社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－NL－167（15）
2005－SLP－ 56（15）
2005／5／27

音声認識率や状況の違いによる音声対話の言語的・音響的特徴の比較

伊藤敏彦　山田真也　荒木健治

北海道大学　情報科学研究科
〒060-0814 札幌市北区北14条西9丁目

人間同士または人間と機械との音声対話において，タスク遂行役の音声認識率、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検索・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た．想定した状況は，音声認識率が100%と約80%の場合，対話相手が人間，応答能力が制限された人間，又は機械の場合，運転中又は停車中の場合である．これらの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24名による実対話音声の収録データに基づいて分析を行なった．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，対話状況の違いと併せて，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見られたほか，応答が自然音声か合成音声かで幾つかの言語・音響的な特徴の差異が明らかになった．

Linguistic and Acoustic Features Depending on Different Situations and Speech Recognition Rate

Toshihiko ITOH, Shinya YAMADA and Kenji ARAKI

Department of Information Science and Technology
Hokkaidou University, Hokkaidou, 060-0814, Japan

This paper presents the characteristic differences of linguistic and acoustic features observed in different spoken dialogue situations and with different dialogue partners: human-human vs. human-machine interactions. We compare the linguistic and acoustic features of the user's speech to a spoken dialogue system and to a human operator in several goal setting and destination database searching tasks for a car navigation system. It has been pointed out that speech-based interaction has the potential to distract the driver's attention and degrade safety. On the other hand, it is not clear enough whether different dialogue situations and different dialogue partners cause any differences of linguistic or acoustic features on one's utterances in a speech interface system. Additionally, research about influence of speech recognition rate is not enough either. We collected a set of spoken dialogues by 24 subject speakers for each experiment under several dialogue situations. For a car driving situation, we prepared a virtual driving simulation system. We also prepared two patterns where we have two dialogue partners with different speech recognition rate (100% and about 80%). We analyzed the characteristic differences of user utterances caused by different dialogue situations and with different dialogue partners in two above mentioned patterns.

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，幾つかの解決すべき課題がある．その一つとして，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 [6]．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 [8] や講義音声 [7] 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 [17] や信頼度を利用した対話制御 [10, 15, 14]，システムの誤認識やユーザの言い直しの検出等に関する研究も行なわれている [3, 2, 5, 11]．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言えない [9]．また，機械操作や音声による対話等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 [16, 12]．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 [13]．我々[4] や itou[1] らは以前に，そういった認知的な負荷や対話相手の違いが言語・音響的な特徴としてどのように影響するのか分析している．しかしながら，我々の以前の研究では実際の対話システムを用いて実験を行ったために，比較項目以外の実験条件をうまくコントロールすることができず，対話相手が人間の場合と機械の場合で音声認識率や応答開始時間などが大きく異なっていた．そのため、正確に比較ができていない要素が幾つか存在した．

本論文では，WOZ 法を用いた擬似音声対話システムを作成し，比較したい要素以外の条件をできるだけ揃える様に実験を行った．また，WOZ 法を用いているため，音声認識率を自由に設定できることから，タスク遂行者側の音声認識率を対話状況の一要因として加えた．そのため、音声認識率の違い，対話相手の違い，認知的負荷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でるかを明らかにすることを目的とする．対話収集のための被験者実験では，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検索・設定を対話タスクとして設定している．

また，対話状況の違いと併せて，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　様々な状況の対話音声収録実験

カーナビゲーションシステムの目的地検索・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検索･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を検索・設定するものである．

本実験で比較対象とした対話状況を表 1 に示す．この対話状況に対して，タスク遂行者の音声認識率が 100%である場合 (実験 1) と，ノイズ等の影響という想定のもので音声認識率が約 80%である場合 (実験 2) の 2 環境で実験を行った．表中の対話相手「機械」とは，本実験用に構築した WOZ 法を用いた目的地検索・設定擬似音声対話システムである．作成したシステムは，あらかじめタスクごとに用意した応答例 (誤応答も用意済み) をマウスで選択すると，対応した合成音声がスピーカを通してユーザ側に出力される単純なシステムである．しかし，相手の割り込みに対してシステム側の発話を途中で止める等，自然な対話現象に対してできるだけ自然に対応できるようにしてある．音声合成システムには日本 IBM 製の ProTalker を使用した．また対話相手「人間」は，タスク内容に精通しさらに自然な対話ができるように事前に訓練した当研究室の学生である．そのため，対話相手「人間・自然音声」は，日常的に行われる電話などの非対面な音声対話と全く同じである．ただ，本システムのタスクが想定している顔見知りでない「オペレーター」との対話を意識させるために，対話相手「人間・自然音声」と「人間・合成音声」に関しては，実在の会社名と所属，偽名 (自然音声と合成音声では別の偽名) を用いて最初に自己紹介を行い，社会的対人関係を明確にしている．さらに,「人間・自然音声」,「人間・合成音声」,「機械」は全て同一人物が対話を担当し，タスク遂行のための対話戦略に違いが出ないように注意した．なお，対話相手「人間・合成音声」と「機械」との差は，ユーザが対話している相手が人間か機械かというユーザの意識の違いだけであり，どちらの場合も WOZ 法の擬似音声対話システムを用いて対話を行っている．

図 1 にユーザと対話相手「機械」との対話例を示す．

表 1: 実験対話状況

対話相手		人間		機械
応答音声		自然音声	合成音声	
運転タスク (DT)	無	O	P	W
	有	O+DT	P+DT	W+DT

本実験における運転操作は，安全性の問題からドライブシミュレーション (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 (100km/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ーを使うことにより実サイズに近づけた．また，運転操作タスクをメインタスクとして集中させる為に，運転操作に失敗した場合は，実験を最初からやり直すことになることを伝えてある．なお，対話実験に用いるドライブ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認知的負担を被験者に与えており，その認知的負担度は，実車を 60km/h で走行させる場合と同程度であることが RRV 法による調査からわかっている．

上述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検索・設定タスクを与えて対話音声収録の実験を行なった．被験者は音声認識率の異なる環境ごとに，

S1 ：目的地を設定して下さい.
U1 ：えー北海道江別市江別市役所.
S2 ：北海道江別市江別市役所でよろしいですか.
U2 ：もう一度言って下さい.
S3 ：北海道江別市江別市役所でよろしいですか.
U3 ：はい，その通りです.
S4 ：目的地に設定します.
S5 ：目的地を設定して下さい.
U4 ：えー北海道留萌市，えー黄金岬.
S6 ：北海道留萌市東奔機材でよろしいですか.
U5 ：いや，違います.
S7 ：目的地を設定して下さい.
U6 ：え，北海道留萌市黄金岬.
S8 ：北海道留萌市黄金岬でよろしいですか.
U7 ：はい，その通りです.
S9 ：北海道留萌市黄金岬でよろしいですか.
U8 ：はい，その通りです.
S10：目的地に設定します.

図 1: 音声対話システムにおける対話例

情報系大学 (院) 生 12 名の合計 24 名であり，音声対話システムに関する知識は全くない.

実験手順としては，最初にドライブシミュレーションを最低 30 分以上使用してもらい，運転タスクへの不慣れによる分析結果への影響を除いた．次に，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，検索条件や目的地などに関しては完全に暗記してもらった．さらに，対話システムに対して，自由なタイミング（割り込み等），自由な発話内容・形態をしても大丈夫なことを事前に伝えた．ただし，タスクに関係ない内容は理解できないことを説明し，タスク外の発話は禁止した．その後，対話相手と運転操作の組み合わせの 6 通り× 2 タスクを順番による影響が出ないように，被験者ごとにランダムな試行順で行った．全ての実験終了後，主観評価用のアンケートに記入してもらった．なお，アンケート項目に「対話相手が機械だと信じていたか」のアンケート項目があり，音声対話システムの性能に対して不自然さを感じた被験者のデータは全て評価データから除き，新たな被験者を用いて実験をやり直している.

音声認識率が約 80%の場合の実験においては，マイクからの入力にノイズを強制的に混入しているために，人間，機械共に聞き間違えることがあることを説明に加えて伝えた.

なお，音声認識率約 80%の実現方法であるが，これは 80:20 の割合で○:×を出すシステムを作成し，その指示に従って，応答内容を選択する方法を用いた．ただ，全被験者が全タスクにおいて，ほぼ同じ音声認識精度になるように，5 発話に 1 回は必ず×が出力され，なおかつ，1 タスク当たりのユーザの発話数を考慮し，1 タスク当たり 5 回以上は×が出ないように制約を加えてある．また，間違える場合は，音響的に近い誤認識・誤理解内容になるように注意した．この方法を用いることで，各タスクの音声認識率は 76%〜80%となり，実験 2 の全タスクでの音声認識率は 78%であった.

北海道大学を出発し，北海道江別市 (えべつし) の友達を拾うために待ち合わせ場所の江別市役所まで迎えに行く．友人が今の時間帯なら夕日がきれいに見えるよと言うので，そこから海沿いの道まで出て道沿いに運転し，北海道留萌市（るもいし）の黄金岬 (おうごんみさき) に行く．きれいな夕日を眺めて気分がよくなったところで，友人が腹減ったと言ってきたので，ちょっと早いけど夕食をとろうと思い黄金岬から近い国道沿いの飲食店を探すことにする.

出発地　：北海道・札幌市・北海道大学
↓
目的地 1：北海道・江別市・江別市役所
↓
目的地 2：北海道・留萌市・黄金岬
↓
目的地 3：黄金岬近く・国道沿い・飲食店

図 2: タスクシナリオ例

3　実験結果

被験者実験によって収集された発話データを言語的・音響的特徴に着目し分析した結果を示す．実験 2 の環境では実験 1 と異なり誤認識が生じるため，誤認識応答に対する訂正発話による発話の増加や言語的・音響的特徴の影響が表れる．その影響をできるだけ除くために基本的な比較では，訂正発話や否定語などの発話は除き，新情報を含んだ発話を分析対象として比較を行った。表 2 に，比較の対象となるそれぞれの状況における発話数を示す.

3.1　音声認識率の違いによるユーザ発話の言語的・音響的特徴

最初にタスク遂行者の音声認識率の違いにのみ着目し，全ての対話状況をまとめだデータによる特徴の分析を行った．表 3 にその結果を示す．これらの結果に対して，F 検定および t 検定による分析を行った．有意差のある言語的特徴としては，音声認識率が約 80%の方は 100%に比べ，平均情報数 (p<0.05) 及び平均新情報数 (p<0.05) が減少し，分割発話 (p<0.05) が増加した．言い換えると，誤認識する相手に対しては，一発話に伝える情報を減らし，細切れに伝えようとする傾向があることがわかる．更に，音声認識率が約 80%の方は動詞省略数 (p<0.01) と間投詞数 (p<0.05) が減少し，平均形態素数 (p<0.01) が増えることから，認識率が悪い相手に対しては余分な不要語はしゃべらずに丁寧な発話になる．また，音響的特徴としては，音声認識率が約 80%の方はパワー平均 (p<0.01)，パワー最大 (p<0.01)，F0 標準偏差 (p<0.01) が増加し，声がより大きく抑揚が強くなることがわかった.

表 2: 新情報発話数

対話状況	O	O+DT	P	P+DT	W	W+DT
実験 1(100%)	236	237	227	227	227	227
実験 2(80%)	245	247	232	243	249	235

表 3: 音声認識率の違いによる言語・音響的特徴

実験	実験 1	実験 2
総タスク	144	144
総発話数	1383	1451
分割発話数	187	265
平均単語数	4.60	5.49
間投詞数	446	341
平均情報数	2.02	1.92
平均新情報数	1.97	1.89
総動詞省略数	489	409
パワー平均 (RMS)	779	1018
パワー最大 (RMS)	3151	4161
F0 平均	75.4	76.5
F0 最小	58.7	57.6
F0 最大	269.4	263.2
F0 分散	16.9	19.9

3.2 対話相手の違いによるユーザ発話の言語的特徴

表 4 には音声認識率 100%(実験 1) における，表 5 には音声認識率約 80%(実験 2) における新情報発話を対象とした対話相手と対話状況の違いによる言語的特徴の結果を示す．まず，誤認識が発生しない環境である実験 1 において，対話相手の違いによってどのようにユーザ発話における言語的特徴が表れるか F 検定及び t 検定で分析した．結果，興味深いことに多くの言語的特徴がタスク遂行者が人間か機械かといった差によって表れたわけではなく，応答が合成音声か自然音声かの違いによって有意な差が表れた．

合成音声では，間投詞が減少し (p<0.01)，実際のタスクを遂行するためには必須ではないが補足的な情報としては有効である省略可能情報が多くなる (O-P:p<0.1,O-W:p<0.01)．また，有意差はなかったが動詞省略が増加する傾向がみられた．これは対話相手に対して認識・理解の助けとなる付加情報を増やしながらも，タスク遂行上には必要ない単語はできるだけ発話はしないようにする意識の表れと考えられる．しかしながら，対話相手「人間・合成音声」は聞き手が人間で応答が合成音声であることを被験者も理解しているにもかかわらず，自然音声と合成音声の間で有意な差が表れた．これは，合成音声が無意識のうちにユーザに「機械」と対話している感覚を与えるためにこのような結果が表れた可能性もあるが，自然音声と合成音声との差である音声の品質（韻律や抑揚も含む）や対話リズムの欠落といった特徴自体がこれらの言語的特徴を引き起こす原因とも考えられる．

次に，約 80%の割合で誤認識が生じる環境である実験 2 において，対話相手の違いによってどのようにユーザ発話における言語的特徴が表れるか F 検定及び t 検定で分析した．表れた言語的特徴は実験 1 とほぼ同じであり，

表 4: 実験 1(認識率 100%) における言語的特徴

対話状況	O	O+DT	P	P+DT	W	W+DT
タスク数	24	24	24	24	24	24
発話数	236	237	227	227	227	229
平均発話数	9.83	9.88	9.46	9.46	9.46	9.54
平均形態素数	4.98	4.56	4.58	4.53	4.58	4.31
平均間投詞数	9.25	9.17	3.83	4.83	5	4.91
平均情報数	1.99	1.97	2.07	1.98	2.07	1.99
平均新情報数	1.97	1.91	2.03	1.96	1.96	1.95
動詞省略数	71	70	86	85	88	89

表 5: 実験 2(認識率 80%) における言語的特徴

対話状況	O	O+DT	P	P+DT	W	W+DT
タスク数	24	24	24	24	24	24
発話数	245	247	232	243	249	235
平均発話数	10.20	10.29	9.67	10.13	10.38	9.79
平均形態素数	6.32	6.03	5.27	5.29	4.84	5.24
平均間投詞数	3.83	3.54	1.67	2.38	1.17	1.63
平均情報数	1.96	1.93	1.96	1.84	1.85	1.98
平均新情報数	1.95	1.89	1.94	1.82	1.81	1.93
動詞省略数	38	41	79	77	93	81

応答が音声合成か自然発声かの違いによって有意な差が表れ，合成音声では間投詞 (O-P:p<0.1, O-W:p<0.01)，平均形態素数 (O-P:p<0.5, O-W:p<0.01) が減少し，動詞省略が増加する (p<0.01) 結果となった．

最後に，音声認識率の違いによって同一の対話相手間でどのような言語的特徴が表れるか分析を行った．有意差のある特徴としては，音声認識率 80%の合成音声の場合，100%の場合に比べて平均形態素数が減少する．また，音声認識率 80%の自然音声の場合，100%の場合と比べて，動詞省略数が減少し，平均形態素数が増加する．これらのことから，音声合成の場合は対話相手が聞き間違える (音声認識誤り) と，より単純な発話に変化していくことが分かる．また，自然音声の場合，対話相手の音声認識率が低いと，より協調的な丁寧な発話になることが分かった．

3.3 対話相手の違いによるユーザ発話の音響的特徴

表 6 には音声認識率 100%(実験 1) における，表 7 には音声認識率 80%(実験 2) における新情報発話を対象とした対話相手と対話状況の違いによる音響的特徴の結果をしめす．

対話相手の違いによってどのように音響的特徴が表れるか F 検定及び t 検定で分析した結果，言語的特徴の場合と同様に，応答が合成音声か自然音声かの違いによって有意な差が見られた．

実験 1 と実験 2 で共通した音響的特徴としては，合成音声の場合，発話開始時間が自然音声の場合に比べて遅くなる傾向があった (p<0.01)．この特徴の要因としては，三つの可能性が考えられる．一つ目は，発話内容を生成する時間が自然音声に比べてかかるためで，これは前述の言語的特徴と組み合わせて考えると，自然音声の場合に比べて理解しやすい文を生成しようとする意識の表れだと思われる．二つ目は，音声合成の韻律情報を貧弱さにより，発話の終わりが予測しにくく，ポーズが開くの

を待ってから発話を開始するために自然音声に比べて遅くなるのではないかと考えられる．最後の要因としては，対話の発話権の移動が自然音声に比べて非常に明確になりやすいため，発話権の確保や発話の意思明示を急がなくても対話が成り立つという意識が働くためではないかと考えられる．

また，誤認識の生じる実験 2 でのみ表れた音響的特徴としては，合成音声の場合，自然音声に比べてパワー平均 (p<0.05) が大きくなった．さらに，パワー最大が大きくなる傾向が見られた．この特徴は対話相手「機械」の方が対話相手「人間・音声合成」より傾向が顕著であった．

表 6: 実験 1(認識率 100%) における音響的特徴

対話状況	O	O+DT	P	P+DT	W	W+DT
発話開始時間 (sec)	0.41	0.31	0.66	0.78	0.63	0.65
平均発話時間 (1) (sec)	1.90	1.87	1.75	1.77	1.86	1.92
平均発話時間 (2) (sec)	1.37	1.43	1.28	1.37	1.38	1.41
平均発話速度 (1) (mora/sec)	7.53	7.52	7.74	7.63	7.64	7.22
平均発話速度 (2) (mora/sec)	10.50	9.65	10.69	9.85	10.26	9.64
パワー平均 (RMS)	606	830	729	919	680	909
パワー最大 (RMS)	2676	3548	2963	3544	2701	3475
ピッチ平均	71.0	78.4	71.9	80.2	72.7	77.9
ピッチ最小	57.2	59.2	58.6	60.5	59.0	58.0
ピッチ最大	245.1	281.3	277.0	251.6	301.1	260.0
ピッチ分散	15.3	17.5	15.7	17.7	17.5	17.4

表 7: 実験 2(認識率 80%) における音響的特徴

対話状況	O	O+DT	P	P+DT	W	W+DT
発話開始時間 (sec)	0.42	0.49	0.75	0.68	0.69	0.63
平均発話時間 (1) (sec)	2.08	1.96	1.74	1.78	1.51	1.75
平均発話時間 (2) (sec)	1.43	1.47	1.26	1.32	1.16	1.33
平均発話速度 (1) (mora/sec)	7.29	7.29	7.59	7.53	7.92	7.49
平均発話速度 (2) (mora/sec)	10.31	9.67	10.24	9.87	10.37	9.66
パワー平均 (RMS)	721	904	1032	1146	1023	1281
パワー最大 (RMS)	3255	3941	3880	4590	3902	5397
ピッチ平均	70.6	74.7	76.9	79.7	77.0	80.0
ピッチ最小	55.7	57.6	57.1	58.6	59.5	57.3
ピッチ最大	299.7	256.4	266.3	231.9	268.8	256.2
ピッチ分散	20.1	20.3	19.1	20.1	19.1	20.5

そして，音声認識率の違いによる同一の対話相手間での比較による音響的特徴の差の分析であるが，音声合成を用いた対話相手の場合で音声認識誤りが発生するような状況ではパワー最大 (P:p<0.05, W:p<0.01)，平均 (P:p<0.1, W:p<0.01) が大きくなった．つまり，音声認識率が非常に高い場合には対話相手の音声が自然音声でも音声合成でも声の大きさには変化がない．しかしながら，音声認識率が悪くなると，自然音声に対しては声の大きさの変化がないにもかかわらず，音声合成に対しては声が大きくなる傾向があることが分かった．さらに，対話相手「人間」で音声認識率が下がるとピッチの分散が広がり (O:p<0.01, P:p<0.01)，抑揚が強くなる傾向があることがわかった．この特徴は有意差は「人間」の場合のみであったが，傾向としては全ての対話相手にいえる．

3.4 認知的負担の有無によるユーザ発話の言語的特徴

認知的負担（本研究では運転タスク）の有無によって表れた言語的特徴はまったくなかった．これは以前の我々の研究の結果と同様な内容となった．文献 [16] 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．本実験の結果でも，この仮説を支持するような結果となった．

3.5 認知的負担の有無によるユーザ発話の音響的特徴

認知的負担の有無によって表れる音響的特徴に関してであるが，認識率が高い環境である実験 1 では，全ての対話相手に対してパワー平均 (p<0.01)，最大値 (p<0.1) が上昇し，ピッチ平均 (p<0.1) が上昇する．対話相手「人間」に対しては，ピッチの分散も広がる (p<0.1) が，対話相手「機械」に対してはそのような傾向はなかった．また傾向として，対話相手「人間」の場合，認知的負担があると発話中に含まれるポーズ（無音部分）が減少し、一文字一文字の発声が伸びることによって話速が落ちる (p<0.2) 傾向が見られた．

認識率が低い環境の実験 2 では，実験 1 とほぼ同様にパワー平均 (p<0.1)，最大値 (p<0.1) が上昇するが対話相手「人間・合成音声」の場合だけ有意差が表れなかった．なお，ピッチに関する特徴は実験 2 ではまったく表れなかった．また傾向として，実験 1 と同様に対話相手「人間」の場合，認知的負担があると発話中に含まれるポーズ（無音部分）が減少し、一文字一文字の発声が伸びることによって話速が落ちる傾向が見られた．それと同時に対話相手「機械」の場合は，ポーズそれほど減少せずに話速が落ちる傾向が見られた (p<0.15)．

3.6 訂正発話と直前ユーザ発話の比較による言語的・音響的特徴

本研究では，実験 2 において各タスク平均 3〜4 の訂正発話が生じる．そこで、参考的な分析として個人差を考慮せず，対話相手が実際に誤認識したユーザの発話とその誤認識に対する訂正発話をそれぞれ比較グループとして，言語的・音響的特徴を簡単に分析してみた．

有意差があったものとしては，対話相手「人間」の場合，訂正発話では一発話に含まれる平均情報数が減少し (p<0.05)，発話速度が上昇する (p<0.05)．ただ，話速が上昇するといっても，一文字一文字の速度自体は変化しておらず，発話中に含まれるポーズ（無音）が無くなることによって話速が上がっている．また，対話相手「人間・

自然音声」の場合では，発話開始時間が減少し (p<0.01)，ピッチの標準偏差も広がった (p<0.1)．なお，発話開始時間の減少は傾向としては全ての対話相手に対して見られた．

3.7 目的地検索・設定タスクの特徴

本研究では，音声インターフェイスとして近年，最も期待されているカーナビゲーションシステムの目的地検索・設定を対話タスクとして用いた．さらに本実験では，あらかじめ全ての目的地や検索条件を暗記させ，全ての目的地を自発的に発話できる状態にした．この状況での目的地検索・設定タスクの発話の特徴として，一つ一つの目的地に関する情報を基本的に一発話で全て入力しようとする傾向が非常に強く見られた．このように本タスクのような伝えるべき情報に曖昧性が少なく，伝える情報量自体もそれほどではない場合，一度に全ての情報を伝えようとする傾向があることが分かった．

4 まとめ

音声認識率・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なった．比較する対話状況としては，音声認識率の高低，対話相手として人間，人間 (合成音声応答) または機械 (音声対話システム) の 3 種類，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見られた．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、応答音声が自然音声か合成音声かの比較と音声認識率の違いによる比較において、幾つかの言語的・音響的な特徴の差異が明らかになった．更に，訂正・再入力発話の分析では，訂正・再入力発話には通常の発話と比べて，幾つかの言語的・音響的特徴な違いがみられた．

今後は，本実験で明らかになった合成音声応答を用いると人間同士の対話とは異なった発話特徴になってしまう原因が，音声の品質が低いことが直接の原因であるのか，対話のリズム (応答タイミング，発話速度の変化，抑揚など）が人間同士の対話と異なっていることが原因であるのかなどを調査・分析し，できるだけ人間同士の対話特徴に近づくような対話ができる音声対話システム構築をしていきたいと考えている． また，逆に今回の分析において有意差が生じたいくつかの言語的・音響的特徴が，音声認識精度や言語理解精度の向上において，優位に働いたり有効に活用できる特徴であれば，その特徴を利用した音声対話システムなどの開発も行っていきたい．

参考文献

[1] K.Itou, K.Fujimura, N.Kawaguchi, K.Takeda, and F.Itakura, Dialogue characteristics in different communication modes, Special Workshop in Maui Lectures by Masters in Speech Processing, 2004.

[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 (2001.3).

[4] 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: 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, 情報処理学会論文誌, Vol.43, No.7, pp.2118-2129 (2002).

[5] 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).

[6] 村上仁一, 嵯峨山茂樹: 自由発話音声認識における音響的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 (1991).

[7] 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).

[8] 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).

[9] 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).

[10] 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 (1996).

[11] 甲斐充彦, 伊藤克亘: 対話システムにおける音声認識, 情処理学会研究会資料, SLP-33-2 (2000).

[12] http://www-nrd.nhtsa.dot.gov/departments/nrd-13/DriverDistraction.html

[13] http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm

[14] D.J.Litman, J.B.Hirschberg and M.Swerts: Predicting automatic speech recognition performance using prosodic cues, In Proc. of the 6th Applied Natural Language Processing Conference, ANLP-NAACL00, pp.218-225 (2000).

[15] D.J.Litman, M.A.Walker and M.J.Kearns: Automatic detection of poor speech recognition at the dialogue level, In Proc. of the 37th Annual Meeting of the Association of Computational Linguistics, ACL99, pp.309-316 (1999).

[16] 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).

[17] T.Schaaf and T.Kemp: Confidence measures for spontaneous speech recognition system, Proc. ICASSP, pp.875-878 (1997).