

同時処理タスクの難易度の変化における言語能力への影響

岩本 善行[†] 伊藤 敏彦[†] 甲斐 充彦^{††}
小西 達裕[†] 伊東 幸宏[†]

音声入力インターフェースを使用する状況において、対話相手が人間又は機械、運転中又は停車中といった対話状況の違いにより発話にどのような特徴の変化があるのかを調べる為に、対話を収集し分析を行った。その書き起こしや言語的・音響的特徴の統計的な分析結果では、運転の有無は発話の言語的特徴に影響を与えないというものであり、我々の仮説とは異なっていた。しかしながら、運転タスクの難易度が低すぎたことによる影響の可能性がある為、運転操作に必要な認知的負荷を変化させた場合の発話の言語的・音響的特徴に関する分析を行った。その結果、発話の言語的特徴においては、ほとんど運転タスクの影響を受けず、音響的特徴に若干の影響を与える事が明らかになった。

Influence of the difficulty of a concurrent task on linguistic competence

Yoshiyuki Iwamoto[†] Toshihiko Itoh[†] Atsuhiko Kai^{††}
Tatsuhiko Konishi[†] Yukihiro Itoh[†]

We investigated the characteristic change of utterances under the different dialogue situations; the situation of using a voice interface of machine versus talking with a human, and the situation of talking with driving versus without driving. The result of a statistical analysis revealed that a driving task does not affect the linguistic features of utterances and the result differed from our assumption. Since this result may be due to a relatively low cognitive load in driving task, we conducted a dialogue experiment under the situation of a concurrent driving task with different difficulty levels. Consequently, it turns out that the concurrent task hardly affects in the linguistic feature of utterance, but has a little influence on the acoustic feature.

1 はじめに

近年、音声入力インターフェースの効果的な応用分野として、カーナビゲーションシステムの音声入力インターフェースが注目されてきており、それらの対話システムの評価方法の研究等が行われている [1]。カーナビゲーションシステムで、車の運転中に画面を注視して目的地設定を行うことは安全上問題があり、音声対話を用いることにより、安全にカーナビゲーションシステムを操作する事が可能となる。しかしながら、現状の音声認識技術では誤認識を回避する事は難しい。さらに、走行ノイズ等が出る車の運転中であればなおさらである。したがって、ユーザビリティを向上させる為の対話戦略や音声認識の精度改善が一層重要となる。また、運転中に音声対話システムと対話をする

という状況が、運転操作に与える影響に関しては幾つか研究例があるが [9]、ユーザの発話にどのような影響を与えるかは明らかではない [2][4][8]。特徴に変化があるのであれば、それをシステムに反映させる事で、システムの対話戦略や音声認識精度の改善が可能となると考えられる。

そこで、我々は様々な対話状況でのユーザの発話の特徴の違いを分析し、比較を行った。具体的には、対話相手が人間又は機械、運転中又は停車中といった対話状況の違いにより、ユーザの発話にどのような影響が現れるかの分析を行った。対話タスクとしては、より現実的な設定として、カーナビゲーションシステムにおけるランドマーク入力を想定した。機械及び人間の両方の対話相手に関して、同等の条件の対話タスクを設定して比較を行った。さらに、運転タスクの難易度を上げることにより、心理的負担を増加させることで、ユーザの発話にどのような特徴の変化が現れるか

[†] 静岡大学 情報学部

Faculty of Information, Shizuoka University

^{††} 静岡大学 工学部

Faculty of Engineering, Shizuoka University

の分析を行った。

2 音声対話データの収録

本節では、対話音声の言語的・音響的特徴の分析を行う為、種々の対話状況の違いにおいて被験者を用いて行った対話データの収集について述べる。

2.1 実験方法

ここでは、運転タスクの有無を含む対話状況の違いによる対話データを収集する為の実験環境について述べる。運転タスクに関しては、実際に車を運転しながら対話実験を行う事は安全性の問題がある為、運転操作の代用としてカーシミュレーションのゲームを用いて実験を行った。実際の運転状況に近づける為、ハンドルとブレーキ・アクセルを用意し、画面はプロジェクターで約 85 インチの大きさに投影した。走るコースは単純な楕円コース（オーバルコース）とし、運転タスクに集中させる為の条件として、コースの左の線に沿って（キープレフト）、正確に 100km/h で走ることを指示した。被験者は、このシミュレーションゲームをプレイしながら目的地までの経路設定を対話的に行う。

対話状況の違いの一つである対話相手に関して、まず、対話相手である「機械」は本実験用に構築した音声対話システムを用いる。このシステムは、ランドマークの入力に対してシステムが理解した内容を明示的に確認するだけの単純なものである。詳しくは 2.2 節で述べる。次に、対話相手である「人間」は実際の商用サービス¹を利用した。詳しくは 2.3 節で述べる。

カーナビゲーションシステムで設定できるランドマークを以下の 3 種類と想定し、タスクの設定を行った。

- (1) 県名+市町村名 (+目的地)
- (2) 県名+高速道路名+IC / JCT 名 (IC: インターチェンジ、JCT: インタージャンクション)
- (3) 県名+鉄道会社名+駅名

用意した目的地設定タスクは 9 タスクである。被験者には、図 1 のような、経路と最終目的地に加え、暗記の手助けとなる簡単な文章を書いた紙を見せ、3 分程度で暗記してもらった。経路/目的地を忘れた時の為に、ヒントの紙を用意し、運転中でも運転の妨げにならない程度に見てもらふことは可能とした。経由地と目的地の設定完了の条件は、経由地に関しては IC か JCT、駅名の設定、目的地に関しては、市町村名+目的地名の設定を設定完了と定義した。

対話収集時の対話状況は、対話相手が機械 (S)、人

¹ コンパスリンク株式会社の商用サービス「コンパスリンク」

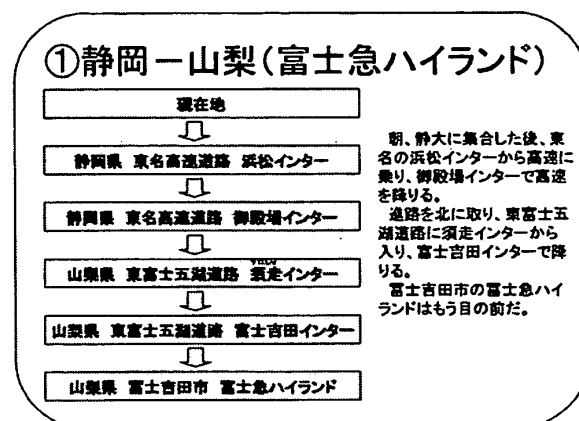


図 1: タスクの暗記用紙

間 (C) の 2 種類と、それぞれに同時処理タスクである運転タスク (G) の有無を組み合わせる。また、システムについての予備知識が無い状態において、ユーザがシステムに対してどのような発話を行うかを調べる為に、どの被験者にも 1 タスク目を、システムについての説明を一切しないで行った (自由対話)。その後、前記の 4 パターンの対話状況をそれぞれ 2 タスクずつ、被験者毎にランダムに順序を変えて行った。また設定タスクに関しても、被験者毎にランダムに変更した。その為、設定してもらった経路は、一人当たり 9 経路である。

被験者は情報系大学 (院) 生 10 名であり、音声対話システムに関する知識はほとんどない。

実験の手順としては、最初に実験の説明を行い、実際に実験で走るコースと走行条件で 1 周走ってもらった。その後、用意した目的地と経路を暗記してもらい自由対話の為の実験を行った。続いて、システムとコンパスリンクの説明をする為に、デモビデオを見せ、システムに入力可能な文法例やシステムの機能 (分割発話機能、部分訂正機能、UNDO 機能等) の説明を行った。その後、残り 8 タスクの実験を実施した。

2.2 目的地設定音声対話システム

本節では、本実験用に構築した音声対話システムについて述べる。本システムはユーザの発話入力に対して、理解内容を明示的に確認するだけの単純なシステムである。認識できる単語は、県名、高速道路名/鉄道会社名、IC 名/JCT 名/駅名/市町村名と、それらの前後に付随する事が許されるいくつかの肯定語、否定語、文末表現である。音声認識部には HTK を用い、音声合成部には日本 IBM 製の ProTalker を使用した。ユーザの音声入力からシステムの応答出力まで

の処理時間は約2秒～5秒である。

システムが受理可能な発話は、詳細化発話、全体訂正発話、部分訂正発話、肯定発話、否定発話、UNDO発話である。詳細化発話とは、前発話の内容を下位レベルにあたる情報を付け加える事でより詳細にする発話である。全体訂正発話は、システムが誤認識・認識棄却を行った場合や、あるいはユーザが誤った入力を行った際に、全ての情報をもう一度再入力する発話である。部分訂正発話とは、システムの認識が部分的に間違っている場合に、その間違っている部分だけを発話し直すものである。肯定発話とは、“はい”等のシステムの確認を肯定する発話のことである。肯定発話がされた場合、システムは、最下位レベルのキーワードが入力されているのであれば、経路/目的地の設定を行い、そうでなければ、未設定項目の設定を促す応答を行う。否定発話とは、“いいえ”等のシステムの確認内容を否定する発話である。その場合、システムの応答は、認識結果から作成されたN-bestの応答候補の中から、使用した応答候補の次の候補を応答する。UNDO発話とは、ユーザの一発話前の発話を無効にし、その直前の状態に戻す為の発話である。

2.3 人間のオペレータとの対話

対話相手が人間の場合の実験では、コンパスリンクと呼ばれる実在のサービスを利用した。コンパスリンクは、情報検索インタフェースとして人間のオペレータを配し、専用オペレータが24時間、情報センターで対応してくれるカーナビゲーションサポートサービスである。このコンパスリンクの「目的地設定・検索サービス」を利用して実験を行った。コンパスリンクは、携帯電話回線を利用し、回線帯域の半分をオペレータとユーザの音声言語による情報交換に使用する。そして、残りの回線帯域をカーナビゲーションシステムとコンパスリンクのサーバとのデータ交換に利用する。音声対話に使用する回線帯域が通常の場合の半分のため、通常の音声対話に比べて、ノイズや音声品質の低下によるオペレータの誤認識が生じやすい。

3 対話状況の違いによる発話への影響の分析

本節では、前節で述べた方法で収録された対話音声データを言語と音響の2つの視点から分析した。

人間相手の場合、目的地設定を行う対話以外に、目的地設定を開始する前の対話部分(オープニング)と、最終目的地設定後の経路全体の設定の確認の対話(ク

表 1: 実験対話状況

運転操作	対話相手		
	機械		人間
	デモ説明無	デモ説明有	デモ説明有
無	SF	S	C
有	SGF	SG	CG

ロージング)があるが、機械相手の対話データとの比較を行う為、対話の整合をとる必要がある。そこで、人間相手での分析データから、これらのオープニング・クロージングは除いている。

3.1 言語的特徴の違いの分析

実験で収集されたユーザの言語的特徴を調べる為に、いくつかの評価項目を想定し、各対話状況毎の全体的な統計をとった。その結果を表2に示す。表中のSF、SGF等の説明を表1に示す。

表中の各項目の定義は次のようになっている。「詳細化発話数」とは直前のユーザの発話に新しく情報を追加した発話数である。「訂正発話数」とはシステムの誤認識に対して、全体訂正もしくは部分訂正を行った発話数である。「発話衝突数」とは対話相手の発話中に、ユーザが発話を開始した発話数である。「間投詞数」とはユーザの発話中の“えーと”等の間投詞の出現数である。「平均形態素数」とはユーザの一発話当たりの形態素数の平均である。「平均情報数」とはユーザの一発話当たりのキーワード数の平均である。ただし、キーワードが発話された発話数だけで割っている。「平均新情報数」とはキーワードが含まれている発話のうち、システムに未入力であるキーワードが入力されたキーワード数の平均である。「動詞省略形発話数」とは体言止め発話や、“～の”といった助詞止めなどの動詞が省略された発話数である。「肯定語のみ発話数」「否定語のみ発話数」とは“はい”や“いいえ”等、肯定語/否定語のみの発話である。「肯定語+情報形発話数」「否定語+情報形発話数」とは肯定語/否定語の後に、引き続きキーワードを発話した発話数である。

表2の結果を見ると、音声認識率や対話戦略の違いによって総発話数が大きく異なり、言語的特徴に関して、適切な比較を行えない。そこで、これらの影響を減らす為、ユーザの発話に新情報が含まれている発話だけを抽出して分析を行った。その結果を表3に示す。

これらの結果を用いて、各評価項目の対話状況の違いによる変化を、F検定及びt検定によって分析した。ここでは、危険率pが5%以下である項目に有意な差があるとした。検定結果から、対話相手の違いに関して、機械相手との対話において特徴的なものとしては

表 2: 対話全体の言語的特徴

評価項目	SF[%]	SGF[%]	S[%]	SG[%]	C[%]	CG[%]
タスク数	4	6	20	20	20	20
総発話数	105	222	578	518	348	398
平均発話数	26.3	37.0	28.9	25.9	17.4	19.9
詳細化発話数	3[2.9]	14[6.3]	42[7.3]	37[7.1]	41[11.8]	43[10.8]
訂正発話数	2[2.0]	4[1.8]	85[14.7]	64[12.3]	8[2.3]	15[3.8]
間投詞数	16[15.2]	15[6.8]	39[6.8]	28[5.4]	91[26.2]	121[30.4]
発話衝突数	15[14.3]	32[14.4]	43[7.4]	33[6.4]	56[16.1]	61[15.3]
平均形態素数	3.1	4.6	3.6	3.6	4.2	3.8
平均情報数	2.0	1.4	1.6	1.6	1.4	1.3
平均新情報数	0.6	0.4	0.6	0.6	1.1	1.0
動詞省略形発話数	41[39.1]	119[53.6]	303[52.4]	271[52.3]	101[29.0]	97[24.3]
肯定語のみ発話数	28[26.7]	54[24.3]	134[23.2]	129[24.9]	128[36.8]	141[35.4]
肯定語+情報形発話数	0[0.0]	0[0.0]	12[2.1]	14[2.7]	32[9.2]	39[9.8]
否定語のみ発話	8[7.6]	12[5.4]	27[4.7]	35[6.8]	2[0.6]	4[1.0]
否定語+情報形発話数	1[1.0]	1[0.5]	48[8.3]	24[4.6]	1[0.3]	1[0.4]
認識率数	49[46.7]	102[46]	175[30.3]	125[24.1]	12[3.5]	22[5.5]
誤認識数	13[12.4]	20[9.0]	127[22.0]	123[23.8]	12[3.5]	16[4.0]

表 3: 新情報発話の言語的特徴

評価項目	SF[%]	SGF[%]	S[%]	SG[%]	C[%]	CG[%]
総発話数	17	41	140	128	132	136
平均発話数	4.3	6.8	7.0	6.4	6.6	6.8
詳細化発話数	0[0.0]	7[17.1]	23[16.4]	24[18.8]	39[28.8]	38[27.9]
訂正発話数	0[0.0]	1[2.4]	7[5.0]	6[4.7]	1[0.8]	0[0.0]
間投詞数	5[29.4]	4[9.8]	19[13.6]	17[13.3]	75[56.8]	92[67.7]
発話衝突数	2[11.8]	7[17.1]	10[7.1]	3[2.3]	20[15.2]	28[20.6]
平均形態素数	5.0	9.0	6.4	6.0	8.1	7.1
平均情報数	2.3	1.4	1.8	1.9	1.5	1.4
平均新情報数	2.3	1.3	1.6	1.7	1.4	1.3
動詞省略形発話数	6[35.3]	34[82.9]	100[71.4]	97[75.8]	82[62.1]	75[55.2]

以下のものが挙げられる。

- (1) 間投詞はあまり出現しない
- (2) 動詞省略発話が多い
- (3) 短い発話になる
- (4) 一つの発話で多くの情報を伝える
- (5) 発話の衝突が少ない

このように、対話相手の違いによる言語的影響が大きいですが、これは、ユーザが機械を人間程の認識率、対話戦略を持っていないと考えているためであろう。そのため、機械に対しては、最低限の発話で出来るだけ多くの情報を伝えようとしている。一方、運転タスクによる言語的特徴の有意な変化は全く見られなかった。運転操作という同時タスクの有無による発話の言語的特徴への影響は小さい結果となった。

3.2 音響的特徴の違いの分析

次に、音響的特徴に関して分析を行った結果について述べる。発話全体に関する分析結果を表 4 に、新情報を含んだ発話に関する分析結果を表 5 に示す。

表中の各項目は次のように定義する。「平均発話時間」とは、一発話の平均時間長である。「平均発話速度」とは、一秒間当たりのモーラ数の平均である。「平均発話開始時間」とは、対話相手の機械又は人間が発話し終えてから次にユーザが発話するまでの時間の平均である。「平均パワー」「最大パワー」とは、それぞ

れ発話パワー (Root Mean Square : RMS) の平均、最大値である。また、表中の (1) は、発話中に含まれるポーズ長を含めた結果であり、表中の (2) は、発話中の無音区間を除いて算出した結果である。

これらの結果を用いて、対話状況の違いによる特徴の変化を F 検定及び t 検定を用いて分析した。機械相手に関して特徴的な結果としては以下のようなことが挙げられる。

- (1) 発話開始時間が遅い
- (2) 発話時間が短い
- (3) 新情報発話に関しては発話速度が速いが、対話全体としては発話速度は遅い
- (4) ピッチの平均値、最大値が低く、ピッチの変化が小さい

機械相手の場合、システムが受理できる文体を考える為か、発話開始時間が遅くなっている。また、言語的特徴で現れたように、必要最低限の情報しか言わない為、発話時間は短くなっている。対話全体として発話速度が遅くなっているのは、訂正発話をうまく認識させようと丁寧にゆっくりと発話するためではないかと思われる。また、新情報発話で発話速度が速くなっているのは、一度にたくさんの情報を伝えようとするからである。また、対話全体に関して、言語的特徴と同様に運転タスクによる影響の特徴はほとんどみられ

表 4: 発話全体の音響的特徴

評価項目	SF	SGF	S	SG	C	CG
平均発話時間 (1)(sec)	2.68	2.16	2.40	2.61	1.96	2.11
平均発話速度 (1)(mora/sec)	4.56	4.73	4.80	4.70	5.44	5.51
平均発話時間 (2)(sec)	1.81	1.43	1.56	1.58	1.54	1.55
平均発話速度 (2)(mora/sec)	6.72	7.09	6.97	7.00	6.64	7.06
平均発話開始時間 (sec)	1.03	1.04	0.96	0.99	0.50	0.49
平均パワー (RMSmean)	16.17	32.47	28.89	28.05	34.17	33.53
最大パワー (RMSmax)	81.07	90.74	143.69	111.90	149.77	149.59
対数化 F0 最小値	4.43	4.44	4.45	4.47	4.47	4.45
対数化 F0 最大値	5.23	5.18	5.21	5.23	5.32	5.32
対数化 F0 平均	4.85	4.86	4.89	4.91	4.94	4.93
対数化 F0 標準偏差	0.16	0.17	0.17	0.18	0.21	0.21

表 5: 新情報発話の音響的特徴

評価項目	SF	SGF	S	SG	C	CG
平均発話時間 (1)(sec)	5.69	3.11	3.77	3.66	3.76	4.06
平均発話速度 (1)(mora/sec)	4.09	5.28	5.11	5.23	4.97	5.16
平均発話時間 (2)(sec)	3.39	2.25	2.64	2.60	2.93	3.00
平均発話速度 (2)(mora/sec)	6.74	6.99	6.81	6.92	6.09	6.59
平均発話開始時間 (sec)	1.47	1.70	1.10	1.41	0.70	0.56
平均パワー (RMSmean)	13.71	32.37	28.23	27.59	33.38	33.17
最大パワー (RMSmax)	56.77	134.54	130.91	102.21	145.73	134.71
対数化 F0 最小値	4.20	4.36	4.37	4.36	4.34	4.33
対数化 F0 最大値	5.41	5.25	5.31	5.31	5.42	5.41
対数化 F0 平均	4.82	4.89	4.89	4.90	4.94	4.94
対数化 F0 標準偏差	0.18	0.21	0.19	0.21	0.23	0.23

なかったが、機械との対話において運転操作を伴う場合に、新情報発話に対する分析ではピッチの標準偏差のみが上昇した。

4 運転状況の違いにおける発話への影響の分析

前節の実験の結果では、運転タスクの有無による発話の言語的特徴への影響はほとんどなかった。しかしながら、設定した運転タスクが実際の運転よりも心理的負担が小さすぎたのではないとも考えられる。ここでは、運転タスクの難易度を複数設定し、実際の運転以上に心理的負担が大きいと思われる運転タスクも用意して対話を収集し、分析した結果について述べる。

4.1 実験方法

用意した運転タスクは前節と同様のカーシミュレーションゲームを用いた。比較を行う難易度は3段階用意し、前節で行った対話実験の際の運転タスクを3段階のうち一番易しい(Easy)レベル、さらに、中間(Middle)レベルと最高(Hard)レベルを用意し、対話実験を行った。Middleレベルでは、常時、運転に意識を集中させる為に直線の少ない山のコースとし、Hardレベルでは、さらに大きな心理的負担を与える為に、ダートコースを使用し、普段の運転ではあり得ない程の心理的負担になるようにした。また、さらにMiddleレベルとHardレベルの運転タスクでは、他の車を追いかける条件も課した。

対話相手としては、被験者一人につき対話相手は機

械か人間かのどちらか一方とした。これは、前節の対話収集実験では対話相手を機械と人間の交互に行っていたが、今回は運転タスクの違い以外の影響を除くためである。対話相手である機械と人間、目的地設定タスク等の実験条件は前節と同様とした。

タスクは、機械相手にゲームをしながら行うタスク(SG)と、人間相手にゲームをしながら行うタスク(CG)の2種類である。設定してもらった目的地タスク数は、Easyレベルを3タスクとMiddle、Hardをそれぞれ6タスクの計15タスクである。

被験者は情報系大学(院)生6名であり、前節と同様の被験者である。

4.2 言語的特徴の違いの分析

実験で収集したユーザ発話の運転タスクの難易度毎の分析結果をそれぞれ表6、表7に示す。新情報発話に関する言語的特徴の分析結果を表8、表9に示す。

これらの結果を用いて、F検定及びt検定による分析を行った。その結果、心理的負担の違いによる発話の言語的特徴の有意な変化はみられず、発話の表層的表現への影響は少ないと言える。この理由として、人間の脳が行う処理において、運転よりも言語処理の方が優先度が高いという仮説が考えられる[9]。

4.3 音響的特徴の違いの分析

実験で収集したユーザ発話の各運転難易度の音響的分析結果を表10、表11に示す。また、新情報発話に関しての音響的分析結果を表12、表13に示す。

これらの結果を用いて、F検定及びt検定による分

表 6: 発話全体の言語的特徴 (人間相手)

評価項目	Easy (%)	Middle (%)	Hard (%)
設定経路数	9	18	18
総発話数	126	263	263
詳細化発話数	12[9.52]	35[13.31]	30[11.41]
訂正発話数	3[2.38]	7[2.66]	10[3.80]
言いよどみ発話数	10[7.94]	23[8.75]	32[12.17]
言い直し発話数	5[3.97]	3[1.14]	4[1.52]
間投詞数	46[36.51]	96[36.50]	103[39.16]
発話衝突数	45[35.71]	58[22.05]	81[30.80]
形態素数	6.27	6.40	6.02
平均情報数	1.67	1.71	1.68
平均新情報数	1.35	1.43	1.22
発話数	14.00	14.61	14.61
動詞省略形発話数	35[27.78]	79[30.04]	65[24.71]
UNDO 発話数	0[0.00]	0[0.00]	0[0.00]
肯定語のみ発話数	35[27.78]	93[35.36]	89[33.84]
肯定+情報形発話数	3[2.38]	20[7.60]	22[8.37]
否定語のみ発話数	0[0.00]	0[0.00]	0[0.00]
否定+情報形発話数	0[0.00]	1[0.38]	0[0.00]
認識棄却数	15[11.90]	19[7.22]	33[12.55]
誤認識数	3[2.38]	10[3.80]	15[5.70]

表 7: 発話全体の言語的特徴 (機械相手)

評価項目	Easy (%)	Middle (%)	Hard (%)
設定経路数	9	18	18
総発話数	119	399	350
詳細化発話数	14[11.76]	43[10.78]	38[10.86]
訂正発話数	8[6.72]	19[4.76]	23[6.57]
言いよどみ発話数	2[1.68]	7[1.75]	5[1.43]
言い直し発話数	0[0.00]	0[0.00]	1[0.29]
間投詞数	4[3.36]	5[1.25]	1[0.29]
発話衝突数	0[0.00]	6[1.50]	2[0.57]
形態素数	3.41	4.15	4.23
平均情報数	1.98	1.91	1.86
平均新情報数	1.75	1.05	1.01
発話数	13.22	22.17	19.44
動詞省略形発話数	66[55.46]	144[36.09]	147[42.00]
UNDO 発話数	0[0.00]	0[0.00]	0[0.00]
肯定語のみ発話数	44[36.97]	126[31.58]	104[29.71]
肯定+情報形発話数	0[0.00]	2[0.50]	2[0.57]
否定語のみ発話数	2[1.68]	25[6.27]	6[1.71]
否定+情報形発話数	2[1.68]	17[4.26]	9[2.57]
認識棄却数	8[6.72]	120[30.08]	86[24.57]
誤認識数	8[6.72]	42[10.53]	44[12.57]

析を行った。その結果、心理的負担が増加すると、最大パワーが大きくなるという変化のみが有意な結果となり、それ以外に有意な差はみられなかった。分析結果では、Middle レベルが最大パワーで一番大きな値をとっているが、F 検定及び t 検定結果では、Easy と Middle の比較 (t 検定で $p < 0.1$) より Easy と Hard の比較 (t 検定で $p < 0.05$) の方が明確な差がでている。

4.4 考察

4.2 節、4.3 節では運転タスクの難易度による言語的・音響的特徴に関して F 検定及び t 検定による分析を行ったが、有意な差はほとんど見られなかった。ここでは、検定結果では有意な差が表れなかったが、分析結果から見られる特徴の傾向を挙げる。まず、言語的分析において心理的負担が増加すると次のような傾向が見られた。

- 対話相手が人間の場合、言いよどみ発話が増加する

表 8: 新情報発話の言語的特徴 (人間相手)

評価項目	Easy (%)	Middle (%)	Hard (%)
設定経路数	9	18	18
総発話数	50	111	105
詳細化発話数	12[24.00]	33[29.73]	30[28.57]
訂正発話数	0[0.00]	1[0.90]	1[0.95]
言いよどみ発話数	5[10.00]	16[14.41]	24[22.86]
言い直し発話数	3[6.00]	2[1.80]	2[1.90]
間投詞数	32[64.00]	69[62.16]	71[67.62]
発話衝突数	14[28.00]	15[13.51]	20[19.05]
形態素数	10.28	9.55	10.52
平均情報数	1.84	1.77	1.95
平均新情報数	1.79	1.71	1.73
発話数	5.56	6.17	5.83
動詞省略形発話数	25[50.00]	66[59.46]	52[49.52]

表 9: 新情報発話の言語的特徴 (機械相手)

評価項目	Easy (%)	Middle (%)	Hard (%)
設定経路数	9	18	18
総発話数	56	135	123
詳細化発話数	13[23.21]	37[27.41]	37[30.08]
訂正発話数	0[0.00]	2[1.48]	2[1.63]
言いよどみ発話数	2[3.57]	4[2.96]	4[3.25]
言い直し発話数	0[0.00]	0[0.00]	0[0.00]
間投詞数	4[7.14]	2[1.48]	1[0.81]
発話衝突数	0[0.00]	3[2.22]	1[0.81]
形態素数	5.24	6.34	6.06
平均情報数	2.05	1.96	1.97
平均新情報数	2.00	1.74	1.86
発話数	6.22	7.50	6.83
動詞省略形発話数	52[92.86]	86[63.70]	89[72.36]

このことから、心理的負担が増加する事で、発話しづらくなるようである。有意な差がでなかった理由としては、暗記したキーワードをそのまま発話するだけでよいタスクであった為、被験者によっては、あらかじめ決めた文章を発声していた可能性が考えられる。もし、文章を考えながら発話するタスクであれば、有意な差がでる事も考えられる。また、音響的分析において心理的負担が増加すると次のような傾向がある。

- 人間相手の場合、発話中の無音区間が減少する
- 機械相手の場合、発話中の無音区間が増加する

人間相手では、心理的負担が増加すると、あまり発話内容を気にしなくても理解される為、頭で考えた内容をそのまま一気に発話してしまえばよい。しかし、機械相手では、常に理解できる文体かどうかを意識する必要があり、言語生成単位毎に、それを意識する為の無音区間が入ることによって、無音区間が増加したと考えられる。

以上のように、分析結果を見てみるといくつか傾向を挙げる事ができる。しかし、これらの傾向はほとんどが言語生成の内容に関係した影響はなく、韻律的特徴に関わる発声器官の特徴に関係した影響である事が分かる。これは、運転タスクが運動に関する認知的負荷のみに関わるタスクであり、言語生成に影響を与えにくいタスクである為と考えられる。仮に、対話中

表 10: 発話全体の音響的特徴 (人間相手)

評価項目	Easy	Middle	Hard
発話開始時間	0.13	0.27	0.15
発話時間 (IP)	3.36	2.90	3.04
発話時間 (EP)	1.63	1.81	1.82
発話速度 (IP)	4.74	4.94	4.81
発話速度 (EP)	9.74	9.95	9.73
最大パワー	2425.33	6748.87	5592.06
平均パワー	674.96	848.45	859.75

表 11: 発話全体の音響的特徴 (機械相手)

評価項目	Easy	Middle	Hard
発話開始時間	1.24	1.00	1.11
発話時間 (IP)	1.45	1.77	1.73
発話時間 (EP)	1.15	1.20	1.35
発話速度 (IP)	7.04	7.28	7.19
発話速度 (EP)	8.88	9.71	9.20
最大パワー	1390.00	2341.44	3266.02
平均パワー	418.35	638.56	923.48

の同時処理タスクがクロスワードパズルのような言語能力に関係したタスクであれば、同様な分析結果が表れるとは限らない。しかし、カーナビゲーションシステムを音声で使用するような状況においては、運転操作の有無はユーザの言語的特徴へは影響せず、音響的特徴へのみ影響する傾向が強い事が分かった。

5 まとめ

本稿では、目的地設定タスクにおいて、様々な状況下での対話を収集し、音声対話における言語的・音響的特徴の違いの分析を行った。対話相手の違いによる対話を分析し、ユーザの発話の特徴に関するいくつかの知見を得た。また、運転タスクの有無は発話にほとんど影響を及ぼさない事が分かった。次に、運転難易度の違いによる発話の特徴の分析も行った。車内で音声入力インターフェースを使用するという状況において、運転の有無や難易度の違いは、発話の言語的特徴においてはほとんど影響を与えず、音響的特徴に若干の影響を与えるであろうという事が明らかになった。

今後は、得られた知見を、対話状況に応じて適切な対話戦略の制御や音声認識を行う手がかりとして利用していく予定である。

参考文献

- [1] 清水, 小島, 脇田, 本郷: 運転中における音声対話システムの評価, 情報処理学会研究報告, SLP-32-17, pp.87-92 (2000).
- [2] 清水, 脇田, 武田, 河口, 板倉: 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, 日本音響学会講演論文集, 3-P-26, pp.215-216 (2001).
- [3] 伊藤敏彦, 岩本善行, 水谷誠, 由浅裕規, 甲斐充彦, 小西達裕, 伊東幸宏: 目的地タスクにおける対話状況

表 12: 新情報発話の音響的特徴 (人間相手)

評価項目	Easy	Middle	Hard
発話開始時間	0.13	0.27	0.15
発話時間 (IP)	6.17	4.83	5.87
発話時間 (EP)	2.93	2.87	3.28
発話速度 (IP)	4.27	4.60	4.36
発話速度 (EP)	8.85	9.38	8.98
最大パワー	2266.89	6016.47	4998.41
平均パワー	668.97	815.60	829.04

表 13: 新情報発話の音響的特徴 (機械相手)

評価項目	Easy	Middle	Hard
発話開始時間	1.24	1.00	1.11
発話時間 (IP)	2.33	2.76	2.49
発話時間 (EP)	1.85	1.86	1.94
発話速度 (IP)	7.04	7.34	7.55
発話速度 (EP)	8.86	9.78	9.65
最大パワー	1318.36	2051.94	2930.87
平均パワー	402.76	638.03	950.47

の違いによる言語的特徴の分析, 日本音響学会秋季全国大会論文集, 2-1-9, pp.65-66 (2001).

- [4] 小島真一, 本郷武朗, 星野博之, 内山祐司: 音声対話の運転への影響評価法の開発, 情報処理学会研究会報告, Vol.99, No.ITS-2, pp.71-75 (1999).
- [5] 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, NLC2000-57, pp.61-66 (2000).
- [6] 伊藤敏彦, 峯松信明, 中川聖一: 音声対話システムの応答部における間投詞の利用とユーザのふるまいの分析, 情報処理学会研究報告, SLP-22-12, pp.61-66 (1998).
- [7] 田本, 川森, 相川: 自律型音声対話システムによる対話データの自動収集, 人工知能学会研究会資料, SIG-SLUD-A001-6, pp.19-24 (2000).
- [8] 阿部: 小特集-声質: 音声言語の多様性に迫る-発話様式のバリエーション, 日本音響学会誌, Vol.51, No.11, pp.882-886 (1995).
- [9] J. D. Lee, T. L. Brown B. Caven, S. Haake, K. Schmidt: Does a speech-based interface for an in-vehicle computer distract drivers?, Proc. World Congress on Intelligent Transport System (2000).
- [10] 甲斐充彦, 伊藤克亘: 対話システムにおける音声認識, 情報処理学会研究報告, SLP-23, pp.7-12 (2000).