

1996 年 6 月

技術研究組合 新情報処理開発機構

1. 概要

RWCP 音声対話データベースは、リアルワールドコンピューティングプログラムの一環として、技術研究組合新情報処理開発機構に設置されていた「RWC データベースワークショップ（音声グループ）」において企画、仕様検討され、監修されたものである。

収録内容：人間同士 1 対 1 対話の音声波形、書き起こしテキスト等
話題「車の購入」24 対話、「海外旅行計画」24 対話

著作権者：技術研究組合 新情報処理開発機構

制 作：三菱電機東部コンピュータシステム（株）

この DVD には、以下のようなデータファイルが格納されている。

- (1) 音声データ
WAV 形式（16kHz、16bit、ステレオ）
- (2) 書き起こしテキスト
仮名漢字表記による音声データの書き起こし
- (3) 発声データファイル
音声波形と書き起こしテキストを対応づけしたデータファイル
- (4) 話者プロフィール
対話収録に参加した話者に関する情報

収録は平成 6、7 年度に実施された。この DVD は平成 6 年度収録の 48 対話についてデータベース化したものである。なお平成 7 年度収録の 13 対話についても、「RWCP 音声対話データベース（97 年版）」として同様にデータベース化されている。

収録作業の仕様と実験環境、ならびにこの DVD に格納されたデータファイルの詳細については、同じくこの DVD に格納されている RWCP-SP96.pdf を参照されたい。

また、下記の Web ページおよび文献もご参照いただきたい（文献[2]については SLP1996.pdf のファイル名で本 DVD に格納されている）。

Web ページ：<http://unit.aist.go.jp/itri/itri-im/spg/DB/DB.html>

- [1] 「RWC データベースワークショップ音声グループ（平成 7 年度）報告書」
新情報処理開発機構（1996.3）
- [2] 田中、速水、山下、鹿野、板橋、岡「RWC 計画における音声対話データベースの構築」情報処理学会音声言語情報処理研究会 96-SLP-11-7（1996. 5）

2. 対話内容および話者

2.1 話題と収録対話数

「車の購入」

顧客が車の購入を計画し、外車販売のショールームに来ている想定

対話数：24 対話

（業者側話者 2 名・各 12 対話、顧客側話者 24 名・各 1 対話）

「海外旅行計画」

顧客が海外旅行を計画し、旅行代理店のカウンターに来ている想定

対話数：24 対話

（業者側話者 2 名・各 12 対話、顧客側話者 24 名・各 1 対話）

2.2 業者側話者（専門家）

「車の購入」

男性話者：1名（現在外車販売会社勤務）

女性話者：1名（外車販売会社勤務経験7年、退職後約1年）

「海外旅行計画」

男性話者：1名（現在旅行代理店勤務）※97年版の話者と同一

女性話者：1名（現在旅行代理店勤務）※97年版の話者と同一

2.3 顧客側話者（質問者）

「車の購入」

男性話者：12名、女性話者：12名

「海外旅行計画」

男性話者：12名、女性話者：12名

3. データファイルの概要

3.1 音声データ

音声データは、DATに収録したものを16kHzにダウンサンプリングした。16bitのサンプルが、Lチャンネル、Rチャンネル交互に記録され、対話開始から終了までを1ファイルとしている。

3.2 書き起こしテキスト

書き起こしテキストは、仮名漢字（UTF-8）により発話単位で記述されている。各発話の先頭には、専門家には「B:」、質問者には「A:」がマークされる。書き起こしテキスト中、無意味語（えー、あの一など）は[]、発話中の相手のあいづちやオーバーラップした発声は{ }、言い直しの区間は（ ）で囲まれている。

3.3 発声データファイル（音声－書き起こしテキスト対応ファイル）

音声データを適当な規則により発声単位に切り出し、その属性などのデータを記述したファイルである。個々の発声単位に対して以下のデータが記述されている。

- 1) シーケンシャル番号（1対話中の発声単位の順序番号）
- 2) フラグ（各発声単位の言語的属性と音響的属性とを表わすコード）
- 3) 話者
- 4) 発話の開始時刻（msecの単位で表わしたもの）
- 5) 発話の終了時刻（同上）
- 6) 仮名漢字記述（発話の書き起こし）
- 7) ローマ字記述（表音記述）
- 8) #（ターミネータ）

3.4 話者プロフィール

収録に参加した話者（専門家と質問者）の年齢、性別、職業、出身地などが記されている。

RWC データベースワークショップ音声グループ委員

（五十音順、所属は当時）：

板橋 秀一 （筑波大学）
鹿野 清宏 （奈良先端科学技術大学院大学）
田中 和世 （電子技術総合研究所）
速水 悟 （電子技術総合研究所）
山下 洋一 （大阪大学）

（禁無断転載）

発行：技術研究組合
新情報処理開発機構

（2012年7月NII-SRC改訂版）