

CIAIR 実走行車内音声データベース

河川 信夫^{†‡} 松原 茂樹^{†‡} 山口 由紀子[†] 武田一哉[†] 板倉 文忠[†]

†名古屋大学統合音響情報研究拠点 〒464-8603 名古屋市千種区不老町 1

‡名古屋大学情報連携基盤センター 〒464-8601 名古屋市千種区不老町 1

E-mail: kawaguti@itc.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, 1999 年より実走行車内における音声や対話を収録してきた。本稿では, 本拠点で収録された車内音声データベースについて報告する。本データベースでは, 実走行する車両を運転中の被験者の音声に加え, 映像, 車両情報などのマルチモーダル情報を収録した点に特徴がある。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, データベースの概要, 車両情報などを用いた言語現象等の解析結果について述べる。

キーワード 音声データベース, ロバスト音声認識, 話し言葉処理, 話し言葉解析

CIAIR In-Car Speech Database

Nobuo KAWAGUCHI^{†‡} Shigeki MATSUBARA^{†‡} Yukiko YAMAGUCHI[†]Kazuya TAKEDA[†] and Fumitada ITAKURA[†]

†Center for Integrated Acoustic Information Research, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan

‡Information Technology Nagoya University Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

E-mail: kawaguti@itc.nagoya-u.ac.jp

Abstract CIAIR, Nagoya University has been collecting the in-car speech database from 1999. This paper reports the various point of characteristics of the database. We have developed the system named Data Collection Vehicle (DCV) which supports synchronous recording of multi-channel audio data from 16 microphones that can be placed in flexible positions, multi-channel video data from 3 cameras and the vehicle related data. In our 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.

Keyword Speech database, robust speech recognition, spoken language processing, spoken language analysis

1. はじめに

名古屋大学統合音響情報研究拠点(CIAIR)では, 実環境における頑健な音声認識の実現や音声対話の高度化を目的として 1999 年より, 実走行車内における音声や対話を収録してきた[1]~[3]. 本稿では, この車内音声データベースについて報告する。本拠点では, 図 1 に示すような専用の音声データベース収録車を構築し, 音声に加え, 画像, 車両操作情報, 車両位置といったマルチモーダルな情報を 800 名に上る被験者に対し収

録した。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, 2000 年度以降の収録では, 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, 対話収録を行った。

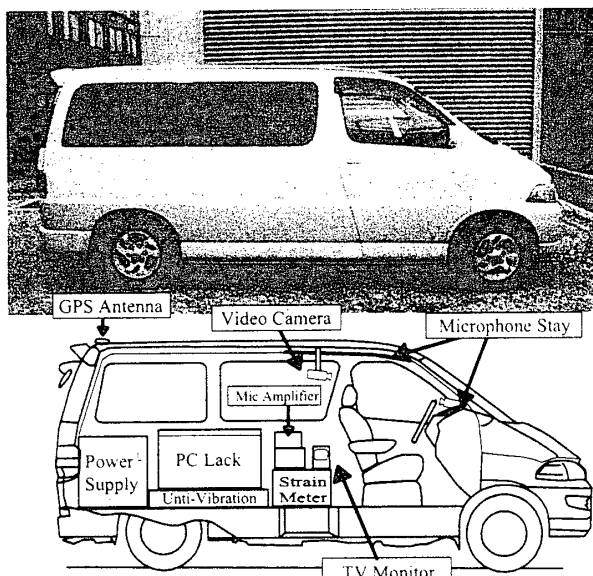


図 1: 音声データベース収録車

音声	16kHz, 16bit, 12ch
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 16bit, 1kHz
位置情報	Differential-GPS (緯度, 経度, 1Hz)

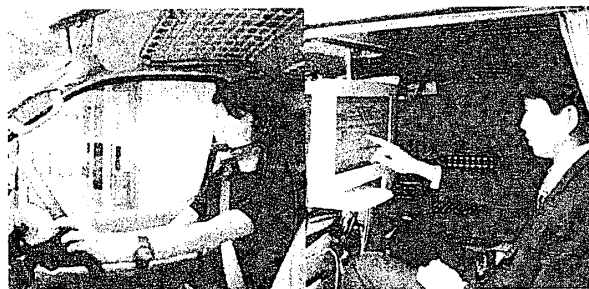


図 2: WOZ を用いた収録

2. 車内音声データベースの構築

本収録では、被験者が実環境の元で実際に車両を運転中のデータの収集を行うことを目的としている。各被験者は1時間程度の運転を行い、同時にデータ収録を行うこととした。そのため、1日あたり収録可能な人数は3名程度となる。1名あたり収集された各セッションについて表1に示す。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[1]を参照されたい。2000年度以降では、より実際の収録を目指し、WOZシステム(図2参照)と音声対話システム[4]を導入し、それぞれ5分間ずつの収録を行った。各被験者はすべての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た。

表 1: 収集されたセッション情報

1999年度の収集		212名
ナビゲータ(人)との擬似対話	11分	
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	30単語	
連続数字	4桁*20	
2000年-2001年度の収集		600名
ナビゲータ(人)との擬似対話	5分	
WOZシステムとの対話	5分	
音声対話システムとの対話	5分	
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	30単語	
連続数字	4桁*20	

各セッションでは表2に示すデータが収録された。対話データに関しては、[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。書き起しに対して、さらにタスク分割や各文の機能を表す意図タグ付与が行われている[5]。マルチメディアデータの収録により、本実験のデータベースは各被験者あたりCDにして4枚近くのデータ量となる。また、各被験者に対しては、実験の前後にアンケートを実施した。実験前には運転経験や音声認識システムの利用経験、カーナビの利用経験、機器操作の得意・不得意についてなど17項目、実験後には実験の感想やシステム利用に対する満足度、改善希望点など7項目の回答を得た。これらの大規模なアンケート情報は実験データと合わせ、様々な解析が可能となる。

被験者の男女比・年齢層を表3に示す。被験者は新聞・アルバイト情報誌等を用いて集めた。可能な限り男女比を均等に、年齢層を幅広く募集したが、収集は平日に実施したため、学生等の若い男性が多くなる結果となった。また車を運転するという実験の性質上、年配の女性被験者の獲得が困難であった。

表 3: 被験者の男女比・年齢層

	男	女	合計
10代	4	0	4
20代	366	162	528
30代	105	85	190
40代	46	35	81
50代	5	2	7
60代以上	2	0	2
合計人数	528	284	812

表4：音声対話セッションの基礎情報

	99HUM	00HUM	00WOZ	00SYS	01HUM	01WOZ	01SYS	合計・平均	
収録時間 (秒)	141822	94692	95300	77922	93465	93862	78169	675232	187.6時間
セッション数	209	294	293	288	295	294	287	1960	
平均セッション時間 (秒)	679	322	325	271	317	319	272		
総発話時間 (秒)	98100	69390	50864	54056	67635	47424	48877	436346	121.2時間
ドライバ	44722	28085	20159	11515	26055	18127	11001	159664	44.4時間
オペレータ	53328	41305	30705	42541	41580	29297	37876	276632	76.8時間
総発話単位数	38760	25251	19585	24944	24178	19993	22904	175615	
ドライバ	20493	12555	9381	10567	11985	9245	10722	84948	
オペレータ	18267	12696	9754	14377	12193	10748	12182	90217	
総形態素数	252289	174848	107010	142674	176915	88459	124018	1066213	
ドライバー	109710	68548	49023	27119	64173	44370	25587	388530	
オペレータ	142579	106300	57987	115555	112742	44089	98431	677683	
発話単位毎形態素数	6.71	7.18	5.64	5.93	7.70	4.60	5.65	6.31	
ドライバー	5.65	5.78	5.31	2.79	5.79	5.24	2.61	4.88	
オペレータ	7.85	8.52	5.97	8.07	9.48	4.10	8.10	7.58	
平均発話速度 (mora/秒)	6.41	6.53	6.12	6.11	6.52	6.01	5.95	6.28	
ドライバ	6.01	6.01	6.07	5.53	6.01	6.06	5.65	5.97	
オペレータ	6.86	7.05	6.18	6.54	7.02	5.96	6.21	6.60	

3. 音声対話セッションの基礎情報

以下では、本音声データベース中の特に音声対話セッションについて、詳細に述べる。表4に音声対話セッションに関する基礎統計情報を示す。812名の被験者に対し、187.6時間、1960セッションの収録を行った。収録総形態素数は106万であり、この種の対話コーパスとしては最大規模となる。形態素の計算においてはフィルタを除いている。この表ではHUMが人との模擬対話、WOZはWOZシステムとの対話、SYSは対話システムとの対話を表し、記号の前の数字は収録年度を表す。各発話は図3に示すように書き起され、フィルタ、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。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「茶筌」[16]を用いて実施した。

3.1. 発話速度

書き起こしテキストの発話時間情報をもとに、ドライバーの発話速度を測定した。全1960セッションを通じての1秒あたりの平均モーラ数は5.97mora/secであり、この値は一般的な対話音声(8.5~12.5mora/sec[6,13])に比べて小さい。また講演調音声(6.5~10.5mora/sec[7,12])と比べても小さく、比較的ゆっくりと発話して

0001 00:01:543-00:10:148 MD:EN:	ちよつと 小腹<H>が ずいたんだけど<H> この 近くに ファーストフード店で<H> あるのかなあ<SB>	& チョット & コバラ<H>ガ & スイタンダケド<H> & コノ & チカケニ & ファーストフードテンテ<H> & アルノカーナ<SB>
0002 00:10:683-00:13:969 FO:EN:	はい マクドナルドと モスバーガーが ございますが<SB>	& ハイ & マクドナルドト & モスバーガーガ & ゴザイマスガ<SB>
0003 00:14:156-00:17:905 MD:NN:	(F あっ)じゃ マクドナルドの 場所を 教えてほしいんだけど<SB>	& (F アッ)ジャ & マクドナルドノ & バシヨオ & オシエテホシンダケド<SB>
0004 00:18:092-00:21:136 FO:EN:	はい マクドナルドは ドライブスルーされますか<H><SB>	& ハイ & マクドナルドワ & ドライブスルーサレマスカ<H><SB>

図3：書き起こしテキストの例

いることがわかる。これは、ドライバーは運転タスクに集中しており、発話タスクには十分な注意が払えないためであることが考えられる。また、収録種別では特に、対話システムとの対話において、発話速度が遅いことが確認できる。これは、音声認識エラー等による誤認識などの動作に対し、ゆっくりした再発音を行っている点や、1発話単位長が短いことが理由である。また、オペレータ(HUM)の発話速度は、6.97mora/secである。WOZやSYSのセッションのオペレータは合成音声であるため、HUMの項だけを対象とした。オペレータもドライバーと同様に標準対話と比較してゆっくりとした発話であるが、これはドライバーに十分情報を伝えるための発話であることが理由である。

表5：話し言葉に特有な現象の出現回数と出現率

項目	ドライバー		オペレータ	
	出現回数	出現率	出現回数	出現率
フィルター	31153	31.24%	20257	29.01%
言い淀み	8475	5.97%	2065	4.55%
言い誤り	4256	2.79%	1621	3.62%

3.2. 発話単位

書き起しの際には、各発話をポーズを用いて発話単位に分割した。発話単位毎の形態素数は収録種別によって大きく異なり、HUMで5.72、WOZで5.27、SYSで2.48である。これは各収録種別の特徴を大きく表しているといえる。人間との模擬対話であるHUMでは、発話単位に形態素が多く含まれるのに対し、対話システムとの対話のSYSでは短い発話が多くなることを読み取れる。収録時に利用した対話システムの性能に影響され、被験者が短い発話を中心とした対話を行ったのではないかと推測される。また、オペレータによって作成された応答を合成音声で出力するWOZとの対話では、発話単位長は短いながらもHUMに近い値となっており、SYSにおいては合成音声よりも、発話理解の性能の低さが対話のスムーズさを低くしているということが読み取れる。

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、フィルター、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。ドライバー発話とオペレ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5に示す。ドライバーの84948発話単位においてフィルターは31153発話単位に出現しており、その出現率は31.2%である。これは、人間同士の対話における従来の調査結果[13,14]と比べて少ない。これは一般の対話と比べ、車内対話では、基本的に発話単位長が短いことに由来すると考えられる。また、オペレータのフィルター出現率が、従来調査[1]と比較して大きくなっているが、これはオペレータの個人差が現れたものといえる。表6にドライバー発話に出現するフィルターの種類、出現順位と累積カバー率を示す。これに関しては従来の調査結果[1,13,14]と同様の傾向を示している。言い淀み、言い誤りの出現傾向も同様の傾向である。

4. 車両情報を用いた分析

本音声データベースの特徴は、音声データのみならず、車両情報や操作情報が同時に収録されている点である。本節では、車両速度、アクセル、ブレーキ、ハンドル情報を用いた分析を行う。収録上の問題により車両情報が付与されているデータは全1960セッション中1741セッションである。

表6：フィルターの出現順位と累積カバー率

種類	個数	累積カバー率
1) あ	6546	21.0%
2) えー	4524	35.5%
3) んー	3832	47.8%
4) あー	1888	53.9%
5) ん	1763	59.6%
6) え	1602	64.7%
7) えーつと	1512	69.6%
8) と	976	72.7%
9) うーん	787	75.2%
10) えーと	755	77.6%
11) あのー	333	78.7%
12) あっ	80	79.0%

なお発話分析を行うにあたり、対話システムとの対話は他の対話との性質が大きく異なるため、以下ではHUMおよびWOZのセッションのみ分析の対象とした。

4.1. 走行状態に関する分析

走行情報に付いては、車両の速度パルス信号より、各時刻での速度やエンジン回転数を車両情報として収録した。車両情報より速度情報を取り出し、各発話中で時速6km/h以上が5秒間以上続いた場合を走行中、それ以外を停車中として分析を行った。全体の62~63%が走行中のデータであった。平均モーラ数は走行中と停車中ではほぼ変わらず、発話単位毎の形態素数に関しても若干、走行中のほうが5%ほど短い傾向があることがわかった。発話単位が短いということは、それだけ比較的単純な発話が多いことといえる。すなわち、停車中と比較して走行中のほうが、運転に対する注意が必要なため、発話に対する意識が低くなっていることが推測できる。

また、表8に、走行状態とフィルター発生率についての調査を示す。[1]での調査と異なり、走行状態とフィルター発生率の間には大きな相関がないことが読みとれる。なお、3節の値と大きな違いが生じる理由はSYSのセッションを除外しているためである。

表7：走行状態に関する分析

	発話時間 (秒)	発話単位数	発話単位毎 形態素数	発話速度		
				(mora/sec)		
走行中	78363	:62.1%	36833	:62.9%	5.17	5.87
停車中	47883	:37.9%	21723	:37.1%	5.44	5.97
合計	126246		58556		5.27	5.91

表8：走行状態とフィルター発生率

	フィルター出現回数	フィルター発生率
走行中	12952	35.16%
停車中	7759	35.72%
合計	20711	35.37%

表 9 : ペダル操作状況に関する分析

	発話時間 (秒)	発話単位数	発話単位数毎 形態素数	発話速度 (mora/s)
アクセル				
操作中	38594 :30.6%	17696 :30.2%	5.30	5.86
非操作中	87567 :69.4%	40860 :69.8%	5.26	5.94
ブレーキ				
操作中	59923 :47.5%	27757 :47.4%	5.29	5.91
非操作中	66314 :52.6%	30799 :52.6%	5.25	5.92
合計	126161	58556	5.27	5.91

表 10 : ペダル操作状況とフィルター発生率

	フィルター出現回数	フィルター発生率
アクセル		
操作中	6421	36.29%
非操作中	14290	34.97%
ブレーキ		
操作中	9840	35.45%
非操作中	10871	35.30%
合計・平均	20711	35.37%

4.2. アクセル・ブレーキペダルに関する分析

アクセル、ブレーキの操作についてはそれぞれのペダルにかかった踏み圧力を圧力センサを用いて収録した。本分析では、ペダルに 0.5kg 以上の力が 0.5 秒以上加わったときを操作中、それ以外を非操作中とした。結果を表 9 に示す。発話時間の約 30% がアクセルを操作中、約 47% がブレーキを操作中であることがわかった。すなわち、被験者は全発話中の約 78% においてペダルによる車両操作を行っていることになる。

また、アクセル操作中は、ブレーキ操作中と比較して、発話速度、発話単位長に対しての影響がやや大きいことがわかる。走行中の発話単位長が短い事に比較して、アクセル操作中の発話単位長には影響が少ないが、発話速度に対しての影響はアクセル操作中のほうが大きいことが読みとれる。

次にフィルターの発生率についてペダルの操作との相関を確認した。結果を表 10 に示す。ブレーキ操作中に比べ、アクセル操作中にフィルターの発生率が若干高くなっていることがわかる。このことから、アクセル操作がドライバーにとって負担になっていることが読みとれる。走行状態は発話単位長に影響を与えるが、フィルター発生率には影響を与えないが、ペダル操作状況においては発話単位よりもフィルター発生率に大きな影響があることがわかる。

4.3. ハンドル操作に関する分析

ハンドル操作については、ハンドルに接続されたギアで可変抵抗を駆動してその位置を取得した。なお、ハンドル操作情報は 2000 年度から収録されているた

表 11 : ハンドル操作に関する分析

	発話時間 (秒)	発話単位数	発話速度 (mora/sec)
直進中	46749 :57.1%	26180 :68.4%	6.04
微調整中	33320 :40.7%	11582 :30.3%	5.81
操作中	1824 :2.2%	507 :1.3%	5.55
合計	81893	38269	5.94

表 12 : ハンドル操作とフィルター発生率

	フィルター出現回数	フィルター発生率
直進中	8353	31.91%
微調整中	5056	43.65%
操作中	263	51.87%
合計・平均	13672	35.73%

め、以下では 2000 年以降の HUM, WOZ セッションのみの分析となっている。ハンドル操作については、発話中に 5 度以内の操作しかしていない場合を直進中、5 度から 180 度の操作をしている場合を微調整中、180 度以上の操作をしている場合を操作中とした。微調整中は、車線変更等の操作を行ったと考えることができる。分析の結果を表 11 に示す。

表 11 より、直進が発話時間中の約 57%、微調整中が約 40%、約 2.2% がハンドル操作中であることがわかる。このデータより、ハンドル操作がドライバーの発話に大きく影響を与えていることが読みとれる。これまでの分析と同様に、特にハンドル操作中といった注意が必要な状況では発話速度が 6% 程度遅くなっている。逆に、直進中では、リラックスしているためか、発話速度が速くなっていることが分かる。

また、表 12 にハンドル操作とフィルター発生率についての相関を示す。発話速度がハンドル操作によって大きく影響を受けると同様に、フィルター発生率もハンドル操作の状況によって大きく変化することが分かる。特に、ハンドル操作中においては、半数の発話にフィルターが現れることから、ドライバーの注意が運転操作によって、発話から離れていることが理解できる。

4.4. 車両情報による分析のまとめ

以上、車両情報やペダル、ハンドル操作による発話への影響を調査した。上記に示したように、ドライバーの発話は、運転状況によって左右されることが確認できた。特にアクセル操作やハンドル操作によって、大きく影響されることが分かった。テレマティクス等による車内情報システムを構築する際には、アクセル操作状況やハンドル操作状況を用いることによって、不注意にドライバーに情報を提示しないような仕組みを導入することが望まれる。

表 13: 本データベースのCD構成

CD1	アイドリング中 連続数字, 音素バランス文, 単語
CD2	走行中 対話音声, 転記テキスト, 車両情報
CD3	走行中 音素バランス文 (運転席, 助手席)
CD4	走行中 対話動画像

本データベースでは, 上記で調査した運転関連情報に加え, 車両の位置情報, 前方や被験者を写した動画像などが含まれている。また, 運転歴やカーナビの使用経験を調査したアンケート情報との相関も調査する必要がある。

5. データベースの構成

本データベースは, 現在最終的な配布に向けて整理・調整を行っている状況である。現在のデータベースは, 各被験者毎に4枚のCDから構成されている。データベースの構成を表13に示す。合計で3000枚を越える容量となるため, 一括の配布には困難を伴う。そのため, 各CDでイメージファイルを作成し, イメージファイルを収録したHDDによる配布や, 必要な部分のみの利用を実現する仕組みを構築することを検討している。

6. おわりに

本稿では, 名古屋大学CIAIRで収録された被験者総数812名の車内音声データベースについて報告した。特に総時間187.6時間, 1960セッション, 収録総形態素数106万の対話コーパスについて, その基礎情報や話し言葉に特有な現象を調査した。また, 本データベースを特徴づける車両情報を用いて, 走行状況, アクセル・ブレーク操作状況, ハンドル操作状況における発話への影響について調査した。本調査により,

- ・ ドライバーの発話速度は通常の対話に比べ遅く, 5.5~6.1mora/secである。
- ・ システムとの対話は人との対話と比較して, 発話単位長が短くなる。
- ・ 走行状態が発話に与える影響は限定的である。
- ・ 車両操作, 特にアクセル操作やハンドル操作を行った場合に, 発話への影響が大きい。

ことが明らかになった。

本データベースは, 被験者数が大きいことに加え, 大規模なマルチメディアデータが存在しており, 今回の調査はその一部に過ぎない。本データベースのさらなる活用によって, 実環境音声対話システムに関する研究がさらに飛躍することを期待する。

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分析について多大な貢献をされた関係者諸氏に心から感謝致します。

文 献

- [1] 河口, 松原, 若松, 梶田, 武田, 板倉, 稲垣: 実走行車内音声対話コーパスの設計と特徴, 情処研報, SLP-34, pp.179-184, (2000).
- [2] 河口, 松原, 武田, 板倉, 稲垣: 実走行車内音声対話データベース, 信学技報, SP-2001-107, Vol.101, No.523, pp.31-36, (2001).
- [3] 河口, 牛窪, 松原, 岩, 梶田, 武田, 板倉: “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, (2001).
- [4] 早川, 磯部, 河口, 武田, 板倉: “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).
- [5] 入江, 松原, 河口, 山口, 稲垣: 意図タグつきコーパスを用いた発話意図推定手法, 人工知能研資, SIG-SLUD-A301-03, pp.7-12, (2003).
- [6] 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).
- [7] 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).
- [8] 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).
- [9] 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 ICSP-2000, III, pp. 362-365 (2000).
- [10] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 松本ほか: 日本語茶筌形態素解析システム「茶筌」version2.0 使用説明書 第2版, Information Science Technical Report NAIST-IS-TR99008 (1999).
- [17] 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).