

インタラクションを通じた人の行動・意図の認識

間瀬 健二

ATR 知能映像通信研究所

〒 619-0288 京都府相楽郡精華町光台 2-2-2

mase@mic.atr.co.jp

概要: インタフェースエージェントが個人とともに行動したり、環境に埋め込まれて、ユーザとインタラクトしながら、ユーザを支援するサービスについて研究を進めてきた。よりよい支援を提供するための1つの手段として、ユーザの行動や意図の認識が、なにをもたらしかを考察する。
キーワード: インタラクション、状況認識、行動認識、意図認識、toy インタフェース、

Recognition of Human Activity and Intention through Man-machine Interactions

Kenji Mase

†ATR Media Integration & Communications Research Laboratories

Abstract: We have been doing the research on the interface-agent which support user's communication and personal activities through various interactions. In this report, we consider how the recognition of user's activity and intention would benefit the support task.

Keywords: toy-interface, interaction, situation recognition, intention recognition

1 はじめに

GUIはプログラム開発、ワープロなどの文書作成、メールやWWWなどのメッセージ通信、表計算のスプレッドシートなどの作業に適しており、オフィスなどで重宝され、大きな発展を遂げた。しかしながら、情報処理技術の進歩は、コンピュータ上で操作管理するデータの膨大な増加をもたらした。個人の認知できる規模を完全に越えてしまっている。いったい、どうやって希望の情報を探し当てればよいのだろうか。それに、コンピュータに伝わる的確なことばで、自分の希望を表現できるのだろうか。そもそも、私たちが欲しいものは、小さなモニタを通して得られるのだろうか。

近年、エージェント的な考え方を導入したインタフェースが注目されている。GUIに対して、そのインタフェースをエージェント指向インタフェース(agent-oriented interface)、その具体的なオブジェクトはインタフェースエージェント(interface agent)[1]と呼ばれる。インタフェースエージェントは、コンピュータが人間に対して、冷淡な機械としてではなく、少し人間らしく、感情をもった機械

として接してくれるオブジェクトになりうる。情報を提示するときに、表情や韻律で含みを伝えたり、人間側の要求や状況を、対話によって積極的に摂取してくれたら便利ではないか。

本文は、このようなインタフェースエージェントが個人とともに行動したり、環境に埋め込まれて、ユーザとインタラクトしながら、ユーザを支援する場面を想定する。そして、よりよい支援を提供するための1つの手段として、ユーザの行動や意図の認識が、なにをもたらしかを考察する。

2 擬人化エージェント

インタフェース・エージェントの最も大きな特徴は、視覚的な擬人性(視覚擬人性、visually anthropomorphic agent¹)といえる。自分の代理であれ、コンピュータの代理であれ、エージェントが視覚的なオブジェクトとして表現されることで、人とのインタラクションがスムーズになる。

¹ Nass& Reeves の The Media Equation によれば、人間は視覚的な要因がなくても全てのものに擬人性を見いだす特性がある [2]。ここでは視覚的な表現としての擬人性を議論する。

身体性をもった擬人化エージェントに動作をつけ、人間の行動や状況を認識する手段を与えることによって、人間同士と同様なノンバーバルコミュニケーションが成立する[3]。よく知られているように、人間は言語 (verbal language) だけでコミュニケーションを行っているのではない。我々はノンバーバルな情報として、視覚的には外観、姿勢、表情、視線、ジェスチャ、距離などを用いている。なかでもジェスチャには、図像 (iconic)、比喩 (metaphoric)、拍子 (beat)、指示 (deictic) などの用法があり、言葉とともに使われることが多い [4]。これらはボディランゲージとも呼ばれる。また、聴覚的には、周辺言語 (para language) と呼ばれる、間の取り方、ピッチ、早さなどの、言葉ならざる言語がある。周辺言語は、対話におけるやりとりをスムーズにしたがり、感情の伝達に役にたつ。

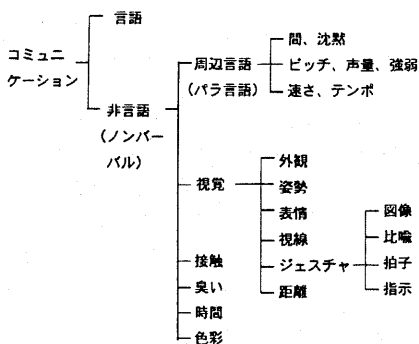


図 1: コミュニケーションにおけるモダリティ

コミュニケーションにおいてはこれらの情報を意識的あるいは無意識に組み合わせて、メッセージを構成し伝達している。これは、コミュニケーションの様相 (モダリティ、モーダル) が複数あるという意味で、マルチモーダルコミュニケーション [5] とよばれる。図 1 に、コミュニケーションにおける多様なモダリティの一覧を示す。

3 知覚すべき状況と手段: 計測、観察それとも対話?

では、どのような情報 (what) をどうやって (how) 知覚させればよいのであろうか。知覚する対象を、ユーザ自身と、ユーザを取り囲む環境とにわけて考

えよう。ユーザを取り囲む環境には、空間とそこに置かれている物のほかに、他の人間も含まれる。さらに、ユーザと環境との関係や相互作用 (インタラクション) も考える必要がある。これを分類すると図 2 のようになる。なお、この図は、ある現時刻のスナップショットの状況を説明している。ユーザの置かれた状況は、過去から未来につながっており、過去の歴史や未来の予測も考える必要がある。

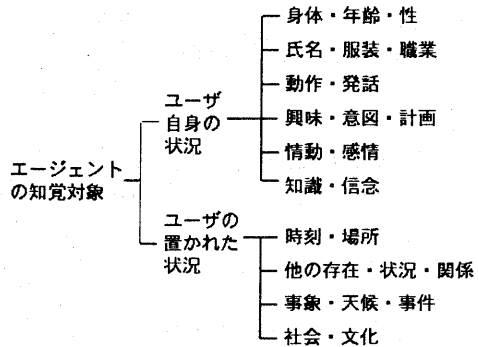


図 2: エージェントが知覚すべき状況

コンピュータエージェントがこれらのユーザの状況をどのように取得すればいいだろうか? 当然、何らかの知覚システムや情報収集システムが必要である。次のようないろいろな方法が考えられる。いろいろな実験システムやサービスが提案されているが、実現するためのコストによって最適の方法を選び利用することになる。

1. ユーザに直接尋ねる (対話システム、GUI、エージェント)
2. ユーザに間接的な問い合わせをして尋ねる
3. ユーザ (身体、行動、メッセージ) を計測する
4. ユーザ (身体、行動、メッセージ) を観察して、識別、認識、理解する (PUI)
5. 第三者 (人) にユーザについて尋ねる
6. 第三者 (機械) にユーザについて尋ねる
7. 周囲を計測、観測し認識する
8. 歴史書、歴史データを調査する
9. 現状から未来を推測する

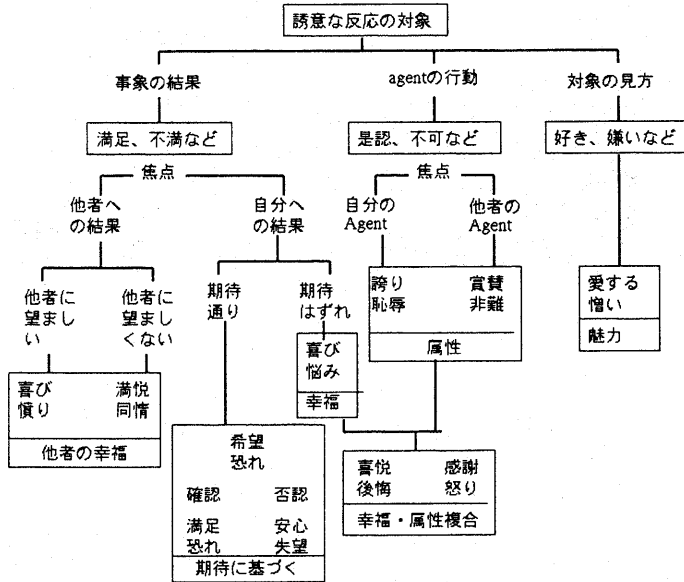


図 3: 感情の認知構造を説明する OCC モデル [6]

4 感情のインタラクション

コンピュータが人間と同じ感情をもてるかどうかは、まだわからないことが多いが、感情的な表現やインタラクションをすることで、人間にとっては、よりスムーズに情報を開示することができるようになる。そのためには、エージェントはユーザの反応をみて状況を認識する必要がある。どのような状況が使えるだろうか。いろいろな段階のモデルがあるが、比較的複雑なモデルとしては、オントニー、クロー、コリンズの OCC 認知モデルがある。彼らは、もともと感情の合成を意図してこのモデルを作ったのではなかったようだが、感情生成の基本的なモデルとしてしばしば引用される（詳しくは文献 [6]、195 ページを参照）。図 3 は OCC 認知モデルの構造である。OCC モデルは、事象（イベント）、エージェント、オブジェクトが産み出すいろいろな状況が、ポジティブまたはネガティブな有意味性を誘発しており（誘発性）、それが感情を起こさせるとしている。これらの状況は確かな軸を規定するので、感情を規定するルールベースが出来上がっている。図に示すように、状況の分岐木をたどっていくと 22 の感情にぶつかる。

5 具体例

いくつかの具体例で、どのような行動や状況認識が望まれ、あるいは実現されているかを紹介する。

5.1 C-MAP とエージェントサロン

ミュージアムや展示会などのガイド、さらに人間同士のコミュニケーションを支援する仲介者として活躍できるエージェントシステムを考えると、ユーザとのインタラクションの中からユーザの興味や背景情報を獲得して、ユーザの内的状況を把握し、自律的に情報フィルタリングをしたりアクションを起こし、コミュニケーションを活性化させる情報を提供する機能が重要となる。

上記のような機能をもつエージェントを実現するため、個人携帯端末により展示の補足情報を提供するガイダンスシステムにおいて、キャラクターの外見をもつインタフェースエージェントを搭載して、エージェントを介した展示者と訪問者のコミュニケーション支援システムを試作した。我々はこのプロジェクトを C-MAP (Context-aware Mobile Assistant Project) [7] と呼んでいる。

C-MAP のシステムは、展示関連情報のデータベースおよびその提供サーバ (WWW)、無線 LAN

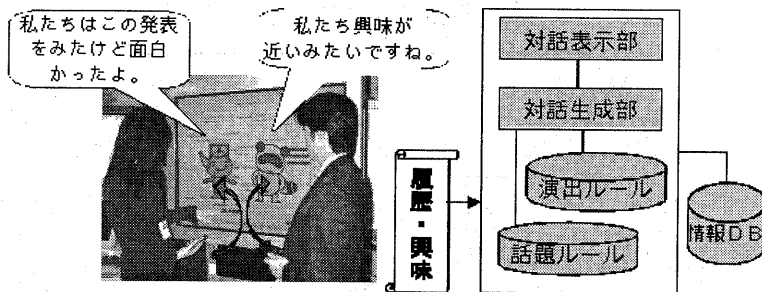


図 4: エージェントサロン

を介してサーバシステムと通信可能な複数の携帯情報端末 (Windows ベースの PC) や小型 PDA を基本構成とする。見学者の位置検出には Active Badge System (以下 ABS) を用いる。ユーザにバッジを着用してもらい、各展示説明のポスターブースにセンサを設置して、バッジとセンサ間の赤外線リンクによりユーザの位置を検出する。ルートプランニングと提案などのガイドサービスをするエージェントは、エージェントサーバで動作する。ジェスチャインタフェースを使って仮想空間を歩き回る VisTA-walk には、C-MAP のエージェントが乗り移ってガイドを始める [8]。システムはエージェントサーバ経由で ABS からユーザ ID を入手し、ユーザに個人化した展示が可能になるようになっている。

エージェントの主な機能としては、ユーザの興味や背景など静的状況の理解、ユーザの動的状況の監視、および環境全体の状況の把握により、見学のガイドとなる見学プランの提案、ユーザ自身の状況把握のための情報提供、および見学者同士のコミュニケーションの支援をすることである。静的状況として、ユーザの興味、背景情報をあらかじめ投入、獲得しておき、それを見学項目の推薦の参考などに行っている。また、ユーザは動的にその興味を変更することもありうる。動的状況としては、ユーザ別の位置および見学経路情報、各項目の見学滞在時間、各展示の混雑状況、見学イベントのスケジュール情報などがある。環境全体の状況としては、全ユーザの分布状況、全会場の混雑状況などが有効な情報として考えられる。

C-MAP のもう 1 つのアプリケーションである、エージェントサロンと呼ぶシステムは、図 4[9] のように、2、3 人のユーザが自分のパーソナルエージェントを乗り移らせて、そこでエージェントとのイン

タラクションをする。この場合、前にたったユーザの人数やエージェントが自分の主人がどこに立っているか、さらに、ユーザがどのような反応や指示をしているかをしることによって、いろいろなサービスをすることができる。例えば、複数のエージェントが画面に登場することになるが、ユーザの立っている位置に合わせて、画面内の場所を変えることによって、ユーザとの一体感を作り上げることを助けられる。

5.2 人形型エージェント

このようなインタフェースと状況認識型のサービスを提供するエージェントは、ウェアラブルや携帯型の PDA に実装することも考えられる [7]。しかし、そのような小型の携帯 PDA は、ポケットや鞆の中、手のひら、机の上などを行き来するのみで、ユーザとの接触の機会は少なくその強度もあまり強くない。ユーザの状況を知るときに、体温や接触強度、ユーザの視点や環境などの重要な手がかりを得るには従来型の PDA では限界がある。

そこで、エージェントを搭載したウェアラブルなシステムとして、携帯 PDA との間ともいえる、人形型を提案している [10]。人形型は PDA に比べユーザの接触機会が多く、よりユーザの感性的な状況認識がたやすくなることが期待される。また、エージェントを搭載したシステムは、機能的に自律性や個性などの擬人性を有することになるが、外見的にも擬人性を有することで、それらの機能へのアクセスが容易になることが想定される。たとえば、エージェントに周囲の状況を見せるときに、顔をそちらに向けてやるなどの動作が自然に生ずるであろう。

実際にセンサー人形を制作した。市販のぬいぐる



図 5: センサー人形の外観

るみ(図5)の中身をくりぬいて、代りにPCとセンサー類をつめている。図6がその内部構造である。初期試作機では、USBビデオカメラ(160x120 pixels)、USBマイクロフォン、標準シリアル接続の3軸角加速度センサおよびサブノートPCを搭載し、スタンドアロン型のシステムを構築し、自動日記の実験を行なった。第2世代の試作機(図6)では、人形とユーザの接触から状況を獲得するために、PCMCIAカード型の16chのAnalog/Digital変換器を導入し、センサーとしてピエゾ圧電素子を利用した圧力センサと曲げセンサ、赤外近接センサ、温度センサなどを人形の各部に装着し、内蔵す



図 6: センサー人形を分解した様子

るコンピュータをカードタイプ(140mm x 100mm x 40mm)のPC(Win2K)として小型化をはかった。接続ポートの都合で、角加速度センサーの代りに1軸の並進加速度センサーを用いた。さらに磁気ディスク(9GB)を搭載して、日記の原データとなるイベントの記憶も可能とした。センサーは合計7種類18個である。無線LANにより外部コンピュータ・インターネットとの接続も可能である。

このセンサー人形システムは、2つの実験用途を考えている。生涯日記の利用[11]においては、持ち歩いてイベントを記録した後、ステーションコンピュータから無線LAN経由でリモートログインして、新規イベントがあればそれをマニュアルで教示して、認識システムである隠れマルコフモデル(HMM)を学習させて次の利用に備える。また、音楽コミュニケーションのための自己表現用途[12]においては、センサーからの状況認識と状況に応じたセンサーから音への変換処理を行ない、人形につけたスピーカをならすようにしている。

6 おわりに

文脈を離れて、画像処理でユーザの動作を認識しているだけでは、なかなか、ユーザの状況や意図はわからない。そういう想いから、インタフェースエージェントに研究の重心をシフトしてきた。コンピュータビジョンにとらわれずに、手にはいるセンサーをいかに使いこなすか。センサーからのデータをどう処理するのか。そこには、パターン認識のテクニックがいかにせる場所が多い。また、ビジョンシステムの人間の行動認識の能力が格段に上がっていることも今後楽しみである。

本文は、思い付くままに、インタラクションや状況において考えるべきことと、我々の現在のトライアルを紹介した。後者の例が、ビジョン研究者のモチベーションを起すヒントになれば幸いである。

ところで、インタラクションの演出も重要な要素である。あるストーリーのなかでユーザ自身が知らないうちに限定された世界に誘導されていけば、そのなかでの選択枝や自由度の制約を不満に思うことは少なくなる。人間の意図する内容は、そのときの状況がよい制約条件を与える。画像処理でジェスチャー認識をしてそれに意味付けをして解釈するまえに、まず状況をシステムが作り出しておくことは非常に有効である。マルチモーダルなインタラクションの SCRIPT を自由に書けるような環境の研究開発に

ついても取り組みが望まれる [13]。

謝辞

日頃ご指導いただき、ATR 知能映像通信研究所中津良平社長に感謝します。また、各研究テーマを進められた、角康之氏、米澤朋子氏、齊藤応志氏ほか、日頃熱心に討論していただく研究員の皆様に感謝します。

参考文献

- [1] 間瀬健二, シドニー フェルス, 江谷為之, アルミン ブルーダリン: “インタフェース・エージェントに関する基礎検討”, 情処研報 HI69-8 情処学会, pp. 55-60(1996).
- [2] バイロン・リーブス, クリフォード・ナス (細馬宏 通訳): “人はなぜコンピューターを人間として扱うか”, 翔泳社 (2001).
- [3] J. Cassel, J. Sullivan, S. Prevost and E. Churchill Eds.: “Embodied Conversational Agents”, MIT Press, MA(2000).
- [4] Justine Cassel et al.: “Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents”, SIGGRAPH94 Proceedings, Orlando, FL, pp. 413-420(1994).
- [5] 宮里勉, 間瀬健二: “ネットワーク利用者を支援するマルチモーダルヒューマンインタフェース”, 情報処理, **38**, 1, pp. 42-47(1997).
- [6] R. W. Picard: “Affective Computing”, MIT Press(1997).
- [7] 角康之, 間瀬健二: “実世界コンテキストに埋め込まれたコミュニティウェア”, 情報処理学会論文誌, **41**, 10, pp. 2679-2688(2000).
- [8] 門林理恵子, 間瀬健二: “実空間でのコンテキストを利用して仮想空間内をガイドするマルチモーダルなパーソナルエージェント”, DiCoMo シンポジウム論文集情報処理学会, pp. 653-660(1998).
- [9] 角康之, 間瀬健二: “エージェントサロン: パーソナルエージェント同士のおしゃべりを利用した出会いと対話の促進”, 電子情報通信学会論文誌, **J84-D-I**, 8, pp. 1231-1243(2001).
- [10] 間瀬健二, Brian Clarkson, 米澤朋子: “幼児期からのウェアラブルと toy 型インタフェース”, 情報処理学会研究報告 HI92-1, pp. 1-8(2001).
- [11] B. Clarkson, K. Mase and A. Pentland: “The familiar: a living diary and companion”, Proc. CHI2001 extended abstracts, Seattle, pp. 271-272(2001).
- [12] T. Yonezawa, B. Clarkson, M. Yasumura and K. Mase: “Context-aware sensor-doll as a music expression device”, Proc. CHI2001 extended abstracts, Seattle, pp. 307-308(2001).
- [13] 間瀬健二, クラウディオ S. ビンヤネス, アーロン F. ボビック: “インタラクティブシステム設計のための時区間表現によるスクリプト記述法”, 情報処理学会論文誌, **39**, 5, pp. 1403-1413(1998).