

Magic Data Chinese Mandarin Conversational Speech Corpus

Identifier: Magic Data

Summary: The corpus by Magic Data Technology Co., Ltd., containing 10 hours of conversational speech data from native speakers of the Mandarin Chinese spoken in mainland China. The transcription accuracy for each sentence is larger than 98%.

Category: Speech

Recording Corpus: Spontaneous conversation on a given topic

Speech Parameter: PCM/Mono/16bit/16kHz

Age Distribution: 4-67 years old

Gender Distribution: M 1: F 1

About this resource:

MAGICDATA Chinese Mandarin Conversational Speech Corpus was developed by MAGIC DATA Technology Co., Ltd.

The contents and the corresponding descriptions of the corpus include:

- The corpus contains **10 hours** of acoustic data, which is mostly mobile recorded data.
- **60** speakers from different accent areas in Mainland China are invited to participate in the recording.
- The transcription accuracy for each sentence is larger than 98%.
- Recordings are conducted in a quiet indoor environment.
- Detail information such as speech data coding and speaker information is preserved in the metadata file.
- The domain of recording texts is diversified, including interactive Q&A, music search, SNS messages, household command and control, etc.
- Segmented transcripts are also provided.
- Transcriptions with complete text contents, times-tamp (start and end time of valid sentences) and speaker identification

The corpus aims to support researchers in speech recognition, machine translation, voiceprint recognition, and other speech-related fields.

The corpus is a subset of a much bigger data (**10566.9 hours Chinese Mandarin Speech Corpus**) set which was recorded in the same environment. Please feel free to contact us via business@magicdatatech.com for more details.

Citation: Please cite the corpus as “Magic Data Technology Co., Ltd.,

“http://www.imagicdatatech.com/index.php/home/dataopensource/data_info/id/101”, 05/2019”.

ABOUT US

Magic Data Technology Co., Ltd. (referred to as Magic Data) was established in 2016. Through our higher-expertise and higher-precision data services, Magic Data has quickly grown into one of the foremost companies in artificial intelligence industry. We strive to provide the most efficient and highest quality one-stop data services for customers in the fields of speech recognition, intelligent imaging and Natural Language Understanding (NLU). Our services include data scheme design, data collection, data annotation/transcription, etc.

CONTACT

- Tel: (+86) 10-82527250
- Email: business@magicdatatech.com
- <http://www.imagicdatatech.com/>