

Overview of “The Corpus of Dialogue Speech in Japanese (CDSJ)”

Provider: TimeHill Inc.

Project:

CDSJ was built as a part of “G-SODA” project for the conversational AI in the future in cooperation with Timehill Inc.

Corpus overview:

CDSJ was designed for spontaneous speech studies of informal/formal dialogue speeches in Japanese. The speech topic, speakers’ age, sex, language backgrounds, the recording settings, data volume and the price are as following.

【Specification】

1. Sampling Frequency: 48kHz
2. Quantization Bit : 16bit
3. Channel: mono \times 4(2)
4. MIC: Shure SM10/SONY C-357/Other
5. File Format: wav
6. Duration of a file: 15~30minutes
7. Number of files: 500~
8. Number of speakers: 200~
9. Speakers’ age: Age 16~70
10. Speakers’ sex: Female/Male
11. Language: Japanese
12. Recording Year: 2000-2022
13. Recording City
Sapporo/Tokyo/Nagoya/Kanazawa/Osaka/Kyoto/Hakata/Okinawa/Other
14. Task/Topic: Guest \leftrightarrow Clerk
15. Unit Price:

* The following is the unit price for one file (one dialogue). After applying for purchase on this site (pressing the "License Agreement" and "Purchase" buttons), you can purchase (DL) one file at a time from the "Stores" site.

For those who wish to order a large volume set, we can also sell them as a set separately, so please contact us at the following. ↓

info@timehill.biz

(1) Speech

¥5,000 (Tax excluded) / \$50.00

(2) Speech and Transcription

¥20,000 (Tax excluded) / \$200.00