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

2005－ＳＬＰ－59（12）
2005／12／21

対話相手の音声の品質を考慮した対話状況での言語的・音響的特徴の分析および様々な観点からの考察

山田 真也　伊藤 敏彦　荒木 健治

北海道大学大学院 情報科学研究科 〒060-0814 北海道札幌市北区北 14 条西 9 丁目

E-mail: {yamaya, t-itoh, araki}@media.eng.hokudai.ac.jp

あらまし　人間同士や人間と機械との対話において，対話相手の違い，対話相手の音声認識率，対話相手の応答音声の品質の違いが発話に与える影響について，対話収集実験から得られた音声対話データを用いて分析を行う．対話タスクはカーナビゲーションシステム上での音声インターフェースを用いた目的地検索・設定タスクであり，そのタスクにおける様々な状況下でのユーザの発話に表れる言語的・音響的な特徴の比較を行った．さらに，各対話状況における対話相手や被験者自身の発話に関する主観的評価のためのアンケート調査を行った．また，機械(対話システム)に対する潜在的意識（先入観）が発話に与える影響を調査するため，アンケートの評価値によりグループ分けを行い，各グループで発話の特徴を分析した．調査の結果，応答音声の品質や発話のリズムによってユーザの発話が変化し，音声の品質は対話相手に対する印象に影響を与えることが分かった．また，機械に対する潜在的意識によりユーザの発話が変化することが確認された．

キーワード　対話分析，対話状況，音声の品質，音声認識率，言語的・音響的特徴

Analysis of Linguistic and Acoustic Features Depending on Different Situations and Discussions from various viewpoints
－The Experiments Considering Voice Quality－

Shinya YAMADA　Toshihiko ITOH　and　Kenji ARAKI

Department of Information Science and Technology,

Hokkaido University, Hokkaido, 060-0814 Japan

E-mail: {yamaya, t-itoh, araki}@media.eng.hokudai.ac.jp

Abstract　This paper presents our analyses of human-human and human-machine interactions and the characteristic differences of linguistic and acoustic features observed in different spoken dialogue situations and with different dialogue partners. The linguistic and acoustic features of the user's speech to a spoken dialogue system and a human operator are compared in several goal setting and destination database searching tasks for a car navigation system. It is said that it is not clear enough whether different dialogue situations, different dialogue partners and different speech recognition rate cause any differences of linguistic or acoustic features on one's utterances in a speech interface system. Additionally, research about influence of voice quality is not enough either. We prepared two patterns where we have two dialogue partners with different speech recognition rate (100% and about 80%). We also prepared three voice patterns (natural voice, synthetic voice and recorded voice). We analyzed the characteristic differences of user utterances caused by different dialogue situations and by different dialogue partners. Furthermore, we analyzed the differences based on user's impression of dialogue partner and a spoken dialogue system using questionnaires. As a result, we found that voice quality and rhythm of partner's utterance affect user's utterances, and an impression of dialogue partner is subject to voice quality. Additionally, we also found that a prejudice toward machines affects user's utterances.

Keyword　Speech Analysis, Dialogue Situation, Voice Quality, Speech Recognition Rate, Linguistic and Acoustic Features

1. はじめに

近年，音声認識技術や音声言語処理技術，コンピュータ性能の向上により，音声インターフェースやタスク指向型音声対話システムが注目を集めている．カーナビゲーションシステムやロボットとの音声対話などといった利用がすでになされており，今後更なる応用が予想される．

しかしながら，実用的な音声対話システムを開発するに当たって，既存の音声認識技術には数多くの解決すべき問題が存在する．その一つとして，音声インターフェースに用いられる音声認識システムの信頼性，頑健性の向上が上げられる．音声認識システムの認識性能は発話様式によって大きく影響されることが指摘されている[1]．そのため，これまでは読み上げ文やニュース音声[2]といったものに対する研究が中心であったが，近年では講演音声[3]や講義音声[4]といった自

然発話の認識が注目されるようになってきている．

これらの研究は，それぞれの発話様式特有の言語的・音響的特徴をとらえた上で、それらの特徴に適した手法などを利用することによりシステム性能の向上を目指している．発話様式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，その要因を明らかにしなければならない．しかしながら，発話に影響を与える要因に関する研究は十分であるとはいえない[5]．

我々[6]や itou[7]らは以前に，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を与えるかについて分析を行っている．先行研究[6]では，対話相手の違い，応答能力の違い，音声認識率の違い，心理的負荷の有無を考慮し，言語的・音響的特徴を比較している．その結果，音声認識率や対話相手の応答能力の違いが発話に大きく影響を与えることが明らかになった．先行研究で取り扱った応答能力には，音声の品質や対話のテンポといった複数の要素が含まれているため，どの要素が発話に大きく影響を与えたかは明らかになっていない．そこで本研究では，WOZ 法を用いた擬似音声対話システムを使用し，応答に録音音声を用いた場合の対話の収録を行った．収録された対話データを分析して先行研究と比較を行うことにより，応答の音声の品質やリズムが言語的・音響的な特徴にどのような影響を与えるかをより詳細に調査する．

また，先行研究と今回の研究では被験者に，対話状況ごとの対話相手や自らの発話に関するアンケート，さらに実験開始前に持っていた機械(対話システム)に対するイメージ(先入観)のアンケートを行っている．その結果をもとにして，主観評価と発話の関係を明らかにするための多角的な分析を行う．

2. 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，様々な対話状況における音声対話を収録する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，目的地のランドマークを検索・設定するもので，以前に我々が行った実験方法[6]と同じである．

本研究および先行研究で対象とした対話状況を表 1 に示す．先行研究は，対話相手の音声認識率が 100%である場合(実験 1)と，ノイズ等の影響があるという想定で音声認識率が約 80%である場合(実験 2)の 2 条件で，自然音声および合成音声を用いて対話収録を行った．本研究では，新たな対話状況として，対話相手の応答に録音音声を使用した場合(実験 3)を追加している．実験 1，実験 2 を用いた調査により運転タスクは言語的特徴に影響を与えないことが分かった．そこで実験 3 では，実際のインターフェースの利用環境に近い運転タスクがある場合のみを対象とし，音声認識率は 100%の場合と約 80%の場合の 2 条件で対話実験を行った．対話音声の収録を行った被験者は大学(院)生 12 名で，音声対話システムに関する知識は全くない．

表1：用意した対話状況

対話相手		人間			機械	
応答音声		自然	合成	録音	合成	録音
運転タスク (DT)	無	O	PS	-	WS	-
	有	O +DT	PS +DT	PR +DT	WS +DT	WR +DT

：認識率 100%(実験 1)　：認識率 100%(実験 3)
：認識率約 80%(実験 2)　：認識率約 80%(実験 3)

3. 実験結果

本節では対話音声収録実験で得られた音声データに対しての言語的・音響的特徴の統計量および検定結果を示す．音声認識率が低い場合には誤認識が発生するため，否定語や訂正発話により発話数の増加や言語的・音響的特徴に影響が表れる．そのため，比較を行う発話は訂正発話，否定発話を除き，新情報を含むものを対象とすることで，音声認識率以外の条件を等しくした．

3.1. 先行研究[6]で得られた言語的・音響的特徴の分析結果

本研究は先行研究との比較を行い，応答能力の要素に関する新たな分析結果を得るものである．ここでは先行研究で得られた言語的・音響的特徴の調査結果について述べる．調査によって明らかとなった特徴を以下に示す．

音声認識率が低い場合の変化
- 自然音声の場合にピッチ標準偏差の増加
- 自然音声の場合に動詞省略の減少
- 合成音声の場合に平均情報数の減少
- 合成音声の場合にパワーの増加

対話相手の違いによる変化
- 合成音声の場合に間投詞数の減少
- 合成音声・高認識率の場合に付加情報数の増加
- 合成音声の場合に発話開始時間の増加

心理的負荷がある場合の変化
- パワーの増加
- 高認識率の場合にピッチの増加
- 発話速度の減少

3.2. 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

まずは実験 3 について，ユーザの対話相手の音声認識率に着目し，その違いによりユーザ発話がどのように変化するか調査を行い，先行研究で得られた結果との違いについて考察する．調査には t 検定を用い，各特徴に対して行った．

表 2に実験 3 の各対話状況における新情報発話の言語的特徴を，表 3にはその音響的特徴を示す．言語的特徴に関しては，音声認識率による比較で有意差が表

れるものはなかった．先行研究では音声認識率の低下に伴い，合成音声の場合に平均情報数が減少している．本研究では，平均情報数の比較で有意な差はないものの，音声認識率が低い場合に減少する傾向が見られ，先行研究の結果に準ずるものとなった．続いて音響的特徴については，「人間」，「機械」の両者で，音声認識率の低い場合にパワー平均値が増加する傾向（$p<0.1$）が表れた．先行研究では応答が合成音声である場合に，音声認識率が低くなるとパワーが増加した．つまり，録音音声を用いた場合の言語的・音響的特徴は，合成音声を用いた場合と同様の傾向を示すことが分かる．

表2：実験3における言語的特徴

対話状況	PR+DT		WR+DT	
音声認識率	約80%	100%	約80%	100%
タスク数	24	24	24	24
総発話数	253	238	257	239
平均発話数/タスク	10.54	9.92	10.71	9.96
分割発話数/タスク	2.71	2.46	3.00	2.08
形態素数/タスク	3.34	3.53	3.27	3.68
間投詞数/タスク	1.25	1.96	1.21	2.46
情報数/発話	1.75	1.83	1.73	1.87
新情報数/発話	1.72	1.80	1.66	1.86
総動詞省略数	128	123	140	120
総付加情報数	26	33	29	32

表3：実験3における音響的特徴

対話状況	PR+DT		WR+DT	
音声認識率	約80%	100%	約80%	100%
発話開始時間(sec)	0.66	0.84	0.71	0.55
平均発話時間(1)(sec)	1.30	1.42	1.38	1.49
平均発話時間(2)(sec)	1.02	1.05	1.00	1.09
平均発話速度(1)(mora/sec)	7.62	7.79	7.39	7.54
平均発話速度(2)(mora/sec)	9.77	10.00	9.83	9.99
パワー平均値(RMS)	1471	1161	1543	1221
パワー最大値(RMS)	4018	3012	4051	3382
ピッチ平均値	140.6	136.0	140.0	134.7
ピッチ最小値	61.3	61.6	63.1	64.0
ピッチ最大値	299.8	308.4	268.9	285.3
ピッチ標準偏差	22.1	22.0	21.4	19.4

3.3. 対話相手の違いによるユーザ発話の言語的・音響的特徴

実験3に対して，対話相手が「人間」である場合と「機械」である場合とで，ユーザ発話にどのような影響が表れるかをt検定により調査した．

言語的特徴，音響的特徴に対してt検定を行った結果，全ての特徴で有意な差は表れなかった．先行研究では，機械であるかどうかという事実によってではなく，合成音声を用いた場合の，音声の品質の悪さや対話のリズムの悪さといった要素が，機械的な印象を与えたことにより，ユーザの発話が影響を受けたとしている．今回得られた結果では，応答能力が同じである場合には差が表れておらず，先行研究の結果を支持している．これらのことは，人と話すような自然な対話を行うためには，対話相手が人間であるかどうかではなく，人間らしさや人間レベルの対話能力を感じさせることが重要である可能性を示している．

続いて実験1，実験2および実験3を相互に比較することで，応答に用いた音声の違い（自然音声，合成音声，録音音声）がユーザ発話に与える影響について分析を行った．先行研究では自然音声と合成音声の比較を行っており，本研究では新たに自然音声と録音音声，合成音声と録音音声の比較を行うことで，音声の品質，リズムや抑揚といった要素が発話に与える影響について調査する．

応答に自然音声を用いた場合と録音音声を用いた場合の比較を行ったところ，録音音声を用いた場合に，言語的特徴では間投詞数($p<0.01$)が減る傾向があり，音響的特徴では発話開始時間($p<0.05$)が長くなった．

間投詞や発話開始時間は話者交代や発話権確保に関係のある特徴である．録音音声と自然音声はどちらも人間の声であり，音声の品質は同じであるにもかかわらず，発話が単純なものに変化して相手の発話の終了を待つ傾向が表れている．これは，録音音声が同じ言い回ししかせず，対話のリズムが単調で発話ターンが明確であるという印象を与えるために，ユーザが発話権確保の必要がないと感じる影響であると考えられる．また，ユーザの割り込みは許すが，対話相手側から割り込みを積極的に行っておらず，そのことも，発話権を確保する必要がないと感じさせる要因と考えられる．

また，録音音声と自然音声の比較，録音音声と合成音声の比較の両方で，録音音声の場合に言語的特徴では動詞省略発話数($p<0.01$)が増加し，形態素数($p<0.1$)が減少した．音響的特徴については，F0平均値($p<0.1$)が大きくなった．これは合成音声を使用した対話システムの場合より，ユーザの発話がさらに単純なものに変化したことを示している．この理由として，人間と同等の音声品質でありながら対話のテンポが非常に悪いというアンバランスさが，ユーザの意識に違和感をもたらして，より機械らしさを感じたためではないかと考えられる．あるいは，文単位で応答を行う録音音声自体が，リズムが一定で決まった応答しかできず，柔軟性がないという機械的な印象をユーザに与えたために，ユーザの発話が機械的なものに変化したとも考えられる．これらのことから，音声の品質をよくするだけでは自然な対話を実現することはできず，対話のリズムを向上させることが重要であることが分かる．

なお，F0平均値について，実験3のユーザの中でF0平均値が他のユーザと大きく異なるものがあったため，そのデータを除いて録音音声と自然音声，録音音声と合成音声の比較を行ったところ，録音音声-自然音声間でのみ有意差が表れ，録音音声のF0平均値($p<0.05$)が大きいという結果が表れた．先行研究では合成音声のF0平均値が自然音声のものよりも大きい傾向が見られており，機械的な発話に変化するとF0が増加する傾向があることが分かる．

3.4. 言語的・音響的特徴の傾向によるグループ分けによる分析

先行研究および本研究から，いくつかの対話状況でユーザ発話の言語的・音響的特徴に有効な変化が観測された．しかしながら，これまでに得られた結果があまりに特異な発話を行うユーザの影響による可能性や，その影響により本来あるべき特徴が消えてしまった恐

れがある．そのため，平均値から大きく外れた特異点となるユーザのデータを除いたものでグループを構成して各特徴の分析を行った．以下では紙面の関係上，表 4に示される有効な結果が得られたグループについてのみ記述する．

表4：各実験での発話の傾向によるグループ

グループ	詳細
A (実験 1)	間投詞数についての特異点となるユーザ 2 名を除外 (全対話状況で使用数多：2 名)
B (実験 1)	動詞省略数についての特異点となるユーザ 4 名を除外 (全対話状況で省略数少：3 名，全て省略：1 名)

グループ A では対話相手が「機械」である場合に付加情報数(O-WS: $p<0.05$)が増加する傾向がある．グループ B では対話相手の応答が合成音声である場合に付加情報数(O-PS: $p<0.1$, O-WS, O+DT-WS+DT: $p<0.05$)が増加する．付加情報とは，目的地の検索・設定を行う際に必ずしも必要ではない情報のことである．先行研究においては，合成音声で付加情報数が増加するという傾向があり，対話相手が機械であるとその傾向は強い．この結果はグループ A および B での結果と一致しており，人は対話相手を機械であると感じると，人間相手では省略するような情報も加えて，情報量的にはより詳細に話そうとする傾向があることを示していると考えられる．さらにグループ A および B では，ポーズを含んだ発話時間が，「人間・合成音声」の場合に減少する傾向が見られた．音響的特徴の傾向によりグループ分けをおこない分析したところ，全発話データを用いた分析で得られた言語的特徴および音響的特徴以外の新たな傾向が見られることはなかった．

3.5. アンケートによる主観評価と言語的・音響的特徴の関連

ユーザの主観評価について検討するために，対話相手に対する印象，自分自身の発話に関する評価，機械(音声対話システム)に対する印象に関するアンケートを行った．ここではアンケートをもとにして行った，ユーザ発話の言語的・音響的特徴に関する分析結果について述べる．

3.5.1. 各対話状況でのアンケート傾向

先行研究および本研究では，対話相手の応答能力の性能がユーザの発話に影響を与えることが明らかとなった．対話相手の音声の品質や対話のリズムは，ユーザの対話相手に対する印象や評価に大きく関わっていると推測され，この印象や評価がユーザの発話に影響を与えた可能性が考えられる．そこで，各対話状況において対話相手の能力や印象，ユーザ自身の発話の主観評価を，アンケートを用いて行い，言語的・音響的特徴との関係を明らかにする．評価は 1 から 7 までの 7 段階で行った．調査項目は対話相手に関するものとして，「対話テンポのよさ」，「音声認識能力」，「意図認識能力」，「応答の正確さ」，「話しかけ易さ」，「目的地の設定し易さ」，「音声の聞き取り易さ」の 7 項目であり，ユーザ発話の評価として「次発話のしやすさ」，「自分の発話の速さ」，「自分の声の大きさ」の 3 項目である．

最初にアンケート結果について，実験ごとに対話状況での比較を行い，対話相手の違いや心理的負荷の有無，自然音声と合成音声との差が主観評価にどのような影響を与えるかを調査した．また，実験 1 および実験 2 と実験 3 を比較し，応答音声の品質やリズムの違いによる主観評価への影響の分析を行った．その結果，以下のようなことが明らかになった．

対話相手の違い，心理的負荷の有無でユーザの主観評価の比較を行ったところ有意差が表れた項目はなく，これらの要素はユーザの主観評価には大きく影響を与えないことが分かった．

表5：合成音声の評価結果(自然音声との比較)

	実験 1	実験 2
対話テンポのよさ	低下($p<0.05$)	低下($p<0.1$: WS のみ)
話しかけやすさ	低下($p<0.05$)	低下($p<0.05$)
音声の聞き取りやすさ	低下($p<0.1$)	低下($p<0.05$: PS+DT,WS+DT)

表6：合成音声の評価結果(録音音声との比較)

	実験 1-3，実験 2-3
対話テンポのよさ	低下($p<0.05$:PS の認識率 80%, $p<0.01$:WS の認識率 100%)
音声の聞き取りやすさ	低下($p<0.1$)

続いて対話相手の音声の品質や対話のリズムといった観点で比較を行った結果について述べる．録音音声は音声の品質は良いが対話のリズムが悪く，合成音声はそのどちらとも悪いものである．表 5は自然音声と合成音声の比較，表 6は合成音声と録音音声の比較を行った結果を示す．自然音声と合成音声の比較，合成音声と録音音声の比較では主に「対話のテンポのよさ」，「話しかけやすさ」，「音声の聞き取りやすさ」に差が表れた．実験 2 の「対話テンポのよさ」については「機械・合成音声」でのみ有意差が得られたが，合成音声を用いた全ての場合で評価が低下する傾向があった．合成音声と録音音声の比較では，合成音声を用いた場合に「対話のテンポ」や「音声の聞き取りやすさ」が低下した．全く同じ内容の発話で，ほぼ同程度の発話速度である合成音声と録音音声の比較で有意差があることから，合成音声の品質が影響を与えたと考えられる．自然音声と合成音声を比較すると，合成音声では「話しかけやすさ」が低下している．つまり「話しかけやすさ」は対話のテンポや抑揚といった対話リズムの要因によって変化する可能性がある．

続いて実験 1 と実験 2 の同じ対話相手間での比較，実験 3 の音声認識率での比較により，音声認識率がユーザの意識に与える影響の調査を行った．

表7：音声認識率が低い場合の評価結果

音声認識能力	低下(p<0.05:WS 以外)
意図理解能力	低下(p<0.05:全て)
応答の正確さ	低下(p<0.1:O 以外)
対話テンポのよさ	低下(p<0.05:O,O+DT, PS+DT,PR+DT,WR+DT)
設定のしやすさ	低下(p<0.1:O+DT, PR+DT,WR+DT)

表 7は音声認識率での比較の結果を示す．「音声認識能力」，「意図理解能力」，「応答の正確さ」はほぼ全ての対話状況で，音声認識率が悪い場合に評価が低下した．有意差が表れなかった状況についても低下の傾向は表れており，これらの項目では，音声認識率が評価に直接影響を与えると考えられる．「設定のしやすさ」は自然音声，録音音声の場合，音声認識率の低下に伴って評価が下がった．低認識率の場合，「設定のしやすさ」の評価値は 3 つの音声間で差はなく，認識率のよい場合には自然音声，録音音声の評価値が録音音声よりも高くなっている．「対話テンポのよさ」は，自然音声と録音音声で合成音声の場合より評価が高く，音声認識率の低下によって，主に自然音声と録音音声で評価が低下する傾向が見られる．これは「設定のしやすさ」と「対話テンポのよさ」の変化の傾向が同じであるということを示している．つまり，設定のしやすさと対話テンポの印象は関連していることが考えられ，対話テンポの印象をよくすることで，ユーザに使いやすさ感じさせることができる可能性がある．

被験者自身の発話に関しては，全ての対話状況で，音声認識率が低い場合には，発話の声を大きく，発話速度も遅くしたという意識傾向が表れた．ユーザは音声認識能力の低い相手に対して，意識的に聞き取りやすいように大きな声でゆっくりと話していることが分かる．

3.5.2. ユーザが機械(対話システム)に対して持っている意識についてのアンケート

これまでの分析結果から，ユーザが機械に対して予め持っている印象(先入観)が，発話に影響を与える可能性がある．そこでユーザが機械に対して持っている先入観を調査するためのアンケートを取り，ユーザの先入観が発話の言語的・音響的特徴にどのような影響を与えるか分析した．アンケートは機械の「知能レベル」，機械の「音声認識能力」，機械の対話の「テンポ」のよさの 3 項目について行った．評価は 1 から 7 までの 7 段階で，人間の能力を 5 として評価している．アンケート結果をもとにして実験 1，実験 2，実験 3 の全ユーザに対し，評価値ごとにグループ分けを行った．グループ分けは評価値の 1-3 を能力が低のグループ，4-5 を能力が中のグループ，6-7 を能力が高のグループとして分類した．

表 8は先行研究および本研究の全ユーザに対してグループ分けを行ったときのユーザ数の分布を示す．括弧の中は左から実験 1，実験 2，実験 3 のユーザ数を表す．ユーザ数の分布から，一般的に機械の能力を人間の能力以下であると評価している人が多いことが分かる．また，知能レベルや音声認識能力といった言語理解能力に関するものよりも，対話テンポという対話を行う能力に対する評価が低く，対話システムに対する対話リズムの印象が悪いことが伺える．

表8：各グループの構成人数

評価	低(1-3)	中(4,5)	高(6,7)
知能レベル	20(3,10,7)	13(7,1,5)	2(2,0,0)
音声認識能力	19(5,5,9)	12(6,5,1)	4(1,1,2)
対話テンポ	24(9,7,8)	8(2,4,2)	3(1,0,2)

次にユーザの意識が機械との対話にどのような影響を与えるのか調査するために，評価値によって分類したグループの「機械・合成音声」の発話データをまとめ，グループ間での比較を行った．比較は評価項目ごとに行った．なお，全ての評価項目で，能力を高と評価するグループは人数が少なく，そのため比較は能力を低と評価したグループと中と評価したグループでのみ行った．以下に「知能レベル」，「音声認識能力」，「対話テンポ」の 3 項目で，ユーザの意識が発話に与える影響について調査した結果を示す．

知能レベルを中と評価したグループと比較して，知能レベルを低と評価したグループでは，平均情報数(p<0.01)，平均新情報数(p<0.01)が減少し，分割発話(p<0.01)が増加した．また，同グループでは，動詞省略(p<0.01)が多くなることが分かった．これは，ユーザが相手の知的レベルに応じて話し方を変化させていることを示し，知的レベルが低いと思っている場合には，情報量を減らして動詞を省略することで，より簡潔に話すということである．知能レベルを低としたグループでは，機械に対する発話で発話時間(p<0.1)は短くなり，有意差はないものの発話速度も遅くなる傾向が見られた．つまり，ユーザは知能レベルが低いと推測している相手に対して，理解しやすいようにゆっくり話す傾向があることが分かる．

続いて音声認識能力についてグループ分けを行い，機械に対する発話への影響を調査する．音声認識能力を低いと評価したグループでは，発話速度が低下する傾向が見られた．詳しく調査を行うために実験 1 と実験 2 でそれぞれグループ分けを行い，個別に比較を行ったところ，実験 2 の音声認識能力を低と評価したグループのみで，発話速度(p<0.1)が低下することが分かった．つまり，実際に機械の音声認識能力が低い場合に発話速度を低下させるということである．これはゆっくり発話することで認識させようとするためであると考えられ，機械の音声認識能力が低いと考えている人には，その傾向が強いものと推測される．

最後に対話テンポの評価が機械との対話に与える影響について調査する．対話のテンポの悪いと評価したグループでは分割発話(p<0.1)が減少し，動詞の省略(p<0.1)が増える傾向がある．機械のテンポが悪いと想像している場合，少ない発話回数でタスクを達成しようとするため，発話の分割を行わず，動詞を省略した単調な発話になるのではないかと考えられる．

続いて各実験で同様に 3 段階の分類を行い，それぞ

れの実験において5名以上で構成されたものを，新たにグループとする．このグループを対象として，対話状況がユーザ発話の言語的・音響的特徴にどのような影響を与えるか調査を行った．表 9は各実験で行った分類により得られたグループを示す．

表9：各実験のグループ

評価	低(1-3)	中(4,5)
知能レベル	実験 2,3	実験 1,3
音声認識能力	実験 1,2,3	実験 1,2
対話テンポ	実験 1,2,3	

まず実験 1 の各グループで対話状況ごとに比較を行い，言語的・音響的特徴を調査した．機械への先入観の影響を受けない(全てのグループで出現)特徴としては，対話相手の音声が自然音声の場合に発話中の間投詞数が多くなり，運転によってパワー平均値が増加する傾向がある．これらの傾向は先行研究で得られた言語的・音響的特徴と一致している．個々のグループで見ると，音声認識能力を中と評価したグループは，応答の音声が自然音声の場合と合成音声の場合の比較ではパワー平均値が変化しなかった．他のグループでは，合成音声の場合にパワーが増加する傾向があるが，先行研究では傾向が表れておらず，このグループの影響によって全体では有意な差とならなかったと考えられる．パワーが変化しないのは，相手の音声認識能力が悪くはないという評価から，大きな声を出なくても認識できるという意識の表れであると考えられる．同グループにおいては，全ての対話状況で心理的負荷による F0 平均値の変化も見られなかった．実験 1 の他のグループにおいては F0 平均値が心理的負荷により増加する傾向を示しており，音声認識能力を中としたグループでは，声の大きさとともに声の高さも，心理的な負荷により変化しなくなっていることが分かる．

続いて実験 2 の各グループで対話状況ごとに比較を行った．機械に対する先入観によらず，対話相手の音声が自然音声であるときに動詞省略発話が少なくなった．また，合成音声の場合には発話開始時間が長くなる傾向が表れた．これらの傾向は先行研究で得られた結果と同じである．音声認識能力を低と評価したグループでは間投詞数，平均形態素数が，応答の音声の品質やリズムによる影響をあまり受けなかった．実験 2 の他のグループおよび先行研究では，これらの特徴が応答の音声の品質やリズムの影響を受けており，音声認識率を低と評価したグループは，他のグループと異なる特徴を持っているといえる．音声認識能力を中と評価したグループでは，対話相手が「機械・合成音声」の場合に，心理的負荷による発話速度の変化が見られなかった．他グループおよび先行研究では，全ての対話状況で，心理的負荷があると発話速度が遅くなる傾向があり，音声認識能力を中としたグループは特殊な発話を行っていることが分かる．

実験 3 においては，実験 3 の全てのグループで音声認識率が低下するとパワーが大きくなる傾向が見られ，実験 3 の全発話データを用いた結果と一致した．

4. まとめ

音声の品質や対話のリズムがユーザ発話に与える影響を調査するため，カーナビゲーションシステムの使用を想定した目的地検索・設定タスクを行い，発話の言語的・音響的特徴の分析を行った．各特徴間の関係を調査するために，対話状況による発話の特徴の変化に着目してグループ分けを行い，各特徴の比較を行った．また，アンケートを用いてユーザの主観評価の分析を行い，ユーザの意識が発話に与える影響について調査した．

応答音声の品質やリズムによる比較を行った結果，音声の品質はよいが，リズムが悪いようなバランスの悪い対話システムの場合，より機械らしい発話になることが分かった．このことから，自然な対話を行うためには音声の品質だけでは不十分であり，声の抑揚や対話のリズムといった要素も重要であることが分かった．また，アンケート調査では対話相手の音声の品質がよいと，評価値が高くなることから，音声の品質はユーザの印象に影響を与えることが示された．さらに，機械に対する先入観により，ユーザの発話が大きく変化することが分かった．

今後は，本研究で明らかとなった言語的・音響的特徴を用いると言語理解精度が向上するのであれば，その特徴を活かした音声対話システムの開発をいきたいと考えている．また，今回の実験では音声の品質によるユーザ発話への影響を明らかにしており，今後は今回明らかとならなかったリズム，抑揚，韻律やイントネーションがユーザ発話にどのような影響を与えるかを，より詳細に分析を行っていくことを検討している．

文　献

[1] 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.

[2] 加藤直人，浦谷則好，江原暉将，安藤彰男，“ニュース音声認識のための(n≧4)-gram を併用する言語モデル”，電子情報通信学会論文誌，Vol.J85，No.6，pp.967-975，2002.

[3] 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”情報処理学会資料，SLP-37-13，pp.73-78，2001.

[4] 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-II，No.11，pp.2473-2480，2000.

[5] 阿部匡伸，小特集-音質“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.

[6] 伊藤敏彦，山田真也，荒木健治，“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”情報処理学会研究報告，SLP-56，pp.101-106，2005.

[7] 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 Procession, 2004.