

特集 2



ロボットの注意機構と発話生成 そして身体表現

今井 倫太

慶應義塾大学

michita@ayu.ics.keio.ac.jp

実世界の情報を介したロボットとの コミュニケーション

本稿では、人間社会で自律的に振る舞うコミュニケーションロボットと人間のコミュニケーションについて取り上げる。ロボットは、コンピュータ上に描かれるCGキャラクタとは違い、現実世界で駆動可能な体を持つ。さらに、コミュニケーションロボットは、活動の場が工場といった閉鎖環境ではなく公共の場や家庭環境であり、一般の人間と直接接する。そこで本章では、実世界の状況を対象としたコミュニケーションに関して取り上げる。

人間同士が対面しコミュニケーションする場合、周囲の状況が話題として頻繁に現れる。たとえば会話の中で「あれ」や「これ」といった指示語が用いられる場合がこの例に相当する。ロボットが現実世界で活動することを考えると周囲の状況に関する話題が人間とロボット間でも取り上げられると思われる。そこで本稿では、周囲の状況を参照する際にロボットにとって必要となる要素技術について説明する。

周囲の状況を参照する際の要素技術には、音声対話の技術だけでなく、さまざまなモダリティ（たとえば、ジェスチャー）の取り扱いが重要となる。本稿では、注意機構および、共同注意、発話生成といった観点から周囲の状況の参照に必要となる技術について具体的に説明する。

インタラクションと注意機構

人間とロボットがコミュニケーションするためには周囲の状況をロボットが知覚していなければならない。特に、自然言語によるロボットとの対話を可能にするためには、意味の状況依存性¹⁾を考慮した自然言語処理が必要となる。たとえば人間は、「あっち行こう」といった発話を容易に理解できる。しかし、この発話を言語情報自体から直接理解することは不可能である。なぜなら「あっち」が何を指すのかは、言語情報だけでなく周囲の状況や文脈に照らし合わせて初めて分かるからである。人間は、会話の中で周囲の状況や文脈を無意識に追っているため理解が可能なのである。一方、ロボットが、同様な発話を理解するためには、センサから獲得される情報からいかに状況を知覚し、言語理解に反映させるかが肝となる。

周囲の状況やジェスチャーといった非言語情報を元に言語情報を取り扱う取り組みは、コンピュータ上の対話システムとして以前から研究されている⁸⁾。これらのシステムでは、実世界の情報を扱うよりもむしろ、コンピュータグラフィクス(CG)を用いた場合の非言語情報を取り扱っている。システムに命令することにより、CGで描かれた仮想世界上で物体を動かせる。システムは命令が与えられると、仮想空間内の物体の配置と命令を照らし合わせ命令を理解する。

代表的な研究としては、Winograd, T.の積木の世界の対話システム(SHRDLU)がある⁸⁾。SHRDLUが動作する世界は、積木と、積木を移動させることができるアー

ムロボットで構成されている（積木もロボットもCGで描かれている）。SHRDLUは、CGの場面と言語表現を結びつけるために次の知識を持つ。

1. 対話の場面に関する知識（テーブルの上にある積木の色および、形、サイズ、積木の位置関係、など）
2. 世界に関する一般的な知識（積木は、ロボットの腕でつかむことができる。別の積木の下にある積木は持ち上げることができない、など）
3. 対話の履歴の記憶

上記の知識を利用することにより、人間からの命令や質問に対して、積木を適切に移動したり、なぜその積木を移動したのかを説明することができる。たとえば、「大きな赤いピラミッドを青いブロックの上に移動しなさい」といった命令をすることができる。この例のほかに、対話の履歴を利用して「そのピラミッドを…」といった表現も解釈可能になっている。最初にあげた命令例の直後にこの命令を行えば、SHRDLUは、対話履歴から当然「大きな赤いピラミッド」を選択することになる。

以上の例からも明らかのように、状況を参照して行わなければならない言語処理とは、省略された事柄や代名詞で置き換えられた事柄までも含めて、言語表現と状況に対応づけることである。しかし、ここで挙げた研究が対象としている状況は、計算機の内部に構成されるCGの世界である。この点が実世界の状況を対象とするシステムと大きく違っている特長である。なぜなら、CGの世界では、積木間の位置関係がすでに計算機内部に蓄えられているからである。つまり、どの状況に注目すべきかといった状況認識の問題なしに言語処理を行ってしまう。この積木の世界での問題解決の話は、ブロックワールドやトイモデルの問題を解いているだけであると人工知能が批判されてきた要因の1つである。一方、実世界で駆動する対話システムを構築していくためには、状況の認識も含めて言語処理をとらえていかなければならないといった難しい課題がある。

実世界の状況をセンサから獲得し対話の処理に利用する研究は、マルチモーダルインタフェースの分野で盛んに行われている⁹⁾。マルチモーダルインタフェースとは、音声や画像といった複数 (multi) の入力様式 (modal) を組み合わせたユーザインタフェースのことを指す。人間からの多様な情報をとらえることにより、人間と機械のコミュニケーションを円滑なものにする技術である。たとえば、人間が機械の前にいるかどうかを画像処理で判断し、人間の有無に応じて音声認識を働かせることが可能である。このことにより、無人のときに起きてしまう雑音による誤認識を防ぐことができる。

マルチモーダルインタフェースの多くはコンピュータ用のユーザインタフェースであり、そのほとんどが、画像認識と音声認識を組み合わせたシステムである。たとえば、人間のジェスチャーをカメラで認識するシステムがある^{2), 6), 10)}。例としてヒューマンリーダー¹⁰⁾を取り上げる。ヒューマンリーダーでは、ユーザが、腕でコンピュータグラフィクスを操作することができ、腕の動きと音声でコンピュータに指示を与えることが可能である。また、タッチパネルを用いたシステムもある⁹⁾。地図が表示されているタッチパネルに触れながら、「ここに行きたいです」といった指示語を用いた情報入力が可能となっている⁹⁾。また、フットマット型のセンサを用いてシステムの前に人間がいることを判断し音声認識を働かせるシステムもある。

以上、自然言語処理システムが非言語情報を取り扱うことの重要性について説明した。また、実際に状況を言語処理に反映させる研究を紹介した。しかし、ロボットのための対話システムを構築する上でさらに解決すべき課題がいくつも存在する。特にロボットは、能動的に実世界を動き回りさまざまな出来事に対処する必要があるため、ただ単に実世界の状況をセンサで獲得するだけでなく次のことに注目しなくてはならない。

- 対話の話題と無関係に起きる状況や、ロボットと人間の行動によって生じる状況までも対話処理に利用する。

この課題を克服するためには、実世界からある情報を選択し注目する機構が必要である。従来の自然言語処理システムで扱っている状況は、話題の対象となっている状況 (SHRDLUでは、積木の世界) であったり、話者の右や左といった言語表現と直接関係があるものであった。また、マルチモーダルインタフェースで主に扱っている情報は、人間のジェスチャーであることから分かるように、人間が相手に何かを直接伝えようと発信している情報である。しかし、システムが扱うタスクによっては、人間が何気なく行っている行動を始めとして多岐に渡った情報がコミュニケーションにかかわってくる。つまり、何が状況として関与するのかを認識し、人間と共通の状況を参照する機構を作る必要がある。

特定の情報に注目する機構は、注意機構と呼ばれ、CGインタフェースを用いた対話システムの研究⁴⁾で取り上げられてきた。これらのシステムでは、特定の非言語情報 (CGオブジェクト) に注意を向ける手法を用いて対処している。たとえば、話題に関連した単語やCGオブジェクトに注意を向けることによって発話内容が限定され、応答文の生成速度を向上させることができる。

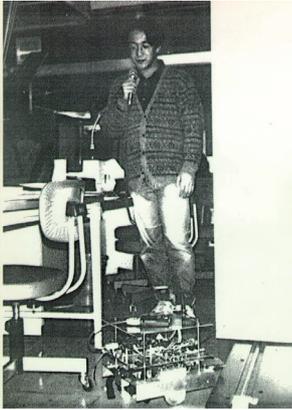


図-1 ロボット用対話システム Linta-II

著者らは、ロボットのセンサ情報の選択に注意機構を発展させ、ロボット用対話システム Linta-II を開発した(図-1)。Linta-II は、自律移動ロボット上に実装されており、ロボットのセンサを利用して実世界の情報を獲得し発話生成を行う。注意機構は、発話生成に用いるセンサ情報を限定する機構である。たとえば、自律移動ロボットが壁に向かって進んでいるときに、ユーザが、ロボットに対し「止まれ」を意図して「ダメ！」と発話したとする。この例の場合、ロボットが前進中に注意機構は前方のセンサ情報へと注意を向けるので、センサ情報が限定される。Linta-II は、センサ情報の限定によって壁の位置を考慮に入れながら発話意図を解析することができるので、ユーザの意図を正しく推論することができる。

発話理解と共同注意

前章で実世界の情報を介した人間とロボットのコミュニケーションには注意機構が重要であることを説明した。しかし、注意機構はあくまでもセンサ情報を選択する枠組みである。つまり対話中にロボットが注目すべき情報を一方的に選択している。一方、実際の対話では同じ実世界の情報に話者と聞き手が注目する。つまり対話では、同じ情報に双方が注目する必要がある、一方的に実世界の情報を選択する注意機構では不十分であるといえる。社会心理学の分野では、同じ実世界の情報に複数の人間が注目することを共同注意と呼ぶ⁷⁾。本章では、人間とロボットのコミュニケーションの実現の観点から共同注意について述べる。

共同注意を実現する上で注意機構に足りないものは次の2点である。

1. ロボットが人間の注意を認識する機構
2. ロボットの注意を表出する機構

これらの機構は、同じ実世界の情報に人間とロボット

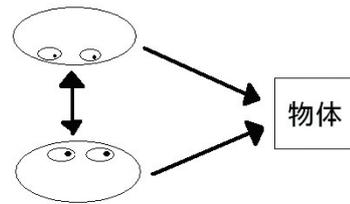


図-2 3項関係の概念図

が注目し共同注意が成立する上で必要不可欠なものである。1.の機構は、実世界のどこに人間が注目しているのかを認識する。つまり、人間が注目している情報にロボットが注意を向けるときに必要となる。2.の機構は、ロボットが注目している情報を人間に気づかせる。この機構により、人間は、ロボットの注目対象に気づくことができる。ロボットの注目の対象に人間が注意を向けることができる。共同注意の成立は、一見、1.の機構だけでも可能だと思える。しかし、実際にはそうではない。2.の機構は、真の意味で共同注意を成立させるためにロボットに欠かすことのできない要素である。特に人間に対してロボットが自発的に発話をする際には必要不可欠である。本章では、共同注意とは何かそして実現に必要な要素は何かについての説明を通して1.と2.の双方の機構が必要であることを述べる。各機構について具体的に行われている研究に関しては、次章で紹介する。

社会心理学や発達心理学の分野では、共同注意を持つ2人と、注意が向けられている対象の3つの間の関係を3項関係という。3項関係は、図-2に示す通り2人の人間と物の間に成り立つ関係である。ただし、3つの要素の間の客観的な関係をただ単純に表しているのではなく、2人の人間から主観的に見たかたちで関係が複雑に絡み合っている。

ある人間の視点から見たかたちで3項関係を説明するために以下では、自分と相手という表現を便宜上を用いる。3項関係では、双方の人間が、自分と物の関係だけでなく相手と物の関係にも気づいている。さらに双方が、相手と物の関係に気づいているので、自分と物の関係に関して相手が気づいていることにも気づくことができる。さらに考察を進めると、相手と物の関係に自分が気づいていることを相手が気づいていることにも気づくことができる。このお互いの関係をお互いが相互に気づいていく関係は無限に続くこととなる。

以上の通り共同注意は主観的な視点が複雑に絡み合っ成り立っている。しかし人間とロボット間に共同注意を成立させる上で重要となるのは、上記の複雑な関係をロボットに理解させるよりもむしろ、3項関係が成立していると人間に思わせることである。3項関係が成立し

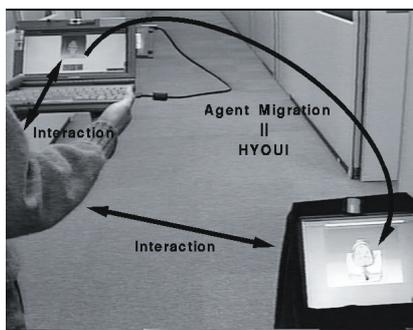


図-3 エージェントの移動による関係の移動

ていると人間に思わせることができれば、ロボットが自発的に発話を生成したとしても、人間は、共同注意の対象を積極的に参照しながらロボットの発話を理解することができる。

以上の説明より分かることは、人間に3項関係を持たせるために、ロボットが注目している対象を人間に気づかせる必要があることである。つまり、2.の機能のロボットによる注意の表出が必要不可欠である。人間は、ロボットの表出からロボットがどこへ注目しているのか意識することができ3項関係の成立へと導かれる。

しかし、実際にロボットが注意を表出し3項関係を構築しようとしても単純には実現できない。この理由は、人間とロボット間の関係にある。本稿では最後に関係性の影響に関して著者らが行った実験について説明する。

人間とロボットの関係は、ロボットの状態（ロボットの行動の意図や知覚している情報）を人間が積極的に推測するかどうか、つまり共同注意が成立するかに深くかかわっている。実験では、ロボットが人に唐突に依頼をした際に、人間に伝わるかどうか調べた。結果を先に述べると、ロボットが人に依頼するためには、お互いの間に何らかの関係が成立している必要があるといったものであった。

以下、この実験について簡単に述べる。実験は、被験者の前に現れた移動ロボットが、「ゴミ箱をどけてください」と合成音声で被験者に依頼するものである（図-5）。

実験では、ロボットが人間の目の前に突然現れても人間とロボット間に関係が形成されにくいことを考慮し、CGエージェントの利用を試みた。CGエージェントは、図-3の左上の携帯端末上のCGキャラクターである。実験では、ロボットとのコミュニケーションが開始される前に、被験者に、携帯端末上のCGエージェントに餌を与えさせるなどCGエージェントに十分慣れ親しんでもらった。

人間とロボットの関係は、被験者が慣れ親しんだCGエージェントを仲介役として形成される。図-3の右下

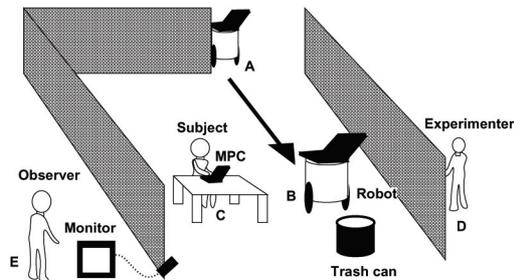


図-4 実験環境



図-5 ロボットの発話を理解できた被験者

に示す通り、実験で用いたロボットはディスプレイを持つ。ロボットが被験者に近づいたときには、このディスプレイに、携帯端末上に表示されているCGエージェントと同一のキャラクターが表示される。また、CGエージェントは、ロボットのディスプレイ上に表示されると同時に、携帯端末上のディスプレイに表示されなくなる。この一連のプロセスは、慣れ親しんだCGエージェントが携帯端末からロボットへ移動したように被験者から見え、被験者とロボットがかかわるきっかけを与える。

実験では、ロボット上のディスプレイへCGエージェントが移動する場合としない場合を比較し、人間とロボット間の関係がロボットから人間への依頼に影響を与えるかどうか調べた。

図-4に実験環境を示す。被験者は、図中のCに座り、携帯端末上のCGエージェントとインタラクションするように教示される。ここで、被験者は、実験の目的がCGエージェントの評価であると指示されており、ロボットの存在はあらかじめ教えられていない。

被験者がCGキャラクターを操作しているときに、ロボットは、図中のAからBに移動し、被験者の前に突然現れる。ここで、CGエージェントは、携帯端末上から消え、ロボットのディスプレイに移動し表示される。また同時に、合成音声で、「ゴミ箱をどけてください」と人に依頼する。この実験は、CGエージェントがロボットに移動しない場合も行われた。

図-5は、CGエージェントがロボット上に移動した場合の被験者の行動を示している。被験者は、CGエージェントが携帯端末上から消えた瞬間戸惑った。しかし、同じCGエージェントがロボット上に表示されているの

で、「ゴミ箱をどけてください」とのロボットの依頼に素直に従い、ゴミ箱をどけた。この実験条件の下、被験者10人のうち、8人がゴミ箱をどけた。

図-6は、CGエージェントがロボット上に移動しない場合の被験者の行動を示している。被験者は、突然現れた見ず知らずのロボットを無視した。この実験条件の下、被験者10人のうち、9人がゴミ箱をどけなかった。

また、CGエージェントがロボット上に移動しない場合の実験では、興味深いアンケート結果が被験者から得られた。CGエージェントがロボット上に移動しない場合、被験者は、ロボットが何を言っているのか理解できていなかったのである。つまり、合成音声を理解できていなかったのである。一方、CGエージェントが移動した被験者は、ゴミ箱をどけたことから明らかなように、ロボットからの依頼を正確に理解していた。

以上の結果を踏まえると、人間は、ロボットとの関係があつて初めて、依頼の理解、つまり、ロボットの意図や知覚している状況を推測することが分かる。ロボットが注意を表出する際も同様に、人間との関係を築きながら行う必要がある。

振る舞いの社会性とロボットの身体表現

コミュニケーションにおけるジェスチャーや視線の中には、アイコンタクトや指差しなど社会的に解釈可能な意味を持つものがあり、人間同士での共同注意の成立に重要な役割を担っている。本章では、社会的振る舞いをするロボットの研究を紹介し、共同注意の成立における身体表現の重要性を説明する。



図-6 ロボットの発話を理解できない被験者



図-7 実験場面。(左)人間とRobovieがアイコンタクトしている。(中)ポスターへの共同注意の成立。(右)Robovieの腕を見る被験者。

前章で説明した通り、人間とロボット間で共同注意を成立させるためには、ロボットが人間の注意を認識すること、人間とロボット間に関係を築いておくことが必要となる。

ロボットが人間の注意を認識する技術の研究は多数行われている^{3), 5)}。これらの研究では、人間の顔の向きや視線から注意の対象を見つけている。MITで開発されたKismet³⁾は、人間の注意の対象を認識するシステムの代表的な研究である。Kismetは、画像から人間の顔領域を識別し、肌色領域の分布から人間の顔の向きを認識する。しかし、人間の注意を厳密に認識するためには、顔の向きと同時に視線を認識する必要がある。CRLで開発されたInfanoid⁵⁾は、人間の視線認識を目指し開発中である。

以上の研究では、人間の顔の向きが人間の注意の方向を表すといった仮定の下で実装されている。一方、人間とロボットの関係形成やロボットの注意の表出の実現では、ロボットのジェスチャーや視線の動きが重要となる。著者らは、ロボットのジェスチャーや視線の動きといった身体表現が、人間の場合と同様に重要な効果を持つかに関して心理実験を行った。

実験では、図-7に示すコミュニケーションロボットRobovieを用いた。また注意の対象を表出するためにRobovieの頭部および腕によるジェスチャーを用いた。さらに人間との関係を構築するためにRobovieの視線を人間に向け、アイコンタクトの成立を試みた。このアイコンタクトは、注意の表出におけるRobovieの振る舞いを人間に気づかせるために、頻繁に行った。アイコンタクトの結果、人間は、コミュニケーションの意図をRobovieから感じ取り、Robovieの注意の対象について気づくことができる。

実験では、共同注意の成立におけるジェスチャーおよびアイコンタクトの重要性を検証した。本実験では、20人の被験者を半数ずつ2つの群に分けた。1つの群(実験群)は、被験者とアイコンタクトを行うRobovieを与えた。もう1つの群(対照群)には、アイコンタクトを行わないRobovieを与えた。腕による注意の表出は、両

	ポスターを見た	Robovieの腕を見た
アイコンタクトあり	10	0
アイコンタクトなし	1	9

表-1 実験結果、アイコンタクトの効果によってポスターを参照した被験者の数 ($U=5, p < .01$)

群共に行った。Robovieが注意を向ける対象は、壁に貼られたポスターである。実験では、注意の表出に従って被験者がどこを見たかを記録した。

実験は、以下の手順で行った。始めに、Robovieが被験者の前を通りすぎ、ポスターの前で止まる。ここで、被験者は、Robovieとポスターの両方を見ることが出来る場所に立っている。次に、Robovieは、被験者の方に振り返り、「これ見てね」と発話しながら、腕でポスターを指し示す。この時点で実験群の被験者には、Robovieが、被験者を見る動作とポスターを見る動作を繰り返す。対照群の被験者には、Robovieが顔を正面に向けたまま動かさない。

表-1に実験結果を示す。被験者は、Robovieがアイコンタクトした場合(図-7左)、ポスターへ視線を向けた(図-7真中)。一方、Robovieがアイコンタクトを行わなかった場合、多くの被験者がポスターを見ずにRobovieの腕先を見た(図-7右)。結果、共同注意の成立にアイコンタクトが影響を与えることが確認された ($U=5, p < .01$)。

実験結果から、実世界の物体を指し示す能力をRobovieが持つことに対照群の被験者は気づかなかったことがうかがえる。つまり、被験者とRobovieの間は、2項関係となっており、第3の物体がそこで考慮されることはない。従来の人間とロボットのコミュニケーションは、1対1のものがほとんどであり、2項関係をベースとしたものであったといえる。一方、アイコンタクトによって、被験者は、Robovieの注意の表出に気づき、ポスターを見ることができた。この実験結果は、被験者および、Robovie、ポスターの間に3項関係が生じ、共同注意が成立したことを示している。つまり、Robovieは、人間とロボットの新しいコミュニケーションを共同注意を元にして実現しているといえる。

人間とロボットのコミュニケーションにおける共感の影響

本稿では、自律的に振る舞うコミュニケーションロボットに必要な要素(注意機構、発話生成、身体表現)について述べた。特に、実世界の情報を介して人間とロボットがコミュニケーションするためには、ロボッ

トが注意機構を持っている必要があること、人間とロボット間に共同注意が成立している必要があることを説明した。さらに、共同注意が成立するためには、ロボットが実世界のどこに注目しているのかを人間に対して表出するだけでなく、人間とロボット間に関係が成立している必要があることを述べた。しかし、本稿で取り上げた共同注意に関する研究はどれも人間とロボットのインタラクションの一部を切り出したものである。今後、ロボットが人間社会で自発的なインタラクションを行い、人間生活をサポートするためには、人間とロボット間で交わされるインタラクションそのものから関係性を築いていく必要がある。著者らは、実世界の物に対して人間が持つ感覚に現在注目している。実世界に関する感覚の共有や共感を用いることにより、インタラクション自体から関係を成立させ、共同注意の成立を試みている¹¹⁾。今後のコミュニケーションロボットの発展のためにも、言語で表現可能な情報のやりとりだけでなく、身体を通じたコミュニケーションの研究開発が望まれるところである。

参考文献

- 1) Barwise, J. and Perry, J.: Situations and Attitudes, MIT Press (1983).
- 2) A Bolt, R.: Put-that-there: Voice and Gesture at the Graphic Interface, Computer Graphics, the Association for Computer Machinery,inc, 14(3): pp.262-270 (1980).
- 3) Breazeal, C.: Designing Sociable Robots (Intelligent Robots and Autonomous Agents, MIT Press (2002).
- 4) Chapman, D.: Vision, Instruction, and Action, MIT press (1991).
- 5) Kozima, H. and Ito, A.: Towards Mindreading by an Attention-sharing Robot, In Proc. of Third Int. Symp. on Artificial Life, and Robotics (AROB III' 98), pp.478-481 (1998).
- 6) Maes, P., Darrell, T., Blumberg, B. and Pentland, A.: The Alive System: Wireless, Full-body Interaction with Autonomous Agents, Perceptual Computing Technical Report 257, MIT Media Lab. (1995).
- 7) Moore, C. and Dunham, P.J.: Joint Attention: Its Origins and Role in Development, Lawrence Erlbaum Associates, Inc. (1985).
- 8) Winograd, T.: Understanding Natural Language, Academic Press (1972).
- 9) 伊藤, 伝田, 中川: マルチモーダルインタフェースと協調的応答を備えた観光案内対話システムの評価, インタラクション'97, pp.135-142 (1997).
- 10) 末永, 間瀬, 福本, 渡部: Human Reader: 人物像と音声による知的インタフェース, 信学論D-II, J75-D-II (2): pp.190-202 (1987).
- 11) 鳴海, 今井: 演出を用いたヒューマン・ロボット・インタラクション, 合同エージェントワークショップ&シンポジウム 2003 (JAWS2003) 講演論文集, pp.438-445 (2003).

(平成 15 年 10 月 29 日受付)