マルチモーダル擬人化エージェントシステム

石塚 満*

1. まえがき

情報環境と我々人間との接する機会の増大につれて、ヒューマンインタフェースの重要性が増してきている。良いヒューマンインタフェースについての真理を極めて簡単に述べると、“Simple is the best.”ということになろう。simple であれば迷うことはない。しかし、情報システムや環境とのヒューマンインタフェースを考える場合、simple だと simple なことしかできず、これだけでは不十分である。ボタン一つのインタフェースは simple で分かりやすいが、それだけではできることは二、三のタスクの範囲内となる。simple でかつ多様なタスクに対応できる形態が必要とされる。

そこで simple というものを物理的 simple さから拡張し、“人間にとっての simple とは”と考える。すると、日常生活の動作は人間にとって特別に覚えたり意識したりしなくても良いという意味で simple である。この性質を利用したメタファ（Metaphor: 比喩）によるインタフェース、具体的にはアイコン (Icon) をマウスで操作する形態の GUI(Graphical User Interface) が現在の主流になっている。擬人化インタフェースは、メディア技術の進展を背景として、この考えを更に進め、より自然のコミュニケーションに近い形態を実現しようとするものである。情報空間の多様化、技術の進展を反映して各種の新しいインタフェースが出現しつつあるが、このようなポスト GUI の有力な形態がマルチモーダルインタフェースであり、その代表的なものが顔、姿を有し音声対話能力をもつ、インタフェースエージェントによる擬人化インタフェースということになる。

本稿では、顔や人物の認識・合成が活用されるこのような擬人化エージェントによるマルチモーダルインタフェースについて紹介する。単にインタフェースだけでなく、新しいマルチモーダルな情報コンテンツとしても有用である、擬人化エージェントを用いるマルチモーダルプレゼンテーション・システムについても紹介する。

* 東京大学 工学系研究科

Key Words: multimodal interface, anthropomorphic agent, lifelike agent

2. 擬人化エージェントインタフェースの経緯

人間に近いロボットはかなり以前から構想されていたので、擬人化エージェントによるインタフェースも潜在的には大分前から考えられていたことになる。例えば、1968 年の S. キューブリック監督の有名な SF 映画「2001 年宇宙の旅」には、会話能力をもち人格も有する HAL9000 が登場している。HAL9000 は読唇（リップリーディング）ができる程に画像認識能力は優れていたが、顔や姿はもたない大型コンピュータであった。

複雑化する情報空間での情報操作の代行者、司書役としてのエージェントについても、以前より潜在的に考えられていた。代表的にはメディアとしてのコンピュータ、先進的インタフェースの提唱者、開発者として著名な Aran Kay は、1984 年にこのようなエージェントについても語っている[1]。しかし、彼の指向はどちらかというと道具（ツール）としてのコンピュータやインタフェースであり、顔や姿をもったインタフェースエージェントについては特に語っていない。

1988 年に Apple 社は将来メディアの形態として Knowledge Navigator の構想を表現し、その中で顔を持つ Phil と呼ばれる案内役の擬人化エージェントインタフェースを登場させた。これはコンセプトビデオであり実際に実現されたものではなかったが、電子秘書、電子執事の役割を果たす擬人化エージェントインタフェースのイメージを示すことになった。B.Laurel は 1990 年の編著で、この背景となる考え方について語っている[2]。

メディア技術の進展を反映して、1980 年代後半に具体的に実現する研究がスタートしている。日本でのこの顔をもつ擬人化インタフェースの研究開発は、世界的にも先行したものであったと言える[3)–9)]。米国ではどちらかというと、キャラクタエージェントを用いるエンター

テイメント的要素の強い研究、生命性を感じるような感情をもつエージェントの生成、教育分野への適用を図る研究等に関心が向けられてきたと言えよう[10)]。1990 年代になると自律性を持ったソフトウェアモジュールとしてのソフトウェアエージェントが注目されるようになったのに伴い、この人間との接面の部分はインタフェースエージェントと呼ばれ、その有力な形態として擬人化インタフェースエージェントが位置付けられるようにもなっている。（米国では擬人性 (anthropomorphism) というよりも、生命をもっているような性質に焦点を当てて lifelike agent と呼ばれることが多い。）

1996 年には B. Reeves と C. Nass（スタンフォード大）は著書"The Media Equation" [11)] の中で、我々人間はコンピュータを含む人工物のメディアに対しても自然に社会性を感じ、人に接するのと同様な対応を示す傾向をもつことを明らかにした。これは人間の（古い）脳は人間との関係により社会性をもつインタラクションを学習して進化してきており、これが今日のコンピュータやメディアに対しても働いていることに起因しているとしている。擬人化インタフェースにしなくても人間は自然に擬人性を感じて対応してしまうという主張であり、全面的に擬人化インタフェースを支持しているという訳ではないが、擬人化インタフェースの設計に参考になることを色々と提示している。

擬人化エージェントインタフェースの実現には、画像メディア部分、音声メディア部分、更にはこれらを統合して動作させることが必要であり、なかなか容易なことではなかったが、1998 年頃よりキャラクタエージェントのツールが登場するようになった。Microsoft Agent[12)] はその代表例であり、音声入出力機能も備えている整備されたツールで、種々の研究に利用されるようになってきている。

3. 特徴と効用

マルチモーダルインタフェースといっても触覚、嗅覚の利用はまだこれからであり、現在のところ音声と Icon より情報量の豊かな（動）画像の効果的利用が中心課題となる。コマンド入力のキャラクタベースのインタフェースは"Remember & Type"型、GUI は"See & Point"型とすると、マルチモーダルインタフェース、あるいは擬人化インタフェースは"Ask & Tell"型として特徴付けられる。また GUI が直接操作 (direct manipulation) を原理とするものであるのに対し、特にインタフェースエージェントという場合は間接操作 (indirect manipulation) 型と特徴付けられる。直接操作できれば好ましいが、複雑で直接操作に直接マッピングできないようなタスクが増えてきている時代背景もある。インタフェースの設計は良い道具（ツール）作りを参考にするという考え方が強くあったが、擬人化エージェントインタフェースの設計については道具ではなく我々の仲間、協力者としての良いパートナを産み出すという考え方への変更が必要とされる。

音声対話インタフェースは 1970 年代より研究が続けられており、着実に進歩しているが、認識性能と共に、機械や単なるディスプレイの外観を呈するコンピュータに話かけるのは、違和感があり心理的な負荷となるという問題がある。正しく理解されたかどうかも分かりにくい。顔や姿を持つことの第一の効用は、ユーザのこのような認知的負荷の軽減である。幾分関係する第二の効用は、音声言語を補助する視線、表情、身振りなどの情報（ノンバーバル情報）もバックグランド・チャネルで伝達し、理解容易度の向上を可能にすることである。人間は脳中に顔認識の特別な機構を有するように発達してきたと言われているように、顔は社会的インタラクションの重要なチャネルである。関係する事項として、音声認識結果が不十分な時には「もう一度お願いします」などと再発声を要求することになるが、単なる機械に要求されると良い気持ちはしないが、姿、表情を有し親切な人格的なものを感じる時には、自然と受け入れられるようになる。第三の効用は、背景あるいは擬人化エージェントの状態により対話の状況、文脈を明らかにし、音声対話の話題の範囲を自然に限定して、認識・理解の性能向上を可能にすることである。

以上のことはコンピュータから人間へのマルチモーダル情報伝達の話だが、逆方向のユーザの状況や表情の認識情報はコンピュータにとり、より気配り思いやりのあるインタラクションの生成に重要となる。またジェスチャ、身振りによるユーザからの情報伝達も可能となる。顔や姿の効用は、海外で電話だけで流暢でない英語で対話する時と、face-to-face の対話との差によってよく理解できるであろう。

顔をもつ擬人化インタフェースは全ての場面で有効であるという訳では必ずしもない。例えば、日常よく使うワープロやメール処理などでは必ずしも擬人化は必要なく、かえって煩わしいということにもなってしまう。しかし、情報空間の広がりと多様化により、我々はしばしば馴染みの薄い多くのシステムや機能に接するようになるが、このような場合、特別な知識がなくても日常生活のメタファでインタラクションを可能にする擬人化インタフェースは有効であり、親しみのもてる情報空間の形成を可能にすることになる。我々と話が通じるパートナが多数存在するような情報空間である。

擬人化エージェントの顔は CG で作成した漫画的な顔

のキャラクタが多く用いられる。3 次元モデル上に実際の顔画像をテクスチャマッピングして動作するようにした、自然感の高い顔も用いられるが、自然感が高く実際の人間に近すぎると、能力も人間並とユーザに過度の期待をもたせてしまうという問題点の指摘もある。

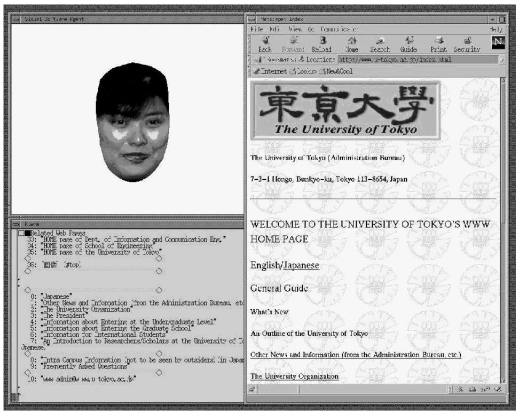
4. システムの研究開発例

国内外の幾つかのシステムの研究開発例と、その特徴を紹介する。

東芝の TOSBURG II[13] はハンバーガ注文の受付をするマルチモーダル擬人化インタフェースであり、音声信号・対話処理部が大きな特徴で、独自開発のエコーキャンセラを組み込んでおり、コンピュータが発声中でもユーザの発話を受け付けられる。顔は漫画的な顔を用いている。東芝ではこの擬人化インタフェースを発展させ、社員のもつ知識、know-how を知識ベース化し、それを共有するシステムの秘書エージェント・インタフェースとして運用している[14]。

ソニー CSL の Talkman[15),3)–5),9)] は会話中の表情の生成の特徴があり、ソニーのコンピュータ製品の質問応答が実現された。顔の 3 次元モデルは K.Waters の DECface を元にしたものを用い、実画像をテクスチャマッピングしている。その後、ソニー CSL では二人の人間の会話を脇で聞いていて、会話の途切れや誤解等が生じているようなら割り込んで、会話が円滑に進むようなアドバイスの発言をする、社会エージェントの試みも行われた[16]。意味、状況理解などの知識処理の部分の充実が必要で、すぐに実現するのは難しいが、将来のパートナとしての擬人化エージェントの可能性を提示している。

東大の我々の擬人化エージェントインタフェースは VSA(Visual Software Agent)[17)–25)] と称する。1990 年前後には並列コンピュータの作製も含めて、3D モデル上にテクスチャマッピングした自然感の高い顔を実時間動作させる等のメディア技術に力を注いだが、1995 年からは WWW ブラウザ（当初は Mosaic、次いで Netscape）との結合が大きな特徴になっている。単なるインタフェースだけだと物珍しいというだけで終ってしまうことになりかねず、システムは有用で魅力あるコンテンツを提供する必要があるが、このような情報コンテンツの作成は容易ではない。膨大な情報コンテンツを含む WWW 情報空間との結合は、VSA の実用性を大きく高めることになり、今後の情報空間の重要なインフラストラクチャになりつつある WWW と人間との間の新形態の接面を提供することになった。同時に、擬人化エージェントを介して提供する（マルチメディア）情報コンテンツを、汎用性のある標準の HTML 形式で作成できるようになった。第 1 図は、この WWW ブラウザと結合し、顔をもつエージェントとの音声会話を通して WWW ブラウザを操作できる VSA システムの画面を示している。



第 1 図　Netscape ブラウザと結合した VSA システム

電総研のシステム[26),27)] は、細やかな腕と指の動きをもつキャラクタエージェントに特徴があり、また登録してあるユーザの顔を認識して「○○さん」とか話しかけてきたりする。この電総研のシステムと東大の WWW ブラウザとの結合技術を用いて、法律知識ベースに基づき法律に関する相談のマルチモーダルインタラクションを行う、ユニークな実験システム Mr.Bengo も作成された[28]。また、RWC(Real World Computing) プロジェクトでは、擬人化エージェントインタフェースをノートパソコンに実装している[29]。

この他、NEC では電子図書館の擬人化秘書エージェントを開発してる[30]。またオムロンでは、感情機能をもつマルチモーダルエージェントを開発しており、音楽を演奏するといったエンターテイメント面の応用を図っている[31]。

海外（特に米国）では、エージェントキャラクタの life-likeness や believability を重要視した研究が多い。（believability は直接には“信頼できる性格”といった意味だが、ここでは“人間味を感じられる性格”といった意味で語られている。）　一つの目標として、アニメータによって生成されるディズニーのアニメーション・キャラクタのような親しみのもてる動作、性格を、自律性を有するエージェントに付与するためのメカニズムの確立が意識されている。

早くから Beliebable Agent を標榜して研究を進めてきたのが、J. Bates（カーネギメロン大）らの Oz プロジェクト[32] である。ボールに目が付いただけの外見のキャラ

クタを用いているが、このCGキャラクタが感情をもっているような不思議な動き、ユーザとの（情緒的）インタラクションをする。

これに続き、現在ではインタラクティブなLifelike Agents研究の中心的存在になっているのが B. Hayes-Roth（スタンフォード大）であり、Virtual Theater プロジェクトを進めている[33]。良く知られた実現例が Spencer's Bar であり、バーのカウンタの環境で個性的な姿をしたユーザの化身であるアバタと自律的なエージェントが、オンラインチャット（従ってユーザからの入力はキー入力）を行うようになっている。

ビジネス面の影響力の点で注目されるのがマイクロソフトの動向である。前述の Microsoft Agent ツール[12] を公開しているように、キャラクタエージェントによるマルチモーダルインタフェースに関心をもち、Persona プロジェクトを進めている[34]。Office97 からユーザが困ったと思われる時に姿をもつディジタルアシスタントが登場するようになっているが、今のところ話し言葉を吹き出し中に文字で出力する形式である。近いうちに音声入出力を可能にすると思われ、広くユーザに受け入れられる様式とタイミングを探っているものと思われる。

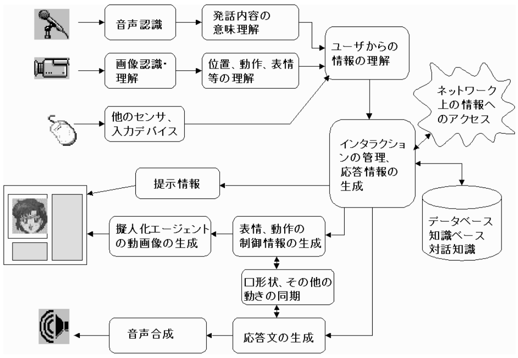
情報コンテンツと結び付いたキャラクタエージェントによるインタフェースとしては、教育分野が進んでいる。この場合のエージェントは教師エージェント(pedagogical agent)と呼ばれる。エージェントを用いる効用は、親しみのもてるインタラクションを可能にすることに加えて、分かりやすく記憶に残るような教示効果をもたらすことである。

J. Lester（ノースカロライナ州立大）らの COSMO システムでは、奇妙な顔、姿のキャラクタエージェントがインターネット・プロトコルについて質問に答え、教示する[35]。

L. Johnson（南カリフォルニア大ISI）らのシステムでは、Steveと呼ばれる教師エージェントが3D環境中で動作し、艦船の機器の操作法や修理法について教示する[36]。このような教示の場合、指示動作と"この"等の指示代名詞の発声とを同期させ、マルチモダリティを効果的に協調させることも重要な課題となる。教育応用に関連するプレゼンテーションについては、6章でも述べることにする。

5. 要素技術とその統合

マルチモーダル擬人化インタフェースを実現するためには、第2図に示すように多くの要素技術が必要となる。また、これらを効果的に統合したシステムとしなければならない。人間の機能に対応させると、耳、口、顔と身体、頭脳、目の機能を実現し、これらを調和させて



第2図 マルチモーダル擬人化インタフェースの構成

動作させる必要がある。

メディアの入出部については利用できるツールが増えてきており、以前に比べて実現の環境は向上しているが、必ずしも目的に適合する機能がなかったり、インタフェースが適合しないことも多い。頭脳に相当する部分はまだ汎用ツールといったものは存在しないので、作成しなければならない。音声認識、音声合成、キャラクタエージェントの動作機能を統合した Microsoft Agent ツール[12]（現在は日本語対応はなし）は、現在のところキャラクタエージェントによるマルチモーダルインタフェース作成に最も適合しており、この提供機能の範囲でなら作成は比較的容易になってきている。

音声認識の性能は進歩してきてはいるものの、制限が少ない自由対話を扱うのは応答の生成も含めて困難な現状であるので、制限を感じさせないようにうまく対話の状況を設定する必要がある。

擬人化インタフェースのための顔画像の合成と認識については、既に■■■■ 6),7) などに解説があり、本特集の別稿にも記されているので、ここでは特に記さず、少しだけ我々の研究について紹介しておくことにしたい。

ユーザとエージェントのインタラクションで親密感を出す上で重要なのが、視線一致(eye-contact)である。ユーザの位置は入力に画像を使うとすると、顔の肌色領域を検出する方法がよく使われる。我々はこの肌色領域の検出に加え、髪領域、動きと履歴情報も利用して、複数の人物がいる環境下でカメラ（従ってディスプレイ）の方を向いている人物（ユーザ）のみを安定して認識、抽出する研究を行った[19]。より簡単なユーザ位置の検出法の開発も行った[37]。

このように検出したユーザの位置の方向へ擬人化エージェントの顔を向ければ、近似的に視線一致が実現されることになる。しかし、固定ディスプレイを使用したのでは、擬人化エージェントの顔、あるいは視線の向き

第 3 図　物理的な視線一致を可能にする 2 軸回転液晶ディスプレイ

は、少しそちらの方向を向いたと思われるような効果しか出せない。そこで、最近、2 軸 (Pan-Tilt) の回転制御可能な液晶ディスプレイ装置（第 3 図）を開発した[38]。顔はディスプレイ上での合成だが、物理的な動きが可能となっている。「眼は口程に物を言う」と称されるように、この動きによりエージェントは関心を向けている対象を示すことで、視線一致を始めとしてインタラクションの新しい情報伝達のバックグランドチャネルをもつことになった。

ユーザの画像認識については、広視野角の固定カメラと狭視野角の回転カメラの 2 視野角ビジョンシステムにより、広視野角カメラで狙いをつけた方向に狭視野角カメラを向けてユーザの口の動きをとらえ、発声区間の検出する方法の開発も行った[39]。これにより、ユーザのシステムへの音声入力はスイッチを押して指示したり、あるいは決められた区間で音声入力する必要があったのを、ユーザの口の動きから音声入力スイッチをオンにすることにより自動化を可能にした。

6. 新形態の情報コンテンツとしてのマルチモーダル・プレゼンテーション

GUI の次の世代の有力なインタフェースの形態としての擬人化インタフェースについて以上に紹介したが、単なるインタフェースとしてだけでなく、魅力あるマルチモーダル情報コンテンツの作成、流通も含めてシステムやツールを考えることも、今後重要になってくる。キャラクタあるいは擬人化エージェントを用いるマルチモーダルな情報コンテンツとして最も有望な形態が、マルチモーダル・プレゼンテーションであろう。

プレゼンテーションは現在の音声認識技術レベルからも好ましい応用の形態といえる。即ち、音声合成については実用的に使えるレベルになっていると言えるが、音声認識については進歩してきてはいるものの自由対話を扱うには必ずしも十分とは言えない。誤認識は音声インタフェースの不自然さ、煩わしさを増し、評価を悪くしてしまう。マルチモーダル・プレゼンテーションでは音声については大部分が出力であり、入力は数％に過ぎないので、現状の音声技術レベルにも適合した応用形態といえる。

現在のプレゼンテーション・ツール（PowerPoint など）は説明用の画面の作成・表示機能を提供し、人間のプレゼンタがこれを操作し、音声と動作でプレゼンテーションを行う形式である。これはマルチモーダルな情報チャネルの使用、そしてインタラクティブ性により、単に文書や本などによる情報提供に比べて情報伝達量が多く、我々人間の理解に適していることから、現代の代表的プレゼンテーション形式になっている。ここでのキャラクタエージェントによるマルチモーダル・プレゼンテーションは、人間のプレゼンタがいなくても近似的に同様な効果を達成するものであり、同時に WWW 上の新しいマルチモーダルな情報コンテンツを可能にしようとするものである。これによって、時間、空間を越えて我々人間にとって分かりやすくて親しみやすいマルチモーダル情報コンテンツの流通、提供を可能にしようとするものである。以下にこのような研究を紹介する。

ドイツ人工知能研究センター (DFKI) では長年、マルチメディアを連携させたプレゼンテーションの研究を行ってきており、その一環としての PPP Persona[40] では、第 4 図のようにヒッピー風のキャラクタエージェントが各種のマルチモーダルなプレゼンテーションを行う。プレゼンテーションの場合にはポインティング動作が重要となるが、ここでの指し棒の使用はなかなかうまい工夫である。WWW での配信の連携も図られているが、ユーザの音声入力による質問等の受け付け機能はもたない。

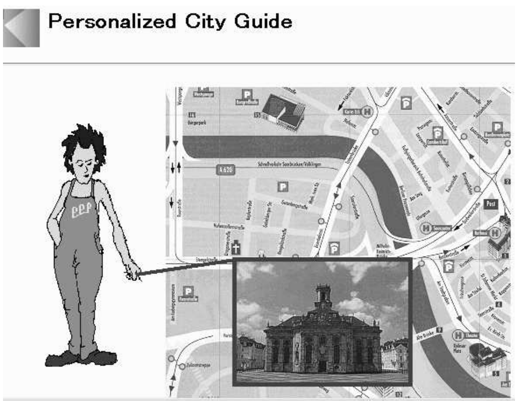
ペンシルベニア大の N. I. Badler らは 10 数年にわたり人物のアニメーションの研究を行ってきており、Jack と呼ぶ 3D 人物アニメーションシステムを作成している[41),42]。この Jack は自然に近い動きをすることが大きな特徴であり、これをテレビの天気予報説明者のように仕立てたプレゼンテーションシステムを作成している[43]。音声出力はあるが、音声入力の受け付け機能はもたない。

これらを含めてマルチモーダル・プレゼンテーション・コンテンツの作成は、個々のシステム毎に固有の方法（固有のコマンド系列、スクリプト言語など）で、多く

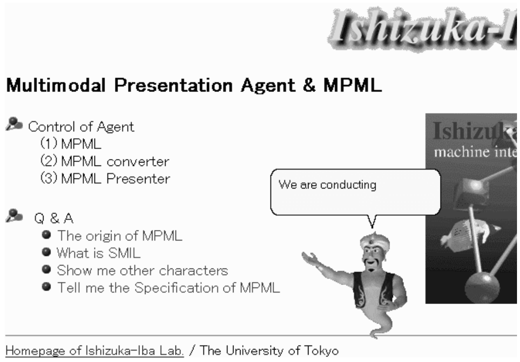
の場合は細部に至るまで記述しなければならず、容易ではない段階にある。我々は多くの人々がHTMLでWebページを作成するように容易にマルチモーダル・プレゼンテーションを記述することを可能にし、かつ各種キャラクタエージェントシステムに共通的に使用できるようにする目的で、記述言語MPML(Multimodal Presentation Markup Language)を開発した[44),45)]。MPMLはXML規格に準拠したマークアップ言語であり、特定のキャラクタシステムに依存せず、エージェントの動作命令などを容易に記述できる。また、SMILに準拠したメディア同期機能(映像コンテンツをプレゼンテーションに同期して提示できる)、音声入力、音声発話によるプレゼンテーション制御機能をサポートしている。第5図はこのMPML記述を変換ツールにより、Microsoft Agentを動作させVBScriptに変換して実行させている、マルチモーダル・プレゼンテーションを例示している。背景となるプレゼンテーション画面の記述には通常のHTMLを使用する。音声による自由な質問や要求に答えるという訳にはいかないので、対応できる質問や要求をユーザに分かるように画面上に明示して対処している。

7. 人工感情

人に接するインタフェースエージェントは人との基本的なコミュニケーション機能をもつだけでなく、愛着を感じさせるような性格のパートナになることが求められる。顔をもつのに機械的反応しかないのでは社会性がなく、長く付き合うのには不満を感じることになってしまう。そこで、愛着を感じられるような生命感(lifelikeness, beliebability)や個性を付与するための重要な要素として、人工的な感情(emotion)機能が注目されること



第4図　PPP-Personaのマルチモーダルプレゼンテーション(独DFKI)



第5図　MPML記述によるMicrosoft Agentを使用してのプレゼンテーション

になる。

感情は喜び(幸福感)、悲しみ、驚き、怒り、嫌悪、恐れ、・・・などの言葉で語られるが、場合によりそのカテゴリ分けは不統一で、根拠も不十分なものであった。これを整理し、今日の多くの感情機能の基礎になっているのが1988年に著書として発表された、認知評価理論(cognitive appraisal theory)あるいは提唱者3名の名前をとりOCCモデル[46)]と称される感情モデルである。簡単に紹介すると、このOCCモデルでは感情をまず何に向けられたものであるかによって整理し、1)事象の結果(consequence of events)に対する感情(これを更に他者に対する結果と自己に対する結果の感情に分ける)、2)エージェントの行動(actions of agents)に関する感情(このエージェントを更に自己と他者に分ける)、3)物への心相(aspects of objects)の感情、に大別する。そして、この分類法に基づき、22種の感情を位置付けている。

感情はOzプロジェクトにおけるように、劇場(theater)といった環境で人とインタラクションする(エンターテイメント性を目的とする)エージェントでは特に重要な要素となっている。これは昨今の電子ペットにも共通している。C. Elliot(デュポール大)の研究は、特にこのようなインタラクティブ・エージェントの感情の動きに焦点を当てており、彼はAffective Reasonerと称している。外界とのインタラクションによって生成される感情の動きを実装する代表的研究が、J. D. Velasquez(MIT)による方法である[47)]。R. Picard(MIT)は1997年の"Affective Computing"の著書[48)]で、このような感情に関係するシステムの研究状況をまとめている。(affectあるいはaffective(必ずしもぴったりしないが訳をつけると情緒、情緒的)は　emotion, emotional(感情、感情的)よりもやや広い概念を表してい

る。）

擬人化インタフェースのインタラクション設計に関係する異なる知見についても付言しておくと、J. Cassell(MIT) はエージェントの感情表現（喜びや困惑の表情）によるフィードバックよりも、彼女が Envelope フィードバックと呼ぶ視線や手の動きや頷き等の頭の動きの方が、より効果的で大切であるという実験結果を示している[49]。

8. むすび

顔と姿を持つマルチモーダル擬人化インタフェースに関する研究について紹介した。同様な趣旨の解説は既に何編か存在しているので、海外の研究状況を多く含める等により、読者の参考になるようにした。

顔や姿とで表される擬人化あるいはマルチモーダル性は、ヒューマンインタフェースの認知に新たな課題を供するようになってきており、単に工学的アプローチだけでなくアートとの融合も重要な課題となる。単なるインタフェースにとどまらず、新形態のマルチモーダル情報コンテンツ、エンターテイメント・コンテンツへの広がりをもつ領域でもある。実用性も高い領域であり、人間の認知、感性に関する洞察に基づき、メディアと知能の技術を駆使、開拓して展開を図ることが求められている。

謝辞: 本稿関係の研究について議論する、VSA の共同研究者である土肥 浩氏に感謝いたします。

(■■)

■ ■ ■ ■

1) A. Kay: Computer Software; *Scientific American*, Vol. 251, No. 3, pp. 41–47 (1984)
2) B. Laurel: Interface Agent: Metaphors with Characters; in *The Art of Human Computer Interface Design* (B.Laurel(ed.)), pp. 355–365, Addison-Wesley (1990)
3) 長尾、徳永： ２１世紀のヒューマンコンピューターインタラクション; 情報処理, Vol. 34, No. 11, pp. 1325–1334 (1993)
4) 竹内彰一： 顔・表情とインタフェース; ビジュアルインタフェース（bit 別冊、平川、安村（編））, 6 章, pp. 117–130, 共立出版 (1996)
5) 長尾確： マルチモーダルインタフェースとエージェント; 人工知能学会誌, Vol. 11, No. 1, pp. 32–40 (1996)
6) 長谷川、森島、金子：「顔」の情報処理; 信学論, Vol. J80–A, No. 8, pp. 1231–1249 (1997)
7) 金子、森島： 顔画像によるアドバンストエージェント; 映像メディア学会誌, Vol. 51, No. 8, pp. 1169–1174 (1997)
8) 速水、竹澤： マルチモーダル情報統合システムの研究動向; 人工知能学会誌, Vol. 13, No. 2, pp. 206–211 (1998)
9) 長尾確： エージェント指向インタフェース; *bit*, Vol. 31, No. 2, pp. 25–34 (1999)
10) C. Elliot and J. Brzezinski: Autonomous Agents Synthetic Characters, *AI Magazine*, Vol. 19, No. 2, pp. 13–30 (1998)
11) B. Reeves and C. Nass: *The Media Equation*; Cambridge Univ. Press (1996)
12) http://msdn.microsoft.com/workshop/imedia/agent/
13) 竹林洋一：音声自由対話システム TOSBURG II—ユーザ中心のマルチモーダルインタフェースの実現に向けて—; 信学論, Vol. J77–D–II, No. 8, pp. 1417–1428 (1994)
14) 河野、屋野、他：ATMS ベースのマルチモーダル入力統合方式を用いたインタフェースエージェントシステム; 人工知能学会誌, Vol. 13, No. 2, pp. 212–220 (1998)
15) A. Takeuchi and K. Nagao: Communicative Facial Displays as a New Conversational Modality; *Proc. ACM/IFIP INTERCHI'93*, pp. 187–193 (1993)
16) K. Nagao and A. Takeuchi: Social Interaction: Multimodal Conversation with Social Agents; *Proc. AAAI-94*, pp. 22–28 (1994)
17) O. Hasegawa, C. W. Lee, W. Wongwarawipat and M. Ishizuka: A Realtime Visual Interactive System between Finger Signs and Synthesized Human Facial Images employing a Transputer-based Parallel Computer, *Visual Computing* (T.L.Kunii(ed.)), Springer-Verlag, pp. 77–94 (1992)
18) H. Dohi and M. Ishizuka: Realtime Synthesis of a Realistic Anthropomorphous Agent toward Advanced Human-Computer Interaction; *Proc.HCI'93*, Elsevior, pp. 152–157 (1993)
19) 長谷川、横澤，石塚：自然感の高いビジュアルヒューマンインタフェース実現のための人物画像の実時間並列協調的認識; 信学論, Vol. J77–D–II, No. 1, pp. 108–118 (1994)
20) O. Hasegawa, C. W. Lee, W. Wongwarawipat and M. Ishizuka: Realtime Synthesis of a Human Agent in response to the User's Moving Image; *Machine Vision and Applications*, (Springer-Verlag), Vol. 8, pp. 122–130 (1995)
21) H. Dohi and M. Ishizuka: A Visual Software Agent connected with the WWW/Mosaic; *Proc. Multimedia Japan'96*, pp. 392–397 (1996)
22) 土肥、石塚：WWW/Mosaic と結合した自然感の高い擬人化エージェントインタフェース; 信学論, Vol. J79–D–II, No. 4, pp. 585–591 (1996)
23) H. Dohi and M. Ishizuka: Visual Software Interface with Realistic Face and Voice-controlled Netscape; *Proc. Int'l Conf. on Computational Intelligence and Multimedia Applications (ICCIMA'97)*, pp. 225–229 (1997)
24) M. Ishizuka and H. Dohi: Multimodal Anthropomorphic Agent Connected with WWW Information Space; *Proc. (CD-ROM) 9th NEC Research Symp. on Human-centric Multimedia Community* (1998)
25) 土肥、石塚：Face-to-Face 型擬人化エージェント・インタフェースの構築; 情処論, Vol. 40, No. 2, pp. 547–555 (1999)
26) O. Hasegawa, et al.: Active Agent Oriented Multimodal Interface System, *Proc. IJCAI–95*, pp. 82–87 (1995)
27) 長谷川、坂上、速水：実世界視覚情報を対話的に学習・管理する人間型ソフトウェアロボット; 信学論, Vol. J82–D–II, No. 10, pp. 1666–1674 (1999)
28) 新田、長谷川、他：論争支援マルチモーダル実験

システム Mr.Bengo; 信学論, Vol. J80–D–II, No. 8, pp. 2081–2087 (1997)
29) 向井、他：マルチモーダルヒューマンインタフェースのノートパソコンへの実装; 1998 年度人工知能学会全大、No. S5–01 (1998)
30) 神谷、呂山、原、宮井：3 次元フォークスルーと CG 秘書を用いた電子図書館インタフェースの開発; 情処 IM 研資料, IM19–5 (1995.1)
31) 牛田、平山、中嶋：自律的行動決定モデルに基づくインタフェースエージェント; 信学論, Vol. J82–D–II, No. 10, pp. 1655–1665 (1999)
32) J. Bates: The Role of Emotion in Biliebable Agents; *Commun. ACM*, Vol. 37, No. 7, pp. 122–125 (1994)
33) B. Hayes-Roth and P. Doyle: Animate Characters; *Automonous Agent and Multi-Agent Systems*, Vol. 1, No. 2, pp. 195–230 (1998)
34) G. Ball, D. Ling et al.: Lifelike Computer Characters: The Persona Project at Microsoft Research; in *Software Agent* (J. M. Bradshow (ed.)), Chapter10, pp. 191–222, AAAI Press/The MIT Press (1997)
35) J. Lester, et al.: Deictic Believability: Coordinating Gesture, Locomotion and Speech in Lifelike Pedagogical Agents; *Applied Artificial Intelligence*, Vol. 13, No. 4–5, pp. 383–414 (1999)
36) J. Rickel and W. L. Johnson: Animated Agents for Procedual Trainning in Virtual Reality: Perception, Congnition, and Motor Control; *Applied Artificial Intelligence*, Vol. 13, pp. 343–382 (1999)
37) 土肥、森川、石塚：擬人化エージェントによるインタフェースのための実時間ユーザ追跡機構; 画像電子学会誌, Vol. 25, No. 2, pp. 121–129 (1996)
38) H. Dohi and M. Ishizuka: Visual Software Agent Interface with a User-tracking Flat Panel Display for Continuous Eye Contact; *Proc. 3rd World Multiconference on Systems, Cybernetic and Information and 5th Int'l Conf. on Information Systems Analysis and Synthesis (SCI/ISAS'99)*, pp. 254–260 (1999)
39) Y. J.Huang, H. Dohi and M. Ishizuka: Man-Machine Interaction using a Vision System with Dual Viewing Angles; *IEICE Trans. Inf. & Syst.*, Vol. E80–D, No. 11, pp. 1074–1083 (1997)
40) E. Andre, T. Rist and J. Muller: WebPersona: A Lifelike Presentation Agent for the World-Wide Web; *Knowledge-Based Systems*, Vol. 11, No. 1, pp. 25–36 (1998)
41) N. I. Badler, C. B. Phillips and B. L. Webber: *Simulating Humans: Computer Graphics, Animation, and Control*; Oxford Univ. Press (1993)
42) J. Cassell, C. Pelachaud, N. Badler et al.: Animated Conversation: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents; *Proc. SIGGRAPH-94*, pp. 413–420 (1994)
43) T. Noma and N. I. Badler: A Visual Human Presenter; *Proc.IJCAI-97 Workshop on Animated Interface Agents: Making Them Intelligent*, pp. 41-51, Nagoya (1997)
44) 筒井、石塚：キャラクタエージェント制御機能を有するマルチモーダル・プレゼンテーション記述言語 MPML; 情処論,（2000 掲載予定）
45) http://www.iv.t.u-tokyo.ac.jp/MPML/mpml.html
46) A. Ortony, G. L. Clore and A. Collins: *The Cognitive Structure of Emotions*, Cambridge Univ. Press (1988)
47) J. D. Velasquez: Modeling Emotions and Other Motivations in Synthetic Agents; *Proc. AAAI-97* (1997)
48) R. Picard: *Affective Computing*; The MIT Press (1997)
49) J. Cassell and K. R. Thorisson: The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents; *Applied Artificial Intelligence*, Vol. 13, No. 4–5, pp. 519–538 (1999)

■■■■

いしづか みつる
石 塚 満 (非会員)
1948 年 4 月 23 日生まれ。1971 年東京大学工学部電子卒。1976 年同大学院博士課程修了、工博。同年 NTT 入社、横須賀研究所勤務。1978 年東京大学生産技術研究所、助教授。同教授を経て、1992 年工学部電子情報工学科、教授。研究分野は人工知能、マルチモーダルインタフェース、知能的ネットワーク化情報空間。情報処理学会、電子情報通信学会、人工知能学会、映像情報メディア学会、IEEE, AAAI の会員。