

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

提供者: 株式会社タイムヒル

プロジェクト:

本コーパスは株式会社タイムヒルの目指す未来のコミュニケーション AI 開発プロジェクト (G-SODA)の一環で構築されたものである。

概要:

本コーパスは、私的又は公的ないくつかの場면을自然に再現した二者間自由発話対話音声データであり、当該研究に供するようにデザインされている。話者の年齢、性別、母語、録音年や都市、録音仕様や分量、および価格(1 対話 1 ファイルあたりの単価)については以下に記す。

【Specification】

1. サンプリング周波数: 48kHz
2. 量子化ビット数 : 16bit
3. チャンネル: モノラル ~~X~~4(2)
4. マイク: Shure SM10/SONY C-357/他
5. ファイル形式: wav ファイル
6. 1 ファイルの時間: 15~30 分
7. ファイル(対話)数: 500 対話以上
8. 異なり話者数: 200 人以上
9. 話者の年齢: 16 歳~70 歳
10. 話者の性別: 女性/男性
11. 話者の母語: 日本語
12. 録音年度: 2000-2022 年
13. 録音都市: 札幌/東京/名古屋/金沢/大阪/京都/博多/沖縄/他
14. タスク/トピック: 客←→店員
15. 価格(単価):

※下記は 1 ファイル(1 対話)の単価です。本サイトで購入申込(「利用許諾契約」と「購入」ボタン押下)後に、“Stores”サイトから 1 ファイルずつ DL 購入いただけます。大量セットをご注文ご希望者様には、別途セット販売も致しますので、下記までお申出ください。↓
info@timehill.biz

(1)音声データ

¥5,000 (税抜) / \$50.00

(2) 音声と書起データセット

¥20,000 (税抜)/ \$200.00