

# 実走行車内音声対話データベース In-Car Spoken Dialogue Database

横尾 祐郁<sup>\*1</sup>  
Masafumi YOKOO

武田 一哉<sup>\*1,2</sup>  
Kazuya TAKEDA

河口 信夫<sup>\*1,2</sup>  
Nobuo KAWAGUCHI

<sup>\*1</sup> 名古屋大学統合音響情報研究拠点(CIAIR)  
Center for Integrated Acoustic Information Research (CIAIR), Nagoya University

<sup>\*2</sup> 名古屋大学大学院工学研究科  
Graduate School of Engineering, Nagoya University

**Abstract** CIAIR, Nagoya University has been collecting the in-car spoken language database for three years. This paper reports the current status of the database construction. In the current collection, each subject has conversations with three types of dialogue systems. One is a human, one is a Wizard of OZ system, and the last is a conversational system. In this paper, we report the specification and the characteristics of the database.

## 1. はじめに

名古屋大学統合音響情報研究拠点(CIAIR)では、音声対話機能を備えた車内情報システムの実現を目標の一つとして、実走行車内における様々な音声、対話データを収集している。本稿では、これまでに収録されている車内音声対話データベースについて報告する。

本音声対話データベースは、実際に被験者が車両を運転中に収集されていることが特徴であり、通常の音声対話とは異なる状況下での対話が収録されている。また、現在の収録では、対話の対象となる車内情報システムとして、機械の役割を果たすよう訓練されたナビゲータ(人)に加え、Wizard of OZ 法に基づくシステム、および音声対話システムを構築し、対話収録を行っている。

CIAIR における車内音声データベースの収録[Kawaguchi 2001] は 1999 年から現在まで続けられており、専用の収録車を構築し、音声に加え、画像や操作情報といったマルチモーダルな情報を大量に収録している。

## 2. 音声対話データベースの概要

### 2.1 音声対話データベースの収集

収集された各セッションについては表1に、データの情報については表2に示す。初年度の 1999 年度は各被験者あたり 11 分間のナビゲータ(人)との擬似対話を収録した。この分析については[河口 2000] を参照されたい。2000 年度には、より実的な収録を目指し、WOZシステムと音声対話システムを導入し、それぞれ5分間ずつの収録を行った。各被験者はすべてシステムとの対話を行うため、タスクの順序が対話に大きな影響を与える。そこで、収録順序はすべての組み合わせを用いて行っている。

WOZ システムはタッチパネル入力と音声合成を用いた位置に基づく情報検索システムである。被験者の発話意図をナビゲータがタッチパネルを用いて WOZ システムに伝え、検索結果に基づき、ナビゲータの指示により合成音声を用いて応答を返す。WOZシステムの応答生成画面を図 1 に示す。画面下部が

検索結果であり、中央には、応答用の文節が表示されている。ナビゲータは各単語を選択することによって、状況に応じた応答を合成音声によって返すことが可能である。

また、本データベース収集のために、車内で実際に移動する音声対話システムを構築した。[早川 2001] 本音声対話システムは、レストラン検索を対話ドメインとした、システム主導の対話システムである。

表1：収集されたセッション

1999 年度	
ナビゲータ(人)との擬似対話	11 分間
音素バランス文(アイドリング)	50 文
音素バランス文(走行中)	25 文
単語収録	50 単語
連続数字	4 桁×20
2000 年度	
ナビゲータ(人)との擬似対話	5分間
WOZ システムの利用	5分間
音声対話システムの利用	5分間
音素バランス文(アイドリング)	50 文
音素バランス文(走行中)	25 文
単語収録	50 単語
連続数字	4 桁×20

表 2：収録されたデータの情報

音声	16kHz, 16bit, 8 ch
画像	MPEG-1, 29.97fps, 3ch
車両制御情報	車速, アクセル, ブレーキ, ハンドル, エンジン回転数
位置情報	D - GPS

連絡先: 横尾祐郁, 名古屋大学統合音響情報研究拠点 (CIAIR), 名古屋市千種区不老町  
Tel. 052-789-4210 Fax. 052-789-4209  
ciair-db@ciair.coe.nagoya-u.ac.jp

図1: WOZの応答入力画面

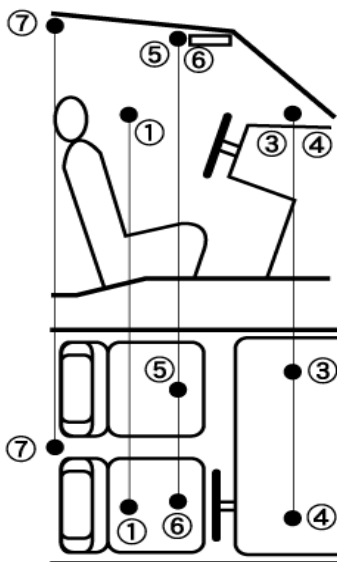
## 2.2 データベースの収録状況

収録は CIAIR で開発した走行車室内音声対話収録システム [河口 2001] を使用している。このシステムは 7 台の PC をネットワークで結合したシステムであり、簡便な操作で複数のチャンネルの音声、画像、車両情報のデジタル収録が可能である。

音声入力用のマイクとして、被験者のヘッドセットマイクには単一指向性の Sennheiser MD-421 を用い、車内 6 箇所に設置した分散マイクロホン用には無指向性の Sony ECM-77B を用いた。(図2)

表2 では 1999 年度、及び 2000 年度に収録した車内音声対話データベースの概要を示す。音声対話データベースは、すべて書き起こしが付与されている。2000 年度は、ナビゲータ (HUM), WOZ, 対話システム(SYS)との対話をそれぞれ別に示してある。形態素解析には、茶筌 ver2.1, ipadic2.4 を用いた。2000 年までの総収録時間は約 140 時間、約 500 名、約 1000 セッションであった。

図2: 収録に用いたマイクの位置



	99HUM	00HUM	00WOZ	00SYS
収録時間(秒)	141810	94692	95300	77922
データ数	209	294	293	288
総発話時間(秒)	97678	69390	50864	54056
ドライバ	44559	28085	20159	11515
オペレータ	53118	41305	30705	42541
総発話単位数	38760	25251	19585	24944
ドライバ	20493	12555	9831	10567
オペレータ	18267	12696	9754	14377
総形態素数	297946	215469	131569	164178
ドライバ	137579	86567	61864	33657
オペレータ	160367	128902	69705	130521
単位毎形態素数	7.69	8.53	6.72	6.58
ドライバ	6.71	6.90	6.29	3.19
オペレータ	8.78	10.15	7.15	9.08

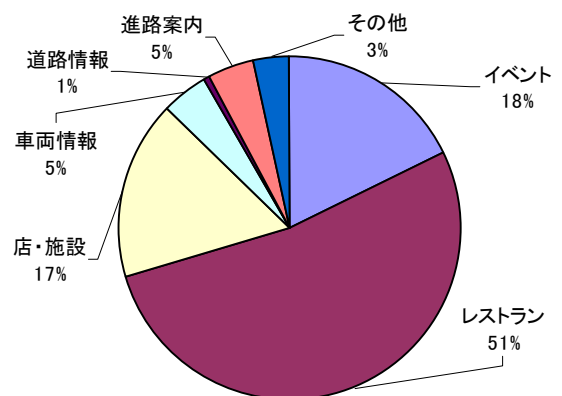
表2: 音声対話データベースの概要

## 2.3 車内対話のタスク

被験者(ドライバ)と車内情報システムとの間で遂行される対話では、実に多くの種類のタスクが想定できる。本データベースの収録では、実環境に近い対話タスクを設定することにより、ドライバの自由発話音声の収集を進めている。

タスクとしては、旅行や催し物などの情報を獲得するイベント対話、飲食店を検索するレストラン対話、デパートや病院などを検索する店・施設対話、給油などの車両情報対話、さらには渋滞情報の獲得を目的とした道路情報対話などがある。1999 年度に収録した 33,885 発話単位の対話タスクごとの割合を図3に示す。全体の約半分はレストラン対話であり、さまざまなジャンルの飲食店検索に関する対話を収録している。

表3: 収録されたタスクの分布



## 3. 音声データの書き起こし

収集した音声データの書き起こし作業は、人手によって行っている。データの分析にあたり、(1)ドライバ発話における話し言葉特有の諸現象を捉えられること、(2)ドライバ発話と車両制御情報との間で同期をとれること、が重要である。このような観点から我々は、日本語話し言葉コーパス(CSJ)の音声書き起こし基準に準拠したタグ付け作業を行うこととした。[小磯 2000] データの言語学的分析として、フィラー、言い淀み、言い誤りなどにタグを付与するとともに、発話をポーズで分割し、各々を発話単位と定め、その開始時間及び終了時間を記録している。図

3 に書き起こしテキストの例を示す。各発話単位の開始・終了時間の右側に、性別（男性 / 女性）、話者役割（ドライバ / ナビゲータ）、対話タスク（道案内 / 情報検索）など雑音状況（有 / 無）に関する情報を付与している。

```
0001 00:04:376-00:07:415 M:D:I:N:
  ちょっと                & チョット
  小腹<H>が                & コバラ<H>ガ
  すいたんだけど<H>        & スイタンダケド<H>
  この                     & コノ
  近くに                   & チカクニ
  ファーストフード店で<H>   & ファーストフードテンテ<H>
  あるのかなあ<SB>          & アルノカナー<SB>
0002 00:08:376-00:08:879 F:O:I:N:
  はい                     & ハイ
  マクドナルドと            & マクドナルドト
  モスバーガーが            & モスバーガーガ
  ございます<SB>            & ゴザイマスガ<SB>
0003 00:10:672-00:10:304 M:D:N:N:
  (F あっじゃ              & (F アッ)ジャ
  マクドナルドの            & マクドナルドノ
  場所を                    & バシヨオ
  教えてほしいんだけど<SB> & オシエテホシンダケド<SB>
```

図3：書き起こしテキストの例

#### 4. むすび

本稿では 2000 年度までに収録したデータベースについて報告した。2002 年 2 月には 800 名の収録が終了し CIAIR ではこれらのデータも随時公開していく予定である。

CIAIR ではデータベースの利用者自身が CD-R による複製作業を行える環境を無償で提供している。

本データベースの利用により車内音声対話システム等、音声情報処理の研究が大きく発展することを切に願う。

#### 参考文献

- [Kawaguchi 2001] Nobuo Kawaguchi, Shigeki Matsubara, Kazuya Takeda, and Fumitada Itakura, "Multimedia Data Collection of In-Car Speech Communication," Proc. 7th European Conf. on Speech Commun. And Tech. (EUROSPEECH2001), Aalborg, Denmark, Sept. 2001.
- [河口 2000] 河口信夫, 松原茂樹, 若松佳広, 梶田将司, 武田一哉, 板倉文忠, 稲垣康善, “実走行車内音声対話コーパスの設計と特徴,” 信学技報, NLC2000-57, Dec.2000.
- [早川 2001] 早川昭二, 磯部俊洋, 河口信夫, 武田一哉, 板倉文忠, “音声対話システムを用いた車内対話の収集,” 音響学会講演論文集, 2001.
- [河口 2001] 河口信夫, 牛窪誠一, 松原茂樹, 岩博之, 梶田将司, 武田一哉, 板倉文忠, “走行車室内音声対話収録システムの開発,” 信学論(D-II), vol.J84-D-II, no.6, 2001.
- [小磯 2000] 小磯, 籠宮, 菊池, 前川, 土屋, 間淵, 斉藤, “「日本語話し言葉コーパス」の書き起こし基準について,” 信学技報, NLC2000-57, 2000.