音声模擬対話における対話制御と快適性について

小林 豊　中島大輔　新美康永
京都工芸繊維大学

人間と機械の音声対話システムを実現するには、人間同士の対話の中で起こっているごく当たり前の現象を一つずつ数え挙げて対処していく必要がある。筆者らは先に、システム発話の内容や形態がユーザ発話に影響を与えることを観察してシステムの対話制御戦略の重要性を指摘し、階層的な対話制御方式を提案した。本研究では、情報機械との音声対話における情報伝達が能率良く、快適であるためには、どのようなシステム応答を生成するのが適切であるかを検討する。そのために音声模擬対話収集支援システムを構築し、音声認識性能や対話制御戦略を模擬的に変化させて43名の話者から延べ67対話を収録し、対話の快適性や情報伝達の能率への影響を調査したので報告する。

Dialog Control Strategy and Comfortableness in Simulated Dialogs

Yutaka Kobayashi, Daisuke Nakajima and Yasuhisa Niimi
Kyoto Institute of Technology

In order to realize a practical human-machine spoken-dialog system, each aspect of human dialog must be well studied and modeled in the system. The authors have pointed out the fact that dialog control strategies affect user's attitude and utterances to the system, and proposed a hierarchical dialog control strategy. Among many factors, in this paper, information transfers and comfortableness in the dialog were studied with respect to system's recognition errors and response strategies. 67 WOZ-dialog from 43 subjects were recorded and analyzed. We found confirmations and rephrasing responses do not decrease the comfortableness of conversations.

1　はじめに

近年、音声対話の研究が活性化する中で、人間同士の音声対話の特性を研究する試みとともに、人間と機械の音声対話に特有の問題を詳しく検討する研究が行なわれている[1]。そのためには、人間同士の対話音声、音声対話システムあるいは人間が真似て行なう模擬対話システムを用いて多量の対話音声データを収集して、対話の特性を解析し、言語現象をモデル化することが行なわれてきた[2, 3, 4, 5, 6]。

人間と機械の音声対話では、人間同士の対話の中で起こっているごく当たり前の現象を一つずつ数え挙げて対処していく必要がある。筆者らは先に、システム発話の内容や形態がユーザ発話に影響を与えることを観察してシステムの対話制御戦略の重要性を指摘し、階層的な対話制御方式を提案した[7]。また、音声認識誤りがある状況での対話制御方式の数学的モデル化の1方式を提案した[8]。

一方、易[9]はユーザの協調性や意欲が、山本ら[10]はタスク、応答方法、応答音声品質などがそれぞれ対話達成に与える影響を検討した。音声対話に限定せず、画像なども加えたマルチメディアインターフェースとの対話に関する研究も行なわれている[11]。このように音声対話システムの様々な側面を総合的に検討して、人間と機械の快適な音声対話を実現しようとする試みが増えている。

本研究では、情報機械との音声対話における情報伝達が能率良く、快適であるためには、どのような対話

制御戦略を採用するのが適切であるかを検討する。そこで、人間と機械との音声模擬対話収集支援システムを構築し、音声認識性能や対話制御戦略を模擬的に変化させて音声対話を多数収録して、対話の快適性や情報伝達の能率への影響を調査した。

2 音声対話収集実験の概要

2.1 模擬対話収集支援システム

京都観光案内システムを含む模擬対話収集実験の環境を図1に示す。ユーザ（被験者）用とシステム用にそれぞれ別の部屋を用意して、システム担当者の存在を気付かれないように注意した。対話は音声のみによって行ない、ユーザはヘッドホンと接話型マイクロホンを利用する。ユーザに臨場感を与えたり、知識を同程度に統一するための地図、対話によって達成すべき目標を明確化するためのプランシートを渡した。一方システム側は、誤認識率といった音声認識性能や確認発話を用いた対話制御戦略を模擬的に変化させながら対話を行ない二十歳前後の学生43名から67対話収集した。

2.2 対話タスク

音声対話システムの構築を考えるとき、比較的小さく限定されたタスク世界における目的指向対話の実現を当面の目標と考えてよい。ユーザ、システム双方が互いに相手から情報を引き出すようなタスクとして、本研究では「京都観光案内システム」を想定した。ユーザはシステムとの対話を通じて京都の1日の観光プランを作成する。

2.3 システム発話制御

音声対話においては意志伝達には、発話テキストの内容に加えて、イントネーション、アクセント、発話速度に代表される韻律情報の果たす役割が大きいことが知られているが、互いに相手の話し方に影響を受けるので、人間と機械の音声対話を考える場合も、ユーザが話し易く快適で、機械にとってはユーザ音声を認識し易い発話制御が望まれる。様々な制御因子を考えることができるが、ユーザ音声の認識誤りを完全になくせないことを容認して、その上でどのようなシステム発話を生成するのが望ましいのか検討する必要がある。

認識誤りをもつシステムではユーザの入力に対して何らかの確認作業が必要となる。そこで2つの方法を考えた。例えばユーザの発話が例1のuのような質問であったとする。確認の手段としてはまず直接的な確認発話をする方法が考えられる(s1)。また、単に答だけを返す(s2)代わりに、認識結果を応答に含ませる間接的な確認発話が考えられる(s3)。これらをそれぞれ直接確認、間接確認と呼ぶことにする。

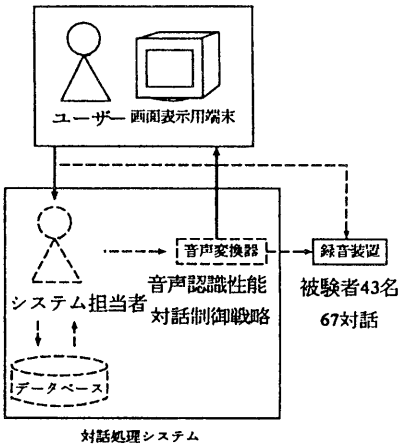

図 1. 模擬対話収集システムの構成

例 1

u：清水寺の拝観料はいくらですか。
s1：清水寺の拝観料ですか。[直接確認]
s2：300 円です。
s3：清水寺の拝観料は 300 円です。[間接確認]

本研究では、ユーザ発話の誤認識率とシステムの直接・間接確認発話の生起確率を種々設定して、模擬対話を収録し、対話の効率と快適性について検討した。

3 収集対話の特徴

限られたタスクの目的指向対話が行なわれる状況のもとで起こる音声対話特有の現象が収集実験の中でいくつか見受けられた。

3.1 詳細化

ユーザの質問内容が不十分な場合、例2のような詳細化のための発話が必要となる。

例 2

s：どういったところを観光してみたいですか。
u：えーと、お寺に行きたいです。
s：どういったお寺がよろしいですか。
u：んーと、あまり有名じゃなくて観光客が比較的少ない所。
s：どの地域がご希望ですか。
u：んーと、洛西の方。
s：洛西周辺には祇王寺、大覚寺、天竜寺などございます。

3.2 文脈の選択

人は無意識の内にその時の文脈、話題の流れといったようなことを踏まえた上で対話を行なっている。種々の知識を最大限に活用して次発話の話題を予測し、可能性の低い解釈を排除するというように、高度な知識処理を用いることによって音声対話における情報交換が能率良く行なわれることは、人同士の対話に照らして考えればごく自然である。この知識処理を用いない情報機械と対話すれば例3のような発話が生じる。人同士の対話ではs',u'に見られるような発話は滅多に生じない。

例3 (宿の風呂の話のあとで)

u：もみじ荘とグランドホテルとではどちらが広いですか。
s'：何の広さをお尋ねですか。
u'：風呂の広さです。
s：風呂はもみじ荘の方が広いです。

3.3 ユーザの誤認識

認識誤りはシステム側に限ったことではなくユーザにもあり得ることである。このときシステムが誤った情報を提供していることにユーザが気付かずに対話は進められてゆくことになる(例4、5)。このような発話が67の収録対話中にどのくらい生じたか表1に示す。システムは正しい情報を与えるものと期待し、答えの部分にのみ注意を向けると生じやすく、現時点ではシステムの誤認識率との相関は見られない。

例4

u：八坂神社への交通機関を教えて下さい。
s：知恩院への交通機関は市バス２０６番に乗り知恩院下車です。
u：あ、はい。分かりました。

例5

u：銀閣寺の近くで昼食をとるとこはありますか。
s：南禅寺周辺には精進料理の店がございます。
u：あ、他はありますか。

4 対話における情報伝達能率

4.1 対話でやり取りされる情報量

先に述べたような音声対話特有の現象を考慮した上で「情報伝達量」なるものを定義しなければならない。情報を自然科学の対象として位置づけ、その性質を明らかにしたC.E.Shannonはこれをビットで表現したが、本研究では対話という立場から個数として捉え確率的概念としては定義していない。そこで以下に示す

表1. システムの誤認識にユーザが気付かない場合
(67対話中9対話)

認識誤り(%)	応答なし(回)
5.3	1
5.6	1
7.4	1
11.8	1
16.0	1
19.0	1
26.8	1
26.9	3
33.3	1
	計 11回

3つの観点から音声対話において伝達される情報量を検討する。

- ユーザが受けとる情報(情報1) — プランシートを埋める為に直接必要な、システムがユーザに対して与える情報。情報1をもつ発話の多くは
  - 「〜は... です」
  - 「... がございます」

  の形となる。システムがこのように発話した時、ユーザは情報を1つ受けとったとする。
- 意思伝達単位の情報(情報2) — システム、ユーザ相互の意思伝達レベルの情報。情報1をもつ発話では基本的に「〜は... です」の形をした発話となるが、情報2をもつ発話はこれ以外に
  - 「... ですか」
  - 「... でよろしいでしょうか」
  - 「... はお分かりですか」

  といったような形をもとりうる。つまり例えば、ユーザの交通機関を尋ねる発話も「ユーザは交通機関を知りたがっている」という情報がシステムに与えられたとする。
- ユーザが受けとるフレーズ単位の情報(情報3) — 上の2つではその発話が情報をもつかもたないかで、情報伝達量を0 or 1としたが、ここでは1発話に含まれるフレーズ単位の情報を考慮し、プランシート作成に意味をもつフレーズがi個含まれていればその発話は情報伝達量iとする。例えば、例6の発話については情報3の定義では情報伝達量3とした。

例6

u：洛西にはどんなお寺がありますか。
s：洛西周辺には金閣寺、竜安寺、妙心寺などございますが。

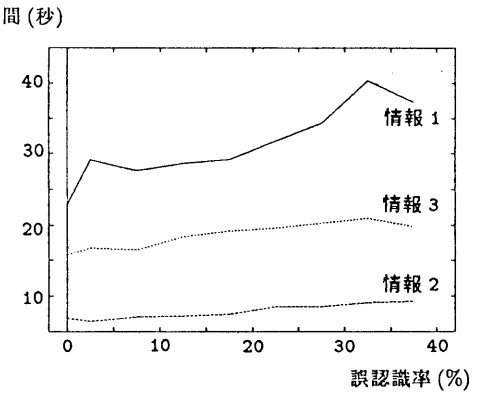

図 2. 単位情報伝達に必要な時間

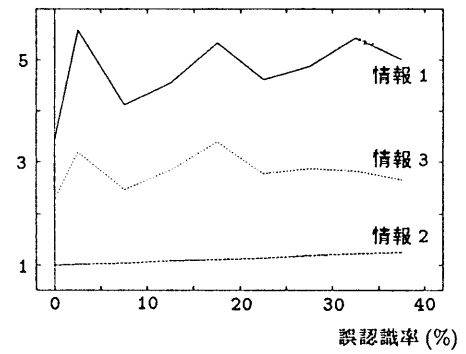

図 3. 単位情報伝達に必要な発話数

4.2 認識誤りと情報伝達能率

情報伝達の能率を計る尺度としては「情報をひとつ得るために必要とした時間, 発話数」を先に述べた3種類の情報について調査した。収録対話を書き起こしたテキストファイルを分析した結果を図 2, 図 3に示す。いずれも情報 2 で最も滑らかな単調増加が現れている。これは音声対話特有の現象、すなわち「詳細化」、「文脈の選択」等によるものと考える。こういった現象内では、たとえシステムが認識誤りを起こさなくとも先に定義した意味での情報 1,3 の流れは現れてこない。一方、情報 2 はコミュニケーションという見方で定義したもので、上のような発話の際でも情報の流れを見ることができる。

今後、対話においてやり取りされる「情報」というものをどのように表現するかさらに詳細に検討する必要がある。

5 対話における快適性

音声対話では、情報伝達の能率と合わせて考えなければならない重要な問題として対話の快適さ、心地よさがある。ここでは快適性の観点から音声対話について検討を行なう。対話の快適性とそれを決めうる幾つかの変量 (説明変量) 同士の間の種々の関係を明らかにし、将来快適な対話制御法を開発する目的で、多変量解析の分析法の一つである重回帰分析を行なった。

5.1 快適性の重回帰式

システム発話制御のパラメータ値と被験者の感じた対話の快適さを各対話について調査し、7 個の変量

$$x_1, \ldots, x_6, y$$

を持つ集団から 51 組の標本を得た。これを用いて、目的変量 y (快適性) を説明変量から予測する重回帰式を求めた。各変量の意味は以下のとおりである。

x_1 : システムの平均発話長 (sec)
x_2 : 対話にかかった時間 (sec)
x_3 : 対話の効率 (items/sec)
x_4 : 間接確認の出現率 (%)
x_5 : 直接確認の出現率 (%)
x_6 : システムの認識誤りの出現率 (%)
y : 対話の快適性 (%)

x_1, x_2は収録対話から測定した値である。x_3の対話の効率とは 1 秒当たりの情報伝達量で「情報 3 の伝達量／対話長」とした。x_4, x_5はともに「(間接・直接) 確認発話出現回数／システム発話数」、x_6は「システム誤認識発話数／システム発話数」である。目的変量 y については、被験者から対話終了後、「対話は長く感じたか」、「確認発話は多くなかったか」といった数個の項目について考えた上で対話の快適性 (0%～100%) をアンケートしたものを変量値とした。求めた重回帰式

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_6 x_6$$

の偏回帰係数と、これから測定単位の影響を取り除くため各変量を平均 0、分散 1 となるように正規化した式

$$Y^* = b_1^* x_1^* + b_2^* x_2^* + \cdots + b_6^* x_6^*$$

の標準偏回帰係数を表 2に示す。

5.2 重回帰式の有意性

求めた重回帰式を用いて目的変量の標本値 y_λ に対するその回帰推定値 Y_λ を求めることができる。求めた推定値がどの程度有効であるのかの検定を行うため各変量の有意性を表す t 値、重回帰式の検定を表す分散比 F、そして測定値 yの全変動に対して予測値 Yの変動の占める割合を表す寄与率 R^2を次式により求めた (表 2)。

表 2. 偏回帰係数および標準偏回帰係数（1）

変量	説明変量 6 個		
	偏回帰係数	標準偏回帰係数	t 値
−	b_0= 5.089	−	−
x_1	b_1= 0.019	b_1^*= 0.009	0.04
x_2	b_2=-0.003	b_2^*=-0.259	**2.10**
x_3	b_3= 0.000	b_3^*= 0.002	0.01
x_4	b_4= 0.024	b_4^*= 0.189	0.84
x_5	b_5= 0.077	b_5^*= 0.123	0.90
x_6	b_6=-0.115	b_6^*=-0.707	**4.45**
分散比	7.29 $> F(6,44;0.01) = 3.243$		
寄与率	0.50		

表 3. 偏回帰係数および標準偏回帰係数（2）

変量	説明変量 4 個		
	偏回帰係数	標準偏回帰係数	t 値
−	b_0= 5.175	−	−
x_2	b_2=-0.004	b_2^*=-0.225	**1.69**
x_4	b_4= 0.021	b_4^*= 0.168	1.25
x_5	b_5= 0.102	b_5^*= 0.166	1.31
x_6	b_6=-0.110	b_6^*=-0.704	**5.49**
分散比	8.45 $> F(4,46;0.01) = 3.757$		
寄与率	0.46		

$$F = \frac{RV/p}{EV/N-p-1},$$
$$R^2 = RV/a_{yy}.$$

ここで目的変量 $y_\lambda(\lambda = 1,2,\ldots,N)$ の全変動を a_{yy}, 回帰推定値 $Y_\lambda(\lambda = 1,2,\cdots,N)$ の変動を RV, 回帰からの残差 $(\varepsilon_\lambda = y_\lambda - Y_\lambda)$ の変動を EVとした。分散比 F によって標本値 y_λの変動がどれだけ多く回帰推定値 Y_λの変動で説明できるか、寄与率 R^2によって回帰による変動 RVが全変動 a_{yy}に対してどの程度の割合を占めているかを知ることができる。

標準偏回帰係数の比較を行なうとつぎのことがわかる。

- 快適性に与える影響は「システムの平均発話長 (x_1)」、「対話の効率 (x_3)」はかなり小さく、「システムの認識誤りの出現率 (x_6)」が特に大きい。
- 係数の正負から「両確認発話の出現率 (x_4, x_5)」は大きくした方が快適性が上がる。

この偏回帰係数の値は t 値にも反映されている。なお表2の太字で示した t 値は、5.0%の危険率でその変量が有意であると判断したものである (t(44;0.05)=1.680)。

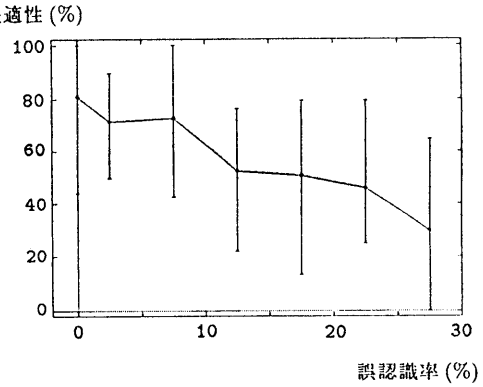

図 4. 誤認識率と対話の快適性

分散比を見ると、この重回帰式が 1.0%の危険率で有意であるといえる。また偏回帰係数の検定（t 値）から、先の重回帰式の中で役割の有意性が特に低い「システムの平均発話長 (x_1)」、「対話の効率 (x_3)」を除いた 4 個の説明変量で重回帰分析した結果が表 3である。

寄与率減少は「説明変量の数を減らすと小さくなり、より 1 から離れる」という性質からやむを得ないこととして、各変量の t 値や分散比は高くなっていることから後の重回帰式はより有意といえよう。これらの重回帰式で予測したときの確からしさが 95%ということにはならないが、F値より余裕をもって大きいので予測に役立つといえる。測定値 y（目的変量）の全変動に対して予測値 Yの変動の占める割合はそれぞれ 50%, 46%で、積極的に有効であるとはいえないようである。最後に図 4に認識誤りの出現率と対話の快適性の関係を示す。

5.3 確認発話と快適性

模擬対話収録後、被験者へのアンケートの結果、確認方法に関して以下のような知見を得た。

- 間接確認は確認作業を含みながらユーザ発話に応えているため、自然な対話の流れが保てるが、認識誤りしたときはユーザに不必要な（誤った）情報を長々聞かせることになる。
- 直接確認は文脈が大きく変わる時は有効であるが、そうでない時に多く出現すると、すぐに煩わしくなる。
- システムが正しく認識していても直接確認が発話されると、その度にユーザは何らかの応答を返さねばならない。

6 まとめ

本研究では、情報機械と人との音声模擬対話収集支援システムを構築し、「京都観光案内」を対話タスクとして、音声認識性能や対話制御戦略を模擬的に変化させながら音声対話を収録した。そして、これらを用いて対話の快適性や情報伝達の能率を分析、検討した。その結果つぎのことが明らかになった。

(1) 限られたタスクの目的指向対話が行なわれる状況のもとで起こる音声対話特有の現象には「詳細化」や「文脈の選択」のための発話がある。

(2) 音声対話における情報伝達能率についての調査で「対話でやり取りされる情報」を様々に定義し、とくに意思伝達のレベルで情報伝達を調べることにより、(1) の現象内での情報の流れを明らかにできる。

(3) 快適性とそれを決めうる様々な変量との相関。(重回帰分析による)

(4) システム発話の直接確認、間接確認が対話の快適性に及ぼす影響。

特に (3) では、対話時間、確認発話の出現率、誤認識率が重回帰式の中での役割が大であること、直接確認、間接確認の出現率増加が快適値を上げること等を明らかにした。

以下に今後の課題を示す。

- 情報の単位として、今回3通りの定義を行ったが、今後 (1) のような音声対話特有の現象を調査しながら、さらに詳細な検討を進めてゆく。
- (3) の重回帰分析では、今後音声資料を増やすとともに、イントネーション、間投詞の挿入など (自然さ)、応答に含む情報の数 (選択肢) など新たな変量についても検討を進めてゆく。なお、システム発話制御パラメータと対話から受ける印象との相関については未だ明らかになっていない部分が多く、この印象を変量とする分析を進めていくことが重要である。
- 上の (4) を踏まえて、「文脈の大きく変わるところで直接確認、ひとつの文脈内では間接確認を主に用いる」というようにそれぞれの異なる特徴を考慮した対話制御戦略の検討が必要。

参考文献

[1] 堂下、"音声・言語・概念の総合的処理による対話の理解と生成に関する研究", 文部省科研費重点領域研究研究成果報告書, 京都大学 (1994.3 および 1995.3)

[2] 竹沢、田代、森元、"音声言語データベースを用いた自然発話の言語現象の調査", 人工知能学会研資, SIG-SLUD-9403-3, pp.13-20 (1995.2)

[3] 田窪、"音声対話の言語学的モデル — 談話管理標識としての感動詞の分析 —", 情処研報, 94-SLP-1-3 (1994.5)

[4] N.M.Fraser and G.N.Gilbert, "Simulating speech systems", *Computer, Speech and Language*, 5(1), pp.81-99 (1991)

[5] 伊藤、秋葉、長谷川、速水、田中、"音声対話システム構築のための実対話データ収集実験", 情処研報, 94(57), 94-SLP-2-6, pp.35-42 (1994.7)

[6] 内藤、黒岩、武田、山本、谷戸、"大規模内線電話受付システムの試作", 信学技報, SP94-90, pp.37-42 (1995.1)

[7] 小林、新美、"対話制御の処理レベルとモデル化", 情処研報, 93-SLP-3-13, pp.59-61 (1994.2)

[8] 新美、小林、"音声認識の誤りを考慮した対話制御方式のモデル化", 情処研報, 95(16), SLP-5-7, pp.47-54 (1995.2)

[9] 易、"音声対話システムに対する被験者反応の一考察", 情処研報, 92-SLP-3-6, pp.30-33 (1993.2)

[10] 山本、中川、"音声対話システムの構成法とユーザ発話の関係の模擬実験による評価", 情処研報, 95(16), 95-SLP-5-9, pp.61-68 (1995.2)

[11] 伊藤、長谷川、他、"音声・視覚・画像をもつインタラクションシステム", 情処研報, 95(16), 95-SLP-5-5, pp.31-38 (1995.2)