

会話ロボットと顔インタフェース

小林 哲 則

(早稲田大学)

1. はじめに：WABOT-2 の思い出

長年音声対話インタフェースをロボット上に実装してきた経験から、音声会話インタフェースにおける「顔」の重要性について述べてみたい。

筆者が最初に関わった会話ロボットは、WABOT-2 という人間形鍵盤楽器演奏ロボットであった。一応形としての顔はついてしたが、機能としての顔があったかどうかは疑問である。話しかけても話し相手を振り向くことはなく、常に譜面台に向かってエレクトーンを引く準備をしている精神的余裕のないロボットであった。たまに話しかけても、認識ミスをすれば何の反応もない。いったい何を考えているのか、こちらの相手をする気があるのかどうかさえ伝わってこない。命令には忠実であって、何かリクエストをすれば演奏を聴かせてはくれたが、一緒に会話を楽しんでいるという気持ちにはなかなかないロボットであった。

2. 音声インタフェースと透過性

WABOT-2 になかったのは、インタフェースとしての透過性に対話相手としての存在感である。ここでは、透過性の問題に限って話を進める。(存在感の問題も顔と関係し、興味深い話題なのではあるが、紙面の関係で省くこととする。)

透過性とは使い易いインタフェースに求められる要件のひとつで、今システムがどういう状態にあるかが利用者に容易に理解できること(状態理解容易性の原則)、そこでどういう操作をすれば良いのかが簡単に想像できること(手順連想容易性の原則)。また、利用者の作用の効果が、すぐに利用者に対しフィードバックされること(フィードバックの原則)をいう。

音声入力においては、本来ならば、思ったことを言葉で伝えれば、相手に指令が届くことが期待されることから、音声は操作手順の連想は容易なはずだ。しかし、実際には個々の場面での操作手順(認識できること)にシステム内部の状態が深く係わる。このため、利用においては状態理解の容易性がより重要になる。しかし、この実現に関して、音声は本質的に不得手である。状態を提示するためには、長々と話さなければならない。状態提示のためのメッセージは、なければ使えないし、あればあったで円滑さを妨げる要因となるという厄介な代物である。また、音は瞬時に消えてしまう性質のものであるから、仮に状態を提示したとしても、それを長く記憶にとどめておくことは難しい。これらの意味で、音声は状態理解の容易性に関してはもともとハンデキャップを背負ったメディアということになる。

3. 会話インタフェースにおける顔の役割

会話において、透過性の欠如を補うのは、一般には視覚的メディアの役割である。単純な意図であれば、

瞬時に伝達可能であり、かつ長期に亘って提示することが可能である。なにより言語情報と並列して提示することができ効率が良い。音を伴わないため、相手の思考を妨げないという利点もある。

我々が通常人同士の会話において、このフィードバック用途に使う視覚的メディアは「顔」である。話を聞く準備が整えば話者を見つめる。話を通じればうなずくし、通じていなければ怪訝な顔をする。こういった内部状態のフィードバックに支えられて、対話は成立している。対話相手がシステムの場合、本来であればシステムからのフィードバックを「顔」に頼る必然性はない。しかし、人同士で顔をこの用途に用いていることから、システムにおいても顔を用いると好都合の場合が多い。最大の利点は、フィードバックの処理を意識に上らせないことである。会話においては、言語情報の伝達が主であって、あくまで内部状態の提示は、言語情報を円滑に伝達するための従たる行為にすぎない。従たる情報処理は、できるだけ主情報の処理を妨げないことが望まれる。新たな視覚的情報に役割を与えて用いるよりは、従来から人同士で使っている流儀を利用することのほうが、人の負担は少ないはずである。

4. むすび：顔インタフェースにとって必要なこと

これまで述べてきたように、透過性の観点から顔が会話インタフェースにおいて重要な役割を演じていることは疑いがない。顔には、ここでは省いた存在感の実現や、親しみやすさの改善など、他にも様々な役割が考えられるが、最も重要なのはここで述べた透過性の問題と筆者は考えている。

さて、この顔インタフェースが「爆発的には広がって行かないこと」を憂い、企画されたのが今回のパネルとのことである。しかし、私見によれば顔インタフェースの立場から、現状を案じることはない。会話インタフェースが重要であれば、それには概ね「顔」がつくと信じる。むしろ問題なのは、会話インタフェースが本当に重要か、あるいは本当に重要な使われ方をされているか、ではなからうか。

これまで述べてきた事柄は、インタフェースにおいて適切なフィードバックが必要であるということと、会話インタフェースにおいては「顔」がその役に適した存在であるということに過ぎない。従たる役割を果たす存在としての「顔」の重要性である。しかし主役が倒れては、脇役がどんなに優秀でも劇は成立しない。

そもそも、情報入力に適したインタフェースの形はアプリケーションに応じて様々であって、インタフェースが会話的に行われなければならない理由はないはずだ。無理な会話的インタフェースの導入事例が多すぎやしないか。そのあたりから検証が必要である。