

# 常時稼働を想定した情報インタフェースとしてのエージェント設計

## Agent Design As An Information Interface Assumed For Full-time Operation

高橋 朋裕

片上 大輔

Tomohiro Takahashi,

Daisuke Katagami

東京工芸大学 工学部 コンピュータ応用学科

Tokyo Polytechnic University, Faculty of Engineering, Department of Applied Computer Science

**Abstract:** In this paper, the purpose of the research is the development of an agent as information interface assumed for full-time operation in human daily life. We prepare the three elements; bodily motion, dialogue component and voice pitch as an appropriate agent design for the concept of full-time operation and evaluate them by impression analysis. In the experiment, we implemented above mentioned three elements based on MMD Agent, and investigated the changes of user impression. As the results, we confirmed the great effects for user impression by adding bodily motion like a nodding for dialogue between agents.

## 1. はじめに

近年、人間とロボット・エージェント間での意思疎通を目的とした研究が盛んに行われている。その内容は多岐にわたり、ロボット・エージェントの発するピープ音による印象実験[1]から、身体動作による印象実験まで[2]と幅広く、エージェントとのコミュニケーションは、次第に人間同士の行うコミュニケーションと遜色のない域へと近づきつつある。

一方、人間対エージェントのコミュニケーションに着目する中で、実際にロボットやエージェントを、日常生活に取り入れて行われている研究は、先に挙げた研究に比べ多くない。

近年、生活の中にロボットやエージェントを取り入れる研究として、商用施設でのロボットによる販売促進の研究[3]や、高齢者との話し相手としてのインタフェースの研究[4]が行われているものの、商用や医療の手段としての目的が強く、我々の生活に密着する形でロボットやエージェントを取り入れる研究はまだまだ発展途上である。

私は人が生活する上で関わりの強い物に着目した。その中で、大抵の人が日常生活と密接に関わりがあると思われるテレビ・ラジオ・インターネットといった、情報インタフェース、及び情報の入手手段にエージェントを取り入れることで、ユーザとエージェントの会話のデータから、ユーザに合わせ効率的に情報の提供を行うことができたり、子供や高齢者な

どにおける、デジタルデバインド緩和の足がかりになるのではないかと考えた。

テレビ・ラジオ・インターネットといった情報インタフェースは、日常的かつ常時利用するものであり、それはエージェントを取り入れた際にも同様であることが望ましいと考える。

本研究は、常時稼働を想定した情報インタフェースとしてのエージェントを設計するに当たって、ユーザがエージェントの音声や動作、会話構成から受ける印象を評価・分析し、日常的にユーザと関わることを想定したエージェントを設計する際の指標を示すものである。

## 2. エージェントにおけるユーザへの印象付加と負荷軽減

### 2.1 エージェントによる情報の印象付加

情報の伝達方法が文字や音声、映像からエージェントを介す方法に変わることによってもたらされる違いのひとつとして、エージェントによって情報へ印象の付加を可能にできることが挙げられる。エージェントが人間に与える印象の研究は数多くあり、複数のエージェント同士の会話における雰囲気生成の報告[5]や、エージェントの会話に身体動作を加えることで高い印象度を与える報告[6, 7]がある。この



図1 MMDAgent 画面

ようにエージェントによる印象効果について報告されているものの、本研究で扱うような 3DCG によるエージェントを用いた検証報告はまだまだ少ない。そのため、主流であるロボット・エージェントとは別の、3DCG を用いたエージェントでは、ユーザに与える印象も変わってくる可能性があると考えている。

## 2. 2 多人数会話における心的負荷の軽減

近年のロボット・エージェントとのコミュニケーションに焦点を当てた研究が盛んになっている。その中でも、複数のエージェントを利用したユーザの意思決定支援が、ユーザの退屈を軽減させる効果があるという報告や[8]、複数のロボット・エージェントにより作られた対話が、聞き手であるユーザの心的負荷を軽減させるという報告[9]など、複数のエージェントによる会話をもたらす効果が数多く報告されている。

本研究では、これらの研究報告から、複数のエージェントによる情報インタフェースにより、日常での常時稼働を想定した際のユーザのエージェントに対する印象の評価や、情報インタフェースから受ける心的負荷を軽減させるための、エージェント設計の指針を示すことを目標とする。

## 2. 3 MMDAgent

MMDAgent (図 1) [10]とは、名古屋工業大学国際音声技術研究所によって開発された、オープンソースの音声インタラクションシステム構成ツールキッ



図2 MikuMikuDance 画面

トである。音声認識、音声合成の機能を搭載しており、設定ファイルに、あらかじめ特定のキーワードと、そのキーワードを取得した際のエージェントが返す返答と動作をシナリオファイルに記述することで、エージェントとの疑似的な音声対話を可能にしている。

後述する MikuMikuDance と互換性があり、MikuMikuDance で使用・編集することのできるモデルデータをエージェントのモデルとして、モーションデータをエージェントの動作として、それぞれ MMDAgent で使用することができる。

本研究では、その拡張性の高さから、本システムのプラットフォームとして採用した。

## 2. 4 MikuMikuDance

MikuMikuDance (図 2) [11]とは、フリーソフトとして一般公開されている 3DCG ソフトウェアである。3D アニメーションの作成に特化しており、描画に Microsoft DirectX を用いることで、作成した映像の描画に長時間のレンダリングを必要せず、作成した動きをリアルタイムで確認できるという点が大きな特徴である。3D アニメーションを作成する上での必要な操作が 1 画面上にまとめられているため、操作が簡単であり、前述の DirectX を使用していることでソフトウェア自体の要求するマシンスペックも低く、初心者でも容易に 3DCG アニメーションを作成することができる。本研究では、実験用デモ作成時にモーションデータ作成ツールとして利用した。

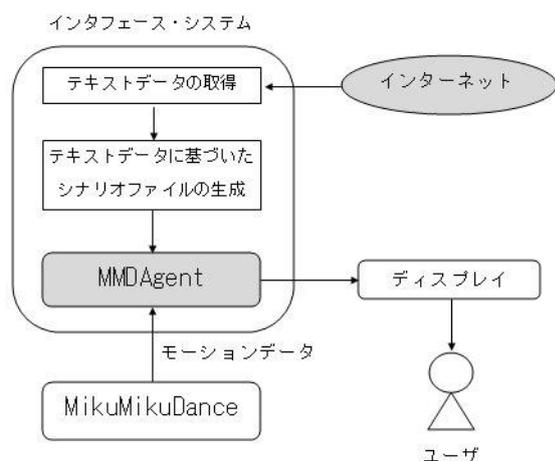


図3 MMDAgentを用いたインタフェースの構成

### 3. エージェントの対話による情報の提供

#### 3.1 コンセプト

エージェント二人によるエージェント同士の対話の場を作り、その対話の場を介することで、ユーザへ間接的に情報の提供を行う情報インタフェースである。日常生活での常時稼働を想定した、新たな情報インタフェースとなることを指標としている。情報の提供役のエージェントが、聞き手役のエージェントと対話することにより、ユーザの情報取得への介入の自由度を持たせ、ユーザ対エージェントという直接的な情報提供の場と比べ、ユーザの情報取得に関する心的負担を軽減させることを目的としている。

#### 3.2 提案する情報インタフェースの構成

提案する情報インタフェースの概要を図3に示す。インターネット上にあるデータ（ニュース等）を取得し、MMDAgentを経由してユーザへ情報を提供するシステムを想定している。具体的には、インターネット上から取得したデータを元に会話を生成、MMDAgentへ生成した会話データを送り、ユーザへ情報の提供を行う。

これに加え、エージェントの対話からユーザの興味のあるコンテンツを提供したり、ユーザが必要としている情報をエージェントに求めるといった利用法が考えられる。

表1 形容詞尺度を用いたSD法アンケート

1	積極的な	1 2 3 4 5 6 7	消極的な
2	人の悪い	1 2 3 4 5 6 7	人のよい
3	生意気でない	1 2 3 4 5 6 7	生意気な
4	ひとなつっこい	1 2 3 4 5 6 7	近づきがたい
5	にくらしい	1 2 3 4 5 6 7	かわいらしい
6	心のひろい	1 2 3 4 5 6 7	心のせまい
7	非社会的な	1 2 3 4 5 6 7	社会的な
8	責任感のある	1 2 3 4 5 6 7	責任感のない
9	そそっかしい	1 2 3 4 5 6 7	そそっかしくない
10	恥知らずの	1 2 3 4 5 6 7	恥ずかしがりの
11	重々しい	1 2 3 4 5 6 7	重々しくない
12	沈んだ	1 2 3 4 5 6 7	うきうきした
13	堂々とした	1 2 3 4 5 6 7	卑屈な
14	感じの悪い	1 2 3 4 5 6 7	感じのよい
15	分別のある	1 2 3 4 5 6 7	無差別な
16	親しみやすい	1 2 3 4 5 6 7	親しみにくい
17	無気力な	1 2 3 4 5 6 7	意欲的な
18	自信のない	1 2 3 4 5 6 7	自信のある
19	気長な	1 2 3 4 5 6 7	短気な
20	不親切な	1 2 3 4 5 6 7	親切な

なお、本実験ではインターネット上のニュース取得からMMDAgentで使うことができるシナリオファイルの形式への返還は手作業で行った

### 4. 実験

#### 4.1 実験目的

本実験ではシステムのエージェントから、ユーザが受ける印象の評価を行い、今後予定している情報インタフェースの長時間稼働実験の際の実験項目及び使用デモを検討するための目安を立てることを目的とする。エージェントから受ける印象の要素を、音声、身体動作、会話構成の3つの項目に仮定し、それぞれ、音声ピッチの高低、身体動作の有無、会話構成の変化を比較できるよう複数のデモを作成し、それぞれの印象の差異を分析する。



図4 実験用デモの様子

## 4.2 実験設定

会話構成と身体動作の変化による印象を評価するため、本システムで動作する「一切の身体動作を行わない会話」と、相槌等に身体動作を交えた会話として、「一人がニュースを読み上げ、もう一方が一文毎に相槌を打つ会話」と、「ニュースをエージェントが一文毎に交替で読み上げていく会話」の計3パターンのデモを用意する(図4)。また、上記とは別に、音声の変化による印象を評価するため、標準で使われている音声のピッチシフトを上げたものと下げたもの(+3.0, -5.0, いずれも OpenJTalk[12]基準による)の各1パターンの、計2パターンのデモを用意する。音声の変化による印象実験の際は、動作による印象操作が行われないようにするため、エージェントは一切の身体動作を行わないものとした。

まず、被験者に身体動作の変化による印象を評価してもらうため、前述の3パターンのデモをランダムに見てもらい、その後それぞれのデモについて、形容詞尺度[13]を用いた7段階のSD法に基づくアンケート(表1)に回答してもらう。

その後、同様に音声の変化による印象を評価するための2パターンのデモをランダムに見てもらい、アンケートに回答してもらう。

被験者には、デモ閲覧に関して、擬人化エージェントの見た目は評価に加えないようお願いした。また、今回の実験では純粋な印象評価を目的としているため、負荷課題等は課さないものとした。

■ 相槌動作なし 平均値 ■ 相槌動作あり 平均値  
■ 交互読み上げ 平均値

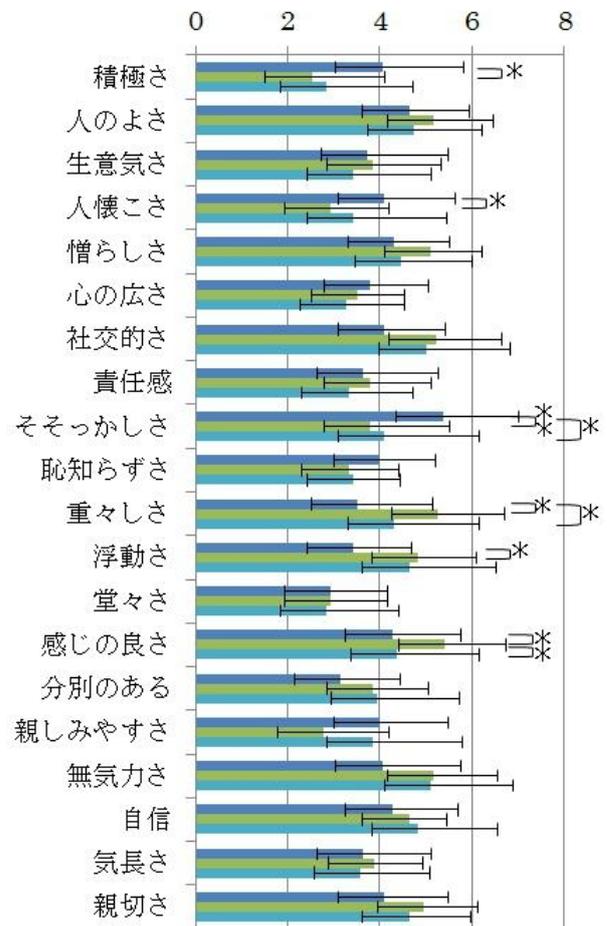


図5 身体動作および会話構成の変化についての印象評価

## 4.3 実験結果

今回の実験では、男性14名、女性5名の計19名(平均21.6歳)に協力をお願いした。

はじめに、身体動作および会話構成の変化について、印象評価の結果を図5に示す。また、各平均値に対して、それぞれt検定を行った。結果、一切の動作を行わない会話の評価の7割の項目の平均値が中間値の4.0付近(±0.5)に留まる結果となり、動作のある他2パターンは明確に差が表れる結果となった。

次に、それぞれの会話パターンとの比較を行った。まず、一切の身体動作を行わない会話と、一方が一文毎に相槌を打つ会話の比較結果について述べる。項目1(積極さ)、項目4(人懐こさ)、項目9(そそっかしさ)の計3項目が、一切の身体を行わない会話の平均値が高く、項目11(重々しさ)、項目12(浮

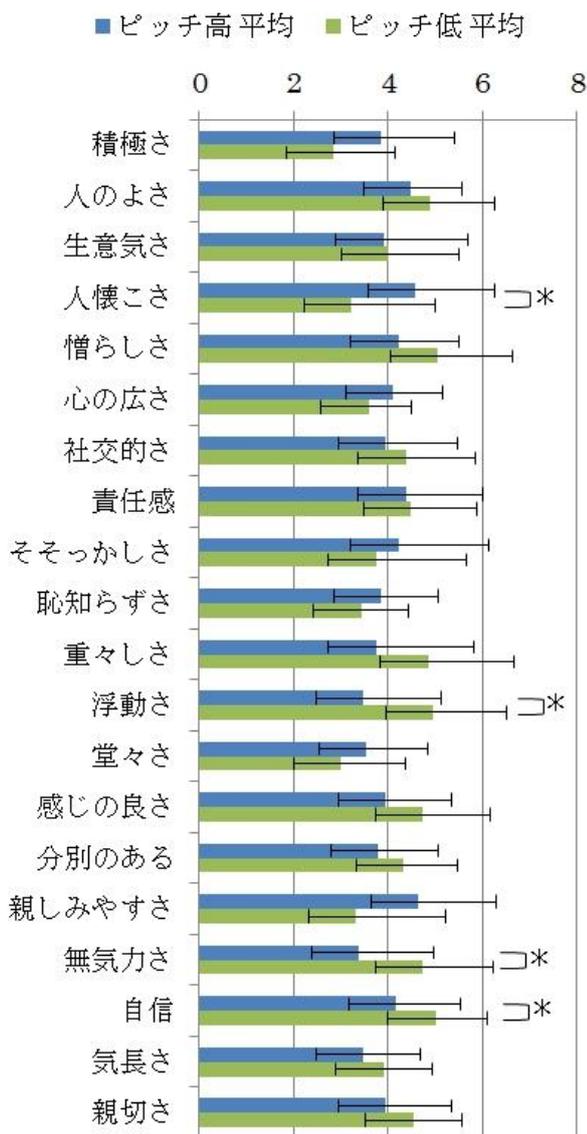


図6 音声の変化についての印象評価

動さ), 項目 14 (感じの良さ) の計 3 項目が, 一方が一文毎に相槌を打つ会話の平均値が高い結果になり, それぞれ有意差 ( $p < 0.05$ ) が見られた. また, その他の項目についても, 一文毎に相槌を打つ会話の方が, 全体的に親しみやすい印象を与える傾向が見られた.

次に, 一切の身体動作を行わない会話と, エージェントが一文毎に交替で読み上げていく会話の比較結果について述べる. 項目 9 (そそっかしさ) が, 一切の身体動作を行わない会話の平均値が高く, 項目 11 (浮動さ) が, エージェントが一文毎に交替で読み上げていく会話の平均値が高い結果になり, 有意差 ( $p < 0.05$ ) が見られた. その他の項目についても, 一方が一文毎に相槌を打つ会話程の差は見られな

ったが, エージェントが一文毎に交替で読み上げていく会話の方が, 全体に親しみやすい印象を与える傾向が見られた.

最後に, 一方が一文毎に相槌を打つ会話と, エージェントが一文毎に交替で読み上げていく会話の比較結果について述べる. 項目 14 (感じの良さ) が, 一方が一文毎に相槌を打つ会話の平均値が高い結果になり, 有意差 ( $p < 0.05$ ) が見られた. それ以外の項目については有意な差は見られなかった.

続いて, 音声の変化についての印象評価の結果を図 6 に示す. また, 各平均値に対して, それぞれ  $t$ -検定を行った.

音声のピッチシフトを上げた際の会話と下げた際の会話の比較結果について述べる. 項目 4 (人懐こさ) が, 音声の高い会話の平均値が高く, 項目 12 (浮動さ), 項目 16 (無気力さ), 項目 17 (自信のなさ) の 3 項目が, 音声の低い会話の平均値が高い結果になり, それぞれ有意差 ( $p < 0.05$ ) が見られた. その他の項目については, ピッチシフトを上げた際の評価が全体的に 4.0 点付近に留まり, ピッチシフトを下げた方が, 全体的に好ましい印象を与える結果となった.

#### 4. 4 考察

実験の結果として, エージェントの動作に関するそれぞれのパターンを見たことによるエージェントに対する印象の変化を見ると, エージェント同士の会話に相槌等の身体動作を加えることで, ユーザへの印象付けに大きく効果があることが確認できた. また, 会話の構成によっても, ユーザに与える印象が変化することが確認できた.

一切の身体動作を行わない会話の印象評価の 7 割の項目が平均 4.0 点付近だった結果の要因の一つとして, エージェントの身体動作がないことによるエージェントそのものに対する印象の判断要素の少なさから, 印象評価の点数が大きく変動しなかったものと考えられる. そのため, 身体動作を交えた会話の印象評価の際, 印象判断の評価となるエージェントの動作が加わるため, エージェントに対する印象の判断要素が増え, 評価の点数に差異が見られるようになったと考えられる. また, このことから, エージェントの動作によって, ユーザへ与える情報の印象付加を行うことが可能であると考えられる.

一切の身体動作を行わない会話と, 一方が一文毎に相槌を打つ会話を比較した際に, 積極さ, 人懐こさ, そそっかしさ, 重々しさ, 浮動さ, 感じの良さの 6 項目について, 後者のパターンの方がユーザに好ましい結果を与えることになった要因として, 身

体動作が加わったことによって人間らしさが表れたためと思われる。また、一方が一文毎に相槌を打つ会話に比べて、一切の身体動作を行わない会話とエージェントが一文毎に交替で読み上げていく会話を比較した際に、大きく差異が表れなかった要因については、お互いが文章を読み上げるだけになってしまうために、動作が相手に話しかける動作のみと、単調になってしまったためと考えられる。

音声の変化による印象評価の結果、ピッチシフトを下げた場合の方が全体的に評価点の振れが大きかった点について、音声の自然さを損なわないために、実験設定の段階でピッチシフトを高くした場合と低くした場合の変化量が均一に設定しなかったことが原因と考えることができる。また、標準の音声元から高めであったこともあり、ピッチシフトを上げた際の変化を被験者が明確に感じられなかった可能性もある。また、今回はMMDAgentの仕様上、女性の音声のみの実験となったが、男性の音声で同様の実験を行った場合、今回の結果と違った結果が表れる可能性も考えられる。

## 5. おわりに

本研究では、常時稼働を想定した擬人化エージェントからユーザが受ける印象の評価を行い、会話構成や身体動作、音声の違いから、ユーザがどのようなポイントに着目して情報を得ているのかを調査した。その結果、会話中におけるエージェントの身体動作が、ユーザの会話内容の解釈や印象付けに大きく影響することがわかった。また、会話中に相槌を挟むことで、ユーザへ与える情報の印象付加が可能であることがわかった。本実験の結果を元に、同一環境下で長時間稼働させた場合、ユーザに実際に与える印象が好意的なものであるか、また、ユーザの心的負荷を軽減させることができているのか調査を行う。

## 謝辞

本研究の一部は 中山隼雄科学技術文化財団の助成による。記して感謝する。

## 参考文献

- [1]小松孝徳, 山田 誠二: エージェントの外見がユーザの情報解釈にどのような影響を与えるのか?, HAI シンポジウム 2006, 20S-A-4, 2006
- [2]中田亨, 森武俊, 佐藤知正: ロボットの身体動作表現と生成される印象とのラバン特徴量を介した定

量的相関分析, 日本ロボット学会誌, Vol. 19 No. 2, pp. 1-8, 2001

[3]村川賀彦: ロボットによる販売促進活動の評価, HAI シンポジウム 2009, 1A-6, 2009

[4]山本大介, 小林優佳, 土井美和子: 高齢者対話インタフェース—対話誘導による問診対話—, HAI シンポジウム 2010, 3C-5, 2010

[5]木村幸士, 湯浅将英, 武川直樹: 多人数エージェントによる会話の雰囲気生成 —文字ばかり読んでないで空気読め—, HAI シンポジウム 2007, 1E-3, 2007

[6]中野有紀子, 村山敏泰, 西田豊明: 会話エージェントにおける情報提供—非言語情報による重要概念の強調—, 社会技術研究論文集 Vol. 2, pp. 159-166, 2004

[7]山崎達也: 擬人化エージェントの動作によるユーザ印象の評価, 映像情報メディア学会誌, Vol. 56, pp. 62-64, 2002

[8]安田淳志, 倉本到, 辻野嘉宏, 水口充: 互いに対立する複数エージェントとの会話によるユーザの意思決定支援手法, ヒューマンインタフェースシンポジウム 2010 論文集, pp. 127-134, 2010

[9]吉池佑太, P. Ravindra De Silva, 岡田美智男: 多人数会話の場に基づくソーシャルインタフェースの提案とその応用, HAI シンポジウム 2010, 1B-3, 2010

[10]MMDAgent: <http://www.mmdagent.jp/>

[11]MikuMikuDance:

<http://www.geocities.jp/higuchuu4/>

[12]OpenJTalk:

[http://www.sp.nitech.ac.jp/demo/open\\_jtalk/](http://www.sp.nitech.ac.jp/demo/open_jtalk/)

[13]堀洋道, 吉田富二雄: 心理測定尺度集Ⅱ 人間と社会のつながりをとらえる<対人関係・価値観>, サイエンス社, pp. 5-8, 2001