実走行車内音声対話データベース

河口 信夫※† 松原 茂樹※‡ 武田 一哉※§ 板倉 文忠※§ 稲垣 康善§

※名古屋大学統合音響情報研究拠点（CIAIR）
†名古屋大学大型計算機センター，‡名古屋大学言語文化部
§名古屋大学大学院工学研究科
〒464-8601 名古屋市千種区不老町１
E-mail: †kawaguti@cc.nagoya-u.ac.jp, ‡matubara@lang.nagoya-u.ac.jp
takeda@nuee.nagoya-u.ac.jp

あらまし 名古屋大学統合音響情報研究拠点(CIAIR)では，実走行車内における音声や対話を継続的に収録している．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の状況について報告する．本データベースでは，被験者であるドライバの対話相手として，ナビゲータ（人），Wizard of OZシステム，および音声対話システムという異なる対象との対話を収録している．本稿では，データベースの構成，および話し言葉の観点からの解析結果について述べる．また，様々な雑音環境下での単語音声データベースによる音声認識結果についても報告する．

キーワード 音声データベース，車内音声対話，ロバスト音声認識，話し言葉処理

.

In-Car Spoken Dialogue Database

Nobuo KAWAGUCHI※†, Shigeki MATSUBARA※‡, Kazuya TAKEDA※§
Fumitada ITAKURA※§ ,and Yasuyoshi INAGAKI§

※Center for Integrated Acoustic Information Research (CIAIR), Nagoya University
†Computation Center, Nagoya University, ‡Faculty of Language and Culture, Nagoya University
§Graduate School of Engineering, Nagoya University
1, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
E-mail: †kawaguti@cc.nagoya-u.ac.jp, ‡matubara@lang.nagoya-u.ac.jp
takeda@nuee.nagoya-u.ac.jp

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. We also report the speech recognition rate of the isolated word database in the different kind of noise situation while car-driving.

Key words speech database, in-car spoken dialogue, robust speech recognition, spoken language processing.

1. はじめに

名古屋大学統合音響情報研究拠点（CIAIR）では，ロバストな音声対話の実現のために実走行車（図１）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，その収録および，観測される言語現象について述べる．

CIAIR における車内音声データベースの収録[2]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．収録方法や収録機材の詳細については[1]を参照されたい．

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ナビゲータとに加え，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．本稿では，また，単語音声データベースとその評価結果についても述べる．

図 1：データベース収録車

2. 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表１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[8]を参照されたい．2000 年度には，より実際的な収録を目指し，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せを用いて行っている．

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, ８ch
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

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２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．

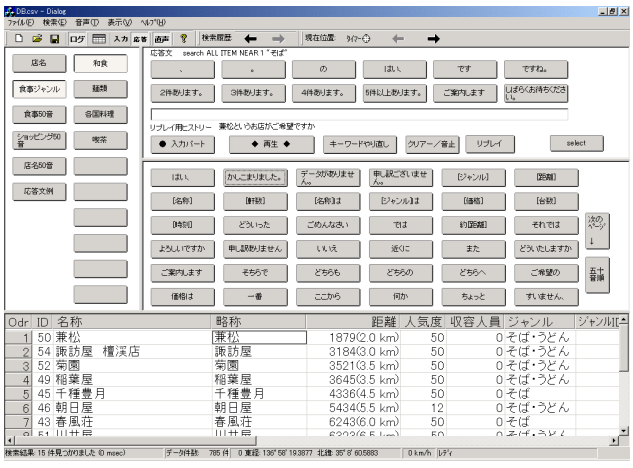

図 2：ＷＯＺの応答入力画面

表３：音声対話データベースの概要

	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

また，本データベース収集のために，車内で実際に稼動する音声対話システムを構築した[9]．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．

3. 車内音声対話データベースの基礎情報

3.1 データベースの収録状況

表３に 1999 年度，及び 2000 年度に収録した CIAIR 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[11]に準拠した書き起し基準に従って，すべて書き起しが付与されている．2000 年度は，人(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，茶筅 ver2.1, ipadic2.4 を用いた．現在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．

この表から，WOZ や対話システムとの対話においては，対人との対話と比較して，全体の発話密度が低くなることが読み取れる．ＷＯＺによる収録は，タッチパネルの操作のため，システムの応答がどうしても遅れがちになり，間の多い対話になる．しかし，ドライバの発話速度や発話単位あたりの形態素数は，対人との対話と大きな違いは無い．一方，対話システムはシステム主導型であるため，どうしてもシステム側の発話が長くなる．そのため，ドライバの発話が少なくなっている．さらに，ドライバの発話が発話単位あたり約 3.2 形態素となり，他の対話形式と比較して短くなる傾向がある．これは，システム主導の対話システムであることから，ドライバは短文での応答が多くなるためである．

3.2 収録した車内対話のタスク

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．

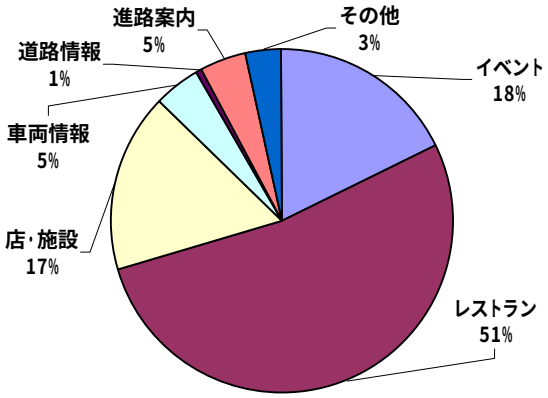

図３: 収録されたタスクの分布

一概に車内対話といっても，その複雑さは対話タスクによって異なる．一般的には，多くのターンテイキングを必要とする会話ほど，タスクの達成に多くのコストがかかっていることを意味する．多くのターン数が必要となるタスクほど，対話上の分岐点が多くなり，それらに適切に対処する必要性が生じる．

各対話タスクのターン数を調べた．タスクの種類ごとの総ターン数，および平均ターン数をそれぞれ表４，図４に示す．車両情報や道路情報などゴールが比較的明確な対話は，少ないターンでそれを達成することができるが，イベント対話のようにゴールがあいまいな場合には，より多くのターンが必要となる．また，レストラン対話のように，ゴールがはっきりしている場合でも多くの検索条件の設定が必要な対話では，ターン数が多くなる．

表４：タスクの種類とサイズ

タスク	対話数	総ターン数
イベント	326	4821
レストラン	1222	14569
店・施設	568	5086
車両情報	179	1271
道路情報	28	165
進路案内	531	1336
その他	99	914
合計	2854	27248

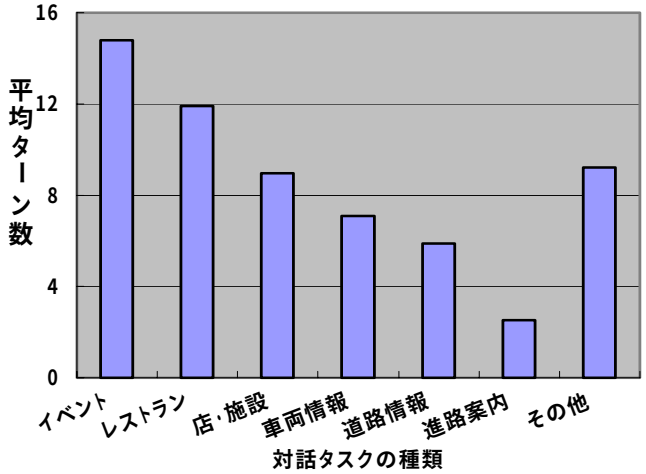

図　4:対話タスクの種類と発話ターン数

タスク達成に要したターン数がその複雑さのひとつの目安になるとするならば，そのデータは今後の車内音声対話システムの開発プロセスに対する１つの指針となる．すなわち，ターン数が少ないタスクの対話システムから多いタスクのシステムへと段階的に開発を進めていくことが考えられる．実際，進路案内や道路情報などのタスクを遂行するための機能は，現在のカーナビゲーションシステムにも備わっており，実用化技術に達している．

3.2 対話音声の係り受け分析

車内音声対話におけるドライバ発話の言語的特徴を明らかにするために，ナビゲータとの間で遂行された 81 対話を対象に係り受け分析を与えた．係り受けの付与は，各発話単位に対して人手で実施し，データの品詞体系や係り受け文法については，基本的に京大コーパス[12]の作成基準に準拠した．ただし，話し言葉に特有な部分については以下のように作業基準を設けた．

- フィラー及び言い淀みは，どの文節とも係り受けの関係にない．すなわち，それらが単独で係り受け構造を形成する．
- 受け文節が省略された場合，係り先がない文節とする．
- 話し言葉に固有の言い回し(「こっから」，「食べてえ」など)については，新たな辞書項目を設け，形態素ごとに品詞を定めた．

24,972 文節からなる 7,784 発話単位に対して，11,148 個の係り受けが存在した．発話単位あたりの平均文節数は 3.21 であり，文章における一文あたりの平均文節数が 10 程度である新聞などの書き言葉と比べてその長さは短い（図５）．

特に，車内対話では，ドライバは運転しながら発話することが多く，対話への集中度が低下するため，発話長は短く，また，その言語構造も複雑になることはあまりない．実際，発話単位あたりの平均係り受け数は 1.43 個であり，図６に示すように 2 個以下の係り受けからなる発話単位は全体の約 80%を占めている．

しかしながら，このことは話し言葉の係り受け解析処理が書き言葉のそれに比べ，著しく簡単であることを必ずしも意味しない．というのも，話し言葉では，あらゆる文節が受け文節をもつわけではなく，係り先がない文節の特定も必要なるためである．実際，全係り受けデータの約 50%に相当する 3,890 発話単位に，受け文節が存在しない文節が存在した．

表 5：係り受け分析データ

対話数	81
発話単位数	7784
総文節数	24972
平均文節数(発話単位あたり)	3.21
総係り受け数	11148
平均係り受け数(発話単位あたり)	1.43

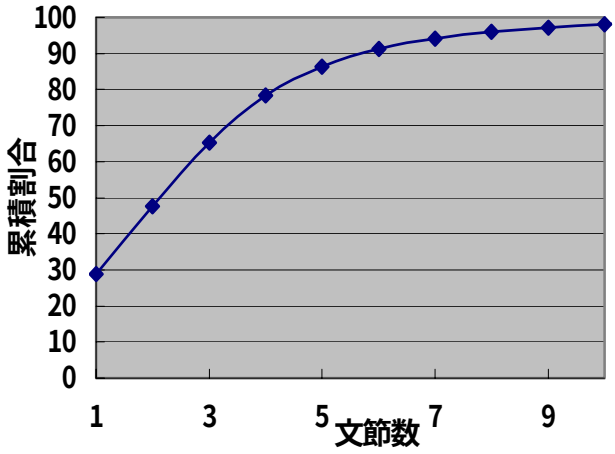

図５:発話単位あたりの文節数と累積割合

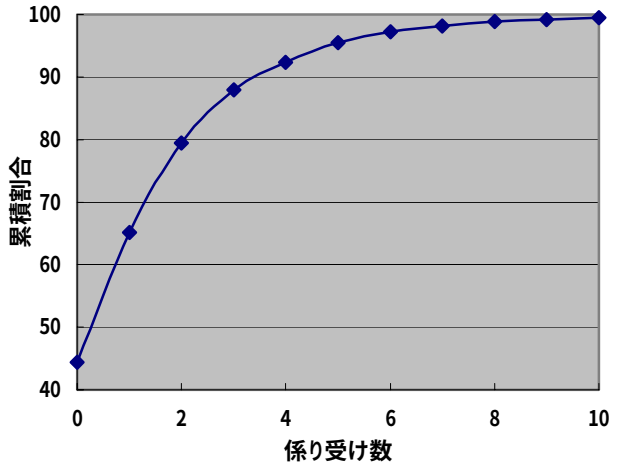

図６：発話単位あたりの係り受け数の累積割合

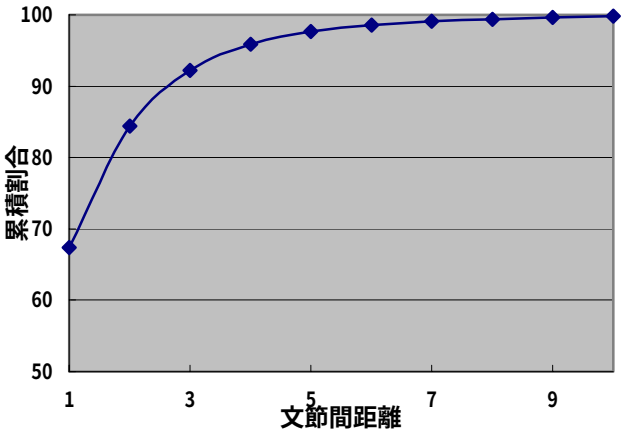

図７：文節間距離の累積頻度

また，係り文節と受け文節の間の距離を調べた．この距離は，文章の理解容易性を計る指標の一つであり，受け文節の位置が係り文節から離れるほど，一般に，理解しにくい表現となる．隣接した文節間の距離を 1 としたときの文節間距離の累積割合を図７に示す．全体の約 3 分の２が隣接する文節間の係り受けであり，また，9 割以上の係り受けが 3 文節以内に係っている．このことからも，ドライバ発話は比較的単純な係り受け構造をもつことがわかる．

4. 単語音声データベース

種々の走行条件下で単語音声データの収録を行った．データベースの主な諸元を表６に示す．データは一つの条件毎に５０単語が収録されており，マイクロホン設置場所のバリエーション（６箇所：図８）を含めると，延べ約１１万単語が収録されている．

当該データベースを用いて，走行条件別に収録された単語発声の認識実験を行った．運転席サンバイザー付近に装着したマイクにより収録したバランス文約８０００文（走行中２５００文，アイドリング中５５００文）により，状態当たり３２混合分布のモノフォンHMMの学習を行った．１６kHz でサンプリングされた音声から，250 Hz 以下の周波数成分をカットし，２４チャネルのメルフィルタバンク分析の結果得られた対数スペクトルから，１２チャネルのＭＦＣＣを求め，これらのデルタ係数とデルタ対数パワーを加えた２５次元の特徴量を認識に用いている．

得られた認識結果を，図 9 に示す．オーディオやエアコンなどが切られた状態であれば，高速走行中でも９０％以上の認識性能を得ることができる．エアコンがＨＩの状態ではＳＮ比を十分得ることが困難であるため，認識性能は著しく劣化するが，エアコンがＬＯの状態では，認識率の劣化はエアコンＯＦＦの状態に比べ大略５％程度にとどまっている．一方，オーディオがＯＮの場合，アイドリング時に比べ走行時においてより高い認識性能が得られることが明らかになった．

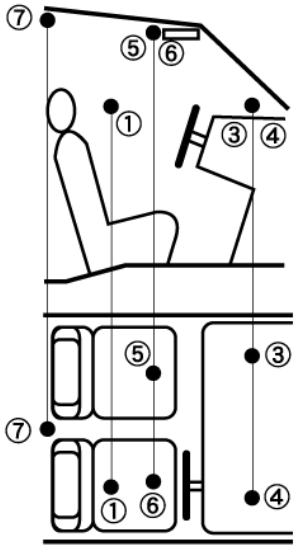

図８：収録に用いたマイクの位置

表６：単語音声データベースの内容

話者	２０名（男性１０名，女性 10 名）
発声内容	５０単語
セッション数	各話者 16 セッション
運転条件	アイドリング，市街地走行，高速道路
収録条件	通常，エアコン(LO)，エアコン(HI)，オーディオオン，ハザード
収録マイク	６個所に設置された無指向性マイク（SONY ECM77B）

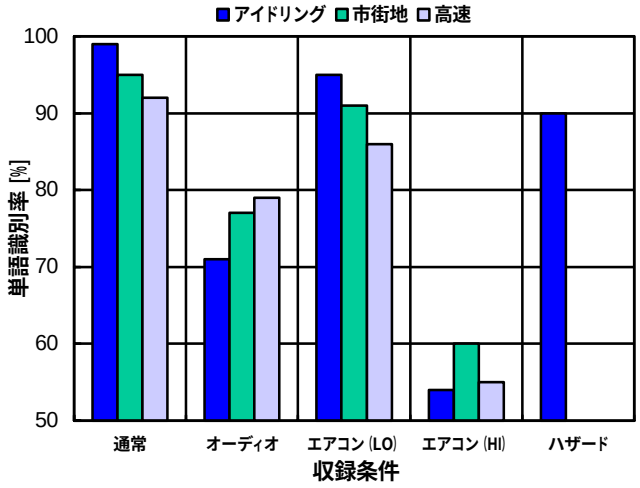

図９：単語音声識別率

5. 関連研究

車内音声対話システムの構築に向けての取り組みは国際的にも進められている．SpeechDat-Car[4][6]は欧州や米国の複数の機関で進められている音声コーパス収集プロジェクトである．このプロジェクトでは，様々な言語で同じ基準に基づいて車内音声が収録されている．ただし，収録は音声が中心である．また，SpeechDat-Car での収録データに基づくハンズフリー車内音声認識の研究も始められている[5].

CU-Move[3]は，コロラド大学で開発されている車内音声対話ナビゲーションシステムである．これは，DARPA の Communicator プロジェクトに参加している研究であり，道案内を行うことを目的にしている．この研究では，携帯電話と WOZ を用いた収録を行っている．我々も携帯電話を用いたデータ収集の試みを行っている[13].

6. おわりに

名古屋大学 CIAIR で収録されている車内音声対話データベースの現状について報告した．2001 年度もデータベースの収集を続けており，本年度終了時には約８００名のデータを利用できるようになる．本データベースに基づく研究についてもデータベースの整理と同時に進められており，コーパスに基づく音声対話や，対話制御などが行なわれている．

５節に述べたように，車内における音声対話の重要性が認められ，さまざまな研究プロジェクトが国際的にも進められている．本データベースがそのような研究を支えることを願う．

謝辞　本研究は文部科学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．データベースの収集や方法について多大な貢献をされた CIAIR スタッフ及び共同研究者の諸氏に感謝致します．

文　献

[1] 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，pp.909-917，Jun.2001.

[2] 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), pp.2027 – 2030, Aalborg, Demark, Sept. 2001.

[3] John H.L. Hansen, Pongtep Angkititrakul, Jay Plucienknoski, Stephen Gallant, Umit Yapanel, Bryan Pellom, Wayne Ward, and Ron Cole, ““CU-Move”:Analysis & Corpus Developmemt for Interactive In-Vehicle Speech Systems,” Proc. 7th European Conf. on Speech Commun. And Tech. (EUROSPEECH2001), pp.2023 – 2026, Aalborg, Demark, Sept. 2001.

[4] Peter A. Heeman, David Cole, and Andrew Cronk, “The U.S. SpeechDat-Car Collection,” Proc. 7th European Conf. on Speech Commun. And Tech. (EUROSPEECH2001), pp.2031 – 2034, Aalborg, Demark, Sept. 2001.

[5] M. Matassoni, M.Omologo, and P.Svaizer, “Use of Real and Contaminated speech for Training of a Hands-Free In-Car Speech Recognizer,” Proc. 7th European Conf. on Speech Commun. And Tech. (EUROSPEECH2001), pp.1569 – 1572, Aalborg, Demark, Sept. 2001.

[6] H. van den Heuvel, J.Boudy, R.Comeyne, S.Euler, A. Moreno, and G.Richard., “The SpeechDat-Car Multilingual Speech Databases for In-Car Applications,” Proc. 6th European Conf. on Speech Commun. And Tech. (EUROSPEECH99), pp.2279 – 2282, Budapest, Hungary, Sep. 1999.

[7] Nobuo Kawaguchi，Shigeki Matsubara，Hiroyuki Iwa，Shoji Kajita，Kazuya Takeda，Fumitada Itakura，and Yasuyoshi Inagaki，“Construction of Speech Corpus in Moving Car Environment,” Proc. 6th International Conference on Spoken Language Processing (ICSLP2000)，pp.362-365，Beijing，China，Oct. 2000.

[8] 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，pp.61-66，Dec.2000.

[9] 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，Mar. 2001.

[10] 松原茂樹，佐藤利光，河口信夫，稲垣康善，“ 統計データに基づく話し言葉音声の係り受け解析，” 情処研報，SLP-36-4，pp.23-28，Jun. 2001.

[11] 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」の書き起し基準について，” 信学技報，NLC2000-57，pp.55-60，Dec.2000.

[12] 黒橋，長尾，京都大学テキストコーパス・プロジェクト，言語処理学会第３回年次大会論文集，pp. 115-118 ,1997.

[13] 清水，脇田，武田，河口，板倉，“ 停車中と運転中のドライバ発話の特徴，” 音響学会講演論文集，Sep.2000.