



Description

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

Feature

- * Real-time long speech recognition: Based on the deep convolutional neural network architecture, through the WebSocket protocol, a long link between the app and the speech recognition engine is built, and real-time recognition of unlimited audio streams is supported; built-in intelligent sentence segmentation function can figure out the start and end time of each sentence, making it suitable for real-time live subtitles, real-time meeting records, real-time court hearings and other scenarios.
- * Speech recognition accuracy: Transcription accuracy is related to language standardization and articulation clarity.
- * Speech recognition speed: Relying on top-notch speech transcription technology, the real-time speech transcription speed is ≤ 300 milliseconds.
- * One-sentence recognition: Support the recognition of speech with a short duration (within 60 seconds), return the non-real-time recognition result with a delay of less than 5 seconds.
- * Support multiple audio codec formats: Real-time speech transcription supports pcm format audio codec algorithms; non-real-time transcription supports mp3, wav, wma, mp4, avi, pcm, m4a and other audio formats; the audio sampling rate only supports 16K and 8K.
- * Text post-processing: The speech transcription private cloud supports intelligent context prediction of the recognition result sentence, provides intelligent sentence segmentation and punctuation prediction, and also supports digital regularization and replacement list capabilities.
- * Recording file transcription: Support calling the recognition service through the http[s] protocol, and transcribing long audio recordings (within 5 hours) into text data, which can be used for interview recording transcription, court trial data entry, meeting minutes summary, call center recording inspection and other scenarios; support functions such as uploading of recordings in pieces, separation of speaker roles, custom hot words and sensitive words, etc.
- * The software recognition language only supports Mandarin.
- * High intelligibility: Provide functions such as smooth speech, hot words, sensitive word detection, number regularization, intelligent punctuation prediction, and intelligent segmentation, effectively improving the readability and intelligibility of texts.