



## Description

Speech recognition service software is a software installed in Kylin system to provide real-time voice transcription function. It provides a complete speech transcription solution to help customers quickly convert speech into text. It has realized a number of core technology optimizations for long speech scenarios, and the recognition accuracy in far distance and noise environments has been greatly improved. It provides two core functions of asynchronous file transcription and real-time speech transcription to meet the different needs of customers. It supports recording file recognition, and customers can upload and convert audio files into text. It supports real-time speech recognition; customers can upload audio streams, and obtain the recognized text stream results.

## Feature

\*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  $\leq 200$  milliseconds

\*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.

\*The software recognizes a mixed language of Chinese and English by default.

\*It supports Mandarin Chinese, English, and a mixture of Chinese and English. The Mandarin Chinese engine supports English words and simple English sentences in Chinese, which basically meets the daily use requirements of Chinese people. The mixed engine of Chinese and English can switch freely between the Chinese and English contexts.