Speech Recognition Service Software V3.017

TS-8611R

Description:

The speech recognition service software is developed with artificial intelligence technology and is a speech transcription product provided to customers in various industries. The product provides customers with a complete set of speech transcription solutions to help customers quickly convert audio into text. The product has made a number of core technology optimizations for long voice scenarios, and the recognition rate in far-field and noisy environments has been greatly improved, with industry-leading technology. The product provides two core functions: asynchronous file transcription and real-time speech transcription to meet the different needs of customers. Recording file recognition supports customers to upload audio files for recognition into text; real-time speech recognition supports customers to upload audio streams and obtain recognized text stream results.

Description:

* Real-time speech transcription: It is to perform real-time speech recognition on the audio stream, which can achieve the effect of "speaking and outputting text synchronously". After the session is initialized successfully, the interface can be continuously called to send audio to the service to obtain recognition results. When sending audio, in order to maintain frame synchronization with the engine, the audio needs to be divided into segments of fixed size for sending, and a certain waiting time is required to ensure that the flow per second is consistent with the sampling rate of the engine. For example, if the engine is 16k 16bit and 1280 audio segments are sent each time, then 40ms should be waited each time.
* Speech recognition accuracy: up to 98%, and up to 95% in speeches and formal meetings.
* Speech recognition speed: Relying on top speech transcription technology, the real-time speech transcription speed is ≤200 milliseconds, and 1 hour of audio recognition can be completed in 6-10 minutes .
* Supports multiple audio codec formats: Currently, real-time speech transcription supports pcm format audio codec algorithm. Non-real-time transcription supports audio in mp3, wav, wma, mp4, avi, pcm, and m4a formats. Currently audio sampling rates only support 16K and 8K.
* Text post-processing: The speech transcription private cloud supports intelligent prediction of the dialogue context of the recognition result sentence, provides intelligent sentence segmentation and punctuation prediction, and also supports number curation and replacement list capabilities.
* Tens of thousands of hours of acoustic model training data significantly improve personalized recognition.
* There is no default language for software recognition. Specify the language pack as needed.
* Other languages and dialects can be purchased. Supported 12 national languages include: Japanese, Korean, Russian, French, Spanish, Thai, German, Vietnamese, Arabic, Bulgarian, Hindi, Italian. Supported 25 dialects include: Sichuan dialect, Cantonese, Shanghai dialect, Hefei dialect, Changsha dialect, Minnan dialect, Nanjing dialect, Taiwanese dialect, Shandong dialect, Tianjin dialect, Suzhou dialect, Northeastern dialect, Wuhan dialect, Henan dialect, Hakka dialect, Shaanxi dialect, Taiyuan dialect, Yi dialect, Nanchang dialect, Guizhou dialect, Hebei dialect, Yunnan dialect, Northern Anhui dialect, Gansu dialect, Ningxia dialect .