音声・映像・運転行動コーパスを用いた車内情報インタフェースの研究
In-car behavioral signal corpus for the advanced human interface technology study

武田 一哉 † **河口 信夫** ‡ **アブット フセイン** ‡ **板倉 文忠** §
Kazuya Takeda Nobuo Kawaguchi Huseyin Abut Fumitada Itakura

1. はじめに

安全かつ快適な移動手段は，現代の日常生活に欠くことのできない社会基盤であり，その重要性は日々増している。さらに今日急速に発展する情報通信技術を，高度な移動技術へ応用する様々な試みが行われている。無線通信，遍在計算，個人認証などの要素技術とならんで，移動環境下で情報システムを快適に利用するためには，インタフェース技術の向上も必要である。車内での携帯電話の利用の例を見るまでもなく，安全性や快適性の確立には，制度的な取り組みと利用実態に即した技術革新が常に必要である。

車内での情報アクセスの有力な手段として「音声認識」が期待されている。すでに多くのカーナビシステムには地名などの音声入力機能が実現されているが，必ずしもその有用性が広く認められるには到っていない。その最も重要な原因は,（開発・評価時と比較して）実際の利用時における音声認識性能が十分でないことに求められる。車内インタフェースのための音声認識技術を発展させるためには，実際の運転環境下で収録されたデータを開発・評価に利用することが不可欠である。本講演では，名古屋大学統合音響情報研究拠点（CIAIR）において行われた，大規模な車内音声対話収集実験について述べる。

2. データ収集の概要

2.1 対話音声

音声データは３年の期間に渡り延べ 800 名の話者について行なわれた。各話者の発声は，音素バランス文（停車中 50 文，走行中 25 文），対話音声（対オペレータ，対 WOZ システム，対音声認識システム），単語発声などであり停車中を除き，全ての発話は運転中に運転席にて行われた。一人あたりの乗車時間は２時間程度，収録音声は１時間程度である。このデータ収集は，名古屋市内の市街地の走行により行った。

2.2 単語音声

上記，対話音声を中心としたデータベースとは別に，多様な走行条件下での音声収集を行うために，孤立単語発声の収録を行った。話者毎に，停車，市街地走行，高速走行の３つの走行条件の下,「エアコンの状態（LO, HI)」,「CD の再生」,「窓明け」,「通常」の５つの車内の状況について収集が行われている。当該発声は 100 名について，上記 15 の条件下で 50 単語について収録されており，延べの発声数は 75000 単語に上る。

†名古屋大学 CIAIR/情報科学研究科
‡名古屋大学 CIAIR/情報連携基盤センター
§名古屋大学 CIAIR/工学研究科

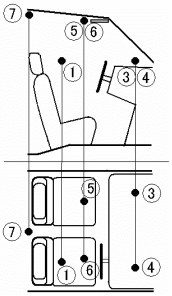

図 1: 分散マイクの設置位置

2.3 収録システム

実験用の収録システムは，音声だけでなく画像や運転に関連する多様な信号を，複数チャネルで同期して取り込めるよう設計した。実験用データ収集車は，音声 16 チャネル，画像３チャンネル，運転データ５チャンネル（アクセル踏力，ブレーキ踏力，ハンドル角，エンジン回転数，車速度），GPS 信号，を全て同期して取り込むことができる。音声は，接話マイク（運転者，オペレータ）の他，車内 6 箇所に設置された分散マイクと，バイザー位置に設置された 4 マイクからなる線形アレイマイクシステム，の計 12 箇所で収録された。分散マイクの設置位置は，図 1 に示すとおりとした。運転行動に関する計測データ例は，図 2 に示すとおりである。

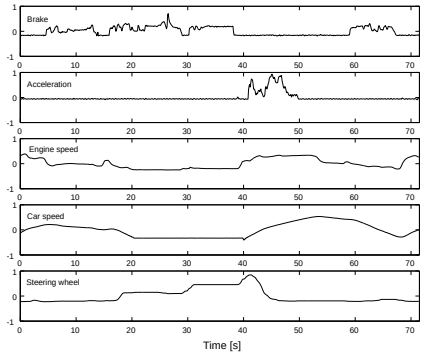

図 2: 運転行動データの記録例：上からブレーキ踏力，アクセル踏力，エンジン回転数，車速，ハンドル切り角

3. 音声認識の基本性能の評価

3.1 単語音声の基本認識性能

単語音声の基本認識性能を図 3 に示す。50 単語の孤立単語認識の結果であり，発話区間の切り出しは手動で

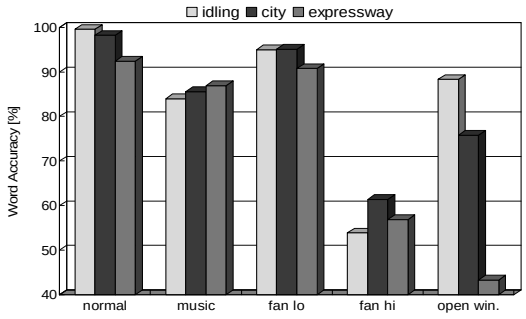

図 3: 単語認識の基本性能

行った。マイク位置 6 番（バイザー付近に設置）で収録した音声を，6000 文（約 200 名による発声）のバランス文で学習した HMM で認識した結果である。学習文のうち、4000 文はアイドリング中に，2000 文は走行中に発声された音声であり，32 混合分布を持つ 500 状態を共有するトライホンを基本認識単位としている。特徴パラメータには，250Hz〜8000Hz の帯域の音声から求めた，12 次までの MFCC に加え MFCC と対数パワーの回帰係数の合計 25 次元の特徴パラメータを用いた。

ファンや CD，窓開けのような車内の環境要因がない限り，バイザー位置にあるマイクと，接話マイクとの間で認識性能の間には大きな差異は見られなかった。

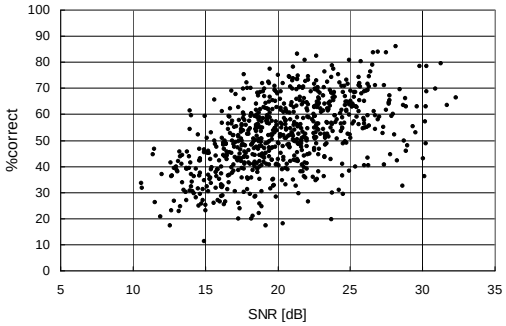

図 4: 対話音声認識の基本性能

3.2 対話音声の基本認識性能

図 4 には，運転者とオペレータとの対話音声,（レストラン案内に関するタスク，語彙約 5000 単語）の認識結果（単語認識率）を示す。横軸は，認識対象音声を収録したバイザー付近のマイク（6 番マイク）における SN 比を，話者毎に平均して表している。認識実験は約 7000 名の音声について行った。音響モデル，言語モデルとも，学習は同一の 700 名のデータを用いて行っており，実験はクローズの条件で行われたものである。話者毎の平均 SN 比の値はおおよそ 15dB から 25dB の間に分布しており，平均的には 5dB の SN 比の向上が 10%程度の単語認識率の向上に対応している。しかしそのばらつきは非常に大きいことが分かる。

現在収集された音声データを用いて，複数マイクロホンを用いた雑音抑圧方法の研究 [2]，対話システムにおける発話文の生成方法 [3] などの研究を行っている。

4. 画像データ・運転行動データの利用

車内の音声インタフェースの信頼度を向上させるためには，音声認識システムの基本性能の向上はだけでなく，運転者の様々な行動データの分析理解を図ることが重要と考えられる。このような研究の一例として，当該データベースに収録された顔画像のデータを利用して，音声の発話区間の検出精度を向上させる方法が報告されている [4]。さらに，運転行動と音声の発話（フィラー密度など）との関係に関する基礎的な検討も行われている [5]。

一方，音声データと同時に収録した，運転に伴う多様な行動データの分析法と利用法の研究も開始している。一例として，運転者の個人性の抽出に関する基礎的な分析を行った結果，運転行動に静的な個人特徴が認められただけでなく，動的な個人特徴についても抽出が可能であることが示唆された [6]。

運転行動信号と音声対話信号の統合的な分析は，情報システムと自動車制御システムのインタフェースを統合するために極めて重要な研究課題であり，今後研究の進展が期待される。

5. むすび

実際の運転環境下で収録された，対話音声・運転行動に関するデータベースを概観し，当該データベースを用いて進められているいくつかの研究を紹介した。

謝辞

本研究の一部は，科学研究費補助金（11CE2005）の補助を受けて行われた。

参考文献

[1] Kawaguchi, N., Takeda, K., et al., “Construction of Speech Corpus in Moving Car Environment”, Proc. International Conference on Spoken Language Processing, pp.1281-1284, 2000 (ICSLP2000, Beijing, China).

[2] T.Shinde K. Takeda and F. Itakura, “Speech Recognizer-based microphone array processing for robust hands-free speech recognition”, Proc. International Conference on Spoken Language Processing, Vol.I, pp.897-900, 2002 (ICSLP2002, Denver)

[3] 村尾浩也，河口信夫，松原茂樹，山口由紀子，稲垣康善『 WOZ システムのログ情報を利用した事例ベース音声対話システムの開発』第 4 回音声言語シンポジウム論文集 (情報処理学会研究報告 (SIG-SLP-44)), pp.135-140, Dec. (2002).

[4] 坂義秀，二宮芳樹，森健策，末永康仁『車載カメラ映像を用いたドライバの発話区間検出』，信学技報 (PRMU2003 March)

[5] 清水司，脇田敏裕，武田一哉，河口信夫，板倉文忠『電話番号案内タスクにおける停車中と運転中のドライバ発話の特徴』音講論集 3-P-26,pp215-216, 2001.3

[6] K.Igarashi, K.Takeda, F.Itakura and H.Abut, “Is our driving behavior unique ?”, proc. Workshop on DSP in mobile and vehicluar systems, (Nagoya 2003 April)