音声言語情報処理 30−12
(2000. 2. 5)

# 実走行車内における音声データベースの構築

河口信夫[1,2] 松原茂樹[1,3] 岩 博之[1,4]
梶田将司[1,5] 武田一哉[1,2] 板倉文忠[1,5]

1) 名古屋大学統合音響情報研究拠点 (CIAIR)
2) 名古屋大学大学院工学研究科 3) 名古屋大学言語文化部
4) 小島プレス工業 (株) 5) 名古屋大学情報メディア教育センター

あらまし 名古屋大学 CIAIR では、ロバストな音声認識・音声対話システムの実現のために、実走行車内において、音素バランス文、連続数字、離散単語の読み上げ音声、およびナビゲーションや情報検索のための音声対話の収録を行っている。本稿では、車内音声データベースの目的、車内でのデータ収集方法、および本データベース構築のために特別に作成したデータ収集車について報告する。

キーワード 実走行車環境、ロバスト音声認識，音声データベース，音声対話コーパス，模擬対話

# Construction of Speech Database in Moving Car Environment

Nobuo Kawaguchi[1,2] Shigeki Matsubara[1,3] Hiroyuki Iwa[1,4]
Shoji Kajita[1,5] Kazuya Takeda[1,2] Fumitada Itakura[1,5]

1)Center for Integrated Acoustic Information Research(CIAIR), Nagoya University
2)Graduate School of Engineering, Nagoya University
3)Faculty of Language and Culture, Nagoya University
4)Kojima Press Industry Co. Ltd. 5)Center for Information Media Studies, Nagoya University

**Abstract** The CIAIR at Nagoya University has been constructing the speech database in moving car. The database consists of (1)phonetically balanced sentences, (2)continuous numerals, (3)separate words and (4)several spoken dialogue used for information systems of navigation and information retrieval. This paper reports the objective of the speech database, the method of the recordings and the developed recording car to construct the database.

**keywords** moving car environment, robust speech recognition, speech database, spoken dialogue corpus, simulated dialogue

## 1 はじめに

近年、車室内における高度情報システムの実現の要求が高まっている [1]。実走行車内において操作可能な情報システムを実現するためには、高騒音下でのロバスト音声認識技術と、自然発話を受理可能な音声対話技術が必要である。文部省中核的研究拠点 (COE) 名古屋大学統合音響情報研究拠点 (CIAIR) では、ロバストな車内音声対話システムの実現を一つの目標と定め、その要素技術の研究開発を進めている。

音声処理技術の発展のために、音声データベースは重要な役割を果たしてきた [2, 3, 4, 5, 6]。しかしながら、従来の音声データベースは防音室や静かな室内で収録された音声が中心であり、実環境における大量データの収録は、行なわていない。走行中の車室内は、ロードノイズや風切り音といった様々な騒音に加え、加速音やバンプ音といった非定常雑音が重畳される環境である。例えば時速 90km で走行している車内のダッシュボードからドライバーの音声を収録した場合、SN 比は -5dB 以下になる [7]。

CIAIR では、このような実走行車内環境での音声処理技術の研究開発を促進するために、大規模な車内音声データベースの構築を行うこととした。音声収録においては複数のマイクロホンを用い、口元だけでなく、車内のさまざまな位置の音を取り込んでいる。これは、実環境下におけるマルチチャネル信号音響処理の研究・評価にも利用可能なデータを収集するためである。ドライバーを被験者として、実走行中の車内において、音素バランス文、連続数字、および離散単語の読み上げを行う。また、自然発話を受理可能な音声対話技術の獲得のために、ナビゲーションやレストラン等の情報検索が可能な車内情報システムとドライバー間の対話を収録し、書き起しによって対話コーパスを構築する。

今回の収録では、車内情報システムを利用している環境を模擬的に実現するために、発話を制限するように訓練されたナビゲータが同乗し、情報システムの役割を果たしている。将来的には、ナビゲータを Wizard of OZ のシステムや実際の対話システムと置き換えて収録していくことを計画している。

また、収録するデータは、音声のみではなく、画像やハンドル操舵角力、アクセル・ブレーキの踏力、GPS による車両位置が同期されたマルチモーダルデータであり、今後の様々な研究に活用できることを配慮している。本データベースの収録のために、複数のモダリティの同期入力を可能とする専用のデータ収録車を作成した。

以下、本稿では、データベース構築の目的、データの収集方法、データ収集車、データの収集状況について述べる。

## 2 車内音声データベース構築の目的

走行車内における音声データベースの構築は、主に以下の項目を目的として行われている。

(1) **走行車内の音響モデル作成**
高雑音下でのロバスト音声認識技術の基礎データ，音環境調査，シミュレーションと実環境の差，多様な道路環境、車内環境での収集（市街地、郊外、高速道路、橋、トンネル、雨、車種の違い）

(2) **対話システムの基本資料収集**
走行車内特有な言語現象、言い回し、語彙、発話の重なり具合の調査、分析、モデル化，対話のタイミング（システムからの発話も含めて），発話内に含まれる重要情報の判断とその割合調査，コマンド対話か、自由発話かの判断材料（効率、使いやすさ、応答も含めて），システム利用の学習過程（ユーザはシステムにどのように適応するか），システムはユーザにどのように適応するべきか

(3) **走行車内のヒューマンインタフェースの研究データ収集**
運転歴、運転状況の違いによる理解能力、発話内容、発話単位、発話速度声質の変化 (ロンバート効果)[8] の測定，車載機器（駆動系）の高信頼操作システム実現のための基礎データ

また、本音声データベースが構築された暁には、以下のような特徴を持つデータベースとなる予定である。

(1) **大規模、実環境、マルチチャネル音声データベース**

(2) **走行車両と話者のインタラクションを記録したマルチモーダルデータベース**
単なる対話のみではなく、画像や車両操作による運転状況等のマルチモーダル情報が統合的に記録されたデータベース

(3) **車内対話に限定したタスク指向データベース**
対話タスクが車内情報システムのみを対象としており、タスク指向対話処理の研究に有効 [9]

(4) **マルチドメイン対話 [10] の解析データ**
車内対話タスクは、ナビゲーション、情報検

表 1: 収集するデータ

| | |
|---|---|
| 停車アイドリング中 | |
| 離散単語の読み上げ | 30 単語 |
| 連続数字の読み上げ | 4 桁 ×40 回 |
| 音素バランス文の読み上げ | 25 文 ×3 箇所 |
| 走行中 | |
| 音素バランス文の読み上げ | 25 文 |
| 情報検索タスク | 20–30 タスク |
| ナビゲーション | 32 交差点 |
| 走行距離 | 約 20km |

少し、喉が乾いてきました。
運転に疲れたので、一休みしたいですね。

これから、彼（彼女）とデートなのに
手持ちの現金がありません。

今日は結婚記念日です。
これから夫婦でお祝いをしたいと思います。

今日は、スポンサーつきなので、豪華に
フランス料理のフルコースが食べたいですね。

図 1: 状況プレートの例

索、車載機器の操作等の複数のドメインを同時に扱うため、ドメイン判断の研究に有効

(5) ナビゲータが機械の代役を担っている
制限されたナビゲータによる発話は、近未来の音声対話システムの方向性を示唆する可能性がある

## 3 データ収集方法

### 3.1 音響モデル用データベース

ロバスト音声認識、音響モデル学習用の音声収集は、主に音素バランス文、連続数字、離散単語の読み上げによって行う。音素バランス文については、特に様々な雑音下での音声収録を目的として、周囲状況の異なる場所での停車アイドリング中、および走行中に読み上げた音声を収録する。各データの詳細については表 1に示す。

#### 3.1.1 音声プロンプタ

ドライバーが実際の運転を行いながら音素バランス文や離散単語を発声してもらうために、携帯 Windows PC 上で音声ファイルを簡便に再生するための音声プロンプタを作成した (図 3) 。音素バランス文は、1 文をそのまま記憶して復唱するには長過ぎるため、1 文を数文節程度の耳から聞いて復唱可能な短い単位に分割し、音声プロンプトを作成した。

D: 夜はフランス料理のフルコースなんか食べたいんだけど どっか高いとこあるかなあ
N: はい フランス料理はこの近くに 6 軒ございます
D: スポンサーがついてるから一番高いところ頼むよ
N: はい ザンビ が近くにございます
D: そこは車停められる
N: はい 車は停められます
D: 値段はどれくらいかな ちなみに
N: はい 夜はコース料理で 8 千円と 1 万円があります
D: 予約しないとだめだよね
N: はい このお店は予約が必要となっております
D: それじゃあ午後 7 時から 4 名で予約しておいてください
N: はい かしこまりました ザンビ に午後 7 時より 4 名様ですね
N: コース料理はどちらのコースになさいますか
D: どちらのっていろいろあるの あ 1 万円だっけ そっちで
N: はい かしこまりました

図 2: レストラン検索・予約タスク

音声プロンプタは分割された音声ファイルをテンキーのみの操作で順に再生したり、繰り返したりする機能を持つ。また、走行車内において、ディスプレイを使わずに現在のプロンプトの状況を把握するために、規則合成音声を用いて現在の状態を発話する機能を持っている。

本音声プロンプタは、運転中のドライバーに対する任意の文や単語、連続数字の発声の指示を可能にし、本データベース構築の一つの要素技術となっている。

### 3.2 車内音声対話データベース

実際の走行車内で車内情報システムを利用する環境を想定し、1 時間強の走行コースを定め、仮想目的地を目指すナビゲーションと、レストランや駐車場などの情報検索や予約を行う情報検索タスクを行い、音声対話の収録を行う。今回の収録では、車内情報システムを模擬的に実現するために、発話を制限するように訓練されたナビゲータが同乗して情報システムの役割を果たした。ナビゲータは事前に走行コースを 4 ～ 8 回程度試走し、ナビゲーションに対する習熟を行っている。ナビゲーションは基本

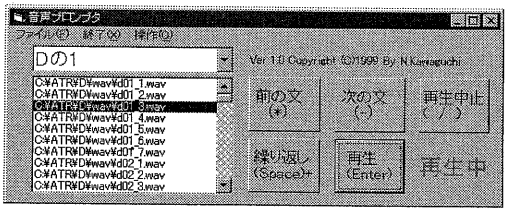


図 3: 音声プロンプタ

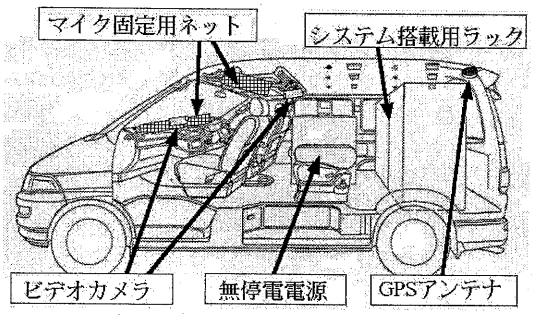


図 4: 収集用車両の概要

的に、車両位置に対するイベントであり、車両が特定の場所に来た時に、特定の指示を行う。また、ドライバーからの問いかけに対しては、随時応答を行う。また、ナビゲータ用に想定対話などを含むマニュアルを作成した。この中では、ナビゲータには視覚情報やジェスチャによる応答を行わないように指示してある。

情報検索タスクに対しては、事前に指示を与えても、運転中にドライバーがそれらを思い出して実行することは困難である。そこで、ドライバーに対する検索タスクの指示のために指示プレート、および状況プレートを用いる。指示プレートには「和食」「中華」「洋食」「喫茶店」「コンビニ」「銀行」「駐車場」といったシンプルな単語が書かれている。運転中のドライバーは、ナビゲータからプレートを見せられた時、その単語に関する要求発話を行うように指示されている。

本収録を行うために、1999 年 10 月に 10 名の被験者によるレンタカーを用いた試行実験を行なった。その結果、指示プレートのみでは、ドライバーはプレートに書かれた単語を単に復唱することが多く、複雑な要求は少なかった。また、似たようなプレートが繰り返されると、異なる発話の生成が困難となり、対話がパターン化する傾向が見られた。

そこで、自然発話を促し、対話に広がりを持たせるために、状況を表す文 (状況プレートと呼ぶ) を作成し、文章をそのまま読み上げても、情報システムに対する指示にならないような工夫を行なった。状況プレートの例を図 1に示す。短い文であるが、走行中に読むことは危険であるので、信号待ち等の余裕がある時にドライバーに見せることとした。また、そのまま読み上げることを防ぐために、プレートを見せた直後には発話せず、プレートを隠した後に、発話することとした。

状況プレートの導入の結果、明らかにバラエティに富んだ対話が増加した。しかしながら、図 2に示す対話に見られるように、図 1の最初の状況プレートの読み上げに近い発話も生じている。ただし、このような発話は、タスクの最初の発話に特徴的であり、以後の対話は自然発話が中心になっている。

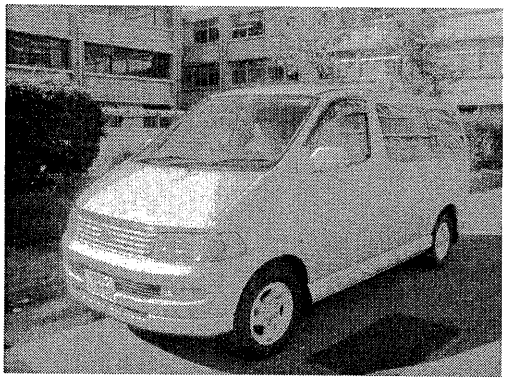

図 5: 収集車両の外観

表 2: 収録機器仕様

| メディア | 仕様 |
|---|---|
| 音声入力 | 16ch,16bit,16kHz–24kHz |
| 音声出力 | 16ch,16bit,16kHz–24kHz |
| 画像入力 | 3ch, MPEG1 (MPEG2) |
| 操舵角力 | ハンドル 16bit,1kHz |
| 踏力 | アクセル・ブレーキ 16bit,1kHz |
| D-GPS | ジャイロ, 加速度センサ付き |

## 4 実走行データ収集車

音声、画像、他の情報を収集するために、本収集実験専用のデータ収集車両を作成した。本車両は音声入出力、画像入力、ハンドル、アクセル、ブレーキ、車両位置などの情報をすべて同期しながら 90

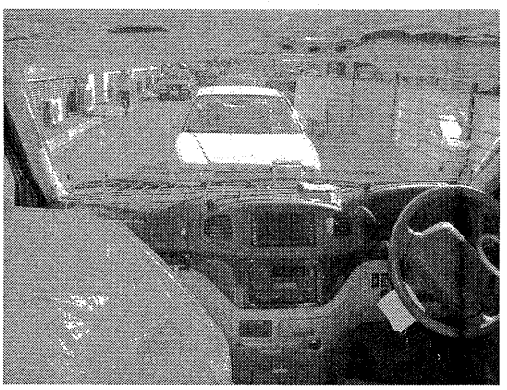

図 6: 収集車両のダッシュボード

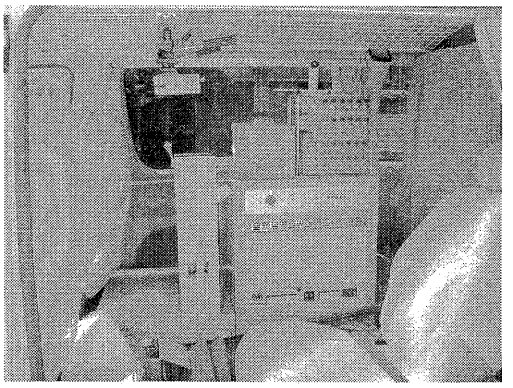

図 8: 後席（無停電電源装置とマイクアンプ

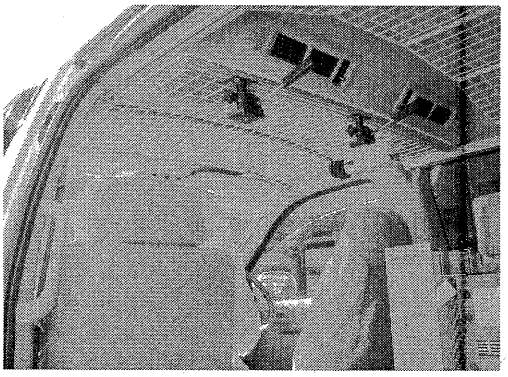

図 7: 収集車室内天井部

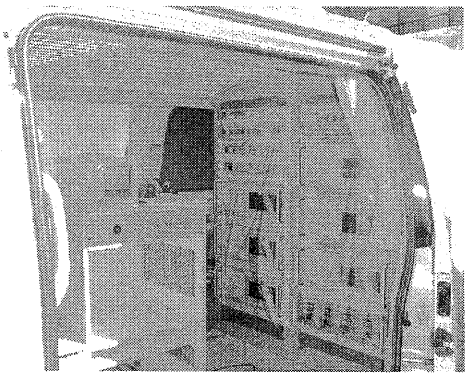

図 9: 収録装置の収納ラック

分間の収録が可能である。収集車両は 6 台の収録用 PC 及び、1 台の制御用 PC を搭載しており、これらは 100Base-T のネットワークで接続されている。データ収録機器の仕様を表 2に示す。これらの装置の稼働は標準の車載バッテリのみでは容量不足が生じるため、専用のオルタネータおよび無停電電源装置を設置してある。車室内の改造により、乗車定員は 3 名となっている。収集車両の概要を図 4に示し、外観を図 5に示す。図 6や図 7に示すようにダッシュボードや天井には、自由な位置にマイクを固定するためのネットが取り付けられている。また、ビデオカメラの取り付け位置を変更可能にするために、天井にはレールが取り付けられている。後席 (図 8) には無停電電源装置やマイクアンプ、TV モニター等が設置され、車両後部 (図 9) には、PC6 台とネットワークを収納するラックが据え付けられている。

## 5 収集データの状況

車内対話音声を書き起したものを図 2,10に示す。ナビゲーションタスクにおいては、ユーザからの確認発話が一部見られるだけで、多くの発話はナビゲータからなされている。一方、レストラン検索・予約タスクでは、ユーザ主導からシステム主導への自然な移行がしばしば見られる。また、運転状況によっては、ナビゲーションタスクと情報検索タスクが混在する場合もあり、複数のタスクドメインの対話が並行して行われる現象は、車内対話の特徴を良く表している。

今後は、大量のデータをさらに解析する必要がある。また、音声対話コーパスの共有化 [11, 12] のために、共有可能なタグ構造を採用することを予定している。

N: 次の交差点 桜が丘を右方向です
N: 信号を右方向へ進み すぐの二股を右方向へ進んでください
D: ここ右ですね
N: はい 右方向です

図 10: ナビゲーションタスク

## 6 まとめ

CIAIR では、ロバストな音声対話認識技術確立のために、高騒音下である実走行車室内の音声データベースの構築をはじめている。今後は、Wizard of OZ 法や音声認識を用いた対話システムを用いた音声対話の収録を進めていく計画である。また、収録した音声や対話データは、整理を行って随時公開していく予定である。

本車内音声データベースは構築を始めたばかりである。意見、要望等があれば、忌憚なく知らせ願いたい。本データベースの構築により、ロバスト音声認識や車内音声対話システム等の研究がより大きく進展することを切に願う。

## 謝辞

本研究は文部省科学研究費補助金 COE 形成基礎研究費 (課題番号 11CE2005) の補助を受けて行われた。データベース収集の目的や方法について熱心に検討を頂いた CIAIR コーパス WG の諸氏に感謝致します。

## 参考文献

[1] 庄境 誠, 中村 哲, 鹿野清宏：音響エコーキャンセラによる車室内音声認識性能の改善, 日本音響学会誌, Vol.53, No.11,pp.877–882(1997).

[2] 武田一哉, 匂坂芳典, 片桐 滋, 桑原尚夫：研究用日本語音声データ・ベースの構築, 日本音響学会誌, Vol.44, No.10, pp.747–754(1988).

[3] 匂坂芳典, 浦谷則好：ＡＴＲ音声・言語データベース, 日本音響学会誌, Vol.48, No.12, pp.878–882(1992).

[4] 川森雅仁, 島津 明, 堂坂浩二, 中野幹夫：対話処理のためのコーパス作成, 信学技法, NLC 97-5, pp.31–36(1997).

[5] 板橋秀一：音声コーパス, 情報処理, Vol.38, No. 11, pp.1012–1018(1997).

[6] Katunobu Itou, Mikio Yamamoto, Kazuya Takeda, Toshiyuki Takezawa, Tatsuo Matsuoka, Tetsunori Kobayashi, Kiyohiro Shikano and Shuichi Itahashi : JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research, J. Acoust. Soc. Jpn.(E),Vol. 20, No. 3,pp.199–206(1999).

[7] 中村 哲, 鹿野清宏：認識技術の進展, 情報処理, Vol.38, No.11, pp.998–1006(1995).

[8] 新田恒雄, 南 重信, 中山 昭, 小野木智宏：高騒音下における音声認識方法の検討, 信学技法, SP 94-20 , pp.45–52(1994).

[9] 松原茂樹, 河口信夫, 外山勝彦, 稲垣康善：発話の同時理解・同時生成に基づく車内音声対話秘書システムの提案, 人工知能学会研究会, 話し言葉と理解, SIG-SLUD-9902, pp.1–6(1999).

[10] Nobuo Kawaguchi, Shigeki Matsubara, Katsuhiko Toyama,Yasuyoshi Inagaki: An Architecture for Multi-Domain Spoken Dialog Systems, Proceedings of the 5th Natural Language Processing Pacific Rim Symposium(NLPRS'99), pp.463–466(1999).

[11] 土屋 俊, 堀内靖雄, 石崎雅人, 前川喜久雄：音声対話コーパスの共有化へ向けて, 人工知能学会誌, Vol.14, No.2, pp.231–242(1999).

[12] 堀内靖雄, 中野有紀子, 小磯花絵, 石崎雅人, 鈴木浩之, 岡田美智男, 仲 真紀子, 土屋 俊, 市川 熹：日本語地図課題対話コーパスの設計と特徴, 人工知能学会誌, Vol.14, No.2, pp.261–272(1999).