

発話マインドに基づく発話交替モデル — 気持ちが読める会話インタフェースを目指して —

湯浅 将英 徳永 弘子 武川 直樹 †

本論文は、擬人化エージェントやロボットなどのコンピュータのインタフェースとの円滑な対話を実現するため、新しい発話交替モデルを提案する。人は日常会話で「話したい／話したくない」という「発話マインド」の表出と他者の発話マインドを推測する能力があると考えられる。提案するモデルには、会話参加者がお互いに他者のマインドを推測する機能を導入する。実際の人の会話の観察／分析実験により、発話交替時に人が発話マインドを表出しながら、お互いにその解釈をしていることがわかった。さらに、擬人化エージェントの視線、表情、体の動きにより発話マインドを表現したときに、人はそれをマインドとして解釈して、発話をするかしないかを決定していることがわかった。これらにより、提案するモデルの妥当性が確認された。人の認知能力を利用した本モデルの概念は、人とコミュニケーションをする擬人化エージェントやロボットの設計論に重要な貢献をすると考える。

A Turn-Taking Model based on Utterance Mind-reading — Towards New Conversational Interface for Heart-to-heart Communication —

MASAHIDE YUASA HIROKO TOKUNAGA AND NAOKI MUKAWA †

In this paper, we propose a novel turn-taking model based on utterance mind-reading and apply to animated agents. In everyday conversation, people use mind-reading ability to infer other's mind to want to speak or not to speak. To investigate the utterance-mind-reading ability, first we observe the actual chatting video scenes and transcribe participants' behavior and utterances. The observations confirm that the utterance mind exists and that we can estimate participant's mind. Finally, by using the agent demo system, we investigate relationship between nonverbal behaviors and user's inferences of minds and a user's actions that start to speak. Experimental results show that participants can read the agent's mind from the agent's nonverbal behaviors. The model and the result will contribute to realization of innovative animated agents and robots that communicate with human in conversation.

1. はじめに

人同士の会話では、会話の発言内容（言語情報）が相手に伝達されるだけでなく、表情や顔の向きなどの非言語情報によって他者の「気持ち」や「思い」をうまく推測し、それらをお互いに適切に通じ合わせている。もし、お互いの気持ちや思いが推測できないと、相手の主張の背景が理解できずに、誤解に基づく間違った行動をすることになる。たとえば、半開きの窓を閉めてほしいときに、窓際の人に困った顔で「窓を」と言ったにもかかわらず、窓を全開にされてしまうような場合、「寒くて困っているので閉めてほしい」という気持ちのやり取りがうまくいっていない。

本研究は、人の持つ「気持ち」や「思い」を「マインド」と呼ぶこととし、人の会話のコミュニケーションが、他者の行動に基づき他者のマインドを解釈し、そのマインドから自分自身のマインドを生成し、さらに相互のマインドを理解することによって成り立つという立場から出発する。ここで「マインド」とは、お互いに読みあうことが可能な意図や感情であり、人が他者の気持ちを理解する過程を説明した「心の理論 (Theory of Mind)」の概念¹⁾に対応する。

本研究は、会話のコミュニケーションとして特に「発話交替」^{2)~4)}に注目する。発話交替は、話者が次々と変わることを指し、同時に話せるのは一人であること、全員が沈黙するのは避けられるなど、会話の交差点としての役割をもつ。発話交替においては、「話したい／話したくない」という「発話マインド」もやり取りされる。相手の話したいマインドを読み取ったら、自

† 東京電機大学 情報環境学部 情報環境学科
Department of Information Environment, School of Information Environment, Tokyo Denki University

分が話すのを控えて相手の主張に耳を傾ける、あるいは交渉時の会話であれば、他者の話したいマインドを遮って話し出すなど、さまざまな行動がとり得る。そのため、発話交替は「マインド」の交差点でもある。発話交替は、その短い時間に微妙な非言語情報のやりとりがされ、「発話マインド」が切り替わり、それ以降の会話の流れや雰囲気を大きく左右する。

本論文では、発話マインドを考慮した発話交替モデルを提案する。そして、そのモデルを擬人化エージェントやロボット等の会話インタフェースへ応用し、気持ちの読める会話インタフェースの実現を目指す。以降では、実際に会話している人の行動を観察して、発話マインドのやりとりを探り、続いて、擬人化エージェントを用いた人との会話インタラクション実験により、提案したモデルを検証する。

次章では、本研究に関連する先行研究について述べる。3章では、発話マインドを考慮した発話交替モデルを提案する。モデルを検証するため、4章では、実際の会話の観測から発話の交替と発話マインドの分類を探り、5章では、擬人化エージェントを用いた実験を述べる。6章で議論し、7章でまとめを述べる。

2. 従来研究

社会心理学や認知心理学において他者の心を理解する過程を述べた「マインド」の研究によれば、人は、表情や視線、頭の動き、体の姿勢などの観察された振る舞いから、他人のマインドをすばやく推測する能力を持つとされている^{6)~9)}。小野らが、人がロボットのマインドを推測することで、不明瞭な音声であってもそれが補完されることを示している⁹⁾。発話マインドについては、Poggiらが、話者の声の音量や韻律などの非言語情報の表現と発話欲求の関連を分類している⁶⁾。しかし、それは主観的に分類したものにすぎず、実際に人がそのように行動し解釈するかは確かめられていない。もし、非言語情報と発話欲求の具体的な関連が分かれば、人の発話マインドの振る舞いを理解したエージェントやロボットが設計できる。また、エージェントやロボットが発話マインドの振る舞いを表出したときに、人がそれを理解できる。本研究では、実際に人間の観察から発話マインドを分類する。さらにそれを実装したエージェントの振る舞いが人に理解できるかを確かめる。

心理学や言語学では、発話交替と視線の振る舞いとの関係が分析され^{10),11)}、会話の話者によって見られた聞き手が、次に話し出す傾向があることが明らかにされている^{3),12),13)}。この知見を利用して、ユーザの

視線の動きを検出し、発話の交替を実現するエージェントやロボットが開発されている^{14)~17)}。たとえば、話者の視線や頭部の動きをカメラや赤外線、または指向性マイクで話者の声の方向を捉える等により、発話を制御するエージェントやロボットがある。しかし、従来までのエージェントやロボットの発話交替は、たとえば「ロボットは、話者を見なければならない」「話者が発話を終えるときに、ロボットを見たら、ロボットは話し出さなければならない」など単純なルールを用いている。このような単純なルールでは、人もそのようなルールで行動することが要請され、話したいときに思いのまま話することができず、活発な楽しい会話の雰囲気を作ることができない。また、お互いの「気持ち」や「思い」が分かりにくく親密さが感じられない欠点がある。このように人同士であれば当然のことが、これまでの会話インタフェースではまったく実現されていなかった。

3. 発話マインドを用いた会話発話交替モデルの提案

3.1 発話交替階層モデル

図1は、我々が提案する発話交替モデルである。簡単のために、二者間のコミュニケーションの場合を表現しているが本論文で扱う三人の会話にも適用できる。図は、発話が外界からの刺激（コミュニケーションの相手の視線や表情、頭の動き、体の動きなど）に対して階層的に判断、応答することを表し、下位層、発話マインド層、上位層以上から成り立つ。下位層は、従来の単純な発話交替ルールに基づいた発話交替に対応するもので、直接的にわかる相手の振る舞い（表情や視線、体の動きなど）の観察から、迅速に発話の行動をする役割を持つ。たとえば、聞き手が話者から「視線や顔がこちらに向けられた」「こちらの方向に話しかけられた」などから、「自分が話すべきである/話すべきでない」と単純にルールとして判断するものである。

中位の「発話マインド層」は、自分が「話したい/話したくない」という発話マインドを表出したり、他者の観察された振る舞いから他者の「話したい/話したくない」の発話マインドを解釈する層である。会話参加者は、話者の発話中（発話開始から次の発話交替まで）に発話マインドを持ち、それを表情や視線、体の動きとして表出する。発話交替時には、話者、聞き手がお互いに表出された発話マインドを解釈し、誰が話すべきかを判断し、会話の衝突と沈黙を回避する。たとえば、話者の会話の終了時に、一人の聞き手が身

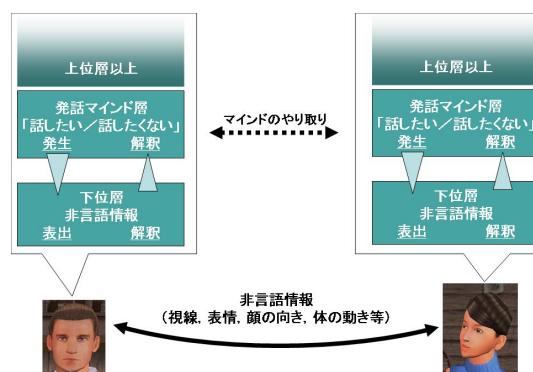


図 1 発話マインドを用いた会話発話交替モデル (図中のキャラクタは TVML⁵⁾ より)
 Fig.1 The turn-taking model using utterance mind (TVML⁵⁾ characters are used

を乗り出して口を開きかけていれば、他の会話参加者は、その人が話したいと予想する。聞き手が話者からの視線を逸らせば、他の人は、その人は話したくないと推測する。そして、人は話したいマインドを示している人には話してもらい、話したくないマインドを示している人には自分の方から話しかける。

上位層は、中位、下位の層からの情報を基に、自分が話すべきか否かを自身の目的や戦略に応じて決めることに相当する。たとえば、「相手にプレッシャーを与えるため、相手が話し続けていたとしても、大きい声で話者に割り込む」「自分が不利にならないように他の人の意見を探る必要があり、自分は何も言わない」など、長期的で複雑な計算に基づく発話交替に当たる。

前章で述べたように、従来までの対話エージェントやロボットは、話者の発話の終了を判断して発話交替するだけの単純なものが多い。進んだものでもカメラや指向性マイクから視線や顔がこちらに向いているかを判断し、単純なルールに従って発話交替を実現する下位層のみからなるものであった。しかし、このような発話交替は、司会者が次話者を指名することによる発話交替や、相手の視線や顔向きによって、必ず話したり沈黙したりしなければならない形式的な発話交替となり、会話に活発性が無く、親密性にも欠ける。我々は、発話マインド層を考慮した発話交替により、相手の気持ちを理解し活発性と親密性をもった会話を実現できると考える。

ここで述べた発話の階層モデルは下記の仮説を検証することにより確かめられる。

- 仮説 1 発話マインドの解釈人は、他者の視線や顔、表情、体の動きから「話したい / 話してほしい」という発話マインドを解釈できる

- 仮説 2 発話マインドによる発話人は、他者の発話マインドを解釈後、適切に発話行動ができる

以下、本論文では、実際の人の会話行動を観察、分析して、上記の仮説を確かめる。さらに視線 / 顔向きなどを変えられる擬人化エージェントと人とのインタラクション実験により仮説を詳細に検証する。

4. 会話映像分析による発話マインドの検証

本章では、実空間の三人会話を撮影した映像を観察、分析することにより、人の発話マインド表出、解釈、発話行動の変化を検証する。ここでは、著者らの研究に基づき、非言語情報による発話交替と発話マインドの関係について述べる。詳細は文献 4) を参照されたい。

4.1 話者の視線と発話交替

人同士の会話の分析には、親友同士の女性三人の会話を収録したビデオ (計 18 分) を用いた。会話分析ツールソフト「CIAO2」¹⁸⁾ を用いて発話内容、顔向き、視線方向、表情、ジェスチャなどのデータを人手により入力した。話者交替が円滑に生じうる区間 (Transition Safety Zone:TSZ) において、話者が発話終了時にどこを見ていたかの視線に注目した。

集計の結果、TSZ 時に話者の視線を受けて聞き手が発話するのが 51 %であったに対して、話者の視線を受けていない聞き手の発話が 24 %、話者の視線が誰も見ていないときの発話が 10 %であった。

よって、従来研究^{3),12),13)} と同様、話者によって見られた聞き手が、次に話し出す傾向があることが確かめられた。一方、話者の視線を受けていなくても、聞き手が話を始めることも多いことが分かった。また、複数の会話参加者が同時に発言し、それらがぶつかる場合や、話す人が誰もいなくなる場合は極めて少な

表 1 発話マインドと定義 (文献⁴⁾ より抜粋)
Table 1 The definition of utterance minds⁴⁾

発話マインド	名前 (分類名)	定義
強	発話立候補	次話者になりたい視線
	発話促進	次話者になってほしいが、むしろ相手の話を聞きたい視線 (相手の話への興味が優先)
	発話懇願	次話者を辞退したいので、相手に話してもらいたい視線 (自信がない、準備ができていない)
弱	視線回避	話者からの視線を避けた視線 (他の人に振ってほしい、自分を見ないでほしい)

かった。これは、話者が話しているときには、次に誰がしゃべりそうで誰がしゃべらないかの発話マインドを会話参加者が非言語情報から解釈し、発話の衝突や沈黙を避けて、会話がスムーズに継続したと考えられる。このときは、話者から向けられる視線は絶対的な役割を果たさず、人はより柔軟に判断、行動していると考えられる。

4.2 発話マインドの表出と解釈の確認

本節では、会話の参加者として特に聞き手に注目し、会話中に聞き手から発話マインドが表出、解釈されている様子を分析する。会話中の聞き手の発話マインドは、その聞き手の行動が他の参加者から観測されることによって解釈されるものと考えられる。他の参加者がどのようにマインドを解釈しているかは映像から読み取ることは難しいため、ここでは、会話映像を観察する三人の評定者を用意し、その評定者が映像シーン中の聞き手の発話マインドを読み取ることによって分析することとした。分析の対象となる映像は、前節と同様の女性三人による会話を用いる。

分析において、特に話者が発話中の聞き手の非言語情報 (視線、姿勢、表情、唇の動き、体の動きなど) に着目し、評定者らは発話マインドを解釈して分類した。分類は、各評定者ら三人の意見交換 (議論) で決定した。表 1 に結果の一部を示す。表中の「相手」とは「もう一人の聞き手」を指す。また、分類の名称については、わかり易さのため一部を変更した。評定の結果、聞き手の持つマインドとして今回の分類実験では、発話意欲の強度に応じた「発話立候補」「発話促進」「発話懇願」「視線回避」の 4 つの発話マインドの存在を確認した。なお、評定者は個々に映像を詳細に観察、評定した。評定後、全員で議論し、観察の対象としていた 196 箇所中 181 箇所 (92%) で分類の合意一致がなされた (議論前の一貫率は 62% であった⁴⁾)。このように、評定者により聞き手の発話マインドが十分に解釈可能であることが確かめられた。

次に、分類した発話マインドに従って、表出された発話マインドが相手に伝わって解釈され、マインドに対応した行動が引き起こされているかを確かめる。マインド通りに行動が起きれば、聞き手の「発話立候補」

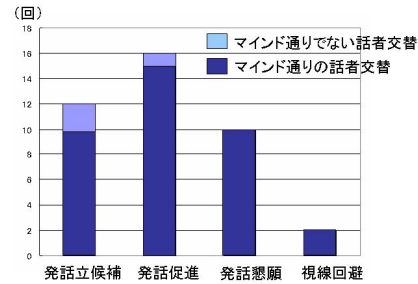


図 2 発話マインド表出と話者交替
Fig. 2 The utterance mind and turn-taking

のマインドを表出した後には、その聞き手が次話者となり、また、「発話促進」「発話懇願」「視線回避」のマインド表出した後にもう一人の聞き手が話し出す、あるいは話者が発話をそのまま継続すると予想される。評定マインドとその後の参加者の発話行動を確かめた結果を図 2 に示す。図より、各発言マインド通りの発話交替がされていることが確かめられた。

4.3 観察結果の考察

前節の結果は、評定者は会話中の視線や顔、体の動きなどから「話したい/話したくない」の発話マインドを解釈できることがわかった。このマインドは実際の会話中に他の参加者から観察され、利用されるものである。また、表出された発話マインドは、他の会話参加者らに解釈されて、参加者らの発話の有無に影響を与えて、マインド通りの話者交替が実現されることが分かった。これらの結果から前章の仮説 1 と仮説 2 が成り立つことが示された。

マインドの評定において、評定者は映像中の言語内容、視線、姿勢、表情、唇の動きなどの非言語情報から読み取っていたと述べている。これは、人間が視線や姿勢などの非言語情報から発話マインドを判断できることを意味する。よって、CG 合成によって構成される擬人化エージェントを用いて、同様のノンバーバル行動をさせても、人はそこから発話マインドとして解釈する可能性がある。

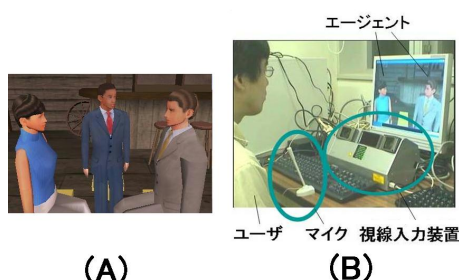


図 3 (A) 三体のエージェント, (B) 二体のエージェントとユーザ
Fig. 3 (A) Three agents, (B) Two agents and a user

5. 擬人化エージェントによる仮説検証実験

5.1 実験方法の概要

本章では、擬人化エージェントのノンバーバル行動によって発話マインドが読み取れるか、また伝わったマインドによって人が発話行動をしようとするかを確かめる評価実験について述べる。

評価を実施するため、擬人化エージェントによる「発話交替シミュレーションシステム (ARABAHIKA)」を開発した¹⁹⁾。本システムは、視線検出装置を用い、複数の擬人化エージェントのうち、ユーザがどれを見ているかを捉えて、自律的に動作するエージェントと発話の交替を体験するシステムである。また、実際の人間の言語ではないが、人間の言語を話しているように聞こえる無意味音声 (ARABAHIKA や UKUJARAH 等) を合成して作成し、エージェントに発話をさせている。これは、エージェント同士、あるいは人とエージェント間での発話の内容そのものによる発話交替を考えないようにするためである。エージェントへの実装には TVML⁵⁾ とその API を用い、さらに、ユーザがどちらのエージェントを見ているかを検出する視線入力装置は NAC 社の Voxer、ユーザからの音声入力は Julius²⁰⁾ を用いた。ARABAHIKA は、下記の二つのモード (A)(B) で動作する。

- 図 3(A) 三体のエージェント会話
画面中に三体が向き合ったエージェントを登場させ、外部からは、三者が話し合いをしているように見えるものである。
- 図 3(B) 二体のエージェントとユーザの会話
(A) の三体のうち、一体をユーザに置き換えたものである。赤外線を用いた視線検出装置を用いてユーザの視線を捉え、ユーザが発話し終わったときにどちらのエージェントを見るかを検出する。二体のエージェントは、それぞれ独立にユーザに見られたかを判断して発話する。

(A)(B) では、エージェントの行動を確率で調整できる。調整により、(B) ではユーザは「ユーザの発話終了時に、視線を向けられたエージェントが話し出す」状況、「ユーザの発話終了時に、視線を向けられなかったエージェントが話し出す」状況、「エージェントらの発話終了時に、ユーザが話に入りにくい」状況などが体験できる。

続いて、ARABAHIKA を用いた発話マインドの表出実験を述べる。始めに検証したい仮説、続いて仮説を検証するための実験方法、結果を述べる。

5.2 仮説

実験では、聞き手の擬人化エージェントの視線、表情、体の仕草といった振る舞いを、ユーザ (実験協力者) が発話マインドとして解釈するか、また、解釈した結果、ユーザがどのような発話マインド (話したくなるか、ならないか) を持つかを調べる。3章の仮説を詳細化した下記の仮説を立てる。仮説の前提として、ユーザの前の画面上には話者のエージェント、聞き手のエージェントがあり、以下の仮説では、聞き手のエージェントの行動に着目する。

仮説 1.1 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら体を上に伸ばすような仕草をすると、エージェントが視線を向けるのみの仕草よりも、ユーザには「エージェントが話し出そうとしている」マインドを持つように見える。

仮説 1.2 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら身を乗り出すような仕草をすると、エージェントが視線を向けるのみの仕草よりも、ユーザには「エージェントがユーザに話してほしい」マインドを持つように見える。

仮説 1.3.1 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら笑顔の表情をすると、エージェントが視線を向けるのみの仕草よりもユーザには「エージェントが話し出そうとしている」マインドを持つように見える

仮説 1.3.2 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら悲しい表情をすると、エージェントが視線を向けるのみの仕草よりもユーザには「エージェントがユーザに話してほしい」マインドを持つように見える

仮説 2.1 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら体を上に伸ばすような仕草をすると、エージェントが視線を向けるのみの仕草よりも、ユーザは「エージェントに話してもらおう」とする

仮説 2.2 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら身を乗り出すような仕草をすると、エージェントが視線を向けるのみの仕草よりも、ユーザは「エージェントに話したそう」とする

仮説 2.3.1 話者エージェントの発話終了時、聞き手エージェントがユーザに視線を向けながら笑顔の表情をすると、エージェントが視線を向けるのみの仕草よりも、ユーザは「エージェントに話してもらおう」とする

仮説 2.3.2 話者エージェントの発話終了時、聞き手エー

エージェントがユーザに視線を向けながら悲しい表情をすると、エージェントが視線を向けるのみの仕草よりも、ユーザは「エージェントに話したそう」とする

次いで、仮説を検証するための実験手順を説明する。

5.3 実験方法

擬人化エージェントに、視線、表情、体の動きなどの組み合わせにより発話マインドを表現する動作を実装した。図4に表現の例を示す。実験は、エージェントが発話マインドを表現した場面の後にエージェントの動作が停止し、そのときにユーザがアンケート評価をするものである。

発話マインドの表現動作は、実際の人の会話映像の観察と、Poggiらによる動作の意味の考察(文献6))を基に検討した。実際の発話マインドの仕草(口をわずかに開く等)と同等のCG作成が困難なものもあったため、様々なCGも用いて予備実験を繰り返すことで表現を絞り、10パターン(表2のP1-P10)を作成した。

なお、予備実験において、人に近いCGのキャラクターは「表情がわかりにくい」と指摘があったため、実験に用いた擬人化エージェントには、目、眉の動きが明確で、表情の動きがはっきりしているアニメ的なキャラクターを選んだ。

実験に用いたエージェントのノンバーバル行動のパターンを説明する。図4(1)の画面中の左のエージェントが話しているが、やがて話が終わり、発話交替すべきときになる。このとき、右のエージェントがユーザの方を見て、図4(2)のP1ようになる。この振る舞いは、視線による典型的な発話交替の動作に対応している。発話交替状況のときに視線を向けられることで、ユーザは自分が話し出すべきか否かを決めることになる。ここまですべて基本的な動作であり、この状況に

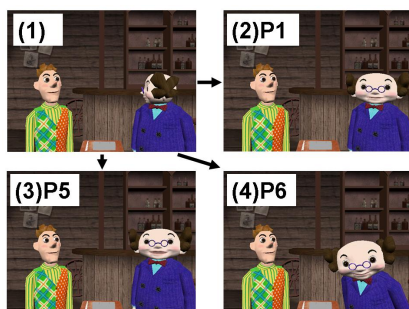


図4 実験で用いた振る舞いパターンの例、P1,P5,P6は表2に対応

Fig. 4 Examples of animated agent's behaviors. P1,P5 and P6 explain in Table2

なった後に画面が暗転し、ユーザに評価するように指示する。暗転をする前までに、視線だけでなく、体の動きと表情の動作を加えたものが表2のP1~P10である(図4(3)(4)も参照)。各パターンにおけるエージェントの行動から、エージェントの発話マインドの解釈、さらにユーザ自身の発話マインドを主観評価により調べる。ユーザは10のパターンを見て、下記の2つの質問に答える。

- 質問1: 映像の最後で、右の男の人は、どうしたい様子だと感じますか?
 - 1. 話をしたいと感じる
 - 2. どちらかといえば話したいと感じる
 - 3. どちらともいえない
 - 4. どちらかといえば話をしてほしいと感じる
 - 5. 話をしてほしいと感じる
- 質問2: 映像の最後で、あなたは、
 - 1. 話をする
 - 2. どちらかといえば話す
 - 3. どちらともいえない
 - 4. どちらかといえば話をしてもらう
 - 5. 話をしてもらう

質問1は、右のエージェントの発話マインドの解釈を尋ねたものである。選択肢番号が小さいほど、エージェントの話したいマインドが高い。質問2は、エージェントの発話マインドを解釈したユーザのその後の発話マインドを尋ねたものである。選択肢番号が小さいほど、ユーザが話し出すことを示す。

5.4 実験結果

理系大学生15人によって実験した。なお、エージェントのパターンは、人ごとに順番をランダムに並び替えて示した。表2の右二列に結果を示す。数値は、上記のアンケートの選択肢数値の平均を表す。表から、エージェントの発話マインド(質問1)は、「(P6)身を乗り出す」が最も値が高く(4.40)、「(P5)笑顔で体を上に伸ばす」が最も値が低い(2.00)と評価された。エージェントの発話マインドはエージェントの行動によって推測結果が明確に異なると考えられる。ユーザの発話行動(質問2)は、「(P5)笑顔で体を上に伸ばす」が最も値が高く(3.73)、「(P6)身を乗り出す」が最も値が低い(1.87)ことが得られ、それぞれについて、ユーザが話し出す/話し出さないの行動選択に顕著な違いが表れた。

単純な視線のみによるマインド表出(P1)と他のマインド表出パターンとの違いを調べるため、t-検定を実施した。発話マインド解釈(質問1)に関しては、P1と有意差があったものは、P3($p<0.05$), P4($p<0.01$), P5($p<0.01$), P6($p<0.05$), であった。ユーザの発話行動(質問2)に関しては、P1と有意差があったものは、P2($p<0.05$)P4($p<0.05$), P5($p<0.01$), で

発話マインドに基づく発話交替モデル—気持ちを読める会話インタフェースを目指して—

表 2 マインド表出パターンと評価の平均値 (N=15)
Table 2 Patterns of utterance mind and average scores (N=15)

エージェントの振る舞い	説明	質問 1	質問 2
(P1) 視線を向ける	ユーザに振り向くのみ、典型的な視線による発話交替	3.33	2.33
(P2) 笑顔で視線を向ける	P1 に笑顔の表情を付加	2.53	3.07
(P3) 悲しい表情で視線を向ける	P1 に悲しい表情を付加	3.93	2.13
(P4) 体を上に伸ばす	ユーザに振り向きながら、体とさらに頭部の位置を上移動する CG 表現で、体をピンと伸ばすような動作	2.13	3.12
(P5) 笑顔で体を上に伸ばす	P4 に笑顔の表情を付加	2.00	3.73
(P6) 身を乗り出す	ユーザに振り向いた後、ユーザに向かって身を乗り出すような前傾姿勢を取る	4.40	1.87
(P7) 笑顔の表情で身を乗り出す	P6 に笑顔の表情を付加	3.27	3.27
(P8) 悲しい表情で身を乗り出す	P6 に悲しい表情を付加	4.27	2.13
(P9) 首をすくめる	ユーザに振り向きながら、体とさらに頭部の位置を下移動する CG 表現で、首をすくめるような動作	3.20	2.67
(P10) 悲しい顔で体をすくめる	P9 に悲しい表情を付加	3.80	2.47

あった。以上の結果、仮説 1.3.1、仮説 2.2、仮説 2.3.2 は支持されなかった。他の仮説はすべて支持された。なお、質問 2 について P1 と P6 で有意差が見られず、仮説 2.2 は支持されなかったが、P5 と P6 では有意差 ($p < 0.05$) があり、エージェントが「体を上に伸ばす」か「身を乗り出す」かによって、ユーザは発話するか否かの選択を変えたといえる。また、二つの質問の選択肢数値には、強い相関 (相関係数-0.85) が見られ、エージェントの発話マインドがどのくらいと解釈するかで、ユーザは発話行動の選択を変えることが分かった。

これらの結果より、聞き手エージェントの視線、表情、体の動きの違いからエージェントの発話マインドをユーザは解釈したことが分かった。さらに、解釈した結果に基づいてユーザが自分が話し出すべきか否かについて行動選択を変えることも分かった。

人のすべての表現機能や動作を備えていない限定された動作の擬人化エージェントに対しても、人は、その表情や姿勢から、発話マインドが読み取れることが示された。

6. 議 論

エージェントを用いた仮説の検証実験では、表情に関しての仮説 1.3.1 と仮説 2.3.2 が棄却された。エージェントの笑顔、悲しみの表情の解釈の分散が大きい理由を予備実験の結果やユーザへのインタビューなどと合わせ考察する。笑顔は、本来うれしいことがあったときに表出されるものである。しかし発話交替時に聞き手エージェントが笑顔を表出すると、エージェントが話者からうれしいことを言われたとの解釈はできるが、その結果、エージェントが話したいのか否かを明確に決めるまでには至らない可能性が考えられる。悲しい表情についても、エージェントが悲しい状態に

あることはわかるが、次にエージェントが話すか否かを解釈することが難しい可能性がある。発話交替時にユーザがエージェントの発話の意欲を感じるためには、喜怒哀楽の表情よりも、「体を伸ばす」「首をすくめる」といった覚醒度や支配性の高低を表すような表現が必要であると考えられる。さらに、たとえば、あらかじめ覚醒度が高いことがわかっている振る舞いをエージェントに用いエージェントの発話意欲を示す、あるいは、支配性の高い振る舞いをエージェントが示しユーザに発話を促す、等の会話制御ができるだろう。今後追加の実験により、発話マインドの非言語表現の詳細を調べる予定である。

今回の映像観察実験では、友人同士の三人の会話を扱っている。通常の友人同士の会話では、会話が和やかな関係を維持することが重要となるため、他人の発話にわざと割り込むことは少ないが、交渉や議論など、調停や争いの場では、他人の発話を妨げたり、無視したりする場合が多くなる。これらの広範な会話タイプの発話交替での発話マインドも調べていく。

今回提案した発話交替モデルは、従来研究で得られているマインドに関する知見により、人がほぼ自動的に発話マインドを推測する能力を持つという前提に基づく。この能力は、相手がたとえ CG の擬人化エージェントであっても、人はマインドを見出し、行動を変化させることが得られた。コンピュータの性能を上げるのではなく、人の認知能力のマインドを利用することは、今後のインタフェース設計に応用できる重要かつ有用な手法であると考えられる。

これまでの対話インタフェースの設計では、人が表出する非言語情報からエージェントやロボットが感情を理解する機能と、ロボットの持つべき感情を非言語情報に表出させる機能とを非言語情報の単純なエンコードとデコードの手順に組み込み、「情報を入力して

出力を返す」という形式のコミュニケーションを目指していた。しかし、人同士のコミュニケーションは単純に情報を入力して出力を返すものだけでなく、気持ちや思いを通じ合わせるものである。本論文では、人によるマインド解釈を述べたが、さらに、コンピュータ側も赤外線カメラやセンサなどを用いて、人の動作から人のマインドを解釈する機能を持つことで、お互いのマインドを理解したインタフェースとなる。本研究のように、ある事象についてのマインドを数種類に分類し、コンピュータ側がいずれかのマインドを表現しながら、人側のマインドが何であるかも解釈することで、単に「コンピュータが反応を返してくれる」だけでなく、「コンピュータと気持ちが通じ合っている」ようなインタフェースが設計できるだろう。本研究は、コミュニケーションにおいて、特に発話交替に注目し発話交替モデルを提案したが、マインドに基づくモデル設計は発話交替に限らず、コンピュータと人との「気持ち」をやり取りする対話設計に重要な指針を示していると考えられる。

7. ま と め

本論文では、マインド推測の概念に基づいた発話マインドを定義し、新しい発話交替モデルを提案した。モデルを検証するため、実際の人間行動観察と擬人化エージェントによる実験を実施した。それらの結果、人が発話マインドを表出、解釈し、またエージェントで表現される発話マインドも解釈し、発話行動していることが分かり、提案したモデルの妥当性が確かめられた。マインドを用いたモデル設計概念は人とコミュニケーションをするエージェントやロボットの設計に重要な貢献をすると考える。

謝辞 本研究の一部は、文部科学省科学研究費補助金基盤研究(B)16300032(2006)、若手研究(B)19700119(2007)、および東京電機大学先端工学研究所重点課題、東京電機大学総合研究所研究 Q06J-14 による。

参 考 文 献

- 1) Baron-Cohen, S.: *Mindblindness*, The MIT Press (1995).
- 2) 石崎雅人, 伝 康晴: 談話と対話, 東京大学出版会 (2001).
- 3) 榎本美香, 伝 康晴: 3人会話における参与役割の交替に関わる非言語行動の分析, 人工知能学会研究会資料, SIG-SLUD-A301, pp.25-30 (2003).
- 4) 徳永弘子, 湯浅将英, 武川直樹: 3人会話の発話交替における視線行動と発話マインド分析-聞き手の立場から見た発話・非発話の戦略-, 信学

- 会 HCS, 映像情報メディア合同研究会, フォーラム顔学 2007 第 12 回日本顔学会 (2007).
- 5) TVML: <http://www.nhk.or.jp/str1/tvml/>.
- 6) Poggi, I., Pelachaud, C. and Caldognetto, E.M.: Gestural mind markers in ECAs, *In Proc AAMAS 2003*, pp.1098-1099 (2003).
- 7) Malle, B.F. and Hodges, S.D.: *Other Minds*, The Guilford Press (2005).
- 8) Sperber, D. and Wilson, D.: *Relevance: Communication and Cognition*, Blackwell Pub (1995).
- 9) Ono, T. and Imai, M.: Reading a Robot's Mind: A Model of Utterance Understanding based on the Theory of Mind Mechanism, *In Proc. AAAI-2000*, pp.142-148 (2000).
- 10) Kendon, A.: Some functions of gaze direction in social interaction, *Acta Psychologica*, Vol.32, pp.1-25 (1967).
- 11) Argyle, M. and Cook, M.: *Gaze and Mutual Gaze*, Cambridge University Press (1976).
- 12) Sacks, H., Schegloff, E. and Jefferson, G.A.: simplest systematics for the organisation of turn-taking in conversation, Vol.50, No.4, pp. 696-735 (1974).
- 13) Vertegaal, R. and Ding, Y.: Explaining effects of eye gaze on mediated group conversations: Amount or synchronization?, *In Proc. 2002 ACM conference on computer supported cooperative work* (2002).
- 14) 松坂要佐, 小林哲則: ROBITA: グループ会話ロボット, 人工知能学会研究会資料 SIG-Challenge-0113, pp.1-8 (2001).
- 15) Scassellati, B.: Theory of mind for a humanoid robot, *Autonomous Robots*, Vol.12, No.1, pp. 13-24 (2002).
- 16) Kojima, H. and Yano, H.: A robot that learns to communicate with human caregivers, *In Proc. of the First International Workshop on Epigenetic Robotics*, pp.47-52 (2001).
- 17) Eichner, T., Prendinger, H., Andre, E. and Ishizuka, M.: Attentive presentation agents, *In Proc. IVA-07*, pp.283-295 (2007).
- 18) CIAO: http://open.nime.ac.jp/software/ciao/index_content.html.
- 19) 湯浅将英, 徳永弘子, 武川直樹: 発話交替シミュレーションシステム ARABHIKA -人間観察に基づく自律的発話交替エージェントの提案-, インタラクシオン 2007 (2007).
- 20) Julius: <http://julius.sourceforge.jp/>.