

2軸のアクチュエータのみを用いた意図伝達モデルの検討 Examination of Communication Model using Only Two-axis Actuators

大澤 博隆*¹ 桑山 裕基*¹ 今井 倫太*¹ 山田 誠二*²
Hiroataka OSAWA Yuki KUWAYAMA Michita IMAI Seiji YAMADA

*¹ 慶応義塾大学
Keio University

*² 国立情報学研究所
National Institute of Informatics #2

In this paper, the authors propose a variation of the Turing test with restricted communication. This modified test uses two-axis actuators for communication instead of the natural language dialogue used in the normal Turing test. This reduction of scope reveals what kinds of features are essential for an imitation game. First, we tried to determine what elements are critical for communication between a user and a robot through a preliminary experiment involving human-human communication. A human manipulator received a video image as input and controlled a "robot box" in a way that would lead a user to put other objects into the box. The results indicated what kinds of behavior are required to show the intention of the manipulator to the user. The results also indicate that there are three distinctive phases to keep interaction. Second, we created a human-computer interaction program embodying the results of the preliminary experiment. The behavior of the robot was programmed using the user's head and hand locations as identified by a motion captures system. Third, we evaluated our program in a communication task whereby the human collected items according to the instructions of the robot box. The results indicated that the program could guide the user correctly and could communicate to them without knowing any background knowledge about them. This suggests that our variation of the Turing test found a low level of intelligence. Hence, we believe it is a way to show the level of artificial "intelligence" that can be achieved by today's technology.

1. はじめに

チューリングテストは模倣ゲームを通してシステムの知性を判定するテストである[Turing 1950]。オリジナルのチューリングテストは自然言語を介して行われている。しかしながら、チューリングテストは評価者の背景知識や状況、前提条件に強く依存している。Cohen はこれらへの依存が知性判定テストを難しくしていると考えた[Cohen 2005]。Hayes らはテストの曖昧性が、チューリングテストから実用的な応用が生まれるのを阻害していると考えた[Hayes et al. 1995]。

一方でバーチャルエージェントやロボットといった擬人化エージェントが社会で使用されつつある[Fong et al. 2003]。これらのエージェントは必ずしもチューリングテストで求められるような知性のレベルを満たす必要は無い。それよりも、ユーザーを惹きつけインタラクションを継続させるだけの振る舞いが要求されている。例えば、ビデオゲームのエージェントは知性的に振舞うことで結果としてユーザーをゲームに没入させる。AIBO や PARO などの動物型のロボットやエージェントは、人間とのインタラクションの継続のために自律的に振舞うことが求められる[Fujita 2004, Shibata 2002]。また、半自律ロボットも知性的な振る舞いを要求される。Glas らは、ロボットが半自律的に行動し、高度な知性を要求される時にのみ人間に操作を渡すことで、結果として複数の擬人化エージェントを人間の要求に合わせて知的に振舞わせ、サービスを継続させられると提案している[Glas 2008]。このような半自律制御は、デジタルパペットのようなバーチャルエージェントの操作負荷を下げつつ、より自律的に振舞うように見せかけるために応用することも可能である[Coutrix 2010]。

このように、現在は古典的チューリングテストが想定したような、人間と同じように振る舞える知的レベルを要求したアプリケーション

は少ないが、それでも、ある程度の知的レベルで振る舞えるだけの自律的行動をシステムに要求しているといえる。この観点から、チューリングテストを改良することを考えると、従来のような自然言語を用いたコミュニケーションを中心とするのではなく、動作やジェスチャ、タイミングなど、もっと低いレベルで達成される、知性を感じさせる反応の質を試すような新しいテストが必要になると考えられる。

言語を除いたチューリングテストはまだそれほど評価されてはいない。自然言語以外のチューリングテストというアプローチはいくつか存在する。例えば、Harnad の提案した Total Turing Test がその一つである[Harnad 1991]。このテストでは自然言語だけでなく、人間自身の行動によって達成されるチューリングテストを仮定している。石黒らは Geminoid と呼ばれる人間を模倣したアンドロイドを作成し、このテストを実際に行なっている[Sakamoto et al. 2007]。アンドロイド研究は人工知能にとって重要な指標ではあるが、このようなテストは自然言語を介したもののよりさらに条件が多くなってしまうため、前節で述べたような要求を満たすテストとは言えない。一方で、Mckinstry らは、与えられた質問に対し正誤のみを返す形でシステムのアウトプットを絞ったチューリングテストのバリエーション、Minimum Intelligent Signal Test を提案している[Epstein et al. 2009]。本試みはオリジナルのチューリングテストに比べ簡略化されているが、自然言語に対する理解を要求しており、前節の要件から考えると設定が複雑すぎる。また、生物の動作を模倣するという関連研究がいくつか行われているが、これらの研究は質問者からのインタラクションを含まず、観測のみとなるため、チューリングテストの要件を満たしていないと考えられる。

以上の関連研究を考え、本研究では二軸のロボットのみを用いた、簡略化されたチューリングテストのバリエーションを提案する。本バリエーションではシステムからユーザーへの出力は2軸のモータのみに限定されている。そして、ユーザーからシステムへの入力ユーザーの顔の位置・向きと両手のそれぞれの位

置・向きのみである。2軸のモータが達成する振る舞いは非常に簡略化されているが、ヒューマンロボットインタラクションの既存研究より、このような簡略化された動作であっても十分に意図を伝えられることが示されているため、我々は2軸の動作がチューリングテストを満たし、十分に簡略化されており、かつ現在のロボットやエージェントの設計に対し、十分に貢献できる条件であると判断した[Yuasa 2010, Yamaji 2010]。

我々はまず、どのような要素が人間と二軸のロボット間のコミュニケーションで重要となるか検討するため、人間がロボットを動かす、人間-人間間のインタラクションを行った。本実験では、操作者である人間が、音声を使わず遠距離からカメラの映像を見ながら、インタラクション対象者の前に置かれた2軸のロボット箱を操作する。この操作によって、ロボット箱の側に置いてある物体を箱の中に入れるように誘導する。本実験結果によって、操作者がユーザーに対して意図を伝えるために、どのような振る舞い、タイミングが意図伝達に必要なものであるか調べる。次に、実験結果を元にして、操作者の動作を真似、ロボット箱を自律的に動かすためのシステムを作成する。最後に、本システムを実際に被験者に体験させ、ユーザーを誘導し、箱の側にある物体を集めさせた。事前実験と本評価を組み合わせることで、人間-人間間と人間-ロボット間のインタラクションを比較したチューリングテストが達成され、本研究ではその成功率を検討する。

2. 事前実験: 2軸のロボットを介した人間-人間間のインタラクション

本事前実験はロボットの操作者とロボットのユーザーの2人の被験者を用いる。操作者として、2人の人間(男女1人ずつ)を用いた。操作者はジョイスティックを使って2軸のロボットを制御する。操作者はロボットのユーザーと隔離されており、音声でのインタラクションも行えないようになっている。

8人の被験者(女性2人、男性6人)がロボット使用の体験を行った。被験者は実験室に入り、10分間のインタラクションを行った。被験者は実験室で何が行われるか、どのような要求が為されるかを事前に一切指示されていない状況であった。また、ロボットが操作者によって操作されていることも事前に知らされていない状態であった。このように説明を行わないことで、ユーザーが実験の目的について予め知識を得てしまうことを防いだ。

我々は2軸のロボットを実験室においた。図1左にロボットの外面の様子を示す。ロボットはそれぞれ0.9度ずつの精度で回転できる2つのサーボモータを持っており、Pitch方向に ± 15 度、Yaw方向に ± 90 度の回転を行うことができる。モータはマイクロコントローラを介し、有線で直接ジョイスティックにつながれており、操作遅延は100ms以下となっている。100ms以下の操作遅延はインタラクションの支障にならないと考えた。実験の様子を図1右に示す。

操作者に与えた指示は以下の通りである。まず、机の上の全ての物を、ユーザーがロボット箱に移動するように誘導する。次にユーザーを椅子に座らせる。次に、机の上の残りすべての物体を、ユーザーがロボット箱に移動するように誘導する。最後に、ユーザーを椅子に座らせたままにする。10分後、ユーザーは自らのビデオを観察し、実験者からのインタビューに応じて実験中に何を考えて行動したか、答えた。また、ロボットの操作者に対しても同じようにインタビューを行った。

2.1 事前実験結果

インタビューの結果より、以下の事実が判明した。まず、操作者の否定の意図は、ロボット箱を左右に振るといふかぶりの動作

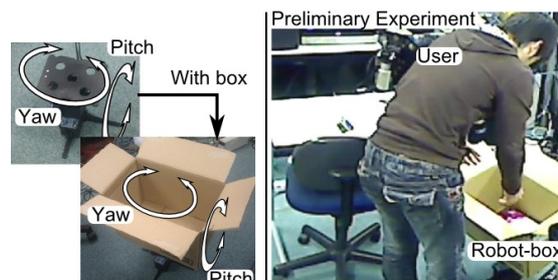


図1. 二軸のアクチュエータによるロボット箱の動作

によって伝えられた。また肯定の意志は、ロボット箱を上下に振るといふなずき動作によって伝えられた。

ユーザーに対する指示動作は、物体の正面をユーザーに向けて行われた。うなずき動作は、ユーザーがロボット箱の正面を同定する手助けを行った。さらに、物体に正面を向けてうなずき動作は、物体のおおまかな角度を示すのに役立った。しかしながら、物体の正確な場所については、ユーザーとの確認を何回か経たあと決定された。箱の上面を対象に向ける動作は、対象を箱に入れてさせるために有効だった。全てのユーザーは箱が物を入れて欲しいと伝えていることを徐々に理解した。また、ロボットが箱に物を入れるように指示したときにも、ユーザーはしばしば箱を観察し続けた。このような理解の齟齬は、ターンの交代を促す表現の必要性を示していると考えられる。

3. インタラクションモデルの検討

3.1 物体の正面同定

心の理論によれば、意図を持つと認識される物体は、正面に当たる側を備える必要があることが期待できる(Baron-Cohen 1995)。例えば動物の正面は顔と頭と認識出来る。またこれとは別に、エージェントは外界に対して働きかけるための側面を持っている。例えば動物は、自身の四足、口、耳、鼻、手を用いて外界に対し働きかける。このような単純化はロボット箱の場合にも有効であったと考えられる(図2)。物体の正面はユーザーに対しロボット箱の注意対象を伝えるのに適切であった。また、箱が外界に対し機能するのは、箱が開いた側であった。

本ロボット箱の場合には、正面はその振る舞いによって決定された。なぜならロボット箱は点対称な形をしており、どちらが正面に当たるか、事前知識の無いユーザーには決定できなかったからである。また、操作者自体も、動かし始めた当初はどちらが正面に当たるかを深く考えていなかった。しかしながら、物体を上下に傾けることにより、操作者とユーザーはともに、Pitchの回転を中心とした側が正面であると理解した。

本発見は次のように一般化できる。もし、物体が正面方向を同定するために特有の形状、テクスチャ、色などを保持している場合、そちらを正面と同定するのは難しくない。飛び出た正面や絵やマークなどがあれば簡単だった、と述べたユーザーもいた。しかしながら、球や円柱や円錐や箱などは、正面同定のためより多くの情報を必要とする。

正面同定の難しさの一方で、機能面を同定することはユーザーにとって難しくなかった。これは、ロボット箱の機能面に当たる部分が、物を入れたり取り出したりする部分しかなく、想定が容易だったためと考えられる。

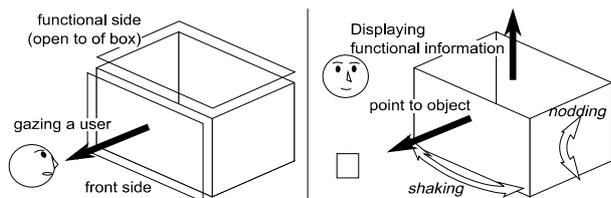


図2. ロボット箱の正面と側面。

3.2 意図とジェスチャ

我々はロボットの振る舞いを意図ジェスチャと制御ジェスチャに分類した。意図ジェスチャとして以下のものがある。例えばうなずき動作とかぶりを振る動作はそれぞれ肯定と否定を表す。また、正面方向を対象に向けることは、その対象に対し何らかの操作を期待していることを示唆した。正面方向をユーザーに対して向けることは、そのユーザーに対して行動要求を行ないたいことを示唆した。機能方向をユーザーに対して向けることは、その機能に関連する動作をユーザーが行うことを期待させた。

うなずき動作は特に、挨拶と誤認識されることが多かった。また、振動するような動作は否定的な意図を持つものとユーザーに伝わるが多かった。このような動作は、ユーザーが現在行っている、あるいは行った行動を止める要求と理解された。

一方で、制御ジェスチャには以下のものが見られた。繰り返し動作はユーザーが動作と意図の関係を発見するための手助けとなった。また、動作中の行動へのロボットの割り込みは、ロボットが次に行う行動をユーザーに想定させるためにより良く働いた。このような動作はロボットの要求に対してユーザーの理解を促進するように働いた。また同じような動作が常に同じ意図を伝えるわけではなかった。いくつかの振る舞いは、それが行われるタイミングに非常に依存していた。例えば、ロボットのうなずき動作は、ロボットの正面がユーザーに対し理解されないと伝わらない。うなずきやかぶりを振る動作は、行動の後に表出されないと、その行動に対する否定的な意図表出であることが理解されなかった。

3.3 インタラクションの3段階

結果を元に、インタラクションを以下の3段階に分類した。

1. 初期化段階: ユーザーはこの段階で、ロボットがエージェントであり、対話可能であることを理解する。この状態では、ユーザーはロボットの正面方向も同定する。ユーザーに理解をさせるための手法として、操作者はユーザーの移動に対し、ロボットの正面方向を追従させていた。このような初期化が充分に行われれば、インタラクションは次の段階に進んだ。

2. 観察段階: ユーザーはロボットからの指示に注目し、その指示がユーザーに対しどのような行動を要求しているか理解する。ロボットはユーザーに対し、ロボットの持っている機能を使うように要請する。このため、操作者はロボット箱の正面を物体に向け、機能面をロボットに対し向けた。また、ユーザーの行動が行われない場合、操作者はロボットの行動を繰り返した。

3. 行動段階: ユーザーは実際に行動を行い、操作者はロボットを介してその行動に対し肯定、または否定の動作を返す。操作者はユーザーの行動を観察しており、もしユーザーが物体を触った場合には、ユーザーに向けて機能面を向ける。一方、もしユーザーが操作者の意図と異なる物体を拾った場合、操作者はかぶりを振って否定の意図を伝えた。正しい物体がロボット箱に入れられた後は、操作者は次の物体への指示を行った。

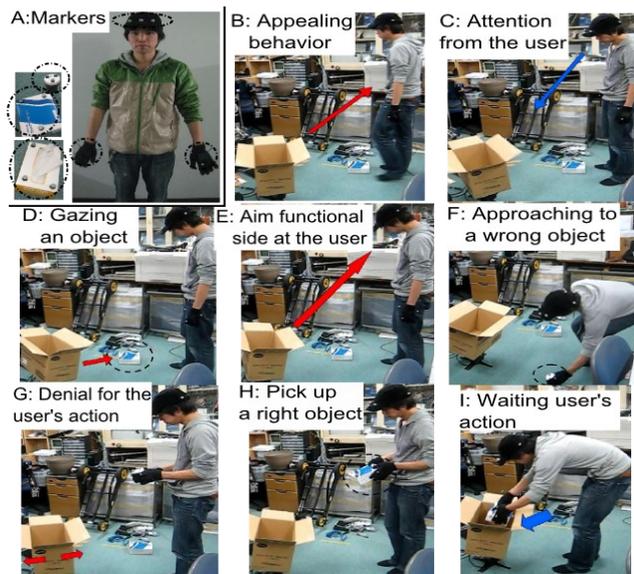


図3. 自動応答システムの動作例

4. 自動応答システムの実装

上記モデルを実現するための自動システムについて説明する。事前実験の結果より、ユーザーの頭と両手の向きがインタラクションに対し大きく影響していることが分かった。この3箇所にはマーカを取り付け、カメラの代わりに位置と向きをモーションキャプチャから得て、物体への接触やロボット箱への接触のデータを取り、人間の操作者を介さない応答を実現する(図3A)。

本システムはロボットの向きとユーザーの頭の向きを計算し、頭の向きが45度より小さいときは、ユーザーがロボットの方向を向いて注目していると判断する。また、ユーザーの手と物体との距離が150mm以下の時に、その物体に接触したと判断する。また、社会的距離の知見を反映し、ユーザーの頭の位置とロボット箱の距離を測定し、500mm以下の時に、ユーザーがロボットに対し働きかけを行ないたいものと推定する[Hall 1966]。

動作例を図3に示す。状態1では、ロボットがユーザーの発見を行い、正面方向をユーザーに向ける(図3B)。次に、ロボットはユーザーに対してうなずき動作を行い、どちらが正面方向であるかをユーザーに提示する(図3C)。ユーザーからの注目が5秒以上続いたときに、ロボットは次の状態に移る。

状態2では、ユーザーがロボットの動きを観察している。ロボットは自らの正面を物体に向け、次に機能方向をユーザーに対し向ける(図3D、図3E)。本動作によって、ユーザーに対し、物体が操作の対象になること、ユーザーが機能方向を使うことを誘導する。また、この動作を繰り返すことで、両情報提示に関連性があることを提示する。物体に近づいたときに、状態2が終わり、ユーザーの手番である状態3に進む。

状態3では、システムがユーザーの動作を観察し、行動が正しいか間違っているかを提示する。ユーザーが間違った物体を取得した場合(図3F)、ロボットはかぶりを振る動作を行う(図3G)。この動作はユーザーに対し、拾い上げ動作が間違っていることを示唆する。もしユーザーが物体を動かす動作をやめたならば、システムは状態2へ推移し、ユーザーに対し再び正しい物体を指定する。一方、ユーザーが正しい物体を取得した場合(図3H)、ロボットは機能面をユーザーに向け、ユーザーが物体をロボット箱に入れるまで動作を停止する(図4I)。物体がロボット箱に入れられたら、システムは再び状態2に戻る。この動作が全ての物体に対し終了するまで、状態2と状態3が繰り返される。

5. 本実験: 2軸のロボットを介した人間-計算機間のインタラクション

我々は3つの物体をロボット箱の前に置いた。ロボット箱は被験者に対し、そのうち1つの物体を箱に入れるように誘導する。被験者は大学院生10人で、女性2人、男性8人。年齢は22~24歳の間である。被験者が事前知識を持つことを避けるため、被験者には部屋に入るという指示のみ伝えた。

被験者が指定の物体を箱に入れた場合、被験者がロボット箱の意図を理解できず、観察しか行わない場合、被験者が物体を拾い上げた後、再び置き、その後行動を行わなかった場合のどれかの条件を満たした時点で、実験は終了する。

実験結果として、4種類の段階の成功例が見られた。A) 被験者が最初の試行で正解の物体を同定し、拾い上げる場合(7人の被験者が該当)。B) 被験者が最初の試行で正解の物体を箱に入れた場合(7人の被験者が該当)。C) 被験者が一度でも箱に正しい物体を入れた場合。(9人の被験者が該当)。D) 被験者が正解の物体のみを箱に入れ、実験が終了した場合(8人の被験者が該当)。

また、かぶりを振る動作は実験中3回確認された。2人の被験者はその動作に従って行動をやめた。

6. 議論

80%の被験者は意図伝達を成功させることが出来た(D)。本結果は我々が提案した3つの動作モデルを元にしたインタラクションがおおむね成功したことを示唆する。この結果は、たとえ2軸のように制約された条件であっても、インタラクションの段階を適切に設計することで、ユーザーの行動を上手く誘導できることを示唆する。

しかしながら、失敗例も確認された。例えば、否定の動作を喜びの動作と誤解した被験者が居た。これらの動作は文化依存である。機能面をユーザーから背ける、など、機能に依存しない表現を使うことで、よりユーザーに対し意図の理解を促進できるのではないかと考えられる。

また、ロボットの知能を過大評価した被験者も見られた。2つの物体をとって交互に見せた被験者がおり、この被験者はロボットがこの2つを比較し選択することを期待していたが、ロボットがこのような能力を持っていないことに気づかなかった。実際的な面から言えば、ロボットのデザインを簡素化することで、このようなユーザの過大評価をある程度防ぐことができると考えられる。一方で、ユーザはタスクに対して最適な知性を不可避的に想定しているのかもしれない。さらなる実験を行って、2軸のアクチュエータがどこまでの知性を想定させるか検討する必要がある。

我々の研究は言語に依存しないチューリングテストの可能性、低いレベルでの知性を評価するやり方と、その応用性を示した。しかし、我々の実験結果は未だ非常に単純な成果の達成に留まっており、他の研究分野の知見を改めて参照する必要がある。

7. 結論

本研究では言語の代わりに2軸のアクチュエータを使い、出力を人間同士が最低限意思疎通可能と見られる範囲まで制限したチューリングテストを提案する。本研究ではまず、人間同士で2軸のアクチュエータのみを介した会話を行い、被験者が対話者を想定するために、どのような要素が本質的であるか検討した。その過程で、ユーザー自身が達成可能な行動を相手に射影するという模倣規則と、ユーザーの行動とアクチュエータの動作のタイミングについての規則を得た。次に本模倣規則に伴い、ユーザーをモニタリングし、行動に合わせてアクチュエータ

の動きを再現し、物体を指示しユーザーに運ばせるタスクを行った。その結果、ユーザーの事前知識なしに、ユーザーに要求された行動を達成させることが出来た。

謝辞

本研究は独立行政法人科学技術振興機構の戦略的創造研究推進事業(さきがけ)領域「情報環境と人」の援助を受け行われました。

参考文献

- [Turing 1950] Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59, 433-460.
- [Cohen 2005] Cohen, P. R. 2005. If Not Turing's Test, Then What? *AI Magazine* 26(4), 61-67
- [Hayes 1995] Hayes, P., and Ford, K. M. 1995. Turing Test Considered Harmful. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 972-977. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence, Inc.
- [Fong 2003] Fong, T. W. 2003. A survey of social interactive robots, *Robotics and Autonomous Systems* 42, 143-166.
- [Fujita 2004] Fujita, M. 2004. On activating human communications with pet-type robot AIBO, Proceedings of the IEEE, 92(11), 1804-1813.
- [Shibata 2002] Shibata, T., Mitsui, T., Wada, K., and Tanie, K. 2002. Subjective Evaluation of Seal Robot: Paro - Tabulation and Analysis of Questionnaire Results -. *Journal of Robotics and Mechatronics*, 14(1), 13-19.
- [Glas 2008] Glas, D. F., Kanda, T., Ishiguro, H., and Hagita, N. 2008. Simultaneous Teleoperation of Multiple Social Robots, *Proceedings of ACM/IEEE 3rd Annual Conference on Human-Robot Interaction*, 311-318.
- [Coutrix 2010] Coutrix, C., Giulio, J., Spagnoli, A., Ma, L., Helin, M., Richard, G., Parisi, L., Roveda, S., and Narula, P. 2010. Engaging Spect-Actors with Multimodal Digital Puppetry, *Proceedings of the 6th Nordic Conference on Human-Computer Interaction*, 138-147.
- [Harnad 1991] Harnad, S. 1991. Other bodies, other minds: A machine incarnation of an old philosophical problem, *Minds and Machines* 1(1), 43-54
- [Sakamoto 2007] Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., Hagita, N., 2007. Android as a Telecommunication medium with Human-like Presence, *Proceedings of 2nd ACM/IEEE International Conference on Human Robot Interaction*, 193-200
- [Epstein 2009] Epstein, R., Roberts, G., and Beber, G. eds. 2009. Parsing the Turing Test.: Springer Netherlands.
- [Yuasa 2010] Yuasa, M., Mukawa, N., Kimura, K., Tokunaga, H., and Terai, H. 2010. An utterance attitude model in human-agent communication: from good turn-taking to better human-agent understanding, *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, 3919-3924.
- [Yamaji 2010] Yamaji, Y., Miyake, T., Yoshiike, Y., De Silva, P. R., and Okada, M., 2010. STB: Intentional Stance Grounded Child-dependent Robot, *Social Robotics, Lecture Notes in Computer Science*, 6414(2010), 71-80.
- [Baron-Cohen 1995] Baron-Cohen, S., 1995. Mindblindness: An Essay on Autism and Theory of Mind. *The MIT Press*.
- [Hall 1966] Hall, E. T., 1966. The Hidden Dimension, *Doubleday & Company*.