

目的地設定タスクにおける対話状況の違いによる 言語・音響的特徴の比較

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

人間同士または人間と機械との音声対話において、対話状況や対話相手の違いによって生じる言語・音響的な特徴の差異に関して実音声対話データの分析結果から明らかにする。機械との対話を扱うため、比較的単純な状況設定としてカーナビゲーションシステムにおける目的地設定タスクを想定し、その音声インタフェースという具体的な状況設定においてユーザ発話に現れる言語・音響的な特徴の差異を比較している。運転操作中の音声インタフェースの使用という状況設定では、最近、運転操作の安全性への影響（ディストラクション）の可能性がいわれている。一方で、設計された音声インタフェースによって生じる言語・音響的な特徴への影響は明らかではない。そこで、相手が人間または機械、運転中または停車中といった対話状況の違いにより発話形態にどのような違いがあるか、被験者 10 名による実対話音声収録実験のデータに基づいて分析を行った。運転操作中の状況設定に関しては、擬似的な運転操作環境を設定した。さらに、対話状況の違いとあわせて、対話相手が誤認識・誤理解した場合の次発話の言語・音響的な分析も行った。その結果、運転操作の有無による言語的な特徴の差異はほとんどないが、音響的な特徴の違いが一部みられたほか、相手が人間か機械かでいくつもの言語・音響的な特徴の差異が明らかになった。

Comparison of Linguistic and Acoustic Features Caused by Different Dialogue Situations in a Landmark-input Task

TOSHIHIKO ITOH,[†] ATSUSHIKO KAI,^{††} YOSHIYUKI IWAMOTO,[†]
MAKOTO MIZUTANI,[†] HIROKI YUASA,[†] TATSUHIRO KONISHI[†]
and YUKIHIRO ITOH[†]

This paper presents the characteristic differences of acoustic and linguistic features observed for different spoken dialogue situations in human-human and human-machine interactions. We compare the acoustic and linguistic features of the user's dialogue speech both for a spoken dialogue system and an actual human-operator service in several landmark-setting tasks for a car navigation system. It is known that speech-based interaction has the potential to distract drivers and degrade safety. On the other hand, it is not clear whether a different dialogue situation causes some acoustic or linguistic differences on their utterances in a speech interface system. We collected a set of spoken dialogue data by 10 subject speakers under several dialogue situations. For the car-driving condition, we prepared a virtual driving simulation system. We analyzed the characteristic differences of user utterances caused by different dialogue situations or the system understanding errors. As a result, we observed that the existence of a car-driving task affects some prosodic features and the difference of human-machine and human-human dialogue conditions also affects the other acoustic and linguistic features, while no significant differences are observed for the other acoustic and linguistic features whether they performed a car-driving task or not.

1. はじめに

近年、音声認識技術や言語処理技術、コンピュータ性能の向上により、音声インタフェースや目的指向型音声対話システムが注目されている。音声インタフェースの応用は、今後様々な環境・対象に広がっていくこ

[†] 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

とが予想されている。しかし、実用化を指向したより高度な音声インタフェースの実現においては、いくつかの解決すべき課題がある。その1つとして、音声インタフェースに用いられる音声認識システムの性能向上と信頼性の確保という問題がある。音声認識システムにおける認識性能は、発話様式によって大きな影響を受けることが指摘されている¹⁰⁾。そのため、これまで音声認識の分野では、読み上げ文を対象としたものが中心であり、近年になって講演音声¹²⁾や講義音声¹¹⁾といった自然発話の認識に注目が注がれるようになってきた。さらに最近では、システムの信頼性、頑健性の向上を目的として、認識結果の信頼度推定³⁰⁾や信頼度を利用した対話制御^{15),20),21)}、システムの誤認識やユーザの言い直しの検出などに関する研究も行われている^{2),3),6),16)}。

これらのアプローチは、それぞれの発話様式特有の言語・音響的特徴を考慮することによって性能の改善を試みている。しかしながら、発話様式のバリエーションに影響を与える要素は、対話タスク、対話相手、発話状況など多岐にわたると考えられ、その要因はまだ明確になっているとはいえない¹⁴⁾。我々は、カーナビゲーションシステムにおける音声対話システムを想定して、このような課題に対する解決法を検討している。カーナビゲーションシステムの音声インタフェースでは運転操作との並行作業が想定されるが、そういった認知心理的な負荷がともなうかどうか対話状況の違いとしてあげられる。たとえば、機械操作や音声による対話などによって起こる注意散漫(ディストラクション)が運転操作の安全性に与える影響が指摘されている^{17),28)}。この問題については、これまで車内での携帯電話利用に関して注目されていたのみであったが、簡単な音声インタフェースの利用でさえも認知心理的な負荷は無視できず、運転操作へ影響を与える可能性があるという報告がある¹⁸⁾。一方で、そういった認知心理的な負荷が言語・音響的な特徴としてどのように影響するのか⁵⁾、また、対話相手が人間か機械かといった違いによって生じるのか、なども明らかではない。これまでの研究では、対話相手が人間か機械かという違いについての比較や⁴⁾、読み上げや講演調などのその特徴の違いが明らかな発話様式に関する分析・比較^{11),12)}などはされているが、認知心理的な負荷に関わる対話状況の違いおよび対話相手の違いについての検討例はほとんどない^{1),13)}。

本論文は、まず対話相手の違いや同時作業の有無といった対話状況の違いによって、言語・音響的特徴にどのような違いが出るかを実験的に明らかにするこ

とを目的とする。そのために、対話相手が機械または人間の場合、および運転操作の有無といった対話状況の違いを想定し、被験者実験を行っている。被験者実験では、現実的な応用の場面に近い知見を得るために、カーナビゲーションシステムにおける音声インタフェースを想定した比較的単純な目的地設定の対話タスクを設定している。その際、対話相手の違いによる影響に関しては、人間のオペレータとの対話と機械との対話の両者において、対話タスクをほぼ同じ設定にして比較評価を行っている。

また、音声対話システムでは、誤認識・誤理解が避けられないため、システム自身が誤認識・誤理解を正しく判断できることが重要となる。これまでの研究では、システムがユーザ発話を誤認識した場合の次発話の特徴や、その特徴から誤ったことを判断する方法も提案されているが^{2),3),22)~27)}、運転操作や対話相手の違いなどの対話状況の違いについては考慮していない。本論文では、対話状況の違いとあわせて、対話相手が誤認識・誤理解した場合の前後でどのような言語・音響的特徴の変化が現れるかについても明らかにする。

2. 対話状況の違いを想定した対話音声収録実験

カーナビゲーションシステムの目的地設定タスクを想定した、様々な対話状況によるユーザの言語・音響的な特徴分析のための被験者実験について述べる。

2.1 実験方法

実験で用いた目的地設定タスクは、カーナビゲーションシステムの使用を想定し、出発地・経由地・目的地のランドマーク(市町村名・インターチェンジ名・駅名など)を設定するものである。設定できるランドマークとしては、以下の3種類を想定した。

- (1) 県名+市町村名[+目的地]
- (2) 県名+高速道路名+IC名 or JCT名
(IC: インターチェンジ, JCT: ジャンクション)
- (3) 県名+路線名+駅名

出発地に関しては市町村名、経由地に関しては駅名またはインターチェンジ名、目的地に関しては市町村+目的地名を設定する(システムに入力するか、人間のオペレータに伝える)こととし、すべて設定できたときに目的地設定タスクの達成となる。

本実験で比較対象とした対話状況を表1に示す。表中の対話相手「機械」とは、本実験用に構築した目的地設定音声対話システムである。構築したシステムは、ユーザの1発話ごとに、そこまでの発話文脈の理解に基づいて明示的な確認を行う単純な音声対話を扱う。

表 1 実験対話状況
Table 1 Dialogue situations.

運転操作	対話相手		
	機械		人間 (オペレータ)
	デモ説明無	デモ説明有	デモ説明有
無	SF	S	C
有	SGF	SG	CG

U1: 愛知県東海道新幹線豊橋駅。
 S1: 愛知県東海道新幹線豊橋駅でよろしいですか?
 U2: はい。
 S2: 目的地に設定しました。次の目的地を設定してください。
 U3: 愛知県の一。 [分割発話入力]
 S3: 愛知県でよろしいですか?
 U4: 東名高速に行きます。 [分割発話入力]
 S4: 愛知県の名神高速道路でよろしいですか?
 U5: キャンセル。 [キャンセル入力]
 S5: 愛知県でよろしいですか?
 U6: 東名の豊川インター。 [分割発話入力]
 S6: 愛知県の東名高速道路の豊田インターでよろしいですか?
 U7: いいえ、豊川インターです。 [部分訂正入力]
 S7: 愛知県の東名高速道路の豊川インターでよろしいですか?
 U8: はい。
 S8: 目的地に設定しました。次の目的地を設定してください。

図 1 音声対話システムにおける対話例

Fig. 1 Example of human-machine spoken dialogue.

ただし、図 1 のシステム動作例に示すように、分割発話入力や部分訂正入力、アンドゥ入力といった入力が可能となっている。詳しくは 2.2 節で述べる。

対話相手「人間 (オペレータ)」は、コンパスリンク⁷⁾と呼ばれる実在のサービスを用いたものである。これはカーナビゲーションシステムに接続された携帯電話回線を利用して、システム操作を人間のオペレータが代行してくれるサービスである。詳しくは 2.3 節で述べる。

本実験における運転操作は、安全性の問題からドライバシミュレーション (ゲーム) を使用し、単純なコース (オーバルコース: 楕円形コース) を線に沿って一定速度 (100 km/h) で走るものとした。ただし、できるだけ実際の運転状況に近づけるため、ハンドル・アクセル・ブレーキを用意し、画面サイズもプロジェクタを使うことにより実サイズに近づけた。

このような対話状況の設定において、複数の被験者に目的地設定タスクを与えて対話音声収録の実験を行った。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まったくない。実験手

出発地は静大浜松キャンパス。今日は今話題の U.S.J. (ユニバーサルスタジオジャパン) へ行く。豊橋に住んでいる友達と豊橋駅で待ち合わせした後、東名の豊川インターから高速に乗る。大阪府の豊中インターで高速を降りたら、U.S.J. まであと少しだ。

出発地 : 静岡県・浜松市

↓
 経由地 1 : 愛知県・東海道新幹線・豊橋駅

↓
 経由地 2 : 愛知県・東名高速道路・豊川インター

↓
 経由地 3 : 大阪府・名神高速道路・豊中インター

↓
 目的地 : 大阪府・大阪市・U.S.J

図 2 タスクシナリオ例

Fig. 2 Example of task scenario.

順としては、最初に被験者に図 2 のようなタスク内容と、実験に関する説明用紙を熟読してもらい、さらに 1 タスクあたり 4~5 地点のランドマークに関しては暗記してもらった。そして、1 タスク目だけは、音声対話システムの性能や機能に関する説明をまったく与えず、ただちにシステムを使用してもらった (表 1 中の SF, SGF, 以下「自由対話」)。この自由対話の終了後、改めて音声対話システムとコンパスリンクのデモビデオを見せて、対話方法、システムに入力できる文型や分割発話入力、部分訂正入力、アンドゥ入力といったシステムに関する機能を例示した。その後、対話相手と運転操作の組合せの 4 通り × 2 タスク (表 1 中の S, SG, C, CG) を、被験者ごとにランダムな試行順で行った。

2.2 目的地設定音声対話システム

実験用に構築した目的地設定音声対話システムは、図 3 に示すように、音声認識部、対話制御部、音声合成部から構成される。このシステムは、目的地が入力されるたびに入力の明示的な確認を行う単純なシステムである。ただし、1 つのランドマークに関して県名、市などを別々の発話で入力できる分割発話入力 (図 1 中の U3, U4, U6)、誤認識したキーワード部分のみを発話しなおすことによって訂正できる部分訂正入力 (図 1 中の U7)、ユーザの前発話をキャンセルして 2 発話前の状態へ戻るアンドゥ入力 (図 1 中の U5 の発話によって、U4 の発話を無視し、U3 の発話直後の状態へ遷移) の機能を本システムでは実装している。

音声認識部には HTK¹⁹⁾ を用い、入力された音声発話から n-best (n=100) の文仮説を出力する。音節モデルには、情報処理振興事業協会 (IPA) の成果物として配布された日本語の 43 音素の triphone モデルを使

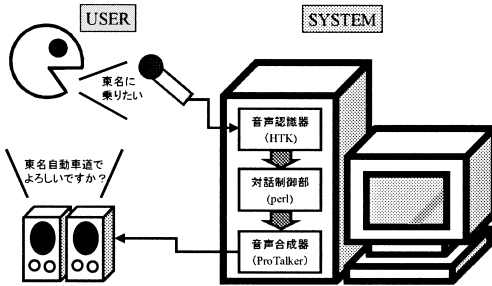


図3 システム構成図

Fig. 3 Configuration of spoken dialogue system.

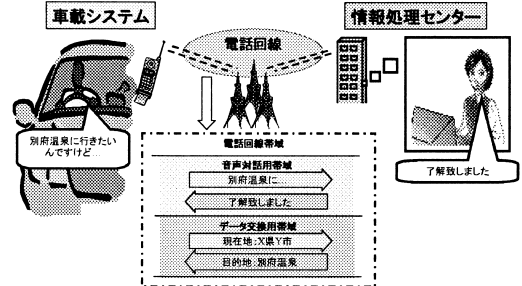


図4 コンパスリンク

Fig. 4 Compass Link.

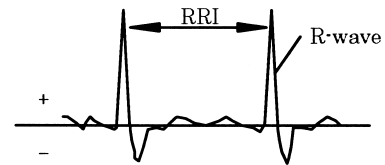
用した⁹⁾。

対話制御部では、音声認識部がユーザ発話の入力に対して地名などのキーワード列からなる n-best 候補を出力すると、その上位候補から順にそれまで入力されたキーワード列の n-best 候補(文脈)の上位候補と組み合わせていき、最初に見つかった地理制約上許されるキーワードの組合せを理解結果として確認応答文を生成する。許される組合せが見つからなかった場合、今回の発話の最上位候補を理解結果として利用する。1つのランドマークの設定は、ランドマークの最下層レベル(駅名やインターチェンジ名など)を含んだ確認応答文が生成・応答され、それに対してユーザが肯定の発話をした場合に完了し、そのとき文脈用の n-best 候補をすべてクリアする。その後、新たなランドマーク設定の対話を開始する。

音声合成部には日本 IBM 製の ProTalker を使用し、対話制御部で作成した確認文を音声によってユーザへ伝える。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。

2.3 人間のオペレータとの対話

被験者に人間のオペレータとの目的地設定タスクの対話を遂行させるために、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(図4)。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。対話相手が専用オペレータであることから、本実験タスクに最適な対話制御で対話が進行することが期待できる。コンパスリンクは、携帯電話回線を利用し、回線帯域の一部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。そして、残りの帯域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。音声対話に使用する回線帯域が通常より少ないため、通常の



$$\overline{RRI} = \frac{1}{N} \sum (RRI)$$

$$RRV = \frac{\overline{(RRI)^2} - (\overline{RRI})^2}{(\overline{RRI})^2}$$

図5 RRIとRRV

Fig. 5 RRI and RRV.

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・誤理解が生じやすい。そのため本実験では、通常の間人同士の対話に比べて、より多くの誤認識・誤理解データを収集することができた。

2.4 実車運転とシミュレーション運転の比較実験
我々は安全性の問題からカーシミュレーションを用いた実験計画を行った。しかしながら、カーシミュレーションを用いた実験では、被験者に与える心理的負担が実車を用いた実験に比べて少ない可能性がある。そこで、実車の運転とカーシミュレーションの運転の心理的負担の違いを測定した。心理的負担の測定には、生体信号による客観的評価値として計測される「心拍ゆらぎ(RRV)」を用いる。一般的に心電波形の隣接するパルスどうしの間隔(RRI)は一定ではなく、つねに変動しており、その変動幅は心理的負担が大きいほど小さくなることが知られている。そのため、一定時間内のRRIの分散をこの変動幅を示す指標RRVとし、心理的負担度を測定する(図5)。すでに柳井らは、このRRVを用いて運転の心理的負担を定量的に評価できることを示している⁸⁾。

RRV測定実験では、まず、被験者3名に運転前のRRVを測定した。次に、対話実験で使用する運転条件でカーシミュレーションを運転してもらい、そのと

表 2 全ユーザ発話の言語的特徴
Table 2 Linguistic features of all user utterances.

評価項目	SF[%]	SGF[%]	S[%]	SG[%]	C[%]	CG[%]
タスク数	4	6	20	20	20	20
発話総数	105	222	578	518	348	398
平均発話数	26.3	37.0	28.9	25.9	17.4	19.9
平均形態素数	3.1	4.6	3.6	3.6	4.2	3.8
分割発話数	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]	5[2.3]	80[13.8]	63[12.2]	9[2.6]	7[1.8]
再入力発話数	43[41.0]	87[39.2]	155[26.8]	119[23.0]	11[3.2]	18[4.5]
発話衝突回数	15[14.3]	32[14.4]	43[7.4]	33[6.4]	56[16.1]	61[15.3]
間投詞数	16	15	39	28	91	121
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
動詞省略発話数	41[39.1]	119[53.6]	303[52.4]	271[52.3]	101[29.0]	97[24.3]
アンドゥ発話数	0[0.0]	0[0.0]	23[4.0]	12[2.3]	0[0.0]	0[0.0]
肯定語のみ	28[26.7]	54[24.3]	134[23.2]	129[24.9]	128[36.8]	141[35.4]
肯定 + 情報	0[0.0]	0[0.0]	12[2.1]	14[2.7]	32[9.2]	39[9.8]
認識棄却数	49[46.7]	102[46.0]	175[30.3]	125[24.1]	12[3.5]	22[5.5]
誤認識発話数	13[12.4]	20[9.0]	127[22.0]	123[23.8]	12[3.5]	16[4.0]
否定語のみ	8[61.5]	12[60.0]	27[21.3]	35[28.5]	1[8.3]	1[6.3]
否定語 + 情報	1[0.8]	1[0.5]	48[37.8]	24[19.5]	1[8.3]	1[6.3]

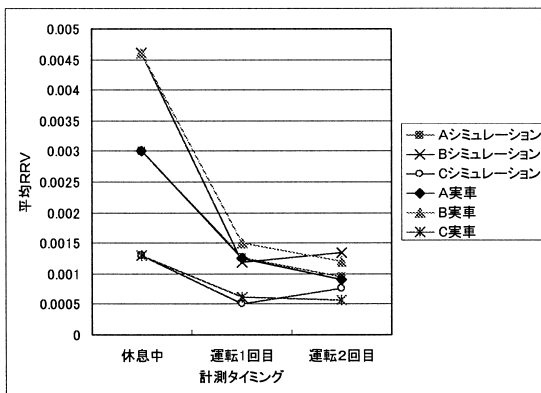


図 6 RRV 変化の比較結果
Fig. 6 Comparison of RRV changes.

きの RRV を 2 度繰り返し測定した。次に、同一被験者、同一手順で走行速度を変化させながら実車運転中の RRV を測定した。RRV の測定値の比較結果を図 6 に示す。

これらの実験結果から、対話実験に用いるカーシミュレーションの運転操作は、実車の運転操作と同程度の心理的負担を被験者に与えており、その心理的負担度は、実車を 60 km/h で走行させる場合と同程度であることが分かった。

3. 実験結果

被験者実験によって得られた発話データを言語・音響的特徴に着目し分析した結果を示す。3.1 節では、対話状況の違いによる言語・音響的特徴への影響につ

いて、3.2 節では、対話相手の誤認識・誤理解に対する被験者の次発話の反応に注目した特徴の変化について、3.3 節では、被験者・目的地設定タスク・慣れなどの要因による言語・音響的特徴への影響の分析結果について述べる。

3.1 対話状況の違いによる影響

実験によって収集された全ユーザ発話に関する言語的特徴を表 2 に示す。1 発話は、機械との対話に関してはユーザの音声入力終了するまで、人間との対話に関しては、発話権が完全に移ったと我々が判断した場所までと定義した。「平均発話数」は、1 タスクあたりの発話数の平均である。「平均形態素数」は、1 発話あたりの平均形態素数である。「分割発話数」は、1 つのランドマークを設定するのに、複数発話で行った回数である。「訂正発話数」は、誤認識・誤理解に対して、訂正を行った回数である。「再入力発話数」は、認識棄却に対して、再入力を行った回数である。「発話衝突回数」は、対話相手の発話と被験者の発話がオーバーラップした回数である。「間投詞数」は、間投詞の出現回数である。「平均情報数」は、1 発話中に含まれたキーワード(県名・市町村名・高速道路名・インター名など)数の平均である。ただし、キーワードが含まれた発話中での割合であり、肯定語のみの発話などは含んでいない。「平均新情報数」は、キーワードが含まれた各発話中で、新出のキーワード数のみの平均を示している。「動詞省略発話数」は体言止め・助詞止めなどのように動詞が省略された発話の出現数である。「肯定語[否定語]のみ」は、明示的な確認発話に対して、

表 3 新情報発話の言語的特徴
Table 3 Linguistic features of new-information utterances.

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
平均形態素数	5.0	9.0	6.4	6.0	8.1	7.1
分割発話数	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]
発話衝突回数	2[11.8]	7[17.1]	10[7.1]	3[2.3]	20[15.2]	28[20.6]
間投詞数	5	4	19	17	75	92
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
動詞省略発話数	6[35.3]	34[82.9]	100[71.4]	97[75.8]	82[62.1]	75[55.2]

肯定語 [否定語] (はい [いいえ], など) のみで応答した発話数であり、「肯定 [否定] + 情報」は、肯定語 [否定語] と何かしらのキーワードを同時に発話した数である。

表 2 の結果を見ると、音声認識システムの誤認識が比較的多いことと対話戦略の違い (人間との対話の場合、発話ごとに明示的な確認はしない) によって、機械との対話での訂正発話・再入力発話が非常に多くなり、総発話数や平均新情報数にその影響がでている。

そこで、これらの影響を除くために、ユーザの発話に新情報が含まれているもの (タスク達成に必要な発話) だけを抽出し、確認への応答 (肯定・否定) 発話などは除外して、分析を行った。その結果を表 3 に示す。

これらの結果から対話状況の違いによる影響の有無に関して、F 検定および t 検定による分析を行った。まず、人間との対話と比較して、機械との対話に特徴的なものとして以下のような傾向が分かった (特に明示しない限り t 検定で危険率 $p < 0.01$)。

- 間投詞はあまり出現しない。
- 動詞省略発話が多い。
- 発話が短い (「平均形態素数」, $p < 0.05$)。
- 1 つの発話で多くの情報を伝える (「平均情報数」)。(「自由対話」ではさらにその傾向が強くなる)
- 明示的な確認発話に対して、否定語のみ発話することがよく出現する (「否定語のみ」)。
- 機械相手でも、誤認識の明示的な確認発話に対しては半分以上は否定語を期待できる (「否定語のみ」, 「否定語 + 情報」)。

なお、発話の言語的特徴に関して、運転操作の有無によって有意な差はまったくみられなかった。また、被験者の運転操作の質に関して、走行線を大きく逸脱することや一定速度 (100 km/h) から一定幅 (5 km/h) 以上変動することもなく、良好であった。

次に、対話状況の違いによる音響的特徴の変化に関

表 4 全ユーザ発話の音響的特徴 (1 発話平均)

Table 4 Acoustic features of all user utterances (means per utterance).

評価項目	SF	SGF	S	SG	C	CG
発話開始時間 (sec)	1.03	1.04	0.96	0.99	0.50	0.49
発話時間 (1)(sec)	2.68	2.16	2.40	2.61	1.96	2.11
発話時間 (2)(sec)	1.81	1.43	1.56	1.58	1.54	1.55
発話速度 (1)(mora/sec)	4.56	4.73	4.80	4.70	5.44	5.51
発話速度 (2)(mora/sec)	6.72	7.09	6.97	7.00	6.64	7.06
パワー平均値 (RMS)	16.2	32.5	28.9	28.1	34.2	33.5
パワー最大値 (RMS)	81.1	90.7	143.7	111.9	149.8	149.6
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

して、全ユーザ発話の分析結果を表 4 に、新情報発話の分析結果を表 5 に示す。表中の「発話開始時間」とは、対話相手が発話を終了してから、ユーザが発話を開始するまでの時間である。ただし、ユーザの発話開始が対話相手の発話と重なる場合は、重なった時間に相当する負の値としている。表中の「(1)」は、発話中に含まれる無音区間 (ポーズ) も発話として含めた結果であり、一方、表中の「(2)」は、発話中の無音区間 (ポーズ) を除いて算出した結果である。

言語的特徴の場合と同様に、評価項目に関して F 検定および t 検定による分析を行った。人間との対話と比較して、機械との対話にみられる音響的特徴として以下のようなことがあげられる (特に明示しない限り t 検定で危険率 $p < 0.01$)。

- 発話開始時間が遅い。
- 「全ユーザ発話」としては、1 発話の発話時間が長い ($p < 0.05$)。
- 「新情報発話」に関しては発話速度が速い ($p < 0.05$) (「全ユーザ発話」としては逆に発話速度が遅くなっている)。
- ピッチの平均値、最大値が低く、ピッチの変化が小さい。

表 5 新情報発話の音響的特徴(1発話平均)

Table 5 Acoustic features of new-information utterances (means per utterance).

評価項目	SF	SGF	S	SG	C	CG
発話開始時間(sec)	1.47	1.70	1.10	1.41	0.70	0.56
発話時間(1)(sec)	5.69	3.11	3.77	3.66	3.76	4.06
発話時間(2)(sec)	3.39	2.25	2.64	2.60	2.93	3.00
発話速度(1)(mora/sec)	4.09	5.28	5.11	5.23	4.97	5.16
発話速度(2)(mora/sec)	6.74	6.99	6.81	6.92	6.09	6.59
パワー平均値(RMS)	13.71	32.37	28.23	27.59	33.38	33.17
パワー最大値(RMS)	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

表 6 誤認識前後における間投詞

Table 6 Filled pauses before and after speech recognition errors.

誤認識前後		SF	SGF	S	SG	C	CG	合計
間投詞無	間投詞無	2	5	65	54	4	3	133
間投詞無	間投詞有	0	0	3	3	0	1	7
間投詞有	間投詞無	0	1	10	7	2	5	25
間投詞有	間投詞有	1	0	3	2	1	4	11

また、運転操作の有無による比較では、機械との対話において運転操作をとまなう場合に、全ユーザ発話に対する分析ではピッチの最大値および平均値が上昇する傾向がみられ(それぞれ、 $p < 0.05$ および $p < 0.01$)、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($p < 0.10$)。その他の音響的特徴の変化はみられなかった。

3.2 誤認識・棄却による影響

誤認識前後の発話における間投詞の出現傾向を表 6 に、形態素数の変化を表 7 に、発話情報数の変化を表 8 に、動詞省略の出現傾向を表 9 にそれぞれ示す。表 8 の“()”内の数字は、認識結果に部分的な誤りがあり、被験者が誤った部分だけを部分訂正したことが原因で情報数が減少した発話数である。“[]”内の数字は、認識結果が部分的に誤ったにもかかわらず、全情報を言い直した発話数である。

これらの結果から、誤認識した発話とその直後の訂正発話との言語的な特徴の違いに関して次のような傾向が分かった。

- 訂正発話は、訂正前の発話より短くなる傾向がある。
- 訂正発話における情報数は減少する傾向がある。
- 情報が部分的に誤認識された場合、ほとんどの場合、誤り部分のみが訂正される。
- 訂正発話において、間投詞や動詞省略に関する特徴はみられなかった。

表 7 誤認識前後における形態素数

Table 7 Number of morphemes before and after speech recognition errors.

	SF	SGF	S	SG	C	CG
誤認識数	10	15	110	102	11	15
平均形態素数(誤直前)	9.40	4.80	4.64	4.19	5.09	8.87
平均形態素数(誤直後)	3.80	3.27	3.88	3.17	3.18	4.67

表 8 誤認識前後における情報数

Table 8 Number of keywords before and after speech recognition errors.

誤認識前情報数	誤認識後情報数	SF	SGF	S	SG	C	CG	合計
1	1	0	4	47	32	5	8	96
	2	0	0	12	3	0	0	15
	3	0	0	2	1	0	0	3
2	1	0	1	15(8)	13(8)	1(1)	3(3)	33(20)
	2	2[1]	0	8[3]	4[2]	0	1[1]	15
	3	0	0	2[1]	2	0	0	4
3	1	1(1)	0	4(2)	3(2)	0	1	9(5)
	2	0	0	2(1)	3(3)	1	0	6(4)
	3	0	1[1]	2[1]	2	0	0	5

表 9 誤認識前後における動詞省略

Table 9 Ellipses of verb phrase before and after speech recognition errors.

評価項目	SF	SGF	S	SG	C	CG
省略無 省略無	4	8	26	20	6	9
省略無 省略有	1	0	14	24	2	3
省略有 省略無	4	4	16	25	2	2
省略有 省略有	1	3	54	33	1	1

次に、ユーザの発話パターン別の音響的特徴の分析結果を表 10 に示す。表 5 の結果では一部の音響的特徴を除いて運転操作による差がほとんどみられなかったため、S と SG、C と CG を結合したデータをそれぞれ「S-AVE」および「C-AVE」として分析を行った。表中の「訂正発話」は、誤認識時の次発話を集めて計算し、「再入力発話」は、発話棄却時(「もう一度言っ

表 10 発話内容別の音響的特徴 (1 発話平均)

Table 10 Acoustic features for different kind of utterances (means per utterance).

評価項目	S-AVE				C-AVE			
	全ユーザ発話	新情報発話	訂正発話	再入力発話	全ユーザ発話	新情報発話	訂正発話	再入力発話
発話開始時間 (sec)	0.97	1.26	0.73	1.00	0.49	0.63	0.25	0.49
発話時間 (1)(sec)	2.50	3.71	1.68	3.51	2.04	3.91	1.73	1.79
発話時間 (2)(sec)	1.57	2.62	1.12	2.24	1.55	2.96	1.48	1.44
発話速度 (1)(mora/sec)	4.75	5.17	6.46	5.10	5.48	5.07	6.85	6.39
発話速度 (2)(mora/sec)	6.99	6.87	7.86	6.72	6.85	6.34	7.62	7.38
パワー平均値 (RMS)	28.47	27.91	31.28	28.24	33.85	33.27	36.06	43.12
パワー最大値 (RMS)	127.80	116.56	88.47	103.95	149.68	140.22	115.29	140.12
対数化 F0 最小値	4.46	4.37	4.49	4.37	4.46	4.34	4.36	4.39
対数化 F0 最大値	5.22	5.31	5.23	5.27	5.32	5.42	5.37	5.37
対数化 F0 平均	4.90	4.89	4.91	4.90	4.93	4.94	4.96	4.99
対数化 F0 標準偏差	0.18	0.20	0.17	0.20	0.21	0.23	0.25	0.24

てください」など)の次発話を集めて計算した結果である。

この結果から、発話内容の違いにより以下の音響的な特徴がみられた(特に明示しない限り t 検定で危険率 $p < 0.01$)。

- 訂正発話の発話開始時間は速い(機械との対話 S において $p < 0.01$, 人間との対話 C において $p < 0.1$)。
- 訂正発話の発話速度は速い(S において $p < 0.01$, C において $p < 0.01$)。
- 機械との対話では、訂正発話のピッチの範囲は新情報発話と比べて狭い(最小値は高く、最大値は低く、標準偏差は小さい)。特に訂正発話のピッチの最小値は他の発話と比べて高い。
- 人間との対話では、「全ユーザ発話」を除いてピッチの標準偏差が大きい。
- 人間との対話では、再入力発話は新情報発話と比べてピッチの平均値が高い。
- 人間との対話では、「全ユーザ発話」と比べてそれ以外の発話はピッチの標準偏差が大きい($p < 0.05$)。
- すべての種類の発話内容に関して、機械との対話は人間との対話に比べてピッチの最大値が低く、標準偏差が小さい。
- 新情報発話と再入力発話に関して、機械との対話は人間との対話に比べてピッチの平均値が低い。
- 訂正発話に関して、機械との対話は人間との対話に比べてピッチの最小値が低い($p < 0.05$)。

なお、以下のような運転操作の有無により有意差がみられた音響的特徴に関連した結果のみを抜粋して表 11 に示す。

- 運転操作ありの対話では、訂正発話・再入力発話に対する応答の発話速度は遅くなる(C において $p < 0.05$)。

表 11 運転操作の有無による音響的特徴 (1 発話平均)

Table 11 Effect of a car-driving task on acoustic features (means per utterance).

評価項目	訂正発話		再入力発話	
	C	CG	C	CG
発話開始時間 (sec)	0.33	0.17	0.40	0.55
発話時間 (1)(sec)	1.39	2.07	1.47	2.01
発話時間 (2)(sec)	1.29	1.68	1.21	1.60
発話速度 (1)(mora/sec)	7.77	5.93	7.04	5.96
発話速度 (2)(mora/sec)	8.23	7.01	7.87	7.05

- 運転操作ありの対話では、発話中の無音増加により訂正発話・再入力発話に対する応答の発話時間は長くなる(C の発話時間 (1) のみにおいて $p < 0.1$)。

3.3 その他の要因による影響の分析

これまでに述べてきた対話状況や発話状況以外の要因によって言語的・音響的特徴にどのような影響が出るか分析を行った。分析対象とした要因は、被験者間による影響(個人差)、目的地設定タスクの種類による影響、実験慣れによる影響である。被験者間による影響の分析は、ランダムに選んだ 2 名の実験結果を検定することで行った。次に、タスクの種類による影響の分析は、ランダムに選んだ 2 つのタスクの実験結果を検定することで行った。最後に、実験慣れによる影響の分析は、1 タスク目と最後のタスクを検定することで行った。なお、検定対象とした評価データは、すべて同一の対話状況の結果だけを用いている。これらの評価データを F 検定および t 検定を用いて分析を行った(特に明示しない限り t 検定で危険率 $p < 0.01$)。

言語的特徴に関しては、以下の傾向がみられた。

- 間投詞・動詞省略発話数には、個人差がある。
- 目的地設定タスクの種類による言語的特徴はみられない。
- 慣れによる言語的特徴はみられない。

音響的特徴に関しては、以下の傾向がみられた。

- 対象とした分析項目すべてに個人差がある ($p < 0.05$) .
- 目的地設定タスクの種類によって、訂正・再入力発話の発話時間 ($p < 0.05$) 発話開始時間 ($p < 0.1$) に差がある .
- 慣れによって、発話開始時間が短縮する ($p < 0.1$) .

4. 考 察

ここでは、まず (1)・(2) で対話相手の違いにのみ依存する言語的・音響的特徴に注目して考察する。続いて、(3) では運転操作の有無による影響を対話相手や発話内容に着目しながら考察する。(4) では、対話相手や運転操作の有無に依存しなかった特徴的な結果について考察する。(5) では、その他の要因による影響を考察する。

(1) 対話相手の違いによる言語的特徴

対話状況の違いによる影響の分析結果から、ユーザは機械が対話相手の場合、タスク達成に必要な情報を言語として成り立つ最低限度の言葉で入力しようとする傾向があることが分かる。また、誤認識した情報に関する確認発話に対して、人間同士の対話ではまず出現しない否定語のみの応答(「いいえ」や「違う」のみ)がよく出現することからも、機械を対等な対話相手というよりはツールとして見なしている傾向がある。これは、ユーザが現状の機械(音声対話システム)をまだ人間と同等に対話ができる相手とは考えていないためと思われる。

もちろん、対等な対話相手と考えにくい理由に、本システムの応答までの処理時間(2秒~5秒)や認識性能による影響の可能性も考えられる。以前に我々は、タスクは異なるが、WOZシステムを用いて人間と同程度の音声認識・理解性能を持つ機械と、応答性能を機械レベル(応答は録音音声、応答時間は約2秒程度)までに制限した人間と、応答を自由かつ即時(約1秒)に行う通常の人間の3種類の対話相手ユーザの発話がどのように変化するか実験を行った⁴⁾。結果として、録音音声の応答に対して実行しにくい行為、たとえば、割込み、相槌などに関しては同じ人間相手でも応答能力を制限されているかどうかで差が現れたが、全体としては応答能力の差よりも対話相手が機械か人間かという意識の影響の方がはるかに大きい結果となった。本実験の言語的特徴の変化も、上述の場合と同様に応答までの時間や認識性能の影響よりも対話相手が機械か人間かの印象の違いによる影響が大きいと考えられる。

今回の実験では、簡単な目的地設定タスクと単純な対話制御を用いたために、よりツールらしさが出てしまったが、この言語的特徴が機械の応答内容や応答方法、対話制御の人間らしさの度合いによってどのように変化するか調査する必要がある。新美ら²⁹⁾は、複雑な京都観光案内をタスクとしたWOZシステムによって、機械らしく対応するモード1と人間らしく対応するモード2による応答の違いによる被験者の振舞いの違いを検討し、結果としては、モード1では「対話全体の長さがあまり変化しないが、被験者の発話の数は少なくなる」、「不要語が少なくなる」、「単語もしくは名詞で終わる句を用いた発話が多くなる」ことを示しているが、単純な対話タスクにおいてもこの傾向があるかどうかは検討を要する。

間投詞に関しては全体の傾向として、機械との対話において出現しなくなるが、一部の被験者(2名)では対話相手に依存せず同程度出現している。被験者間の分析結果からも間投詞と動詞省略の出現頻度に関しては個人差が大きいことから、これが被験者の音声対話システムに対する意識の違いによるものか癖によるものか検討する必要がある。

(2) 対話相手の違いによる音響的特徴

音響的な特徴に関して、機械相手の場合、ユーザが発話するまでの時間が長い。これも、機械に対する自由な発話は受理(認識)されないという印象のために、発話内容を生成する(どのような発話をすれば受理してもらえるか考える)ための時間がかかっているためと思われる。また、以前の実験⁴⁾では、ユーザは機械に対して発話速度を低下させ、認識されやすいように丁寧に発話しようとしている傾向があった。本実験でも「全ユーザ発話」でみると、ユーザは機械への発話速度を低下させ、認識されやすいように丁寧に発話しようとしている。しかしながら、言語的特徴の方で述べたが、本実験の目的地設定タスクではシステムに対して一度にたくさんの情報を伝達しようとする傾向があるため、新情報発話の発話速度に関しては人間相手に比べて速くなっていることが以前の結果とは異なっている。

また、平均的なピッチやその変化は、機械との対話において小さくなる傾向がみられた。このことから、訂正発話の検出などのために韻律的な手がかりを用いる場合でも^{3),6)}、対話タスクや対話状況を考慮する必要があることを示唆している。

さらに再入力発話に関して、対話相手が機械か人間かの違いによって音響的特徴の傾向が異なっていた。つまり、機械相手の場合の再入力発話は新情報発話の

音響的特徴に類似し、人間相手の場合の再入力発話は訂正発話の音響的特徴に類似していた。この原因としては、対話制御方法の違いが大きく影響していると考えられる。我々が用いた音声対話システムは被験者の発話を棄却をした場合、つねに「すみませんが、もう一度言ってください。」とユーザに再入力を要求する。しかしながら、人間がこのような応答をすることはほとんどなく、何かしら現在の理解状況の情報を追加して再入力を要求する。そのため、ユーザの受ける印象として、機械への再入力は最初からまた新規に入力をやり直している感覚が、逆に人間への再入力には直前の発話を訂正しているような感覚が働くために、このような違いが現れると考えられる。そのため、機械との対話では訂正発話とそれ以外の発話との違いがより明確に表れる傾向があった。このことは、機械との対話という対話状況や対話制御によって、かえって有用な音響的手がかりをもたらず可能性を示唆している。

(3) 運転操作の有無による言語的・音響的特徴

次に、運転操作の有無による発話の言語的特徴の違いは、予想に反してまったくみられなかった。文献(28)では、音声インタフェースの使用による運転操作への影響(ディストラクション)の理由として、運転操作のような脳内の処理に優先して言語能力を使用する影響の表れではないかと述べている。本実験の結果は、この仮説を支持するような結果となった。一方で、運転操作レベルの作業タスクでは、対話の言語的特徴へ影響を与えるほどの心理的負荷ではなかった可能性もある。今後、もっと心理的負担の高い作業タスクを並行タスクとしたときの言語的・音響的特徴への影響を分析する必要があると考えている。

また、運転操作の有無による発話の音響的特徴もほとんどみられなかったが、音響的特徴の中の韻律的特徴のいくつかのみに、運転操作の有無によって違いがみられたことは興味深い。1つの仮説として、本実験で使用した同時処理タスクは運動制御に関連したタスクであり、似た制御に属する発声器官制御の中で最も繊細な制御が必要な韻律制御へ影響したのと考えられる。

訂正発話・再入力発話での運転操作による影響を見てみると、対話相手が人間の場合だけに発話速度と発話時間に関してその有無の影響がみられた。これに関しては、さらなる詳細な分析が必要であると考えているが、以下のような仮説が考えられる。すなわち、機械相手ではそもそも発話開始時間が遅いので、発話開始前の時点で、何を言い直すか訂正するかを思考済みであるために差がでにくい。しかし、対人間では、とりあえず間をおかずに発話を開始する傾向があり、運

転タスクへの注意を失わないようにしていると長時間の思考を妨げやすくなり、発話を延ばすことによって対処しようとするため差が出るのではないかと考えている。

(4) 対話状況に無関係な言語的・音響的特徴

訂正発話における言語的特徴に関して、いくつかの傾向が対話状況に依存せずにみられた。まず、音声認識の部分的な誤りに関しては、全体をもう一度言い直したり、新たな情報を追加して訂正したりすることはほとんどなく、誤った部分のみを訂正する傾向がある。また、訂正発話は誤認識前より情報数を減らし、平均形態素数を少なくすることによって、訂正の要点のみを伝えようとするのが分かった。さらに、誤認識に対する確認発話に対しては、半分以上において否定語が出現しているが、この結果は、平沢らの音声対話システムの誤解に対するユーザ応答の分析結果とほぼ一致する²⁴⁾。

さらに、訂正発話において発話開始時間の短縮や発話速度の上昇がみられた。これは、誤認識・誤理解はできるだけ早く修正しようとする人間の意識の現れだと考えられる。

(5) その他の要因による影響

言語的特徴に関しては、目的地設定タスクの違いや慣れによる影響はみられなかったが、音響的特徴には目的地設定タスク間や慣れによる影響が若干みられた。しかしながら、今回の実験では、すべてのタスクをランダムな順番で実施しているため、分析結果への影響はないと考えられる。

5. ま と め

対話状況の違いによる言語的・音響的特徴の違いを分析するため、目的地設定タスクを想定して対話音声収録実験を行った。比較する対話状況としては、対話相手として機械(音声対話システム)または人間(オペレータ)の2種類の状況の違いによるものと、運転操作の有無に関する対話状況の違いを想定した。

収録した対話音声の分析結果として、運転操作の有無による言語的な特徴の差異はないが、韻律的な特徴には違いが一部みられた。これらのことから、運転操作程度の同時処理タスクの有無は、言語生成能力へほとんど影響することはないことが分かった。また、相手が人間か機械かの比較においては、ほとんど同一のタスク設定であったがいくつかの言語・音響的な特徴の差異が明らかになった。機械相手の場合は、システムを自然な対話相手とは認識できず、ピッチ変化が小さくなるなど特有の言語・音響的特徴を持った発話に

なる傾向がみられた。さらに、訂正・再入力発話を他の発話と比べた分析では、対話状況に依存しないいくつかの言語・音響的特徴のほか、対話状況に依存した特徴もみられた。

今後は、対話状況や発話内容の違いに関して明らかになった特徴をさらに詳細に分析し、音声対話システムへ応用することを検討する予定である。また、対話タスクや入力文体がより複雑になった場合や同時処理タスクの難易度が異なる場合に、言語的・音響的特徴がどのように変化するか調査する必要があると考えている。さらに、音声インタフェースとディストラクションとの関係についてはまだ明確になっていないため、その分野で明らかになる知見を考慮しつつ音声インタフェースの検討を進めていく必要があると考えられる。これまでの実験で対話相手の違いによる特徴の差異が認められたが、機械相手の音声インタフェースであっても対話相手を機械的か人間的かと感じる意識の違いによっても同様な違いが生じることも予想される。したがって、知的なインタフェースを設計していくためには、ユーザ意識や対話状況の違いなどを考慮することが今後重要と考えられる。

謝辞 コンパスリンクを利用した実験、RRV 測定実験に関し、便宜を図っていただいた日産自動車総合研究所車両交通研究所の岸則政様、大野健様に深謝いたします。

参 考 文 献

- 1) 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).
- 2) 角谷, 北岡, 中川: カーナビの地名入力における誤認識時の訂正発話の分析と検出, 情報処理学会研究報告, SLP-37-11, pp.61-66 (2001.7).
- 3) 山肩, 河原: 音声対話システムにおける訂正発話の韻律的特徴の分析, 人工知能学会研究会, SIG-SLUD-A101-3, pp.5-12 (2001.3).
- 4) 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問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).
- 5) 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).
- 6) 甲斐充彦, 石丸明子, 伊藤敏彦, 小西達裕, 伊東幸宏: 目的地設定タスクにおける訂正発話の特徴分析と検出への応用, 日本音響学会秋季全国大会論文集, 2-1-8, pp.63-64 (2001).
- 7) <http://www.compasslink.ne.jp/cplweb/>
- 8) 柳井達美, 毛利 宏, 管沢 深, 岸 則政: 生体情報を用いた車両運動性能評価の一検討, 日本機械学会第 2 回交通・物流部門大会講演論文集, pp.535-538 (1993).
- 9) 河原ほか: 日本語ディクテーション基本ソフトウェア(99年度版)の性能評価, 情報処理学会研究報告, SLP-31-2, pp.9-16 (2000).
- 10) 村上仁一, 嵯峨山茂樹: 自由発話音声認識における音響のおよび言語的な問題点の検討, 日本音響学会音声研究会資料, SP91-100, pp.71-78 (1991).
- 11) 西村, 伊東: 講義コーパスを用いた自由発話の大語彙連続音声認識, 信学論(D-II), Vol.J83-D-II, No.11, pp.2473-2480 (2000).
- 12) 奥田浩三, 中嶋秀治, 河原達也, 中村 哲: 講演音声の音響的特徴分析と音響モデル構築方法の検討, 情報処理学会研究会資料, SLP-37-13, pp.73-78 (2001).
- 13) 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, 信学技報, SP2000-105, pp.61-66 (2000).
- 14) 阿部匡伸: 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).
- 15) 新美, 小林: 音声認識の信頼性に基づいた対話制御方式, 信学技報, SP96-30, pp.75-80 (1996).
- 16) 甲斐充彦, 伊藤克亘: 対話システムにおける音声認識, 情報処理学会研究会資料, SLP-33-2, pp.7-12 (2000).
- 17) <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/DriverDistraction.html>
- 18) <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm>
- 19) <http://htk.eng.cam.ac.uk/>
- 20) Litman, D.J., Hirschberg, J.B. and Swerts, M.: Predicting automatic speech recognition performance using prosodic cues, *Proc. 6th Applied Natural Language Processing Conference, (ANLP-NAACL00)*, pp.218-225 (2000).
- 21) Litman, D.J., Walker, M.A. and Kearns, M.J.: Automatic detection of poor speech recognition at the dialogue level, *Proc. 37th Annual Meeting of the Association of Computational Linguistics (ACL99)*, pp.309-316 (1999).
- 22) Oviatt, S., Levow, G.-A., MacEachern, M. and Kuhn, K.: Modeling hyperarticulate speech during human-computer error resolution, *Proc. ICSLP*, pp.801-804 (1996).
- 23) Levow, G.-A.: Characterizing and recognizing spoken correction in human-computer dialogue, *Proc. COLING-ACL98*, pp.736-742 (1998).

- 24) 平沢純一, 宮崎 昇, 相川清明: 質問-応答連鎖からの音声対話システムの誤解の検出, 信学技報, SP2000-115, pp.121-126 (2000).
- 25) 平沢純一, 宮崎 昇, 中野幹生, 相川清明: 音声対話システムの誤解に体するユーザの応答の分析, 情報処理学会研究会資料, SLP-29-27, pp.157-162 (1999).
- 26) 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).
- 27) Krahmer, E., Swerts, M., Theune, M. and Weegels, M.: Problem spotting in human-machine interaction, *Proc. Eurospeech*, Vol.3, pp.1423-1426 (1999).
- 28) Lee, J.D., Brown, T.L., Caven, B., Haake, S. and Schmidt, K.: Does a speech-based interface for an in-vehicle computer distract drivers?, *Proc. World Congress on Intelligent Transport System* (2000).
- 29) 新美, 小林: 音声認識を容易にする対話制御方式, 文部省重点領域研究“音声対話”公開シンポジウム予稿集, pp.49-58 (1993).
- 30) Schaaf, T. and Kemp, T.: Confidence measures for spontaneous speech recognition system, *Proc. ICASSP*, pp.875-878 (1997).

(平成 13 年 11 月 27 日受付)

(平成 14 年 4 月 16 日採録)



伊藤 敏彦 (正会員)

昭和 46 年生。平成 11 年豊橋技術科学大学大学院工学研究科博士後期課程電子・情報工学専攻修了。同年静岡大学情報学部情報科学科助手に赴任。音声言語処理研究に従事。博士(工学)。

士(工学)。



甲斐 充彦 (正会員)

平成 8 年豊橋技術科学大学大学院博士後期課程修了。同年豊橋技術科学大学工学部助手。平成 11 年静岡大学工学部講師。平成 12 年同助教授。音声認識を中心とした音声言語

処理と対話処理に興味を持つ。博士(工学)。日本音響学会, 電子情報通信学会各会員。



岩本 善行

昭和 52 年生。平成 12 年静岡大学情報学部情報科学科卒業。平成 14 年同大学大学院情報学研究科情報学専攻修了。同年(株)日立製作所入社。在学中は音声対話に関する研究

に従事。



水谷 誠

昭和 53 年生。平成 13 年静岡大学情報学部情報科学科卒業。現在, 同大学大学院情報学研究科情報学専攻に在学中。音声言語理解に関する研究に従事。



由浅 裕規

昭和 55 年生。平成 14 年静岡大学情報学部情報科学科卒業。現在, 同大学大学院情報学研究科情報学専攻に在学中。音声対話システムに関する研究に従事。



小西 達裕 (正会員)

昭和 62 年早稲田大学理工学部電子通信学科卒業。平成 4 年同大学大学院博士後期課程修了。平成 3 年早稲田大学理工学部情報学科助手。平成 4 年静岡大学工学部情報知識工学科助手。現在, 同大学情報学部情報科学科助教授。博士(工学)。知的教育システム, 知的対話システム等に興味を持つ。電子情報通信学会, 人工知能学会, 教育システム情報学会, 日本認知科学会各会員。



伊東 幸宏 (正会員)

昭和 55 年早稲田大学理工学部電子通信学科卒業。昭和 62 年同大学大学院博士後期課程修了。同年同大学理工学部電子通信学科助手。平成 2 年静岡大学工学部情報知識工学科助教授。現在, 同大学情報学部情報科学科教授。工学博士。自然言語処理, 知的教育システム等に興味を持つ。電子情報通信学会, 人工知能学会, 日本認知科学会, 言語処理学会, 教育システム情報学会各会員。