

## 1 はじめに

[Bolt80] など、80 年代初頭頃から開始されたマルチモーダルインタフェース研究は、近年の音声・画像情報の認識・生成技術の進展と、Post GUI としての期待の高まりにより活発化し、今までにも数多くの試作システムが開発されている [Maybury93]。しかし、その多くは、主にマルチモーダル入力された参照表現の照応解決 [Kobsa86, Koons93] や、その生成 [Wahlster91] などを課題とし、マルチモーダル入力の統合あるいは生成に関心を置くものであって、ユーザとの自然なインタラクションの実現には至っていない。

また、Post-GUI 実現へのもう一つの有力なアプローチとして、[Apple87] などに端を発する擬人化エージェント技術を挙げることが出来る。これは、(1) 表現力が高く、(2) より自然であり、(3) アプリケーションも広い、といった特徴を持つ [Wahlster97]。そのため、ユーザとの自然なインタラクションの実現に大きく寄与し、また、マルチモーダリティを HI に導入する意義を与えるものでもある。

これまでも擬人化インタフェース技術を用いた様々なシステム [竹林 94] [神尾 94] [鈴木 96] [知野 97, 河野 98] を開発し、研究を進めてきた。本研究は、この流れに沿ったものであり、非言語メッセージの活用によって、次節で説明するメタコミュニケーション能力を実現した新しいインタフェースを提案する。

## 2 メタコミュニケーション

メタコミュニケーションとは、「コミュニケーションを成立させるためのコミュニケーション」を意味する [黒川 94]<sup>1</sup>。人間同士の対話では、あいづちや、うなずき、視線一致などによって、会話の開始や発話交替、あるいは話者聴者といった役割の認識がなされたり、あるいは情報の伝

<sup>1</sup>メタコミュニケーションには、ゲームのルール説明や機器の操作方法の説明などを指す場合もある [Yvonne89]。

達状態の通知、確認、あるいは問い返しなど、通信路の確立、維持、解除などが実現されていると考えられる。こういったメタコミュニケーションは、対話能力の中で大きくかつ本質的な部分を占めるものであり、そこでは非言語メッセージ [Vargas87]<sup>2</sup> が重要な役割を担っている。ところが、従来のインタフェースでは、例えばユーザとシステムとのインタラクションの開始、中断、あるいは終了などは、*a priori* とされ、メタコミュニケーションに関してまで考慮されたものはほとんど無かった<sup>3</sup>。

本研究では、このメタコミュニケーション能力の欠如こそが、現在のヒューマンインタフェースの重大な問題点であるとの立場を取る。そして、ユーザとのインタラクションに於いて、非言語メッセージを積極的に活用することにより、HI にメタコミュニケーション能力を与えることで、自然でより使いやすいインタフェースの実現を目指すものである。

## 3 “GAZE To TALK” システム

従来の音声インタフェースの大きな問題点の一つに、「誰に向かって話されているのかが分からない」ことが挙げられる。そのため、システムへの入力を意図していない発声 (*e.g.* ユーザの隣にいる人に対する発声、一人言) や、あるいは周囲雑音などによって誤動作が発生してしまっていた。これはメタコミュニケーション能力の欠如の一つであると言える。

他方、音声入力の受け付け可否を、従来の Push-To-Talk 方式<sup>4</sup>の様に、例えばボタンやマウス操作などによってユーザに制御させる様になると、(1) 非接触、(2) *hand-busy* 時にも使用可などと

<sup>2</sup> コミュニケーションに於いて授受されるメッセージの中で文字として書き表せないものの総称。アクセント、イントネーションといった韻律から、表情、身振り、姿勢、さらに年齢や、服装までもが含まれる。

<sup>3</sup> 音声認識技術の問題点を、HI への応用を含んで広く検討した [嵯峨山 94] においても、この観点の指摘はなされていない。

<sup>4</sup> ボタン等を押すなどユーザの身体的な操作によって音声入力を可能とする方式。