2006－SLP－61 (2)
2006／5／11

運転時における発話の音響的特徴を利用した対システム発話の推定

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

様々な対話状況で収録した対話音声の音響的特徴を用いた機械学習により対システム発話の推定を行い，その有効性を調査する．対話音声には，我々がすでに調査・分析を行ってきた 2 者対話実験で収録したものと，今回新たに追加した 3 者対話実験で収録したものを使用した．収録した対話の状況設定はカーナビゲーションシステムを使用することを想定した目的地検索・設定タスクである．2 者対話実験ではタスクの遂行役としての人間，または対話システとの対話で，3 者対話実験では同乗者の人間とタスク遂行役の対話システムとの対話で，対人間発話と対システム発話を収録した．対話状況が異なる発話での比較により，様々な発話条件を考慮した識別も行った．実験の結果，対話相手の音声認識率や親密度が識別の性能に影響を与えることが分かった．

Identification of Utterances Made to a System by Using In-Car Speech Acoustic Features

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

This paper presents usefulness of identifying user's utterances made to a spoken dialogue system using machine learning which uses acoustic features of user's utterances recorded in various situations. We have already performed dialogue experiments with two speakers (human-human or human-machine patterns) in several situations and we newly performed the experiments with three speakers (human-human-machine). The dialogue task simulates voice control of a car navigation system, where we made users perform goal settings or look the goal up in destination database. We prepared a spoken dialogue system for all experiments and prepared a human operator for the experiment with two speakers. We used the dialogue data achieved from the experiments and identified user's utterances made to the spoken dialogue system. Additionally, by comparison with utterances which were collected from different situations, we researched the influence of various conditions on performance of identifying utterances.

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の音声対話システムの実用的応用が注目されている．しかし，現在，音声インターフェースが一般的なマンマシンインターフェースとして一般的に普及しているとは言えず，カーナビゲーションシステムにおいて，安全性の観点から実用化されているのみである．ところが，車内での運転者と同乗者の雑談のようなシステムへの発話（命令）以外の音声や発話区間の誤検出の問題から，現在のカーナビゲーションシステムでは一般的に入力時にボタンを押してから発話を開始する方法が用いられている．これはハンズフリー，アイズフリーという安全性の観点から最も有効な音声インターフェースを用いているにもかかわらず，その音声インターフェースの操作にためにボタン操作という運転上の危険性のある操作を行うという大きな矛盾を含んでいる．

これらの解決方法としてカーナビゲーションシステムにおける完全なハンズフリーの音声インターフェースを実現する必要があるが，その実現に向けた最初の一歩として，音声のみを用いたシステムへの発話（命令）の識別を検討する．

音声のみを用いたシステムへの発話を識別する方法として，音声スポッタ[1]のように意識して変化させることができる音声中の韻律的特徴や言語的特徴を利用し，通常の発話では用いない状態にまで大きく変化させることにより識別する方法も考えられる．しかしながら，これはユーザに自分の発話を不自然に感じさせ，また発声時に発声方法を強く意識する必要があるため負担がかかる．

我々は，よりユーザ満足度の高いインターフェース

を目指すために，同じように音響的特徴や言語的特徴を利用することでシステムへの発話を識別するとしても，できるだけユーザに負担をかけない方法によって自然に変化した音響的特徴や言語的特徴を利用しての識別を目指す．例えば，相手がシステムであることにより自然に生じる音響的・言語的特徴が存在するならば，逆に相手を強く意識させることなどによってより識別が容易な特徴をユーザにとって自然に誘発させることができる可能性がある．

我々はこれまでにカーナビゲーションシステムの使用を想定したタスク指向対話において，様々な条件を設定した異なる対話状況での音声対話を収集し，対話データの音響的・言語的特徴の分析を行っている[2]．その結果，

- 対話相手が人間か機械かという事実自体は発話の変化にそれほど影響を与えない
- 機械的な発話を行うような対話相手に対して，人間は自然な発話（対話）ができない
- 対話のリズムや音声の韻律（感情を含む）などの応答能力を人間に近づけることが自然な発話（対話）促すために重要
- 対話相手の音声認識・言語理解率も発話に変化をもたらす
- システムに対する先入観も発話に変化をもたらす

といった知見を得ている．

これらの知見から我々は音声インターフェースに関して，二つの方向性を検討している．一つは，人間同士のコミュニケーションをできるだけ模倣したシステムを構築し，人間が機械とのコミュニケーションであることをできるだけ意識しないで使用できるインターフェースを開発すること．もう一つは，機械的なシステムに対する発話やコミュニケーションが，人間同士のものとは異なることをうまく活用し，システムの精度を向上させたり，新たなインターフェースを構築することである．この方向性での研究として，杉本らは発話の特徴を利用した発話の識別を行っている[3]．

本研究は後者の方向性の一環として，カーナビゲーションシステムへの発話がオペレータや同乗者への発話と異なることを利用し，その特徴を用いてシステムへの発話を識別することを検討した．また様々な状況下での発話データを比較し，システムへの発話を識別しやすい条件，状況などが存在するかなども検討した．なお，今回は予備的な実験として，発話の音響的特徴のみを識別に利用している．

2. 対話相手の違いによる発話の言語的・音響的特徴の変化

我々はこれまでに，様々な対話状況でのタスク指向対話のユーザ発話に関する分析を行ってきた[2]．本節では，分析で得られた対話状況の違い，特に対話相手の違いに関する分析結果を中心に述べる．

対話相手の違いによる発話への影響ついて，発話相手が人間あるいは機械であるという事実が発話に変化を与えることはほとんど無かった．一方で，対話相手の応答能力に関しては，応答に自然な音声を用いた場合と合成音声・録音音声といった応答能力を制限した場合では発話に大きな違いが表れた．後者では，間投詞の減少や発話開始時間の増加など，機械的な発話への変化が見られた．これらの結果より，人間と機械で発話の変化を生じさせることは可能なため，発話の推定が可能であると考えられる．

対話相手の音声認識能力については，対話相手の応答能力と関係があることが明らかになった．対話相手の認識能力がよくない場合（認識率がおよそ 80%）には，応答が自然音声の場合には丁寧に抑揚を付けて話すようになり，合成・録音音声の場合には声を大きくして抑揚を付けて簡潔に話すようになった．これら結果からも，音声認識率が発話の識別に影響を与える可能性が高いと考えられる．

続いてユーザの機械に対する様々な先入観に関して，システムの知的レベルが低いと考えている場合には発話を細かく区切る傾向が見られ，システムの対話のテンポが悪いと考えている場合には，単純で一度に多くの情報を伝えるという変化が見られた．音声認識能力に関しては，先入観による影響は見られず，対話を行うことにより得られた実際の対話相手の認識能力にのみ影響を受けることが分かった．これら先入観による発話への影響はそれほど大きなものはなかった．しかしながら，全体的な発話の傾向が見られるため，発話識別の性能の向上に活用できると考えられる．

3. 発話音声の収録方法

対話タスクとして想定した目的地検索・設定タスクは，カーナビゲーションシステムを使用し，目的地のランドマークを検索・設定するものである．本研究で識別実験に用いた対話音声には大きく分けて 2 種類あり，それぞれ異なる実験により収録されている．一つは人間（オペレータ）または機械のどちらか一方と対話を行い，目的地の検索・設定する 2 者対話より収集した発話音声である．そしてもう一つは，助手席に同乗者を乗せて雑談を行いながら，カーナビゲーションシステムで目的地のランドマークを検索・設定するという状況を想定した 3 者対話での発話音声である．

表 1に音声を収録した対話状況を示す．

表 1：実験対話状況

実験の種類		2 者対話		3 者対話		
対話相手		人間	機械	人間		機械
対話の種類		タスク指向		雑談		タスク指向
熟練度		低	-	低	高	-
応答音声		自然音声	合成音声	自然音声		合成音声
音声認識率	100%	O2-100	W2-100	H3	H4	W3-100
	80%	O2-80	W2-80	-		W3-80

2 者対話実験については，先行研究で収録した音声データを用いた．音声認識率は対話相手の音声認識率である．被中の対話相手「機械・合成音声」は WOZ 法を用いたシステム（以下 WOZ システム）であり，その操作は対話相手「人間・自然音声」と同一の人物が行っている．WOZ システムでは，想定される入力に対してあらかじめ応答内容の候補を用意しておき，それらの中から適切なもの（音声認識率 80%の場合，不適切なものも含まれる）を選択するという操作を行っている．「人間・自然音声」（以下オペレータ）は日常行われる電話などの非対面の対話と同じである．

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むかえに行く．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，友人が腹減ったといってきたので，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 1：2 者対話実験のタスクシナリオ例

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 ：北海道留萌市東幸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 2：2 者対話実験におけるシステムとの対話例

運転タスクには安全性の問題からドライブシミュレータ（ゲーム）を用いた．タスクの内容は単純なコースを線に沿って時速 100km/h で走行してもらうというものである．このタスクの認知的負荷は RRV 法により実車を 60km/h で走行させるものと同程度である．

音声収録実験に用いた被験者は対話状況ごとに 12 名の男子大学（院）生であり，音声対話システムに関する知識は持っていない．

続いて本研究で新たに収集した 3 者対話実験について述べる．この実験では同乗者がいる状況での車の運転を想定しており，被験者と同乗者と音声対話システムの 3 名が発話者となる．対話は被験者と同乗者間および被験者と音声対話システム間でのみ行われ，同乗者の役割は実験者本人が務めている．音声対話システムには 2 者対話と同様の WOZ システムを使用しており，被験者には，人に対する発話とシステムに対する発話を見分け，システムに対して話しかけた場合にのみ反応するように設計されていると説明している．

被験者と WOZ システムの対話におけるタスクは，目的地の検索・設定である．被験者と同乗者の対話は雑談および目的地の選択・決定である．これは，設定されたシナリオに沿って，同乗者があらかじめ用意しておいた行動計画候補の中から，相談をしながら目的地の選び出す作業のことを指す．シナリオは 2 人でドライブに行くという設定であり，システムに対する入力には，ドライブの目的地となるものを 2 つ，食事をする場所を 2 つの合計 4 つの検索・設定が含まれる．また，運転タスクは 2 者対話実験と全く同じ操作を，同環境下で行わせた．

また，3 者対話実験では被験者と対話システムの組み合わせを 2 パターン用意した．一つは被験者が実験者と親しい間柄にある知人および友人で，WOZ システムの音声認識率が 100%の組み合わせである．もう一方は被験者と実験者が初対面あるいはそれほど親しくなく，音声認識率が 80%の組み合わせである．被験者は各パターン 12 名で，全て男子大学（院）生であり，音声対話システムに関する知識は有していない．

実験手順は，まず被験者にドライブに行くというシナリオのみを伝える．シナリオには具体的な行動予定が含まれてはおらず，同乗者と相談の上で行動を決定すること，目的地を 2 つと食事を 2 回行う予定であることだけ伝えている．その後，目的地と食事場所を交互に 2 回検索・設定してもらう．ただし，目的地の設定タスクの前には必ず雑談を交えている．

4. 対システム発話の識別

我々は対話音声収録実験により，様々な対話状況での対話を収集した．我々は，これらより得られた対人発話および対システム発話の中から，対システム発話を識別する．識別には音響的特徴を属性とした機械学習を利用する．

4.1. 発話データ

表 2：2 者対話の発話データ数

対話相手		オペレータ		WOZ	
音声認識率		100%	80%	100%	80%
DT	無	237	326	229	347
	有	240	327	232	334

表 2に 2 者対話から抽出した各対話状況における発話データ数を示す．表中の DT は運転タスクを意味しており，運転タスクが有る場合と無い場合での発話数をそれぞれ示している．2 者対話実験では，全ての対話状況で同数の目的地設定タスクを遂行している．音声認識率の違いによって見られる発話数の差は，否定発話や訂正発話によるものである．また，発話はすべてタスクに関連したものとなっている．

表 3：3 者対話の発話データ数

親密度	高	低
音声認識率	100%	80%
対人	300	2353
対 WOZ	-	548

表 3に 3 者対話から抽出した各対話状況における発話データ数を示す．新密度が低い対話の同乗者への発話，システムへの発話はすべて抽出した．親密度が高いものに関しては 300 の対人発話を対話の中からランダムに抽出した．なお，3 者対話の対人間発話には笑い，相槌などの発話も含まれている．

4.2. 学習に用いた音響的特徴

本研究では対システム発話の識別に音響的特徴のみを使用した．表 4は判別に利用した特徴を示している．パワーの情報としては RMS（Root Mean Square）を用い，ピッチとして F0 の値を使用した．今回，音響的特徴は全て静的な特徴量を用いており，発話中の動的な変化量などは用いていない．発話時間は発話の開始から終了までのポーズを含めた時間であり，有声発話時間は発話の中でピッチが存在する時間を合計することにより算出している．パワーの平均値は発話時間での平均値により求めており，ピッチの平均値は有声発話時間での平均値により算出している．

表 4：判別に利用した発話の音響的特徴

時間情報	発話時間
	有声発話時間
パワー情報	RMS 平均値
	RMS 最大値
	RMS 分散
ピッチ情報	F0 平均値
	F0 最小値
	F0 最大値
	F0 分散

4.3. 学習方法

対システム発話の識別には，発話の音響的特徴を属性とした機械学習を用いている．機械学習のツールとしては，決定木学習ツールの C4.5 を使用した．

5. 実験結果

本節では対システム発話の識別結果を述べる．評価方法として，データの偏りによる影響を防ぐために 10 分割交差検定を用いた．

また，識別には学習・評価に用いる発話データの対話状況などの条件を組み替えることで，様々な観点から考察を行った．

5.1. 実環境に近い状況での識別性能

実際のカーナビゲーションシステムを操作する状況にもっとも近い，雑談をしながら音声認識率 80%の WOZ システムを使用してもらう 3 者対話実験の対話データを用いて，実環境を想定した場合での対システム発話の識別を行った．

表 5：全発話を用いた識別結果

対象発話	対人	対 WOZ
適合率	0.89	0.70
再現率	0.95	0.49
F 値	0.92	0.58

まずは 3 者対話の対話データから取り出した発話の全てを用いて識別を行った．用いた発話データの数は，対人発話が 2353 発話で対システム発話は 548 発話である．表 5はこれら全てのデータを用いた交差検定による結果を示している．

対人発話の識別精度が非常に高い一方で，対システム発話は精度が低く，特に再現率では 0.5 を下回っている．一つの原因として発話データの数に非常に大きな開きがあるということが挙げられる．本実験で用いた C4.5 では，学習時に最も正解の数が多くなるような決定木が作成されてしまうため，データ数に差がある場合には，データ数の多いほうに判定されやすい．したがって今回の識別では対システムの再現率が低下したと考えられる．

そこで，対人発話から対システム発話と同数の 548 発話をランダムに抽出し，新たに交差検定を行った．また，同数の発話で学習を行い，発話全てを評価データに用いて識別を行った．表 6，表 7はそれぞれの識別結果を示している．

同数の発話データを用いた結果では対人発話，対システム発話ともにほぼ同じ識別精度となった．また，同数の発話で行った学習で全発話を識別した結果，全発話による学習に比べて精度が上昇した．これらの結果より，車内で同乗者と対話を行いながらカーナビゲーションシステムを操作するという実環境においても，ある程度の精度で，システムに対する発話を識別でき

る可能性があることが示された．

表 6：同数の発話を用いた識別結果

対象発話	対人	対 WOZ
適合率	0.78	0.77
再現率	0.77	0.78
F 値	0.77	0.78

表 7：同数発話の学習による全発話識別結果

対象発話	対人	対 WOZ
適合率	0.97	0.54
再現率	0.82	0.89
F 値	0.89	0.67

5.2. 音声認識率の違いによる識別の影響

表 8：音声認識率の違いによる識別性能

対 O	認識率 100%				認識率 80%			
対 W	認識率 100%		認識率 80%		認識率 100%		認識率 80%	
対象発話	O	W	O	W	O	W	O	W
適合率	0.63	0.56	0.76	0.85	0.79	0.68	0.61	0.66
再現率	0.55	0.64	0.78	0.83	0.80	0.66	0.68	0.59
F 値	0.59	0.60	0.77	0.84	0.80	0.67	0.64	0.62

各対話状況における 2 者対話の対システムと対オペレータ発話を組み合わせ，音声認識率の違いによる発話の識別の影響を調査した．表 8はそれらの組み合わせによる識別性能を示している．表中の対 O および対 W は，それぞれ対オペレータ発話，対システム発話の意味であり，認識率は対話相手の音声認識率である．

対システム発話の識別という観点から見ると，実システムの環境に最も近いと考えられる認識率 100%の場合の対オペレータ発話と認識率 80%の場合の対システム発話の組み合わせが最も高精度である．一方，対オペレータ発話の観点では，認識率 80%の場合の対オペレータ発話と認識率 100%の場合の対システム発話の組み合わせが最良である．この結果より，認識率が異なると一方の発話の音響的特徴が大きく変化するため識別の性能が向上することが分かる．つまり，認識率が低い対話相手への発話の識別の方がより容易になるということが明らかとなった．また，認識率 100%の対オペレータ発話と認識率 80%の対システム発話の組み合わせでは，どちらの識別とも高精度であることから，認識率の低下による発話の変化は対システム発話でより顕著であることが推測される．

5.3. 先入観の違いによる識別への影響

先行研究において，我々は 2 者対話実験を行う前に被験者が対話システムに持っている先入観に関してアンケートで調査している．アンケートの項目は「知的レベル」，「音声認識能力」，「対話テンポのよさ」の 3 つである．本実験では，アンケートの評価値に基づいて被験者のグループ分けを行い，各グループで発話の識別をすることで先入観と識別性能との関係を調査する．

表 9：知的レベルのグループ別の識別結果

グループ	人間以下		人間同程度	
対象発話	対人	対 WOZ	対人	対 WOZ
適合率	0.61	0.58	0.64	0.59
再現率	0.44	0.73	0.62	0.61
F 値	0.51	0.65	0.63	0.60

表 10：音声認識能力のグループ別の識別結果

グループ	人間以下		人間同程度	
対象発話	対人	対 WOZ	対人	対 WOZ
適合率	0.63	0.60	0.60	0.61
再現率	0.52	0.70	0.61	0.60
F 値	0.57	0.64	0.60	0.60

表 11：対話テンポのグループ別の識別結果

グループ	人間以下		人間同程度	
対象発話	対人	対 WOZ	対人	対 WOZ
適合率	0.67	0.59	0.54	0.60
再現率	0.49	0.76	0.65	0.48
F 値	0.56	0.67	0.59	0.53

識別を行ってみたところ，どの項目に関しても特に目立った変化は見られなかった．我々が以前に行った分析では，ユーザの先入観による発話の変化があったことを報告したが，これらは言語的特徴に関するものであり，音響的特徴には変化が表れていない．今回の識別では言語的特徴を利用していないため，識別結果に影響が表れなかったと考えられる．

5.4. 同乗者との親密度の違いによる識別への影響

3 者対話実験では被験者として親しい間柄の 12 名と，初対面あるいはそれほど親しくない 12 名の 2 パターンで対話を収録している．これら 2 つの被験者グループでは雑談の盛り上がり方がかなり異なるので，これらの比較により，親密度が発話の識別に影響を与えるかどうかを調査する．

学習には親密度が高いグループ，低いグループそれぞれの対人発話，および対システム発話を用いるべきであるが，3 者対話実験では，前者のグループで用いた WOZ システムの音声認識率が 100%であり，後者のグループではおよそ 80%の WOZ システムを使用している．これらの認識率の差が同乗者との対話に影響を与えることはないが，WOZ システムに対する発話には影響があるため，識別する対システム発話は同じ条件のものを用いる必要がある．そこで，今回は対システム発話に 2 者対話実験より得られたものを使用した．また，対システム発話についての認識率も 2 パターン用意することで，認識率を識別性能の比較条件として加えた．対人発話は両グループともランダムに抽出した 300 発話であり，対システム発話は認識率 100%で 232 発話，認識率 80%で 334 発話である．

表 12：親密度が高い場合の識別結果

対システム	認識率 100%		認識率 80%	
対象発話	対人	対 WOZ	対人	対 WOZ
適合率	0.73	0.70	0.80	0.81
再現率	0.80	0.61	0.79	0.82
F 値	0.76	0.65	0.79	0.82

表 13：親密度が低い場合の識別結果

対システム	認識率 100%		認識率 80%	
対象発話	対人	対 WOZ	対人	対 WOZ
適合率	0.84	0.76	0.90	0.83
再現率	0.81	0.80	0.79	0.92
F 値	0.82	0.78	0.84	0.87

表 12, 表 13はそれぞれ親密度が高いグループ，親密度が低いグループについて，2 パターンの対システム発話を用いた識別結果を示している．

対システムの識別性能に関して，両グループともに音声認識率が低い場合に高い識別精度を示した．このことにより，親密度の違いによる対人間発話の音響的特徴の変化は，音声認識率低下による対システム発話の音響的特徴の変化に比べて識別に有効に働いていないことがわかる．

親密度の観点からは，全ての発話の識別で親密度が低い場合に高い識別精度を誇っている．この要因として，親密度が低い場合には通常よりも丁寧な口調で話すことが影響を与えた可能性がある．一方で，対システム発話は短く分割された単調な発話になりやすい．これら傾向により，親密度が低いグループでは音響的特徴の差が大きくなり，識別性能が向上したと考えられる．

5.5. 短い発話による識別への影響

本研究では収録された全ての発話を対象にして発話の推定を行っている．しかしながら，肯定発話や相槌など発話長の短い発話は音声の変化が乏しく，また正確にピッチやパワーを測定することができない恐れがある．そこで，これらの発話の影響を調査するために，発話長の短い発話を除いて識別を行った．

今回除外する対象は発話時間が 0.5 秒未満の発話とした．対象とした発話は最も識別精度のよかった，親密度が低い，システムの音声認識率が 80%のものである．短い発話を除外する前の対人発話は 300 発話，対システム発話は 334 発話であり，除外後は対人発話数は 240 発話，対システム発話は 220 発話となった．

表 14：発話長の短いものを除いた識別結果

対象発話	対人	対 WOZ
適合率	0.85	0.77
再現率	0.77	0.85
F 値	0.81	0.81

表 14に短い発話を除外した場合の識別結果を示す．結果より，全ての性能が若干ながら低下していることが分かる．この原因を調査するために，学習により得られた決定木を分析した．発話長の短いものを含めて識別した場合では，パワーの平均値，ピッチの平均値が非常に有効な判定基準とされていた．一方で，発話長の短いものを除いた場合には，パワーやピッチ平均値に加え，それらの分散が判別に有効に働くことが分かった．しかしながら，発話時間が有効に働くことを示すような学習結果は見られず，識別の精度が低下した原因を特定することはできなかった．

6. まとめ

様々な対話状況で収録した対話音声の音響的特徴を利用して対システム発話の推定を行い，その有効性について調査した．対象とする発話は，カーナビゲーションシステムを使用するという設定の下，我々がすでに発話の特徴について分析を行った 2 者対話実験による対話データと，今回追加した 3 者対話実験で収録した対話データである．3 者対話実験では発話者として同乗者を新たに加えた．これらの対話データの組み合わせで発話の識別を行うことで，様々な条件の発話識別への影響も調査した．条件としては，対話相手の音声認識能力，被験者の対話システムに対する先入観，被験者と同乗者の親密度を想定した．

発話の識別実験を行った結果，対話相手の音声認識能力が異なる発話同士の識別では，識別の性能が向上した．特に，認識能力が低いシステムと高い人間の場合で最高の性能を示した．このことから，人間は相手の認識能力が低い場合，特に相手が機械であるときに発話が変化し，認識能力の差がある相手の識別は容易になることが明らかとなった．また，親密度の低い人間に対しての発話は親密な人への発話と異なり，機械との判別に有利であることも明らかとなった．

文　献

[1] Masataka Goto, Koji kitayama, Katunobu Itou and Tetsunori Kobayashi, Speech Spotter: On-demand Speech Recognition in Human-Human Conversation on the Telephone or in Face-to-Face Situations, Proceedings of the 8th International Conference on Spoken Language Processing, pp.1533-1536, (2004)

[2] 山田真也，伊藤敏彦，荒木健治，“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”電子情報通信学会技術研究報告，2005-SLP-059, pp.67-72. (2005)

[3] 杉本夏樹，北岡教英，中川聖一，“音響特徴を用いた対システム発話と対人間発話の識別”，電子情報通信学会，総合大会，D-14-9, pp.133 (2006)