音声言語情報処理 7−14
（1995. 7. 21）

音声対話におけるエージェントの態度と人間の発話の関連

八木 正紀†, 平栗 覚‡, 伊賀 聡一郎†, 安村 通晃†

慶應義塾大学大学院政策・メディア研究科†
慶應義塾大学環境情報学部‡

Abstract

音声認識・合成技術の発達により、音声を用いたエージェントベースのインタフェースの増加が予想される。人と機械がより自然なインタラクションを行なうためには、人間とエージェントが音声を用いて対話する際の特徴、人間同士の対話との共通点や相違点などについての知見が重要となる。本研究では、人間とエージェントの音声対話の際に、エージェントの態度が変化することによって人間の発話にどのような影響があるのかを検証するため、Wizard of Oz 方式による疑似音声対話システムを構築し、対話実験を行なった。さらに実験によってえられた対話例およびアンケートの結果に基づいて分析・考察を行なった。

Relationship between the attitudes of agents and human conversation in a speech dialogue system

Masanori Yagi†, Satoru Hiraguri‡, Soichiro Iga†, Michiaki Yasumura†

Graduate School of Media and Governance, Keio University†
Faculty of Environmental Information, Keio University‡

Abstract

An agent-based computer interface which use speech dialogues is largely expected by developing speech recognizer and voice synthesizers. The speech dialogue between users and computers must be clarified in comparison with human-human interaction to make the interaction between user and computer more natural.

In this research, we have designed an experimental system by the Wizard of Oz method to investigate the effects of attitudes of agents on ones of human beings in a speech dialogue. Then we analyzed the dialogues taken from this experiment and the result of questionnaire performed.

1 はじめに

近年、人と機械 (コンピュータ) との新しいインタフェースとして音声認識・音声合成を用いたものが注目されている [1][2]。しかし、機械と会話をするという行為には少なからず抵抗を感じることがあるだろう。そこで、コンピュータ内にエージェントを仮定することで、スムーズなインタラクションを可能にしようという研究も見られる [3]。

今後多様な表現力を備えたエージェントによるインターフェースが増えてくることが予想されるが、それらが人間にどのような影響を与えるかについてはあまり考慮されていない。

そこで、本研究では、人間とコンピュータ内のエージェントが音声によって対話を行なうことを想定し、その特徴や人間同士の対話との共通点・相違点などを明らかにするため、Wizard of Oz(以下WOZ) 方式による模擬音声対話システムを構築し、対話実験を行なった。特にエージェントからの情報提示の変化 (冗長度や口調) によって、人間の発話にどのような影響が表れるかに注目して実験を行っている。

2 予備実験と仮説

本実験に先立って、今回と同様の、エージェントとの音声対話による情報検索というタスクを用いて予備実験を行なった。

その際の対話例では、被験者の多くが単語に近いレベルの発話しかしていなかった。

この原因について分析の結果、次の 2 点が大きく影響しているものと推定された。

(1) エージェントの発話量が多い

タスクの説明など、エージェントが被験者に情報を提示する際の発話は人間同士の対話と比較してかなり冗長なものになっている。それに対し、被験者側はエージェントに対して必要十分な情報を与えるだけで対話が成り立ってしまうため、単語レベルに近いものになってしまったのではないかと思われる。

(2) エージェントの口調が丁寧

エージェントの口調がかなり丁寧なものに設定してあったため、それに影響されて被験者が緊張してしまい、発話量が制限されてしまったのではないか。

これらの推測に基づき、次のような仮説を設定した。

「エージェントの口調や態度が変化することで、被験者側の発話の冗長性が変化する」

この仮説を検証するため、以下に述べる実験システムを構築し、対話実験を行なった。

3 実験システムについて

実験システムは、WOZ 方式を用いた模擬的な音声対話システムである。図 1にシステムの構成を示す。

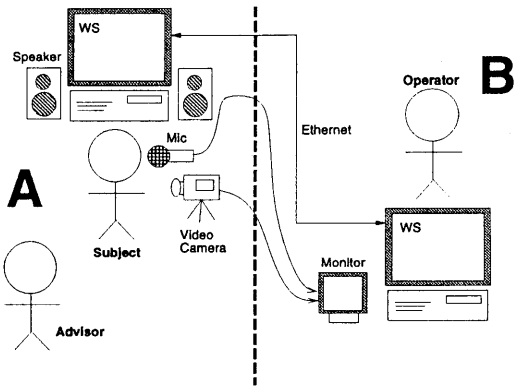

図 1: システム構成図

部屋 A に被験者とアドバイザー、部屋 B にオペレータがおり、被験者の実験の様子と音声はビデオカメラとマイクによりオペレータに送られる。被験者にはオペレータの存在は告げられず、あくまで機械による音声認識を行っていると説明しておく。

オペレータは被験者がマイクで発話した音声を元に予め用意した返答を選択し、部屋 A の端末へ送信する。送信された返答は WS 上の音声合成ソフトウェアによって処理され、合成音声によって被験者に提示される。

実験のパラメータを音声情報のみに限定するため、ディスプレイには簡単な説明と、被験者に発話のタイミングを促す簡単なアイコンを表示しておくだけで、その他にグラフィカルな情報は提示していない。アドバイザーは実験の説明と問題が生じたときの手助けをするのみで、実験中は被験者とのインタラクションは行なわないようにした。

4 エージェントのモード

実験を行なうにあたり、エージェント側の発話の違いが被験者に対してどのような影響を及ぼすかを調べるため、発話の口調や冗長性が異なる3つのモードを用意した。各モードの特性は表1の通りである。

表 1: モードによる変化

モード	冗長性	口調
A	高い	丁寧
B	低い	丁寧
C	低い	ぞんざい

モード A は、予備実験の際のエージェントの発話とほぼ同一である。それに対しモード B およびモード C は、情報伝達に不要な語句を日本語として不自然でない程度に削っている。また、モード A・B の丁寧な「ですます調」に対し、C はぞんざいな口調になっている。各モードにおける対話例を以下に挙げておく。
(被:被験者の発話　シ:システムの発話　を示す)

(モード A)

シ: あなたの知りたいのはあいだかずお先生についてですか。

被: はい、そうです。

シ: あなたの知りたい情報は何でしょう。

被: あいだ先生の今期の担当科目はありますか。

シ: あいだ先生の担当の講義には、企業と会計、経済管理と会計情報があります。あいだ先生について他に知りたい情報があればいってください。

(モード B)

シ: えとうじゅん先生ですか。

被: はい。

シ: 知りたい情報は何でしょう。

被: 研究会のテーマについて教えてください。

シ: 翻訳論です。他に知りたい情報があればいってください。

(モード C)

シ: ふるかわこういち先生のことかな。

被: そうです。

シ: 何を聞きたいんだね。

被: 先生のオフィスアワーを教えてください。

シ: それなら、月曜十五時から十六時三十分だ。他に知りたいことがあったらいってほしい。

5 タスクについて

被験者に与えるタスクは、実験のために作成した当大学湘南藤沢キャンパス (SFC) 内の教員およそ 120 名ついての簡単なデータベースを、音声を用いて検索してもらうというものである。各教員について、以下の 5 項目の情報を対話的に検索することができる。

- 研究室の番号
- 内線番号
- オフィスアワー
- 研究会のテーマ
- 担当講義

6 実験の方法

実験に先立って、被験者はアドバイザーから以下のような説明を口頭で伝えられる。

- SFC の教員に関する情報を検索してもらう
- 操作はすべて音声でコンピュータと会話することによって行う
- 細かい操作方法や質問できる項目は画面や音声によって提示される
- 実験中はアドバイザーは原則として質問などには答えない

これらの説明の後、実際に実験が開始される。最低一人の教員について検索してもらったら、一旦プログラムを終了する。ここでアドバイザーは、エージェントについての印象などを被験者に簡単に述べてもらう。

その後、異なるモードについて同様に実験を繰り返し、それぞれに対する印象、他のモードとの違いについてインタビューする。3つのモードすべてについての実験が終了した段階で、被験者にアンケートを記入してもらい、実験全体が終了する。

被験者は、比較的音声対話システムに関する知識の少ないSFCの学生(8名)である。なお、モードの順番による影響を減らすため、被験者によって提示するモードの順番を変更した。

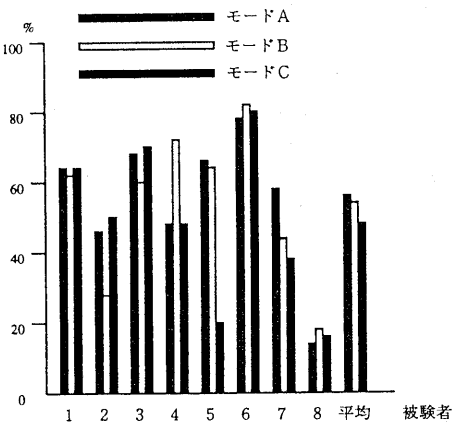

図 2: 全語数における不要語の占める割合

7 不要語について

被験者の対話を分析するにあたり、不要語の概念を定義する。

まず、被験者からエージェントに対して送られた発話を単語ごとに分割する。それらの単語のうち、情報伝達に必要でない単語を不要語と定義する。

(例) えー/じゃあ/X/先生/の/オフィスアワー。

上の発話において、エージェントはすでに被験者が「X先生」についての情報を得ようとしていることを知っているものとする。この時、エージェントにとって本当に必要な単語は「オフィスアワー」のみである。したがって、全単語6語のうち5語が不要語となる。

図2は、実験で得られた対話の不要語を被験者・モード別に集計したものである。縦軸は、被験者・モード別のそれぞれの対話での、全語数における不要語の占める割合を示している。

8 アンケート

以下の項目からなるアンケートを、実験終了後に被験者に記入してもらった。

1. 音声認識についてどの程度知っていますか?
2. コンピュータの声は聞き取り易いと思いますか?
3. 音声による入力は快適だと思いますか?
4. 実現したらこのシステムを使いたいと思いますか?
5. このシステムに好感が持てますか?
6. 3つの中でどれが一番対話しやすかったですか?
7. 3つの中でどれが一番人に近いと思いましたか?
8. その他、このシステムに対するご意見、ご要望をお願いします。

1〜5は5段階評価、6、7は3つのモードの中から一つだけ選んでもらった。アンケートの結果を表2に示す。

9 考察

(1) アンケート結果・収録対話について

アンケート結果および収録対話の分析をもとに考察を加える。

合成音声による対話はあまり聞き取り安いとはいえないものの、対話システムとしては概ね好意的に受けとられていると言えるだろう。

表 2: アンケート結果

被験者	回答						
	1	2	3	4	5	6	7
1	2	3	4	4	4	B	C
2	2	4	3	4	4	C	C
3	3	2	4	1	3	A	A
4	3	4	2	1	3	C	C
5	2	3	5	5	5	B	B
6	1	3	4	5	5	B	A
7	1	2	2	4	4	A	C
8	1	2	2	1	3	A	A
平均	1.9	2.9	3.3	3.1	3.9		

実験前、「どのモードが最も対話しやすかった (6 番)」や「どのモードが最も人に近かったか (7 番)」の項目はある程度偏りが出るのではないかと考えていた。6 番がほぼ均等に分散しているのに対し、7 番の方はモード C が 5 人とやや他よりも多くなっている。「話し易さ」と「人との近さ」は一致するのではないかと予想していたが、一致するものとしないものはほぼ半々といえる。

次に収録対話中の不要語数の割合を見る。図 2 によれば、被験者によって、モード間の変化のパターンが大きく異なっている。しかし、モードごとの平均を見てみると冗長性の異なるモード A、B がほぼ同程度なのに対し、口調がぞんざいなモード C は不要語の割合がやや少ない。冗長性よりも口調の方が大きな影響を及ぼしているという傾向が見える。

(2) エージェントのモードについて

被験者の発話への影響を見るためにモードを設けてエージェントの態度に変化をつけた。

モード A では「○○先生のオフィスアワーは月曜 16 時 30 分から 18 時です」というところをモード B では「それは月曜 16 時 30 分から 18 時です」といいかえるなどの方法で、A の方を B よりも冗長性を高く設定したが、被験者にとってモード A・B 間の区別を認識することが困難であったことがアンケートからわかる。

モード C の口調は、モード A・B の「です・ます調」の丁寧なものではなく、ややぞんざいなものに設定した。口調を親しみやすいものにすることで、被験者の発話の不要語が増加するのではないかと考えたが、図 2 によれば、モード C は A・B よりも不要語の占める割合は少なくなっている。

(3) エージェントの人格

アンケートの回答の中に、「コンピュータと認識してしまうので、“人” を想定することはできなかった」というものがあった。このことからもわかるとおり、現在のシステムにおける音声対話のみのエージェントに対して被験者が人格を認めるのは難しいようである。対話をより人間らしいものにするためにはなんらかの形でエージェントに人格があるかのように思わせる必要があるだろう。

結論としては、エージェントの発話の冗長性は被験者に認識されにくく、被験者側の発話の冗長性には影響しない。また、エージェントの口調が丁寧なものに比べて、ぞんざいなものの方が被験者側の発話の冗長性は低くなるということが言える。

10 今後の課題

以下に、今後の実験に向けての課題を挙げる。

(1) タスクについて

被験者により積極的に実験に取り組んでもらう目的で、なるべく身近な情報として、教員情報の検索というタスクを設定した。

しかし、このタスクには、以下のような問題点が考えられる。

- コンピュータ内の情報を人間が引き出すのみ
- データ間に関連性が乏しい

コンピュータ内にどんなデータが蓄積されているかは、あらかじめ被験者に知らされているため、被験者はエージェントの問いかけに応じてその中から自分にとって必要と思われる情報を選んでいくだけで良い。そのため、単語のみの発話で充分目的が達成されてしまう。実際、被験者によってはほとんどすべての発話を必要最低限の単語のみで行っている場合もあった。

また、実験で得られる情報は極めて断片的であり、相互に関連を持っているとはいい難い。被験者

がある情報を得たとして、それからなんらかの連想を行って別な情報の検索を行う、という状況が生まれにくいのである。

結局、「コンピュータ側の情報を人間側に送る」という行為を断続的に繰り返しているだけになってしまい、「人間同士の対話」とは異なった対話が行なわれてしまっている。

今後の実験では、より自然な会話が可能になるタスクを設定したい。

(2) 「人間同士の対話」とは

今回の実験は、態度の異なるエージェントの中でどれが人間同士のものに近い対話を作り出せるかという点に注目して行った。しかし、人間同士の対話というのはどのようなものであるか、ということについては明確にしていなかった。

漠然と人間同士の対話に近付くほど「被験者側の発話の冗長性が増すのではないか」という予想を立てていたが、分析結果は逆に「冗長性が減少する」という結果が出ている。しかし、それ以上に個人差の影響も無視できないほど大きい。

対照実験として人間同士の対話実験を行ない、比較する必要がある。

(3) その他

その他の主な課題を挙げておく。

- 人間同士による対話実験との対照を行なう
- モード間の差異をより明確にする
- エージェントに人格を付与する
- 発話のタイミングを明確にする
- 被験者数を増やす

これらを踏まえ、新たなタスクを用いて実験を行なっていく。また、現行のタスクを生かした実験も行なう予定である。

11 おわりに

人間とコンピュータ内のエージェントが音声によるインタラクションを行う際、エージェントの態度によって人間の発話がどのような影響を受けるかを調査するため、疑似音声対話システムを構築し実験を行った。実験によって得られた対話データにもとづき、不要語の占める割合という観点から分析を試みた。そして、エージェントの発話は、冗長性よりも口調の変化が大きな影響を及ぼすという結果が得られた。今後も継続的に実験を行っていき、エージェントの態度が人間に与える影響を明らかにしていきたい。

謝辞

本研究を行なうにあたり協力をいただいた以下の方々に、感謝の意を表する。

WS上で動作する音声合成ソフトウェアを貸していただいた(株)リコー殿。収録対話を分析する上でのアドバイスをいただいた(株)リコーの望主雅子さん。実験を行っていく上で数々の協力をしていただいた安村研究会のみなさんに感謝したい。

参考文献

[1] 亀山晋他,"カーナビゲーションにおける音声対話", 日本音響学会講演論文集 平成6年3月, pp.21-22, 1994.

[2] 吉岡理, 南泰浩, 鹿野清宏,"電話番号案内を対象としたマルチモーダル対話システムにおける音声入力の評価", 日本音響学会講演論文集 平成6年3月, pp.43-44, 1994.

[3] 内藤剛人, 竹内彰一,"Situated Interface:社会的インタラクションに向けて", 日本ソフトウェア科学会 WISS'94,pp.37-45,1994.

[4] 八木正紀, 伊賀聡一郎, 安村通晃, "人と機械の音声による対話実験", 慶應義塾環境情報研究所,KEIO-IEI-RM-95-002,1995

[5] 安村通晃, 伊賀聡一郎, "マルチモーダル・インターフェイスの試み, 情報処理学会第35回プログラミング・シンポジウム, pp.93-100, 1994.

[6] Hauptmann, A., and Rudnicky, A., "Talking to Computers: An Empirical Investigation.", *International Journal of Man-Machine Studies, 28(6)*, pp.583-604, 1988.