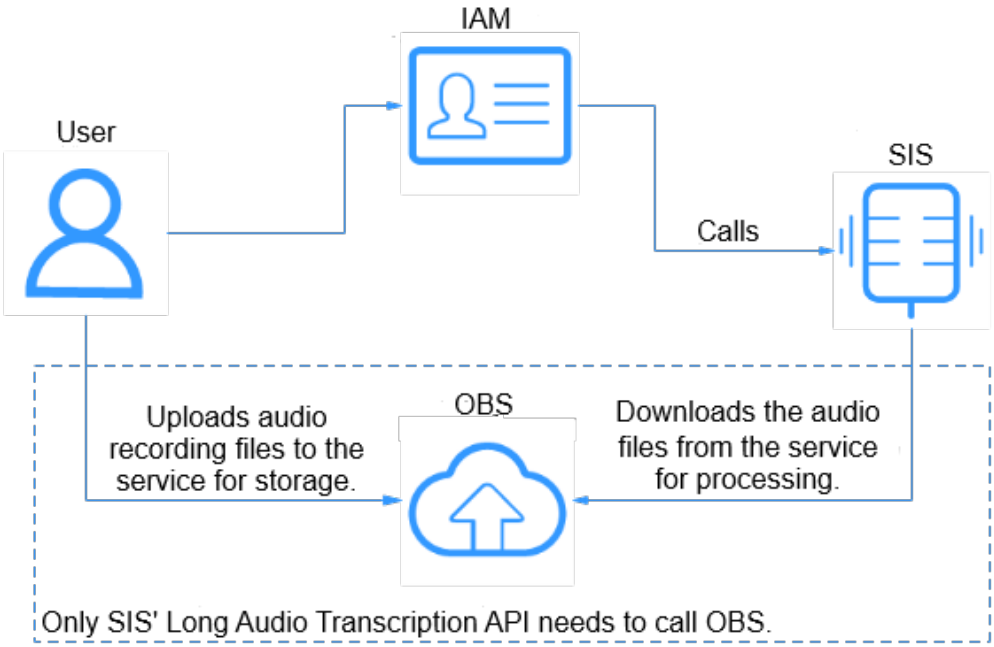
- **Pay-per-use**

  Pay-per-use billing refers to tiered pricing based on the number of successful API calls. For detailed charges, refer to the **Price Calculator**.

5 Related Services

Figure 5-1 shows the relationship between SIS and other services.

Figure 5-1 Relationship between SIS and other services



Identity and Access Management (IAM)

IAM provides user authentication and authorization for SIS.

6 Restrictions and Limitations

6.1 RASR

- This service is currently available in the **AP-Singapore** and **ME-Riyadh** regions.
- The audio sampling rate is 8 kHz or 16 kHz, and the sampling bit depth is 8-bit or 16-bit.
- Mandarin Chinese, English, and Arabic are supported.
- The endpoint detection parameters (vad_head, vad_tail, and max_seconds) influence the sentence segmentation results and may introduce some errors. They are effective in continuous and single-sentence modes but not in streaming mode.

6.2 Short Sentence Recognition

- Short Sentence Recognition supports various formats, including pcm16k16bit, pcm8k16bit, ulaw16k8bit, ulaw8k8bit, alaw16k8bit, and wav.
- The audio duration cannot exceed 1 minute.
- This service is currently available in the **AP-Singapore** and **ME-Riyadh** regions.

6.3 Real-Time Text To Speech

- This service is currently available only in the **ME-Riyadh** region.
- Arabic and English are supported, with text not exceeding 10,000 characters.
- The supported synthesis sampling rates are 8 kHz and 16 kHz.

