実走行車内音声対話コーパスの設計と特徴

河口 信夫[1 2]　松原 茂樹[1 3]　若松 佳広[5]　梶田 将司[1 4]
武田 一哉[1 5]　板倉 文忠[1 4]　稲垣 康善[5]

1 名古屋大学統合音響情報研究拠点 (CIAIR)
2 名古屋大学大型計算機センター
3 名古屋大学言語文化部
4 名古屋大学情報メディア教育センター
5 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
kawaguti@nuie.nagoya-u.ac.jp

あらまし　本稿では、名古屋大学音響情報研究拠点 (CIAIR) で構築中の実走行車内音声対話コーパスの設計と特徴について述べる。道案内や店情報検索をタスクとする 162 対話を対象とした特徴分析の結果、(1) ドライバーの発話速度は通常の対話音声に比べて遅く、5 〜 7(mora/sec) である、(2) ドライバーの発話におけるフィラーの出現頻度は、1 発話単位あたり 0.33 個、1 秒あたり 0.174 個であり、通常の人間対人間の自由対話に比べて少ない、(3) 車両の走行中と停止中とでは、発話速度や話し言葉に特有な現象の出現に関して差がない、(4) 停止中に比べ走行中の発話には、感動詞、及び、文発声途中でのポーズの出現頻度が高い、ことなどが明らかになった。

キーワード　音声コーパス、音声対話、車内対話、実走行環境、ドライバー発話

Design and Characterization of In-Car Speech Corpus

Nobuo Kawaguchi[1 2]　Shigeki Matsubara[1 3]　Yoshihiro Wakamatsu[5]　Masashi Kajita[1 4]
Kazuya Takeda[1 5]　Fumitada Itakura[1 4]　Yasuyoshi Inagaki[5]

1 Center for Intergrated Acoustic Information Research, Nagoya University
2 Computation Center, Nagoya University
3 Facutly of Language and Culture, Nagoya University
4 Center for Information Media Studies, Nagoya University
5 Graduate School of Engineering, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, 464-01, Japan
kawaguti@nuie.nagoya-u.ac.jp

Abstract　This paper describes the design and the characterization of the in-car speech corpus which CIAIR at Nagoya University has been collecting. The investigation of 126 spoken dialogues has indicated the following characteristic features: (1) The speed of the driver's speech is 5 〜 7 (mora/sec), more slowly than that of the usual conversational speech, (2) The occurrence of fillers is, less than that of the spontaneous speech, in a ratio of 0.33 to a speech unit, (3) The speech speed and the filler occurrence have little difference between the speech in a moving car and that in a stopped one, and (4) The interjectional expressions and the pauses occur more frequently in the driver's speech in a moving car environment.

key words　speech corpus, spoken dialogue, in-car speech, moving car environment, driver's speech

1 はじめに

音声対話処理の研究では，対話で生じる諸現象を詳細に把握することが重要であり，対話コーパスはそのための貴重な資料となる [13]．また，有用なコーパスを作成するためには，利用目的に即したリアリティの高いデータ収集環境を構築することがポイントなる．

文部省中核的研究拠点 (COE) 名古屋大学音響情報研究拠点 (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定め，そのための要素技術の研究・開発を進めている．システムには，高騒音下での音声をロバストに理解すること，運転中のドライバーとの間で円滑に対話を遂行すること，などが求められる．我々の目的に合致したコーパスを構築するには，実走行環境下でのドライバーとの対話を収録することが望ましい．

CIAIR では，現在，実走行車内音声データベースの構築を進めている [4, 5]．これは，ロバスト音声認識を目的とした車内音声コーパスと，自然発話の理解を目的とした車内音声対話コーパスの 2 種類からなる．このうち，車内音声対話コーパスでは，道案内や店情報検索などをタスクとするドライバーとナビゲータとの間の対話を収録している．収録音声の書き起こしテキストだけでなく，アクセルやブレーキ踏力などのドライバー情報，エンジン回転数やスピード，現在位置などの車両情報，車室内外画像など，多様な情報を同期的に備えたマルチメディアデータベースとなっており，車内対話研究のための基礎データとして活用できる．

本稿では，車内音声対話コーパスの設計と特徴について述べる．まず，音声データの収録方法と書き起こし作業の概要について説明し，次に，162 対話を対象とした特徴分析の結果について報告する．分析では，(1) 車内対話におけるドライバー発話の特徴付け，及び，(2) ドライバーの走行中発話と停止中発話の比較，という 2 つの観点から，発話速度，発話文の長さ，話し言葉に特有な言語現象，品詞の出現傾向などの項目について調査した．

2 実走行車内音声対話コーパス

実走行環境下で自然な対話を遂行可能な車内情報システムの実現のため，道案内や店情報検索をタスクとするドライバーとシステムとの対話を収集している．本コーパスにより，走行車内特有の言語現象や発話の重なり具合の分析が可能となり，車内対話をモデル化するための基礎資料となる．対話データだけでなく，画像や車両状態のマルチモーダル情報を統合的に活用でき，走行状況・運転状況とドライバー発話との関係の分析に有用なデータとなる．図 1に，ブレーキ踏力，アクセル踏力，エンジン回転数，スピードに関する制御情報のサンプルを示す．なお，収録方法や収集

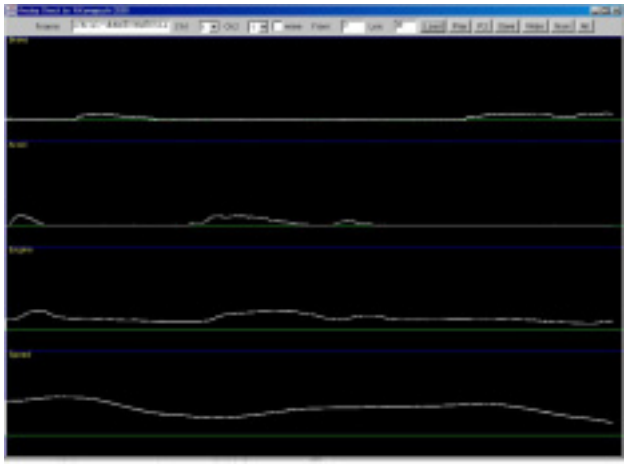

図 1: 車両制御情報 (ブレーキ，アクセル，エンジン，スピード)

システムの詳細については，文献 [4, 5] を参照されたい．

2.1 車内対話データの収集

機械との対話において発声されるドライバー発話を収集するために，Wizard of OZ(WOZ) システムや実際の対話システムでの収録が考えられるが，システムの性能の限界でもって，ドライバー発話の多様性を大きく制限することは望ましくない．そのため，我々は，収集の第一段階として「人間対人間」の対話を収録することとした．ただし，十分に訓練されたナビゲータがシステムの役割を担当することとし，さらに，ジェスチャ応答を行わない，ドライバーと目線を合わせて会話をしない，など，WOZ システムとの対話に近い形態を採用している．また，情報検索タスク対話では，ドライバーに発話行為のモチベーションを提供するため，「和食」「コンビニ」等の対話トピックや対話状況を表記したプレートを提示し，発話のきっかけを与えている．

2.2 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) ドライバー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) ドライバー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 [7] に準拠したタグ付け作業を行うこととした．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 2に書き起こし

```
0001 00:01:543-00:10:148 MEKENC
    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て <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 00:10:683-00:13:969 FXKENC
    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
    ございますが <SB>                 &  ゴザイマスガ <SB>
0003 00:14:156-00:17:905 MEKENC
    (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                            &  バショオ
    教えてほしいんだけど <SB>         &  オシエテホシンダケド <SB>
0004 00:18:092-00:21:136 FXKENC
    はい                              &  ハイ
    マクドナルドは                    &  マクドナルドワ
    ドライブスルーされますか <H><SB>  &  ドライブスルーサレマスカ <H><SB>
```

図 2: 書き起こしテキストの例

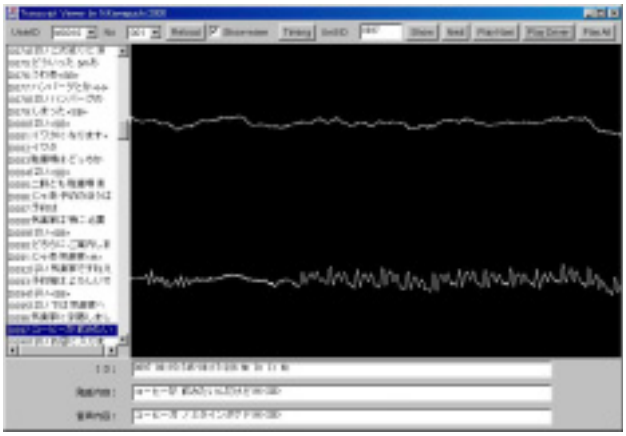

図 3: 音声対話分析支援ビューア

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（ドライバー / 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．音声対話ビューア (図 3) を作成し，対話分析支援ツールとして利用している．

3　車内対話の特徴

実走行車内対話の特徴を明らかにするために，収録された対話音声データの書き起こしテキストに基づいて対話データの特徴分析を試みた．具体的には，ドライバーの発話を対象に，発話速度，発話の長さ，話し言葉に特有な現象の出現傾向の調査を行った．なお，形態素分析は，形態素解析ツール「茶筅」 [12] で採用されている品詞体系に準じて実施した．

現在までに書き起こし作業が完了している 162 対話を分析の対象とした．ドライバー役を担当したのは

表 1: 車内対話の分析に使用した対話データ

項目	数値	割合 (%)
対話収録時間 (秒)	135,864	
平均対話収録時間 (秒)	839	
総発話時間 (秒)	81,064	
ドライバー	27,858	34.3
ナビゲータ	53,205	65.7
総発話単位数	31,539	
ドライバー	14,580	46.2
ナビゲータ	16,959	53.9
ドライバー発話文数	19,009	
ドライバー	7,250	38.1
ナビゲータ	11,759	61.9
総形態素数	214,079	
ドライバー	76,356	35.7
ナビゲータ	137,723	64.3

112 人であり，うち 50 人 (男 33 人，女 17 人) が各 1 対話 (計 50 対話) を，残りの 56 人が (男 39 人，女 17 人) が各 2 対話 (計 112 対話) を担当した．ナビゲータは作業に熟練している女性 3 人が行った．分析対象とした対話の収録時間，発話時間，発話単位数，発話文数，形態素数について，ドライバー発話とナビゲータ発話が占める割合とともに表 1に示す．対話収録時間に対する総発話時間は約 6 割であり，実走行車室内で行われるため，疎らな対話となっている．

3.1　発話速度

転記テキストの発音形部分と発話時間情報をもとに，ドライバー発話の発話速度を測定した．1 秒あたりの平均モーラ数は，6.158 であり，この値は一般的な対話音声 (8.5 〜 12.5(mora/sec))(例えば [1, 9])，さらには講演調音声 (6.5 〜 10.5(mora/sec))(例えば [2, 8]) と比べても小さく，比較的ゆっくり発声していることがわかる．この理由として，ドライバーは自動車の運転というメインタスクに注意の多くを注いでいるため，発声への注意度が低下し，発声速度が落ちることが挙げられる．また，対話相手であるナビゲータの平均モーラ数は 5.778(mora/sec) であり，ドライバーがナビゲータの発話調に影響を受けている可能性もある．なお，ナビゲータの発話速度が遅い原因の多くは，ドライバーからの検索要求に対して，情報検索作業を進めながら応答を行っていることによるものと考えられる．

表 2: 話し言葉に特有な現象の出現回数と出現率

	ドライバー		ナビゲータ	
項目	回数	出現率 (%)	回数	出現率 (%)
フィラー	4,842	33.1	670	4.0
言い淀み	952	6.5	692	4.1
言い誤り	526	3.6	289	1.7

(出現率は 1 発話単位あたりり出現する割合)

3.2 発話単位の長さとポーズ

話し言葉における発話単位を言語的に判定することは難しいため，コーパスの書き起こし作業では原則としてポーズで区切られた部分を一つの発話単位として定めている．ドライバー発話の形態素分析を通して求めた 1 発話単位あたりの形態素数 (発話単位の長さ) は 5.237 であった．発話単位の長さは，区切りとなるポーズの長さの設定に依存するため，他のデータと比較することは難しいが，ナビゲータの発話 (発話単位長は 8.121) と比べ，かなり短くなっている．

一方，コーパスでは，文末を判定する言語分析もあらかじめ行っており，1 発話文あたりの形態素数は，ドライバー発話において 10.53 であるのに対して，ナビゲータ発話では 11.71 とその差は大きくない．すなわち，1 発話文あたりの発話単位数は，ドライバー発話で 2.011，ナビゲータ発話で 1.442 であり，このことはドライバー発話の文中でポーズが発生する頻度が高いことを意味している．

3.3 話し言葉に特有な現象

話し言葉に特有な現象として，フィラー，言い淀み，及び，言い誤りを取り上げ，その出現頻度及び種類について調べた．ドライバー発話とナビゲータ発話における諸現象の出現総数と 1 発話単位あたりの出現個数 (出現率) を表 2に示す．

3.3.1 フィラー

ドライバーの発声による 14,580 の発話単位中に出現したフィラーの総数は 4,872 個であり，1 発話単位あたり 0.331 個出現している．これは，人間同士の対話に対する従来の調査結果 [9, 10] と比べると少ない．この原因として，ナビゲータの発話形態の自由度が低く，結果的に機械との対話の側面が多く存在していること，及び，3.2 節でも指摘したように，そもそも短い発話が多いということが挙げられる．表 3に，フィラーの種類，及び累積カバー率を出現順にならべて示す．個人差により分布に偏りが生じているものの，上位 5 語でフィラー全体の 62.3% を占めており，これは従来の調査結果 [9, 10] と同様の傾向を示している．出現位置については，全体の 74.3% のフィラーが発話単位の文頭に現われている (ナビゲータ発話では，68.3%).

表 3: フィラーの出現順位と累積カバー率

	ドライバー	
項目	個数	累積カバー率 (%)
1) あ	945	19.5
2) んー	819	36.4
3) えー	509	46.8
4) あー	448	56.1
5) ん	299	62.3
6) うーん	188	66.2
7) えーっと	156	69.4
8) と	153	72.5
9) え	134	75.3
11) えーと	107	77.5
10) あのー	102	79.6
12) あっ	80	81.3
13) あの	80	82.9
14) えっと	48	83.9
14) えーっとー	48	84.9

一方，ナビゲータ発話のフィラーは，1 発話単位あたり 0.040 個であり，ドライバー発話と比べても極めて少ない．これは，発声形態の自由度が大きく制限されていることや，話者がナビゲータとしての役割を十分に習熟しているということ，を裏付ける結果となっている．

3.3.2 言い淀み・言い誤り

言い淀みは，ドライバー発話に 952 回，6.5% の発話単位に出現し，言い誤りは，526 回，3.6% の発話単位に出現した．一秒あたりの出現回数はそれぞれ，0.034 回，0.010 回であり，従来の調査結果と比べても大きな違いは認められなかった．

4 車両制御情報を用いた走行中発話の特徴分析

車両の走行中と停止中とでは，ドライバーの精神状態が異なることが予想され，これがドライバーの発声に影響を及ぼす可能性がある．そのような影響の有無と程度を明らかにするために，ドライバー発話の車両走行中と停車中での違いを調査した．調査には，前節で分析の対象とした 162 対話のうち，車両速度情報が得られている 46 対話 (ドライバー役 46 人 (男性 31 人，女性 15 人)) を用いた．表 4にその概要を示す．

表 4: 車両制御情報に基づく分析に用いた対話データ

項目	値	割合 (%)
ドライバー発話時間 (秒)	8,571	
走行中	5,373	62.7
停止中	3,198	37.3
ドライバー発話単位数	4,015	
走行中	2,622	65.3
停止中	1,393	34.7
ドライバー発話文数	3,614	
走行中	2,369	65.6
停止中	1,245	34.4
総形態素数	22,625	
走行中	14,213	62.8
停止中	8,412	37.2

4.1 発声速度に関する比較

車両の走行中と停止中とでは，ドライバーの音声対話行為への注意力が減少し，発声速度に影響を及ぼすと予想される．影響の程度を調べるため，ドライバー発話の一秒あたりのモーラ数を，走行中と停止中とで測定した．平均モーラ数は，走行中では 5.491，停車中では 5.464 であり，両者の間に有意な差は認められなかった．

4.2 発話単位の長さに関する比較

車両の状態と，ドライバー発話のポーズの頻度との関係を明らかにするため，走行中，及び停止中における発話単位の長さを調べた．発話単位あたりの形態素数は，停止中の 6.039 に比べて走行中は 5.421 と短くなる傾向がある．一方，発話文あたりの発話単位数 (走行中:1.107，停止中:1.119) については有意な差は見られない．停止中に比べ，走行中で発話単位が短くなるものの，ポーズの出現率についてはあまり差がないことがわかる．

4.3 話し言葉に特有な現象に関する比較

次に，走行中と停止中での話し言葉としての違いを調べるため，フィラー，言い淀み，言い誤りの出現頻度を調べた．走行中及び停止中における現象の出現回数，及び発話単位あたりの出現回数を表 5に示す．両状態間で特に有意な差は認められなかった． なお，タウンページ検索をタスクとした車内音声対話におけるフィラー出現率を調査した実験 [11] においても，車両の走行による影響は認められていないが，今回の調査でも同様の結果となった．

フィラーの種類別の出現頻度順を表 6に示す．フィラーの種類に関する出現傾向は，本来，個人差に大きく依存することが分かっており，状態間での大きな差はなかった． フィラーが発話単位の先頭に出現する割合についても，走行中発話の 73.6% に対して停止中発話が 71.3% と，有意な差は見られなかった．

表 5: 話し言葉に特有な現象の出現頻度の比較

	走行中		停止中	
項目	回数	出現率 (%)	回数	出現率 (%)
フィラー	1,074	41.0	610	43.8
言い淀み	209	8.0	110	7.9
言い誤り	103	3.9	67	4.8

(出現率は 1 発話単位あたりり出現する割合)

表 6: フィラーの種類と出現頻度の比較

出現	走行中		停止中	
順位	種類	割合 (%)	種類	割合 (%)
1	あ	17.5	んー	22.3
2	んー	32.2	あ	37.4
3	あー	41.8	あー	47.9
4	ん	50.3	えー	55.4
5	えー	58.2	ん	61.6

(割合は全フィラー総数に対する累積カバー率)

4.4 品詞の出現傾向に関する比較

対話データ中に現れた形態素の品詞ごとの出現回数と出現頻度を表 7に示す．走行中対話と停止中対話との間で，品詞出現頻度の違いが大きいものとして，感動詞を挙げることができる．相対頻度は停止中の 3.5% から走行中の 5.5% へと約 1.54 倍に増加している．感動詞の多くは「はい」であり (走行中の感動詞発声の 67.4(%)，停止中の感動詞発声の 52.0(%) に相当)，走行中の出現率は停止中の 2 倍に達している．これは走行中のドライバーは，ナビゲータからの問いかけに対して応答が単純かつ単調になりやすいことを示している．感動詞「はい」は一つの品詞のみで一つの発話単位を構成する傾向があり，このことは走行中において発話単位の長さが短くなるという 4.2 節で示した特徴と関連している．

5 おわりに

CIAIR で構築中の実走行車内音声対話システムの設計と特徴について述べた．本稿では，(1) 車内対話におけるドライバー発話の特徴付け，及び，(2) ドライバーの走行中発話と停止中発話の比較，という 2 つの点を中心に，発話速度，発話単位の長さ，フィラー

表 7: 品詞の出現頻度の比較

品詞	走行中 回数	走行中 割合 (%)	停止中 回数	停止中 割合 (%)	比率
動詞	1,638	11.5	998	11.9	0.97
形容詞	515	3.6	300	3.6	1.02
助動詞	1,342	9.4	756	9.0	1.05
接頭詞	103	0.7	85	0.7	0.72
名詞	4,463	31.4	2,646	31.5	1.00
助詞	4,089	28.8	2,571	30.6	0.94
連体詞	100	0.7	59	0.7	1.00
副詞	504	3.5	310	3.7	0.96
感動詞	777	5.5	298	3.5	**1.54**
接続詞	436	3.1	239	2.8	1.08
その他	246	1.7	177	2.1	0.82
総数	14,213		8,412		

(割合は総形態素数に対するもの，比率は停止中の割合に対する走行中の割合の比)

表 8: 感動詞「はい」の出現頻度の比較

品詞	走行中 回数	走行中 割合 (%)	停止中 回数	停止中 割合 (%)	比率
「はい」	524	3.7	155	1.8	2.00

などの諸現象の出現頻度，及び，品詞の出現傾向を調査した結果を報告した．本調査により，

- ドライバーの発話速度は通常の対話音声と比べて遅く，5 〜 7(mora/sec) である．
- ドライバーの発話におけるフィラーの出現頻度は，1 発話単位あたり 0.33 個，1 秒あたり 0.174 個であり，通常の人間対人間の自由対話と比べて少ない．
- 車両の走行中と停止中とで，発話速度，フィラーなどの話し言葉に特有な現象の出現率の差は小さい．
- 停止中に比べて走行中での発話には，感動詞，及び，文発声途中でのポーズの頻度が増加する．

ことが明らかになった．

本稿で示したデータは人間との対話の収録によるものであり，機械との対話とでは，異なる結果となる可能性がある [3, 6]．次段階として，Wizard of OZ 方式に基づく対話データの収録を現在進めており，その調査結果については稿を改めて報告する予定である．

謝辞　車内音声対話コーパスの書き起こしは、日本語話し言葉コーパス (CSJ) の書き起こし基準に準拠致しました。基準策定に携われた方々に感謝致します。本研究の一部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。

参考文献

[1] 広瀬，阪田：対話音声と朗読音声の韻律的特徴の比較, 信学論， J79-D-II(12), pp. 2154–2162 (1996).

[2] 籠宮，菊池，小磯，前川：大規模話し言葉コーパスにおける発話スタイルの諸相 – 書き起こしテキストの分析から –，音講論 (秋)， pp. 107–108 (2000).

[3] 上條，秋葉，伊藤，田中: 音声対話データの分析と発話理解への応用, 人工知能研資， SIG-SLUD 9402-6, pp.31–36 (1994).

[4] 河口，松原，岩，梶田，武田，板倉：車内音声対話コーパスの構築, 情処研報, SLP-30, pp. 57–62 (2000).

[5] Kawaguchi, N., Matsubara, S., Iwa, H., Kajita, S., Takeda, K., Itakura, F. and Inagaki, Y.: Construction of Speech Corpus in Moving Car Environment, *Proc. of ICSLP-2000*, III, pp. 362–365 (2000).

[6] 黒岩，武田，井ノ上，山本: 機械との対話における発話分析, 信学技報， SP 94-30, pp.57–64 (1994).

[7] 前川，籠宮，小磯，小椋，菊池： 日本語話し言葉コーパスの設計，音声研究， 4(2), pp. 51–61 (2000).

[8] 峯松，片岡，中川：講演調の話し言葉に対する分析，情処研報, SLP-8, pp. 39–46 (1995).

[9] 村上，嵯峨山：自由発話音声における音響的な特徴の検討, 信学論, J78-D-II(12), pp.1741–1749 (1995).

[10] 中川，小林：自然な音声対話における間投詞・ポーズ・言い直しの出現パターンと音響的性質, 音響学会誌, 51(3), pp.202–210 (1995).

[11] 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, 音講論 (秋)， pp. 105–106 (2000).

[12] 松本ほか：日本語茶筅形態素解析システム「茶筅」 version2.0 使用説明書 第 2 版, Information Science Techinical Report NAIST-IS-TR99008 (1999).

[13] 山本：音声対話データベース構築の現状, 音響学会誌, 54(11), pp. 797–802 (1998).