様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月　 ５ 日現在

機関番号：１３７０１
研究種目：基盤研究(C)
研究期間：2011～2014
課題番号：２３５００３２７

研究課題名（和文）心を読むことによるコミュニケーションの創発

研究課題名（英文）Emergence of communication through mind-reading

研究代表者
伊藤　昭（Ito, Akira）

岐阜大学・工学部・教授

研究者番号：４０３０２３０１

交付決定額（研究期間全体）：（直接経費）　　2,500,000円

研究成果の概要（和文）：心を読むことに基づくコミュニケーションを計算機に実装可能なアルゴリズムとして検討した。主要な成果は、次のとおりである。
１．心を読むコミュニケーションの発生要件を「非零和ゲーム状況＝利害が完全には一致しないが協調を必要とする状況」と定式化し、人工的にその状況を生成することで、嘘やだましを含む心を読むことによるコミュニケーションを創発させた。
２．人が（人工物を含む）対話相手に心属性を付与する条件を、外見要因、行動要因の２面から調査した。また、心を読むことによるコミュニケーションの創発におけるメタ信号を役割を、身振りをコミュニケーションメディアとして用いて分析した。

研究成果の概要（英文）：We investigated "communication by mind-reading" as an computer algorithm. The main results are as follows:
1. The necessary conditions for the emergence of mind-reading communication is formulated as "a non-zero-sum game situation", i.e., the situation where the conflict of interests exists, but cooperation is necessary for good performance. Mind-reading communication with lies and betrayals emerged in an artificial situation satisfying the above condition.
2. How humans attribute "mind to (human or artificial) agents is investigated from appearance and behavioral factors. The role of meta-signals in the emergence of mind-reading communication is investigated using gestures as communication media.

研究分野：認知科学　人工知能

キーワード：心を読む　コミュニケーション　非ゼロ和ゲーム　創発　HAI

１．研究開始当初の背景

人と機械が言葉を使って対話をすることは、人類の長年の夢であった。現在、対話をする機械はカーナビから Siri まで、すでに実現されているようにも見える。しかしながら、それらの機械は対話内容を「理解」しているわけではないし、ましてや我々が日常行っている対話のように、相手の心（意図）を読んで対話しているわけではない。

我々は、これまで「心を読む」ことのアルゴリズムの解明と、計算機への実装を目指した研究を行ってきた。心を読むことの最終目的は、相手の行動を予測することである。我々のこれまでの研究で、人が行っている「心を読む」ことによる予測と、機械（計算機）が得意とする統計的予測との共通点・差異を明らかにしてきた。

簡単に言えば、心を読むということは、相手が心を持つと仮定し（意図スタンス）、心的概念を用いて相手の行動予測を行う（心の理論）ことである。これにより探索空間が制限され、相互に相手の行動予測が容易になり、結果として効率の良いコミュニケーションが可能となる。

しかしながら、心を読むことによる行動予測が成功するためには、相手も心を読む主体であることが必須である。このため対話者は、対話に先立って、相手に心があるかどうかを決定しなければならない（心の帰属問題）。

上記の知見をもとに、今回の研究では心を読むことによる人と機械とのコミュニケーションの実現のための、メカニズムの理論的解明を行う。

２．研究の目的

心を読むことによるコミュニケーションを人と機械とのインターフェースに導入することは、これまで「人にやさしい」など感性的な目標が掲げられて研究されてきた。確かに、我々の意図を汲んで行動してくれる機械があればうれしいと思う。でも、「心を読むことによるコミュニケーション」はそのような人の身勝手な願望をかなえてくれるものではない。

我々は、心を読むことによるコミュニケーションを、人がある種の機能を実現するために、必要に迫られて導入した手段である考えている。その機能とは、嘘をつく能力とモチベーションを持った者同士での、意味のある（有効な）コミュニケーションの実現である。

「自己の利益のためには嘘をつく可能性のある相手とはコミュニケーションが成立しない」と考えるのも、逆に「普通人は嘘をつかない＝嘘をつくのは、コミュニケーションに於いて例外的である」と考えるのも正しくない。利己的な自由人同士が、コミュニケーションの必要を感じた時、行動の自由を確保しながら、互恵的な情報交換を成立させるために編み出した手法が「心を読むことによるコミュニケーション」である。

この目的を達成するためには、参加者には高度の情報処理（計算）が要求される。心を読むことによるコミュニケーションは、人にとっても決して「使いやすい、人にやさしい」コミュニケーション手段ではない。

本研究では、上記のスタンスの下、人の心を読むことによるコミュニケーションの成立要因、創発の仕組みを解明することを目的とする。その際、嘘やだましなどの心を読むことに伴う言語現象が観測されることが、心を読むことの発現の一つの指標となる。

３．研究の方法

研究内容は、次の大きく二つに分けられる。

(1)心を読む能力を持った者同士で、どのように心を読むことによるコミュニケーションを創発するのか、また、そのための要件は何かを実験的に調査する。

人のコミュニケーションでは、自然言語はもとより、表情、身振り、視線など、あらゆるメディア（媒体）が必要に応じて動員される。またその意味も、歴史的、文化的に確立されてきたものであり、多くの研究者が様々な側面から研究を行ってきている。

我々は、そのような歴史的、文化的なしがらみにとらわれずに、「心を読む」メカニズムに焦点を合わせた調査を行うため、あらかじめ意味が共有されていない人工的な信号のみを用いて、人がコミュニケーション方法を確立する過程を分析する。

具体的には、コミュニケーションに使用する信号を予め意味の付与されていない単色信号や、単音信号に限定し、二人の被験者に協調を必要とするタスクを課すことで、そのコミュニケーション行動を観測する。得点構造を、完全に利害の一致する場合、協調が必要だが利害の対立が存在する場合、と条件を変えて行う。

その中で、心を読むことがどのような役割を果たしているのか、また自然言語に見られるような、嘘やだましなどの言語現象が人工的なメディアであっても観測されるのか、実験的に調査を行う。

(2)「心を読む」ことの前提として、相手に心を帰属する必要があるが、人はどのような人工物に心を帰属するのか。今回は、外観と振舞いの二つの要因について、実験的に調査する。

具体的には、コイン当てゲームにおける相手の行動予測、最後通牒ゲームにおける公平性への配慮、ちょっとした振舞いの違いによる信頼度の変化、ロボットの性能評価に対する見かけと行動の影響、など、様々な状況を作り、人の人工物への「心帰属」行動を観測する。

人が人工物に心を帰属したかどうかの判断には、アンケートによるもののほか、どのような行動予測をするのか、相手への信頼行動、逆にずるい行動、嘘やだましなどの行動、が発生するのかを、評価指標として用いる。

４．研究成果

研究の方法で述べたことに対応する形で、主要な研究成果を以下に述べる。

(1)様々な協調型タスクにおけるコミュニケーションの創発行動を観測した。いずれの場合にも、試行錯誤により与えられた信号に意味付与を行い、大多数の被験者ペアで、タスクを成功させるのに必要となる信号を確立させることに成功している。しかしながら、完全に利害が一致するタスクと、部分的に利害の対立のあるタスクでは、信号の使われ方に明確な違いが生じている。以下に、具体的な課題を用いた実験結果を説明する。

・単音信号を用いた迷路探索課題

これは、二人のプレーヤーが空間（迷路）を探索し、アイテムを取得して（ゴールで）出会う、という課題である。アイテムを取得してゴールすると、アイテムに応じた得点が、また歩数に応じたペナルティ（負の得点）が加算される。完全協調型タスクでは二人の総得点が、部分的に利害の対立する課題では、各自の得点を最大化することが目的となる。ただし、課題自体は一緒にゴールに到達することなので、相手との協調なくしては、高得点は望めない

適切なタイミングで相手と出会うためには、相手との情報交換が必須である。必要な信号は、自己の移動・位置情報、アイテムの取得状況、ゴールの発見などの情報の伝達である。完全協調型課題では、一方のプレーヤーが特定の意味で使用した信号が他方のプレーヤーでも模倣され、急速に信号の意味が確定していく。具体的には、相互に現在位置の確認、アイテム発見、ゴール発見などの信号を適切に交換することで、効率よくゲームをクリアしている。

一方、部分的に利害の対立のある課題では、プレーヤーの行動はそれほど単純ではない。様々な信号が、試行錯誤的に用いられるものの、相手からの信号の意味が推測できた場合でも、それが必ずしも模倣されるわけではなく、自分からは同じ意図を別の信号を用いて伝達することが生じる。結果として、完全協調問題に比較して、多様な信号が用いられる。

ペアの中には、協調性が高く、早い段階で信号の共有を果たすもの、逆に最後まで完全な信号の共有を実現できないものもいる。しかし、後者の場合でも、一定程度の信号共有を実現し、コミュニケーション無し条件よりは、良い結果を得ている。

さらに興味深いのは、利害の対立がある条件下では、次の二つの現象がしばしば観測されたことである。一つは、曖昧な意味の信号の使用、場合によっては相手に誤解を引き起こすような信号の使用があり、自然言語でいえば、嘘やだましの萌芽的現象だと考えられる。

もう一つは、強調、抗議、喜びなどのような「メタ信号」の出現である。これらは、観測された信号の交換が、それ自身の持つ情報の伝達にとどまらず、相手の意図の推測・制御、また自己の心的状態の伝達にまで及んでいることを示し、「心を読む」ことのコミュニケーションが創発している証拠であると考えている。

・単音信号を用いた非ゼロ和ゲーム対戦

2x2 利得表で表現される多数の非ゼロ和ゲームを、単音信号を唯一のコミュニケーション手段として、被験者ペアに対戦してもらった。用いた非ゼロ和ゲームは、協調を必要とするが、相手を出し抜くことで高得点を得られるもの、双方の手を同期させる必要のあるもの、相手の善意に頼るしかないものなど、様々な利得表の組み合わせになっている。こちらは、迷路問題に埋め込まれた非ゼロ和ゲームと異なり、相手の意図は、直ちに相手の行動（手の選択）として確認できる。

この場合交換すべき情報は、次に自分が・相手が何を出すかという２択情報のみである。ほとんどの被験者ペアでは、２択情報の信号の共有が実現した。

約半数のペアでは、全体で一番有利な行動を選択するという協調的行動が見られ、コミュニケーション行動は比較的単純であった。一方、残りのペアでは、非協調的行動がしばしば見られ、結果として多様な信号の交換が出現した。ただ後者の場合でも、完全にコミュニケーションが崩壊するわけではなく、行動の一致だけが必要な問題については、適切に情報交換が行われていた。

この問題では、迷路問題以上に相手の意図が行動に明確に表れるため、強調、修正要求、抗議、喜び、など様々なメタ信号の観測されている。しかしながら、１時間弱の実験時間の中で、相手側の非協調行動を変えさせられたペアは少なく、逆に相手の非協調行動に対しては、自らも非協調行動で対抗する、という戦略がとられることが多かった。

以上二つの実験をまとめると、たとえ利害の対立があっても、協調が必要だと分かれば、利用できる信号を使って試行錯誤的に信号の意味を確立し、必要なコミュニケーションを実現することができる。しかしながら、その結果発現する信号体系は、曖昧性、多義性のあるものであり、しばしば相手を誤解させるもの、意図的に行動と異なる信号を出すものなど、嘘やだましの現象が観測された。

さらに興味あるのは、強調、修正要求、抗議、喜び、など様々な「メタ信号」の出現である。これらは、そのままではゲーム遂行に有用な情報ではないが、自己の心の状態を伝達することで、相手の心の変化を促すなど、「心を読む」コミュニケーションに特有の現象である。

予め意味の付与されていない通信システムを用いても、利害が完全に一致しない非ゼロ和ゲーム状況下では、人は「心を読むことによるコミュニケーション」をその場で作り上げることができる、ということを示したのは、今回の研究の最大の成果であると考えて

いる。

(2)人工物（ロボット）の様々な「見え」や「振舞い」が、人の「心帰属」行動にどのような影響を与えるのかを、様々な状況・課題を用いて、実験的に調査した。
・ペニーマッチングゲームにおいて、インターフェース画面に何を表示するかで、プレーヤ（被験者）の行動を制御できることが分かった。人の顔画像を表示すると、人形の画像を表示するよりも、人は対戦相手（同じ計算機プログラムである）の「意図」を読もうとする傾向がある。逆に人形の場合は、相手の手のパターンを見つけようとする傾向がある。つまり、相手を機械と思えば予測に強化学習を用い、心があると思えば、意図を読もうとする。
・最後通牒ゲーム課題では、インターフェースに簡単な線画表情を提示したところ、表情を適切に制御することで、被験者をより公平に行動するよう促したり、逆により利己的に行動させられることが分かった
・人（被験者）が情報を持たない課題について、ロボットにアドバイスを求めるとき、ロボットの微かな「ためらい行動（反応時間が遅い、動作が機敏でない）」が、人の信頼行動に影響を与えることが分かった。微かなためらい行動は、人がロボットのアドバイスを採用する確率を大きく減少させる。
　これらすべての実験は、被験者に相手が人間であると思わせるようなものではないにもかかわらず、ある種の「見え」「動作」は被験者が対象人工物に「心属性」を付与する傾向を引き起こし、反応の差を生み出しているものと思われる。
・完成された外見のロボットと、配線など内部構造がむき出しのロボットについて、その機能、親しみやすさなどを評価してもらうと、評価は外見の完成された方が高くなる。ロボットには、いくつかの対人的振舞いが実装されており、ロボットとのインタラクションの後では、インタラクション前ほど、被験者の評価はロボットの外見に影響されなくなる。
　コミュニケーションとの関連でいえば、「心を読む」ことによるコミュニケーションは、両対話者が心を読みあうこと、特に相手に自分の心を読ませることが重要である。そのためには、対話している相手の機械にも「心がある」と人に思わせる、「心属性の付与」が重要になる。今回の一連の知見は、そのような、人が「心を帰属しやすい」人工物開発へ向けての大きなヒントになっている。

５．主な発表論文等
〔雑誌論文〕（計６件）

１．伊藤 昭、寺田 和憲、人工知能と心の理論ー心を持つロボットへの試み、査読有、臨床発達心理実践研究 Vol.9 (2014) pp.16-20.
２．寺田 和憲，山田 誠二，小松 孝徳，小林 一樹，船越 孝太郎，中野 幹生，伊藤 昭、移動ロボットによる Artificial Subtle Expressions を用いた確信度表出、査読有、人工知能学会論文誌 Vol.28, (2013) pp.311-319.
３．寺田 和憲，深井 英和，竹内 涼輔，伊藤 昭、振舞いに対する予測可能性が生物性と意図性の知覚に及ぼす影響、査読有、電子情報通信学会論文誌(D)，Vol.J96-D(2013) pp.1374-1382.
４．寺田 和憲、岩瀬 寛、伊藤 昭、Dennett の論考による３つのスタンスの検証、査読有、電子情報通信学会論文誌(A)、J95-A(2012) pp.117-127
５．寺田 和憲，山田 誠二，伊藤 昭、ボーナス付きマッチングペニーゲームにおける人間からエージェントへの適応プロセスの解明、査読有、人工知能学会論文誌、」Vol.27(2012) pp.73-81.
６．寺田 和憲、伊藤 昭、人間はロボットに騙されるか?--ロボットの意外な振舞は意図帰属の原因となる、査読有、日本ロボット学会誌 Vol.29 (2011)pp.43-52

〔学会発表〕（計９件）

１．Takakazu Mizuki, Akira Ito and Kazunori Terada, The sharing of meta-signals and protocols is the first step for the emergence of cooperative communication, The Second International Conference on Human-Agent Interaction(HAI2014),2014 年 10 月 29 日~ 31 日 筑波大学(茨城県つくば市)
２．K. Terada, Y. Imamura, H. Takahashi and A. Ito, A Fixed Pattern Deviation Robot that Triggers Intention Attribution, The Second International Conference on Human-Agent Interaction(HAI 2014), 2014 年 10 月 29 日~ 31 日 筑波大学(茨城県つくば市)
３．Kazunori Terada, Seiji Yamada and Akira Ito, An Experimental Investigation of Adaptive Algorithm Understanding, 35th annual meeting of the cognitive science society, 2013 年 07 月 31 日~8 月 03 日, Berlin, Germany.
４．Kazunori Terada, Chikara Takeuchi and Akira Ito, Effect of Emotional Expression in Simple Line Drawings of a Face on Human Economic Behavior, The 22nd IEEE International Symposium on Robot and Human Interactive Communication (Ro-man2013), 2013 年 08 月 26 日~ 29 日, Gyeongju, Korea.
５．Akira Ito, Shota Sobue and Kazunori Terada, How humans establish communication in non-zero-ssum game, The 22nd IEEE International Symposium on Robot and Human Interactive Communication (Ro-man2013), 2013 年 08 月 26 日~ 29 日, Gyeongju, Korea.

６．Akira Ito and Kazunori Terada, Mind-reading communication under the conflict of interests, The 1st International Conference on Human-Agent Interaction (iHAI2013), 2013年08月07日~９日, 北海道大学,(札幌市、北海道)
７．Kazunori Terada, Seiji Yamada and Akira Ito, Fixed Pattern Deviation Hypothesis of Intention Attribution, 2012 International Workshop on Human-Agent Interaction, (iHAI 2012), 2012年12月07日~09日 kyoto, Japan
８．Akira Ito, Yuki Goto, Kazunori Terada, Establishing communication in an artificial interaction environment, 21st IEEE International Symposium on Robot and Human Interactive Communication, 2012年09月09日~13日 Paris, France.
９．Akira Ito and Kazunori Terada, The Sharing of Meanings of Signals Through Limited Media in Two-player Games, 20th IEEE International Symposium on Robot and Human Interactive Communication (Ro-man2011), 2011年10月1日-3日 Atlanta, U.S.A.

〔その他〕
ホームページ等
http://www.elf.info.gifu-u.ac.jp/index.html

６．研究組織
(1)研究代表者
伊藤　昭（Akira Ito）
岐阜大学・工学部・教授
研究者番号：40302301

(2)研究分担者
寺田　和憲（Kazunori Terada）
岐阜大学・工学部・准教授
研究者番号：30345798