

## 2. エッセイ集

# 1 対話型ロボットの研究

石黒 浩 (大阪大学/ATR)

### 対話型ロボット

ディープラーニングによって最も実用的になった技術の1つが音声認識である。特に、GoogleやAppleのスマートフォンの音声認識は、従来の音声認識に比べて格段に精度が良くなった。この音声認識の技術を背景に、新たな人工知能ブームが生まれ、その人工知能ブームは人と知的にかかわる対話型ロボットのブームを生みつつある。

近年の対話型ロボットの中で、最も成功しているのはAmazon Echo<sup>☆1</sup>であろう。ただ、Echoはロボットと言うよりもスマートフォン用の音声認識デバイスという方が正しい。

そのEchoをもう少しロボットに近づけたのがMITのBreaseal教授らによって開発されたJibo<sup>☆2</sup>になる。3自由度しかないロボットであるが、工夫された自由度の配置によって、感情豊かな表現ができるようになっていく。発想の背景にはBreaseal教授らのAffective Computingの研究がありそうだ。

これらのロボットはどちらも米国で生まれ、現状のメディアを無理なく延長したものとなっている。ロボットというより入出力デバイスが、独立したものと考えた方がよい。すでに一般的に利用されている、Bluetoothスピーカやマイクの機能を拡張したものと考えられる。そのためかAmazon Echo等はすぐに人々に受け入れられた。

一方、日本から発信される対話型ロボットは、人間型が多い。音声認識技術が進歩して最初に登場したのがソフトバンクグループ(株)のPepper<sup>☆3</sup>である。Pepperは当初20万円程度の価格で販売することが計画されていた。パソコンと同様の価格で、パソコンと同様に誰もが手軽に利用できるパーソナルなロボッ

トとして普及させることを目標としていた。今ではもう少し高い値段で売られているが、それでもロボット

の性能から考えれば十分に安い。

その20万円程度の値段で売り出されているのが、シャープ(株)のRoBoHoN<sup>☆4</sup>である。Pepperとは異なり、手で持ち運べる小さいロボットである。携帯電話程度の大きさで、もちろん携帯電話としても使うことができる。バッテリーで動作できる時間の短さが問題ではあるが、小さいながらも、歩いたり話したりすることができ、人間型ロボットの基本的な機能をおおむね備えている。

PepperとRoBoHoNのちょうど中間的な大きさを持つのが、ヴィストン(株)のSota(図-1右)である。PepperやRoBoHoNは、ユーザがソフトウェアを入れ変えることで、多様な機能を実現できるパーソナルロボットとしての普及を目指しているが、Sotaも同様である。SotaはPepperやRoBoHoNに先駆けて、開発システムをオープンにしており、研究機関を中心に徐々に使われ始めている。一方で、PepperとRoBoHoNはどちらも自律的に動き回ることを想定してデザインされているが、Sotaはテーブルトップで利用するロボットであり、移動機能は持たない。その分、コストが安く10万円程度の値段で売られている。

Sotaに似たロボットに、JST ERATO 石黒共生ヒューマンロボットインタラクションプロジェクトで開発されたCommU(図-1左)がある。CommUはSota



図-1 CommU(左)とSota(右)

☆1 <https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E>

☆2 <https://www.jibo.com>

☆3 <http://www.softbank.jp/robot/consumer/products/>

☆4 <https://robohon.com/>

と同様のテーブルトップロボットであるが、視線制御、瞬き、口の開閉等があり、より表情豊かに対話できるようになっている。CommUも近い将来の実用化を目指して開発が進められている。

## メディアとしてのロボット

このような対話型ロボットは、物理的作業を行うロボットとは異なり、人との間で情報のやりとりを行う新たなメディアとして登場してきた。

Amazon Echoは単なる音声デバイスの延長であり、その意味で純粋音声メディアである。Jiboはそれにディスプレイと3つのモータを組み合わせて、人らしい体ではないものの、感情的表現を用いながら人とかわることができるメディアとなっている。Pepper, RoBoHoN, Sota, CommUは、人間と同様の体や動きを持つ人間型ロボットメディアである。

我々人間は常に人間に似たメディアを求めてきた。たとえば、Windowsの質問に答えるエージェントなど、CGを使ったディスプレイ上のエージェントである。しかしながらこれらのエージェントはいつの間にか姿を消した。ユーザが人らしいエージェントを求める一方で、パソコンやスマートフォン等では、その使用目的が決まると、エージェントは邪魔になって使われなくなる。

しかし、人間の脳が人間を認識する多くの機能を持つ以上、人間にとって理想的なインタフェースは人間であり、我々は新たな人間らしいメディアを再び求める。その結果登場したのが、Pepperをはじめとする人間型メディアとしてのパーソナルロボットである。まだ、爆発的な普及に至っていないため、真に世の中を変革するメディアとなるかどうかは分からないが、その可能性は高そうである。

## 対話とは何か？

対話型ロボット、パーソナルロボットの研究開発において遅れている研究は、実は対話研究である。エージェント研究においては、ある程度研究されているが、まだまだ十分ではない。ロボットの対話においてボトルネックになっているのが、音声認識が重要であるという考えを持つ研究者が多かったためか、「対

話とは何か？」という問いに答える研究がほとんどなされてこなかったという点である。

対話に関する研究テーマとしては以下のようなものが考えられる。

まず感情の使い方である。人間型ロボットは表情や身ぶり手ぶりで感情を表現でき、それによって瞬時に他者と情報を共有できる。「感情で何を伝え、言語で何を伝えるか？」という問題について、対話の設計方法を確立する必要がある。

次に、対話における理解の問題がある。「対話において相手はどれほど理解をしているのか?」「対話とはどれほどの深い理解を必要とするのか?」実際に筆者の研究グループでは、ロボット2体と人間1人、合計3人で構成される社会的状況においては、ロボットは人間の発話を理解しなくても、人間との間で簡単な対話を続けることができることを発見し、その対話のルールを特許申請している。

さらには、「対話とはそもそも発話をする必要があるのか?」という疑問もある。物理世界における対話では、相手の存在感を感じ、自分の存在をアピールしながら対話することは重要であるが、そのために、必ずしも自分が発話をする必要はない。対話のストーリーを自らの意思で選択できれば、誰の声で話していても、その対話を自分のものとして受け入れることができる。

そして最後に重要なのは、「人は対話においてそのストーリーを自分で作り出しているか?」という問題である。多くの場合、誰かが生み出したストーリーの一部を単に記憶し、再生しているに過ぎなかったりする。ならば、「ロボットはどのようにストーリーを取り込み、管理し、人に提示すべきなのだろうか?」

こうした対話の本質にかかわる問題は、音声認識の問題がある程度解決できるようになった今では、対話型ロボットの実現において、より重要な問題となってきた。

(2016年6月27日受付)

石黒 浩 (正会員) ishiguro@sys.es.osaka-u.ac.jp

1963年滋賀県生まれ。大阪大学大学院基礎工学研究科システム創成専攻教授(特別教授)・ATR石黒浩特別研究所客員所長(ATRフェロー)。工学博士。2011年大阪文化賞(大阪府・大阪市)受賞。2015年文部科学大臣表彰受賞。