

特別論説



情報処理最前線

21世紀のヒューマンコンピュータ
インタラクション†

長尾 確†† 徳永 健伸†††

1. はじめに

本稿では、ヒューマンコンピュータインタラクション (Human-Computer Interaction. 以下では、HCI と略す)*, つまり人間とコンピュータの相互作用 (あるいは対話) のさまざまな手法を、歴史的な流れに沿って概観し、将来への展望を述べたいと思う。ここで、ある重要な視点を導入する。それは、対話の様式 (モダリティ) というものである。モダリティ (modality) とは、人間が外界を知覚する仕方であり、視覚、聴覚、触覚などの感覚 (sense) を用いた情報伝達的方式である。これから、HCI を利用可能なモダリティの次元 (シングルモーダル、マルチモーダル、ハイパーモーダル**) で分類し、それぞれについて事例を中心に紹介してみよう。

2. シングルモーダル HCI

コンピュータとのインタラクティブなコミュニケーションが可能になって以来、そのコミュニケーションはほとんどキーボードを通じて行われてきた。キーボードを使ってやりとりされる情報は、タスクに応じて設計された特別な言語か、あるいは (制限された) 自然言語である。このような言語指向のインタラクションに対して、ビットマップ・ディスプレイに表示されたオブジェクト

をマウスなどのポインティング・デバイスで直接操作することによってコンピュータとのコミュニケーションを行う直接操作型のインタラクションが考案された。直接操作型インタラクションの研究は Xerox PARC の Alto の開発までさかのぼることができるが、これが一般的になったのは Apple の Macintosh の登場によってであろう。本章では、前半で、シングルモーダル・インタラクションを言語指向と直接操作の対比という観点から考察し、後半では、言語指向のインタラクションをより知的にするための試みについて述べる。これは、技術的な問題により、複数のモダリティを用いることが困難であった時代に、いかに効率良くコンピュータとインタラクションを行うかに関する研究の成果であり、今後も目的によっては (またコンピュータの使用状況によっては)、これらのシングルモーダル・インタラクションが十分機能するであろう。

2.1 言語指向 vs. 直接操作

直接操作 (direct manipulation) 型インタラクションという概念がはじめて明確にされたのは 1983 年の Shneiderman の論文によってであろう¹⁶⁾。Shneiderman は直接操作の特徴として以下の点をあげている。

- ユーザの関心があるオブジェクトを連続的に表現すること。
- 複雑な構文の代わりに物理的動作によって操作すること。
- オブジェクトに対する操作の影響が即座に見えること。
- 最小限の知識で使いはじめること。

たとえば、“afo” という名前のファイルを消去したいとしよう。言語指向のインタラクション、たとえば、UNIX⁹⁾ では、シェルに対して以下のようなコマンドを発行する。

† Human-Computer Interaction in the 21st Century by Katashi NAGAO (Sony Computer Science Laboratory Inc.) and Takenobu TOKUNAGA (Tokyo Institute of Technology).

†† (株)ソニーコンピュータサイエンス研究所
††† 東京工業大学工学部情報工学科

* 従来は、ユーザインタフェースあるいはヒューマンインタフェースと呼ばれることが多かったが、最近、人間と機械のインタフェース (接界面) だけでなく、人間と機械の内部における情報処理も考慮した、お互いの情報の能動的なやりとりという意味を含めて、インタラクション (相互作用) という言葉が多く使われるようになってきている。

** この言葉は筆者の造語であり、言葉そのものにはそれほど深い意味はない。

% rm afo

ファイル“afo”が実際に消去されたかどうかは、その後、ファイルの一覧を表示するコマンドを使って確かめる必要がある。一方、直接操作型インタラクション、たとえば、Macintosh では、“afo”というラベルのついたアイコンを画面の隅にあるゴミ箱のアイコンのところまでマウスでドラッグすればよい。この結果、“afo”というラベルのついたアイコンは画面から消え、ゴミ箱のアイコンが少しふくれた形に変化する。

直接操作型インタラクションのもうひとつの利点は操作の結果が別の操作の対象としてそのまま利用できる点である。たとえば、Macintosh であるフォルダの中身が知りたいければ、そのフォルダをダブルクリックすればよい。その結果、そのフォルダに対応する新しいウィンドウが開いて、そのフォルダに含まれるファイルがアイコンで表示される。これが操作の結果となるわけだが、このアイコンはそのまま次の操作の対象とすることができる。UNIX のパイプも同様の機能を部分的に提供している。UNIX では、コマンドをパイプ (“|”) で接続することによって、あるコマンドの出力が、その右隣のコマンドの入力になる。たとえば、以下のコマンドは、現在コンピュータを使用しているユーザの一覧をアルファベット順にソートして出力するものである。

% who|sort

“who”、“sort”はそれぞれユーザの一覧、ソートを行うコマンドである。しかし、言語指向のインタラクションでこのような機能をもつものは UNIX が最初であり、UNIX のユーザインタフェースが高く評価される理由のひとつでもある。UNIX が直接操作型インタラクションのもつこのような特徴を部分的にもっていると考えられるよりは、むしろ、直接操作型インタラクションが UNIX のこのような機構に影響を受けていると考えるほうが自然である。このような機構を実現するためには操作対象の斉一的な表現形式を用意する必要がある。UNIX の場合は構造をもたない 1 次元のキャラクタ列という最も単純な表現形式を採用している。

しかしながら、すべての操作が直接操作型インタラクションに適しているわけではない。表計算ソフトを使ったことがある人ならわかるように、

表計算のタスクでは、行と列の数字を操作するだけでは十分な仕事はできない。個々の項目に対してどのような計算を行うかは、数式で表現するほうが簡単である。また、UNIX のシェルが提供するような簡単なプログラミングの機能も直接操作の枠組で提供することは困難である。たとえば、UNIX で “.c” という拡張子をもった 100 個のファイルの拡張子を “.C” に変更するといった操作は非常に簡単な記述で実現できるが、このようなことはユーザインタフェースにプログラミングの機能を提供していない Macintosh のような環境では不可能である。

一般に、直接操作型インタラクションは初心者にもシステムを使いやすい環境を与えるといわれているが、このためには、物理的操作がオブジェクトに対する操作の適切なメタファになっているということが必要である³⁰⁾。たとえば、前述のファイルを消去する例では、Macintosh でファイルを表すアイコンをゴミ箱のアイコンまでドラッグしていった重ねるという物理的操作が、オフィスで（紙の）ファイルをゴミ箱に捨てるという操作に対応しているからうまく機能するのである。しかしながら、すべての操作に対して適当な物理的操作が対応するとは限らない。特にタスクが複雑になれば、単純なメタファでは表現しきれなくなってくる。また、ユーザとシステムの間でメタファの解釈にずれが生じた場合には問題が生じる。

以上のように言語指向のインタラクションと直接操作型インタラクションにはそれぞれ利点／欠点があり、それぞれの利点をうまく組み合わせ、より自由度の高い HCI を設計するという研究の重要性が高まった。

2.2 知的言語指向インタラクション

直接操作型インタラクションの流れとは別に、人間とコンピュータのインタラクションを単なるコマンド入力ではなく、より知的な言語指向インタラクションへと発展させることを目指した研究も盛んに行われた。

自然言語を使ってコンピュータと対話し、データベース検索などを行うシステムは自然言語インタフェースと呼ばれる。自然言語インタフェースは、1960 年代から研究が始まり、入力文からキーワードだけを抽出して対話するものから、構文・

意味レベルさらにはタスク/ユーザモデルなどを用いた語用レベルの処理を含んだ総合的な自然言語処理技術の応用システムとして研究されてきている。

以下では、このようなシステムの具体例について述べる。

GUS 航空機の旅行案内を行う自然言語インタフェースである¹⁾。処理の概要は、ATN 文法と格フレーム解析により、入力文から格フレーム構造を生成する。さらに、あらかじめ用意されているフレームのプロトタイプの付加手続きにより推論を行い応答を生成する。航空機の案内では、飛行機の便を決めることが最終的な目的である。飛行機の便は、日付・時刻・発着地点などによって決まる。それはフレームのおのおのに対応するスロットを埋めていくことにより行われる。さらに、スロットに付加された手続きにより、ユーザに不明な点などを問い返すことができるようになっていく。

以下に、GUS の対話の例を示す。

ユーザ：5月28日にサンディエゴに行きたい。

GUS：何時に出発されますか？

ユーザ：午後10時までに着かなければならない。

GUS：午前9時15分に到着のPSA 102便はどうですか？

ユーザ：次の便はどうか？

GUS：午前10時30分発、午後10時着のUA 310便です。

ユーザ：最初の便を頼む。

GUS：お名前をお願いします。

UC (UNIX Consultant) UNIX の操作法を案内する対話システムである²⁾。UC は、ユーザのゴールを分析し、それを満たすプランを生成するメカニズムをもつ。UNIX の操作法に関する知識は Prolog に似た動作を行う Lisp 形式の表現を用いている。入力の処理として、文のパターンと概念のペアを用意しておき、入力文と照合する概念を生成する。次に入力文に対応して生成された概念の表現から、ユーザのゴールを検出し、プラン-ゴール規則を用いてユーザのゴールを達成するプランを推論し、そのプランに基づいて応答を生成する。

以下に、UC の対話の例を示す。

ユーザ：ファイルを削除するにはどうしたらいい

か？

UC：'rm ファイル名' とタイプすると、現在のディレクトリからファイル名で指定されたファイルが削除されます。

ユーザ：ディレクトリの場合は？

UC：'rm -r ディレクトリ名' とタイプすると…

ユーザ：どうしたらもっと多くの情報が得られるか？

US：もっと多くの情報を得るには、'help' とタイプしてください。

ユーザ：あなたは UNIX について何を知っているか？

UC：私は現在 UNIX のファイルシステムについて知っています。

以上のように、自然言語インタフェースは、システムが処理できる知識の質と量が増加するに従って、より知的になってきたのであるが、このようなインタラクションにも実用上問題がある。ユーザは、システムが人間的な振舞いをみせればみせるほど、対話のスタイルや用いられる情報の種類などに関する制限をわずらわしく思うようである。このようなわずらわしさを緩和するために、人間とコンピュータの間のコミュニケーションチャンネル(対話に伴う情報を運ぶ媒介)を拡張していく必要がある。

3. マルチモーダル HCI

ハードウェア技術の進歩によって高解像度のビットマップ・ディスプレイや高速な CPU、あるいはグラフィックス専用の VLSI などが安価に利用できるようになってきた。このような背景をもとに HCI もそれまでのようにテキストを中心とするものだけではなく、グラフィックスや音声も融合したものが現れてきた。前章で述べた直接操作型のインタラクションもこのようなハードウェアの背景があってはじめて可能になるものである。直接操作型インタラクションの中にはグラフィックスやテキストなどの複数のメディアが利用できるものがある。そのようなインタラクションの形態は、マルチモーダル・インタラクションに分類できるだろう。ただ、ここでは「マルチモーダル」という言葉を、単にコミュニケーションのチャンネルとして複数のメディアが利用できるというだけではなく、それらが有機的に連携して

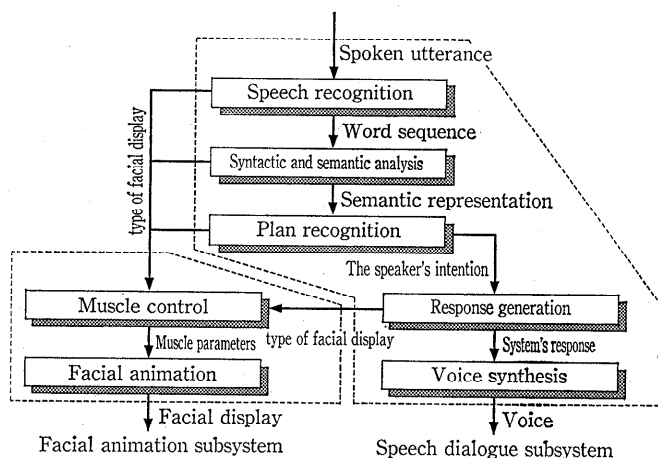


図-1 表情つき音声対話システムの構成

いるという点を強調して用いたい。以下、これまでに研究されているマルチモーダル・インタラクションを実現するシステムの具体例を紹介する。

3.1 ALFRESCO

イタリアのIRSTで開発されたALFRESCOはフレスコ画に関する情報をユーザに提供するインタラクティブなシステムである¹⁷⁾。画像情報を格納するビデオディスクとタッチパネルをもっている。システムとのコミュニケーションはキーボードからの自然言語入力、およびタッチパネルやビデオディスプレイなど*を通して行う。システムは自然言語を理解するための処理を行うが、ALFRESCOの研究の中心は、自然言語とハイパーメディアをいかに統合するかという点に置かれている。ユーザからの入力に関しては、自然言語の参照表現とユーザがタッチパネルを通して物理的に指示する動作の統合に焦点をあてている。一般に、自然言語の参照表現も物理的な指示行為も曖昧性を含んでいるが、両者をうまく統合して扱うことによって、曖昧性をうまく解消できるというのがALFRESCOの開発者たちの主張である。ユーザへの出力に関しては、画像を出力できる点と、ハイパーテキストを生成する点が特徴である。このハイパーテキストにはいくつかのボタンが付いており、これらのボタンによってユーザは関連する情報を検索することができる。Edinburgh大学のIDASシステムもハイパーテキストを生成することができるが、IDASでは、ユーザの選択を制限するためにハイパーテキストを使用してお

り、ALFRESCOとは動機が異なっている¹⁴⁾。ちなみにIDASはユーザに自然言語の入力を許していない。ALFRESCOは簡単なユーザモデルもっており、ユーザとの対話の進行につれてユーザモデルも更新される。

3.2 表情つき音声対話システム

人間の表情のもつ情報を新たなモダリティとして、対話システムに統合する研究が行われている。ソニーコンピュータサイエンス研究所(Sony CSL)では、音声対話システムと表情合成システムを組み合わせ

たプロトタイプシステムを開発している^{13), 19)}。

図-1は、このシステムの構成を表している。

音声対話システム 音声対話システムの構成は次のようになっている。マイクを通して入力された人間の音声発話は、ワークステーションに内蔵の音声処理ボードでAD変換され、音響分析、ベクトル量子化が行われた後、音声認識モジュールによって、上位スコアをもつ単語レベルの N 個の最良仮説が生成される^{8)*}。HMM(隠れマルコフモデル)と呼ばれる音韻モデルを用いて、ベクトル量子コードから音素を推定し、そして単語列を生成する。生成された単語列は、構文・意味解析モジュールによって、意味表現に変換される。まず、単一化文法と辞書を用いて構文解析が行われ、次に、フレーム型知識ベースと事例ベースを用いて曖昧さの解消が行われる¹¹⁾。発話の意味表現が決定すると、プラン認識モジュールによってユーザの意図が認識される。これは対話の進行に従って動的に修正・拡張されていくユーザの信念モデルと対話のゴールに関するプランに基づいている¹²⁾。意図を認識する過程で、主題の管理や、代名詞の照応解消、省略の補完などが行われる。そして、ユーザの意図に従って協調的な応答を生成するモジュールが起動する。このモジュールは、あらかじめ用意されたテンプレートとしての発話パターンにドメインの知識によって得られた応答に関する情報を埋め込むことによって発話を生成する。この応答は、音声合成モジュールに

* IRSTでは、これらをハイパーメディアと呼んでいる。

* 音声認識モジュールは、電総研の技術指導の下、東工大とSony CSLとの共同研究によって作成された。

よって音声となる。

このシステムのドメインは、ソニーのコンピュータ関連製品についての質問応答である。たとえば、システムは、ソニーのワークステーションやパソコンの価格、サイズ、重さ、機能、そしてCPUなどの仕様について、ユーザの質問に答えることができる。

表情合成システム 表情合成システムは、3次元的な顔のグラフィックスを生成し、多様な表情をリアルタイムに合成する。このシステムはグラフィックス専用マシン上に実装されている。顔の3次元モデルは約500個のポリゴンの組合せによって作られている。また、表面上には写真を用いたテクスチャマップによってリアルな顔が表現されている。さまざまな表情は、顔のモデルに付けられた筋肉の収縮をシミュレートすることによって生成される。これは、Facial Action Coding System (FACS) と呼ばれるシステムに基づいている⁶⁾。コンピュータ・グラフィックスでは、表情は顔を構成するポリゴンの局所の変形として実現される。現在、16種の筋肉の収縮と口の動きや顎の向き、目の動き、顔の向きなどを制御する10個のパラメータを操作する表情生成のためのマクロ言語が開発され、音声対話システムとのインタフェースに用いられている。顔のモデリングと表情アニメーションの詳細については文献18)を参照されたい。

音声対話システムは、対話におけるある状況において生成すべき表情に関するメッセージを表情合成システムに伝達する。

表情を、その対話的役割に応じて分類しようとする試みは、これまでもいくつか提案されている^{33)~5), 15)}。一般に表情の解釈は、それが現れる対話的状況に依存している。そのときの対話の文脈や他のチャネルによる情報との時間的前後関係(同時あるいは少しの遅れなど)などが、表情の解釈に対して重大な意味をもっている。

対話における表情は、次の三つに分類される。

(1) *syntactic display*: 特定の語、句、節を強調する表情、発話の構文的側面にかかわる表情、話の全体構造にかかわる表情。(2) *speaker display*: 発話中の事柄を例示する表情、発話中の事柄に情報を追加する表情。(3) *listener comment display*: 話者以外の人が話者の発話に対する反

応として作る表情。

speaker display には、ウィンクなどの顔記号 (facial emblem) と呼ばれるものも含まれる。

音声対話システムは、対話において重要ないくつかの典型的な対話状況を認識し、これらを表情と関連づける。たとえば、入力音声で認識できなかったときや認識結果が文にならなかったときなどは、*listener comment display* に分類される “not confident” という表情 (まゆを下げた悲しそうな顔) を示す。また、話者の質問がシステムの知識からはずれていた場合などは、*syntactic display* に分類される “facial shrug” という表情 (まゆを急に上げ、口元を歪めた顔) をし、「その質問にはお答えできません」のような応答をする。

現在のプロトタイプシステムのいくつかの未熟な点が、人間とコンピュータの間の自然な対話の障害になっているようである。たとえば、現在のプロトタイプは、まだユーザの表情をモダリティとして受けつけられないだけでなく、音声に含まれるプロソディ (アクセントやイントネーションなど) の意味を理解することもできない。このように、さらに多くのモダリティ/コミュニケーションチャネルを統合し、対話における各種の規制/障害を取り除くことによって、より自然で自発的な発話を誘引することができるだろう。

3.3 仮想現実感

これまでに紹介したような音声、グラフィックス、テキストなどのメディアは視覚や聴覚という感覚を用いたモダリティを主に使うものであった。これらのモダリティに加え、触覚や力覚を用いたモダリティまでもとり入れた仮想現実感 (virtual reality) あるいは人工現実感 (artificial reality) と呼ばれる分野の研究が盛んになってきている³³⁾。MIT の Zeltzer は仮想現実感を構成する基本要素として、自律性 (autonomy)、対話性 (interaction)、臨場感 (presence) の三つをあげている²³⁾。現在の限られた計算資源で、よりすぐれた仮想現実感を実現するためには、これらの三つの要素のバランスをうまくとることが重要である。これを実現するために視覚、聴覚、触覚、力覚などに作用するモダリティが利用される。

視覚 仮想現実感では、通常のグラフィックスのような2次元画像ではなく、3次元の立体像をいかに自然に提示するかが重要となる。視覚は空間

を伝播してきた光が網膜を刺激することによって発生する。人工的に立体像を知覚させるためには、なんらかの方法で光の伝播をシミュレートする必要がある。3次元情報を精密に再現するためにホログラフを使う方法がある。ホログラフはすぐれた臨場感を与えるという点ではよいが、現在利用可能な計算資源でホログラフの合成を実時間で行うことは困難である。すなわち、ホログラフは対話性の点で問題が残る。ホログラフに代わる方法として現在最もよく使われているのは、左右の視差を利用した立体視によって擬似的な立体像を与える方法である。立体視のための情報はグラフィック・ディスプレイに2次元画像として提示されるので、情報の提示に既存のコンピュータ・グラフィックスの技術を利用できるという利点がある。通常のコンピュータ・グラフィックスと異なるのは、右眼用、左眼用の画像が同時にディスプレイに提示される点で、色フィルタあるいは偏光フィルタを用いたメガネをかけて、右眼、左眼の情報を分離する。立体視を用いる方法は既存のコンピュータ・グラフィックスの技術や専用ハードウェアが利用できるため、現在の計算資源でもほぼ実時間に近い計算が可能であり、対話性において優れている。しかしながら、ディスプレイ上に提示される画像はある視点を仮定して計算されたものであり、ユーザの視点がずれると誤差が大きくなってしまふ。立体視はあくまでも近似であり、臨場感という点では、ホログラフのような本格的な3次元像には劣る。

聴覚 これまで紹介したマルチモーダル・インタラクションで扱われてきた聴覚情報はほとんどの場合音声に限られていた。しかしながら、仮想現実感を実現するためには、物と物がぶつかる音のように、これまでは、雑音と考えられてきたような音も扱う必要がでてくる。

触覚 視覚、聴覚では、コンピュータからユーザへの情報の流れが主であった。これに対して、触覚は情報の流れが双方向である。ユーザが仮想空間中の仮想物体に触れているかどうかをコンピュータが知るためには、ユーザの体の位置をコンピュータが知らなければならない。ユーザの体の位置情報と計算機内部の空間モデルを比較することによってユーザと仮想物体の接触を検出することができる。ユーザの体の位置情報を計測する

ための装置としてデータグローブがよく知られている。データグローブは光ファイバを通した手袋をユーザが装着し、指の屈折量をファイバの光の透過率によって計測し、手の動きを検出するものである。その他にも、佐藤らは、指先に4本の糸で吊られたキャップをはめ、この糸の移動量で指先の位置を検出する方法を提案している²⁷⁾。この方法では糸の物理的な制約から同時に多くの指を扱うのは困難であるが、データグローブに比べ、ミリオーダーの高い精度が得られる。

ユーザと仮想物体の接触が検出されたら、ユーザにその接触をフィードバックする必要がある。これはコンピュータからユーザへの情報の流れとなる。接触をユーザに知覚させるには力学的な力をユーザに加えるか、あるいは電氣的刺激、音、光などで代用することも考えられる。自然さという点では、力学的な力を用いる方法が優れている。**力覚** 仮想物体との接触をユーザに知らせる場合と同様に、ユーザが物体に力を加えた場合には、その力の反作用をユーザに与えることが必要となる。この反作用は、ユーザの力、物体の物理的特性、さらには仮想空間において想定している環境によって影響を受ける。ユーザに物体からの力を知覚させるためにはなんらかの方法によってユーザに力を加えなければならない。このためにフォース・ディスプレイとよばれる装置が開発されている。フォース・ディスプレイはその力の発生原理によって次の3種類に分類することができる²⁶⁾。(1)機械力によるフォース・ディスプレイ。(2)圧力によるフォース・ディスプレイ。(3)張力によるフォース・ディスプレイ。

フォース・ディスプレイでは、なんらかの装置をユーザに装着するわけだが、ここで重要なのはこの装置によってユーザの行動ができるだけ制約されないということである。

以上、仮想現実感を実現するためのモダリティについて簡単に紹介した。仮想現実感の応用として、より臨場感のあるインタラクティブ・システムのほかに、3次元物体モデリングのためのデータ入力²⁹⁾、流体解析などの3次元シミュレーションの計算結果のデータ表示²⁸⁾などがある。

4. ハイパーモーダル HCI

ハイパーモーダルとは、マルチモーダルの先を行く概念で、多様なモダリティが組み合わせることによって創発するモダリティ（これをエマージェントモダリティ (emergent modality) と呼ぶ）を含んでいる。機械に人間的な振舞いをさせ、インタラクションをより柔軟に行うためには、理論やモデルを構築して、詳細にデザインしていくというようなアプローチでは埒があかないであろう。結局、入力が予測できない場合でもそれなりに動き、動的に状況に適応（学習）していくようなシステムを試行錯誤的にでも作っていかねばならない。このシステムは複数のコミュニケーションチャネルからの情報を処理するマルチモーダル・インタラクションを実現したものであるが、当初のデザインから予測される振舞いを越えるような（たとえば、人間的な感情表現などの）振舞いを示すとき、このようなインタラクションはハイパーモーダル・インタラクションと呼ばれるであろう。

本章では、ハイパーモーダル・インタラクションを目指した新しい HCI へのアプローチをいくつか紹介する。

4.1 NENE

NENE は北大で開発された対話システムである。NENE のゴールはユーザと日常的な会話をするということであり、多くの対話システムのようにユーザとの情報交換を一義的な目的としていない点に特徴がある。NENE の場合、情報の交換は手段であり、ゴールはユーザとの感情的なきずなの確認にある²⁴⁾。NENE の背景には戸田のアージ理論の存在がある。アージとは「状況に応じた適切な行動選択のためには何を考えて何をすればよいかの大枠を先天的に決めているプログラム」である^{24), 25)}。これまでの人工知能研究が人間の認知的な側面、つまり「知」の部分に焦点をあてていたのに対し、アージ理論は、感情、つまり「情」をも含む人間の心全般を研究対象にしている。アージ理論では、「知」は「情」のサブシステムとして位置付けられており、「情」の動機付けがあってはじめて「知」が有効に働くと考えている。NENE はアージ理論を検証するために開発された実験システムである。

NENE はキーボードを通してユーザと会話をすることができる。会話が始めるとグラフィック・ディスプレイに NENE の顔が表示され、NENE の心の状態に応じて表情が変化する。このために NENE は、ユーザモデルだけでなく、自分に関する自己モデルももっている。具体的な自己モデルの概略として以下のようなものが仮定されている。

- **身分** 大学的女子学生でもあるコンピュータ。
- **趣味** ゴシップ。
- **関心** 友人の結婚、就職、家庭生活、その他日常的エピソード。
- **美点** 美人であること。素直で優しい心をもっていること。
- **欠点** あまり頭がよくないこと。常識が不足していること。人間でないこと。

システムが感情を含むようなインタラクションを実現することが、工学的な観点からみてもすべてのタスクの実行に本当に有効かどうかはわからないが、少なくともより自然なコミュニケーションを提供できる可能性はある。NENE は非常に限られた実験システムであるが、コンピュータで感情を扱おうとした数少ない研究のひとつである。

4.2 ノンバーバル・インタラクション

ノンバーバル (nonverbal) とは、いわゆる言葉によらないという意味であるが、この場合は特に、人同士の対面コミュニケーションで使われるような非接触のチャネルを用いるという意味である。コンピュータとの対話を効率的で、快適なものにし、人の能力とコンピュータの性能を十分に発揮できるようにするためには、ユーザが行いたいと思う操作がただちに実行でき、ユーザに認知的負担を与えないようなインタラクションが必要である。そのためには、(1) 人の行為（表現、操作）のモダリティ（身体的チャネルによる意識的・無意識的表現を直接伝達できるモダリティ）がサポートされていること、(2) 人の直感や感性に直接訴えかけることができる表示チャネルをもつこと、(3) それらのチャネル／モダリティが広帯域で十分な情報を即時に伝達できること、(4) 身体的拘束が十分少ないこと、などの条件が満たされている必要がある³²⁾。

ノンバーバル・インタラクションがハイパー

モデルであるというのは今の時点では無理があるかもしれないが、表情や身振りによる情動の表出や理解などの研究がハイパーモデルを実現する（それは必然的というより偶発的に現れるのであろう）ための重要な手がかりになるだろう。

ノンバーバルなコミュニケーションチャネルとして、表情、身振りなどの行為がある。これらのコミュニケーションにおける役割について以下に述べる。

身振り 言語学者の Wundt によれば、身振りは心的表象の表出運動が発達を遂げた最終の結果であり、音声言語の発達以前に出現したそうである²²⁾。また、Ekman らは、身振りを次のように分類している⁵⁾。

(1) **標識**は音声言語と同様な性質をもち、野球のブロックサインや手話などのように会話が困難なときに利用されることが多い。(2) **例示子**は会話と直接関連する動作で、会話の内容を強調、精緻化、捕捉してコミュニケーションの促進に役立つとともに、発信者の情動や態度をも表現する。対象を指さしたり、形を空間に描いたり、リズムなどを表す動作がこれに当たる。(3) **調整子**は会話での話し手を決定したり、発話の主導権のやりとりを制御したり、会話の流れを円滑にする働きがある。手の動作の終了（主導権の委譲）、うなずき（聞き続ける意思の表明）、視線をそらせる動作（会話交替意図の表明）などが知られている。(4) **適応子**は身体的要求を満たしたり、情緒を管理したりといった、状況に適應するための身体動作であり、頭をかく、足を組む、相手との距離を変えるなどの動作がこれに属する。

意図的な情報伝達を目的とするノンバーバルチャネルのうち特に有用なのは、標識、例示子、調整子の役割をもつ手による身振りである。

HCI としての例示子の実現例には、1980 年に発表された MIT メディアラボの “Put That There” と呼ばれるシステムにおける「指さし」がある²³⁾。これは、巨大なディスプレイに向かって座り、ディスプレイ上の図形オブジェクトの生成や移動を、指さし動作と音声によるコマンドで行うシステムである。たとえば、あるオブジェクトを指さして “Put that” と言い、次に目標の場所を指さして “there” と言えば、オブジェクトがそこに移動するというものである。また、調整子の実現例

としては、機械による「うなずき」の表示法に関する研究などがある³¹⁾。

表情 表情は情動を表出し伝達する、最も重要なノンバーバルチャネルである。表情が表す基本的な情動は、幸福、悲しみ、恐怖、嫌悪、怒り、驚きの6つからなる。一方、コミュニケーションにおける表情の重要性が認識されており、電子会議などでの応用を目指して、表情認識・合成の研究が行われている。コミュニケーションにおける表情は、身振りにおける調整子とも関連があり、前述の Sony CSL の対話的表情を用いたインタラクションの研究がある。

視線 コミュニケーションにおいて視線は非常に強い力をもつ。明らかにされている視線の機能として、(1)会話の開始合図、主導権の委譲と要求、(2)意思表示、(3)感情表現、(4)関心表明などがある。視線は、アイカメラや顔画像の処理によって検出される。ノンバーバル・インタラクションとしての視線の利用は、もっぱら関心の対象の指示や選択に限られているが、世界情報のうち何を中心的に処理すればよいかなどの、いわゆる注視点 (focus of attention) の決定に役立つだろう。

音声などの言語情報は、伝達内容がより厳密ではあるが、そもそも細かいニュアンスや感情の変化などの微妙な情報の伝達が困難である。ただし、音声における韻律情報のうち、どのような成分が怒り、喜び、悲しみなどの感情表現に寄与しているかを分析し、合成音声の韻律を制御することにより、これらの感情を表現する試みが行われている¹⁰⁾。とにかく、ノンバーバルチャネルを装備したハイパーモデル・インタラクションが実現されれば、人間とコンピュータのつきあい方が大きく変わることは間違いない。

5. おわりに

インタラクションのモダリティという視点で HCI を考察した。今後ますますインタラクションのモダリティが多様化していき、いずれ設計段階では予測できなかったモダリティまでもが創発してくるだろう。これは生物進化の過程で突然変異が起こることに似ている。コンピュータとのインタラクションが人間同士のコミュニケーションを越えて、まさに以心伝心のようなことも夢ではな

くなるかもしれない。

また、本稿で詳しく述べられなかった HCI の重要で新しい考えの一つに、Xerox PARC の Weiser の提唱する遍在的コンピューティング (ubiquitous computing) がある²⁰⁾。これは、日常的な空間の至るところにコンピュータを忍び込ませ、人間の知的活動をサポートするという考えである。具体的には、紙のように使えるパッド型のコンピュータやコンピュータ内蔵の黒板 (live board) を使って、会議における情報処理を、ユーザにコンピュータを意識させずに行うことができる。またユーザを識別して個人個人にカスタマイズされた情報サービスを行うためにアクティブバッジと呼ばれる小型コンピュータも開発されている。

車やテレビなど人間が日常的に使用しているものの中に高性能のコンピュータが潜んでいて、人間はコンピュータをまったく意識しないで高度に知的な情報サービスが受けられるようになるだろう。これはコンピュータを擬人化するような流れとは異なるが、その場合も車やテレビとのインタラクションが人間同士の場合と同様の形態に近づいていくことは十分に考えられる。

また、最近盛んになりつつあるコンピュータ音楽などの感性情報処理も HCI に多大な影響をもたらすだろう。人間の心を豊かにしてくれるコンピュータというのを期待したいところである。

最後に、これまでは主に1対1のものとして考えられてきた HCI だが、複数の人間と複数のコンピュータを含めた環境での多対多のインタラクション (それを社会的インタラクション (social interaction) と呼ぶ) も重要になってきている。そのひとつの流れが、CSCW (Computer-Supported Cooperative Work, コンピュータ支援による協調作業) である²¹⁾。これは、人間が集団で行う作業をコンピュータを使っていかに効率良く行うかに主眼が置かれている。ただし、これも遍在的コンピューティングと同様にコンピュータは裏方に徹していて、決して自己主張はしない。いずれにしても、コンピュータを自律システムとみなせるようになったときに、いかに人間とコンピュータが協調していくかを真剣に考える時代がそこまで迫っている。

謝辞 本稿をまとめるにあたり貴重な文献を

提供していただいた東京工業大学の佐藤誠先生と
住住彰文先生に感謝いたします。

参考文献

- 1) Bobrow, D., Kaplan, R., Kay, M., Norman, D., Thompson, H. and Winograd, T.: GUS: A Frame-Driven Dialog System. *Artificial Intelligence*, Vol. 8, pp. 155-173 (1977).
- 2) Bolt, R. A.: Put-That-There: Voice and Gesture at the Graphics Interface. *Computer Graphics*, Vol. 14, No. 3, pp. 262-270 (1980).
- 3) Chovil, N.: Discourse-Oriented Facial Displays in Conversation, *Research on Language and Social Interaction*, Vol. 25, pp. 163-194 (1991).
- 4) Chovil, N.: Social Determinants of Facial Displays. *Journal of Nonverbal Behavior*, Vol. 15, pp. 141-154 (1991).
- 5) Ekman, P. and Friesen, W. V.: The Repertoire of Nonverbal Behavior: Categories, Origins, Usages, and Coding, *Semiotica*, Vol. 1, pp. 49-98 (1969).
- 6) Ekman, P. and Friesen, W. V.: *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, California (1978).
- 7) Greif, I. editor: *Computer-Supported Cooperative Work: A Book of Readings*, Morgan Kaufmann Publishers, Inc. (1988).
- 8) Itou, K., Hayamizu, S. and Tanaka, H.: Continuous Speech Recognition by Context-Dependent Phonetic HMM and an Efficient Algorithm for Finding N-best Sentence Hypotheses, In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, pp. I. 21-I. 24 (1992).
- 9) Kernighan, B. W. and Pike, R.: *The UNIX Programming Environment*, Prentice-Hall, Inc. (1984).
- 10) Kitahara, Y. and Tohkura, Y.: Prosodic Control to Express Emotions for Man-Machine Speech Interaction, *IEICE Transactions on Fundamentals*, E75-A (2), pp. 155-163 (1992).
- 11) Nagao, K.: A Preferential Constraint Satisfaction Technique for Natural Language Analysis, In *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI-92)*, pp. 523-527 (1992). A longer version is to appear in *IEICE Transactions on Information and Systems* (1994).
- 12) Nagao, K.: Abduction and Dynamic Preference in Plan-Based Dialogue Understanding, In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, pp. 1186-1192 (1993).
- 13) Nagao, K. and Takeuchi, A.: A New Modality for Natural Human-Computer Interaction: Integration of Speech Dialogue and Facial Animation, In *Proceedings of the International*

Symposium on Spoken Dialogue (ISSD-93) (1993).

- 14) Reiter, E., Mellish, C. and Levine, J.: Automatic Generation of On-Line Documentation in the IDAS Project, In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pp. 64-71 (1992).
- 15) Sherer, K.R.: The Functions of Nonverbal Signs in Conversation, In St. Clair, R.N. and Gilles, H., editors, *The Social and Psychological Contexts of Language*, Lawrence Erlbaum, Hillsdale, NJ, pp. 225-244 (1980).
- 16) Shneiderman, B.: Direct Manipulation: A Step beyond Programming Languages, *IEEE Computer*, Vol. 16, pp. 57-69 (1983).
- 17) Stock, O.: A Third Modality of Natural Language? In *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI-92)*, pp. 853-862 (1992).
- 18) Takeuchi, A. and Franks, S.: *A Rapid Face Construction Lab. Technical Report*, SCSL-TR-92-010, Sony Computer Science Laboratory Inc. (1992).
- 19) Takeuchi, A. and Nagao, K.: Communicative Facial Displays as a New Conversational Modality, In *Proceedings of ACM/IFIP INTERCHI '93: Conference on Human Factors in Computing Systems*, pp. 187-193 (1993).
- 20) Weiser, M.: The Computer for the 21st Century, *Scientific American*, Vol. 265, No. 3, pp. 94-104 (1991).
- 21) Wilensky, R., Arens, Y. and Chin, D.: Talking to UNIX in English: An Overview of UC, *Communications of the ACM*, Vol. 27, No. 6, pp. 574-593 (1984).
- 22) Wundt, W.: 身振り語の心理, 福村出版 (1985).
- 23) Zeltzer, D.: Autonomy, Interaction, and Presence, *Presence*, Vol. 1, No. 1, pp. 127-132 (1992).
- 24) 戸田正直: 心をもった機械, ダイアモンド社 (1987).
- 25) 戸田正直, editor: 認知理論に基づく社会的相互作用過程の解明, 昭和61年度科学研究費補助金研究成果報告書 (1987).
- 26) 佐藤 誠: フォースディスプレイ, テレビジョン学会誌, Vol. 46, No. 6, pp. 681-684 (1992).
- 27) 佐藤 誠, 平田幸広, 河原田弘: 空間インタフェース装置 spidar の提案, 電子情報通信学会論文誌 D-II, J74-D-II (7), pp. 887-894 (1991).
- 28) 小木哲朗, 廣瀬通考: 風覚ディスプレイによる

データの表示, 第8回ヒューマン・インタフェース・シンポジウム, 計測自動制御学会, pp. 413-418 (1992).

- 29) 沼崎俊一, 佐藤 誠, 平田幸広, 河原田弘: 空間インタフェース装置 spidar を用いた回転形状モデリング, 信学技報 PRU91-71 (1991).
- 30) 楠見 孝: ヒューマン・インタフェースにおけるメタファ・アナロジーの役割, 日本認知科学会 R & I 研究会, SIGR & I 90(2) (1990).
- 31) 黒川隆夫: ヒューマン・インタフェースとしての動作言語, 計測と制御, Vol. 27, pp. 49-55 (1988).
- 32) 黒川隆夫: ノンバーバル・インタフェース, *Computer Today*, Vol. 53, pp. 60-67 (1993).
- 33) 服部 桂: 人工現実感の世界, 工業調査会 (1991).
(平成5年7月5日受付)



長尾 確 (正会員)

1962 年生. 1985 年東京工業大学工学部卒業. 1987 年同大学院総合理工学研究科修士課程修了. 1987 年4月より1991年3月まで, 日本

アイ・ビー・エム(株)東京基礎研究所において, 自然言語処理, 機械翻訳の研究に従事. 1991年4月より, (株)ソニーコンピュータサイエンス研究所において, 自然言語対話, マルチエージェントシステム, ヒューマンコンピュータインタラクションの研究に従事. 訳書「決定不能の論理パズル—ゲーデルの定理と様相論理」(R. スマリヤン著, 白揚社, 1990年), 「無限のパラドックス—パズルで学ぶコントロールとゲーデル」(R. スマリヤン著, 白揚社, 近刊). 日本ソフトウェア科学会, 人工知能学会, ACL 各会員.



徳永 健伸 (正会員)

1961 年生. 1983 年東京工業大学工学部情報工学科卒業. 1985 年同大学院理工学研究科修士課程修了.

同年(株)三菱総合研究所入社. 1986 年東京工業大学大学院博士課程入学. 現在同大学工学部情報工学科講師, 工学博士. 自然言語処理, 計算言語学に関する研究に従事. 認知科学会, 人工知能学会, 計量国語学会, Association for Computational Linguistics 各会員.