

音声対話エージェント実用化への課題

荒木雅弘

京都工芸繊維大学 工学部 電子情報工学科

1. はじめに

1998年10月に行われたパネル討論「擬人化インタフェースは二年後に果たして普及しているか？」において、山本吉伸氏は「人型と擬人化を区別する」という提案を行っている。著者はこの提案に賛成する。そして、音声対話エージェントとして、人型・擬人化それぞれに求められる機能を分析した。結果として、パネルの前提になっていたと思われる「擬人化エージェントの要素技術は出揃った」という考えが誤りであったことを指摘したい。

2. 人型エージェントと擬人化エージェント

人型エージェントは「その外見が人の形をしている」ものを指し、擬人化エージェントは「その振る舞いが人に準じている」ものを指すものとする。山本氏は前者を機能、後者を現象と位置付けていることから、これらの概念は直行するように解釈することができる。

例えば、イルカの外見を持ったヘルプエージェントは人型をしていない擬人化エージェントであると言える。しかし、エージェントの範囲をユーザとインタラクションを行う音声対話エージェントに限定すると、人型であってかつ擬人化でないエージェントというものは考えられない。従って、音声対話エージェントにおいては、人型は擬人化の部分集合であるということが言える。

それでは、音声対話エージェントが人型をしている利点は何かであろうか。

擬人化エージェントにおける「人に準じた振る舞い」を具体的に定義すると、(1)自然言語によるユーザからの要求を受け付け、(2)何らかの知的処理を行い、(3)自然言語などによって応答を返し、(4)そのインタラクションの内容を憶えていることであると言える。このうち、人型という特徴が何らかの貢献をし得るのは(3)であり、さらに特定すると、(3)の応答内容のうち、非言語情報に限定できる。感謝や謝意の表現、理解状態の表出(うなづきや首をかしげるなど)、内部状態の表出(認識中・検索中)などの非言語情報が、自然なインタラクションの実現に有効である。ユーザがエージェントの表出する非言語情報を解釈する際に、人間同士のインタラクションプロトコルに関する知識が援用できるということが人型の利点であると考えられる。そうであれば、人型で非言語情報を表現するときに顔に限定するのは必ずしも最良の方法であるとは言えず、全身を使ってよりわかりやすく表現するほうがよいのではないだろうか。さらに人型という枠をはずした場合、アニメーションキャラクターによる様々なデフォルメした動作や効果音なども利用可能である。

従って、非言語情報の伝達性能は顔のクオリティや伝達手法のノウハウの問題ではなく、伝達媒体を比較検討してその伝達の正確さや快適さを検証すべきであり、既に要素技術として確立されたものであるとは言えない。

3. 音声対話エージェントに欠けている要素技術

次に、(1)のユーザからの要求入力について考察する。現在、VoiceXMLなどで記述した音声対話システムがあまり利用されていないことから、一塊の情報をキャッチボールのようにやりとりする対話スタイルに疑問を持っている研究者も多く、自然な対話を目指した研究がいくつか見られる。部分発話の理解、相槌、システムからの割り込みなど、自然な対話を実現するための必要な機能の多くは研究途上である。

(3)に関して、人に準じた機能を実現するためには、エージェントはタスク遂行に必要な知識を持つ必要があり、さらに知的情報交換を行うためのインフラ整備も必要になる。Web サービスやセマンティック Web はこのような背景でのネットワークサービスインフラの提供であると位置づけることができる。セマンティック Web に関するある文献¹では、以下のような質問に応えるような知識表現体系の構築を目指していると紹介している。

"Tell me what wines I should buy to serve with each course of the following menu. And, by the way, I don't like Sauternes."

求められる擬人化エージェントは、このような発話を認識し、理解する必要がある。しかし、このような長い発話をターンとすることは非現実的であるから、もっと小さい単位のインタラクションを積み重ねて、ユーザの要求を受け付けなくてはならない。ユーザの発話の間に打たれるシステムの相槌は、その部分発話がタスク遂行に関与していることを判断した上で打たれるはずである。しかし、どのようにしてタスク知識をインタラクションの制御に反映させることができるのだろうか。

いくらサービス側が高度な機能を提供しようとも、また、インタフェース側が自然なインタラクションを実現しようとも、このサービスとインタラクションのギャップを埋めなければ擬人化エージェントは実現できない。この問題の解決には、状態記述をベースとするか、あるいはもっと高レベルのタスク記述をベースとした対話記述言語の開発が鍵となるであろう。

(4)に関しては、ユーザのコンテキストの保存も重要な要素である。特にそのコンテキストを保持しているものをエージェントとして実体化することはユーザにそのコンテキストを積極的に使わせるきっかけを与えることにつながる。これもあまり複雑な記述にならないように対話記述言語に組み込むという解決方法が考えられる。

4. おわりに

本稿では、音声対話エージェントへの要求をリストアップし、本当に必要なものがまだ要素技術として確立していないことを指摘した。今後は ISTC(音声対話技術コンソーシアム)などでこれらの課題に取り組まれることを期待する。

¹ <http://www.w3.org/TR/owl-guide/>