

II 擬人化エージェントの生成 Generation of Life-like, Anthropomorphic Agents

間瀬 健二
Kenji Mase

(株) エイ・ティ・アール知能映像通信研究所
ATR Media Integration & Communications Research Laboratories

1 はじめに

人間同士のコミュニケーションにおいては、協同で作業をしたり作業の依頼や命令という目的のタスクを実行する際に、創造的発想、相互理解、自己理解、話題の整理および話題の展開などの重要な側面がある。コミュニケーションにコンピュータが介在して、これまで人間同士では達成できなかった密でクリエイティブなコミュニケーションを実現しようとする、このような側面の活動を支援するコンピュータシステムを開発することが必要となる。そのアプローチとして、自律性をもったコンピュータ・エージェントが介在するコミュニケーション形態を考えることができる。

このエージェントは、当然のことながら人間とのインタフェースもこなすことになりインタフェース・エージェント [1] と呼ぶことができる。そのエージェントに擬人性をもたせることはいろいろな面で有効となる。例えば、エージェントの作業空間や状態を可視化するとき擬人的なキャラクタを用いると理解度が増すことは容易に想像できる。本文では、コンピュータグラフィクス (CG) によるキャラクタの生成技術を中心に擬人化エージェントの生成、制御技術を解説する。なお、擬人化エージェントの実現には、少なくとも CG によるアプローチと人間 (動物) 型ロボットによるアプローチがあるが、ここでは CG によるアプローチだけを考えることにする。

2 擬人化エージェントの合成

擬人化エージェントの合成の基本はコンピュータグラフィクスによる人物像の生成技術である。従来からの人物像の生成技術は図 1 にまとめることができる。

2.1 擬人化とエージェントの機能

エージェントをなぜ擬人化するのだろうか。まず、インタフェース・エージェントに要求される機能から考えてみる。石田 [2] によれば、エージェントに要求される技術は、自律性、知性、コミュニケーション、擬人性、

パーソナリティなどである。インタフェースにエージェント的考え方を持ち込むことによって、さらに次のような側面が強調される。

- エージェントに対する認知性：振舞い (behavior) の予測や作業 (action) の可視化が行われ、エージェントの作業や振舞いに対する認知性がある。擬人性 (anthropomorphy) や生きいき性 (life-likeness) が助けになる。
- エージェントの操作性、エージェント空間の身体性：大規模な情報検索、パターンマッチング、情報提示などに於ける膨大なデータを直接見せられても人間であるわれわれは操作できないし処理しきれなくなってしまう。しかしエージェントを介在することによって、これらの情報処理指示、過程および結果の操作が容易になる。たとえば、エージェントにメタファー的あるいはアイコン的なジェスチャ [3] をさせることによって、メッセージの理解や記憶が容易になる。
- エージェント行動の社会性、適応性：エージェントは知性を備えて人間とのインタラクションに対応する必要がある。例えば暗黙的な動作や状況を察知して、それに合わせた対応をするための動作認識理解を必要とする。また、個々の状況や個別の人間の性格にあわせて適応的に行動するためには学習能力が不可欠である。
- 親密性：ツールとして親密性が上がり、慣れ親しんだり、操作性を向上させ、創造性を刺激することは重要な要素である。
- 自律性：エージェントは指示された範囲での自律的な活動が認められ、そのための知性を有することが望まれる。自律的な活動により、ユーザが煩わしさから開放される可能性がある。

このように、インタフェース・エージェントに擬人性を付加することによって多くの恩恵を受けることができる可能性がある。

	質感	形状・構造	動き	相互作用	
部位	身体	テクスチャマッピング 関節でのしわ、変形	スケルトン メタボール表現 サーフェス表現 インプリシット面	物理モデリング インバースキネマティクス 弾性体表現 モーショントラッキング	物理モデリングによる衝突・反射
	顔	テクスチャマッピング 皮膚微細モデル	サーフェス表現	表情 (モデル変形、筋肉モデル、スプリングモデル、表皮多層モデル) FACS (表情コード) 発声同期 (TTS, STS)	
	頭髮	異方性反射	三角柱モデリング 糸状モデル ソリッドテクスチャ	群制御によるたなびき	衝突
	内部		ボリュームレンダリング		
感情個性			振る舞いの表現 動作の誇張 (周波数解析など) 音楽との同期	人間とのインタラクション	
衣服	テクスチャマッピング	サーフェス表現 自動裁断	物理モデリングによる変形	身体との接触 (有限要素法) 階層的衝突検出	
入力	スキャナ、カメラ	3次元スキャナ入力 磁気センサ CT	赤外線トラッカ 磁気センサ データグローブ		

図 1: 人物像生成技術の概観

3 擬人化エージェントの動きの生成

インタフェース・エージェントの特徴である擬人性や生きいき性 (life-likeness) をもたせるには CG (コンピュータ・グラフィックス) キャラクタを作成して、エージェントが決めた行動に合せて動作させる必要がある。最近では CG キャラクタは市販されているモデリングツールで比較的容易に生成できるようになったり、モデルそのものも市販されるようになった。また、動きのデータもモーションキャプチャ (コンピュータで動きを生成するのではなく、人間の動作を光学的あるいは磁気センサなどを使って入力する方式。手足の関節などにマークをつけたりして入力する) されたデータが使えるようになったが、必要な動きがあればその都度、実際の人間に動いてもらう必要がある。エージェントの個性やタスクにあわせて、動きをもっと自由に簡単にデザインできるようにするのはいまだに困難である。

動きの生成について両極端の例を紹介する。第 1 は、最近話題になった Disney のフル CG アニメーション映画「Toy Story」である。そこでは主人公のキャラクタは 712 個の制御点を全部アニメータがコンピュータ上の手作業で動かしていた。毎秒 30 フレーム (映画は 24 フレーム) の写真を 1 枚ずつ、少しずつ動きをつけながらアニメーションにする作業は膨大なものになる。実際は、すべてのフレームについて絵を書く作業はセルアニメーションの時代の話で、コンピュータ化によってこの作業はずいぶん省力化された。2 つの特徴的なポーズ (キーフレームと呼ぶ) から動きを補間することで数フレームから数 10 フレームの動きは自動的に生成することができるようになったからである。それでもフレーム数は、10 分間で 18000 枚、1 時間なら 108000 枚にもなる。時間とお金をかけて、1 本の完成された映画を作る

ときはこれでも良いが、本文の目的であるコンピュータのインタフェース・エージェントとして、コンピュータ上でいろいろな動きができるようにするには問題がある。

一方、モーションキャプチャを使う方法がある。最近 CG で作られたアイドル歌手がデビューしたが、これもモーションキャプチャした動きを使っておどっている。ぬいぐるみをかぶる代りに、CG でできた衣装 (と身体) を着ておどっているようなものである。CG アイドル歌手がもしテレビの生番組に出演したら、後ろでだれかがぬいぐるみのようなモーションキャプチャのセンサを身につけている光景が思い浮かぶ。インタフェース・エージェントを動かすためにいちいち人間が後ろで動かなければいけないとしたら本末転倒である。

しかしながら、これら両極端の技術は擬人化エージェントを動かす際にいくつか参考になる点がある。前者はアニメータの力を借りて誇張されてはいるが自然な動き (むしろ「それらしい」動きといったほうが正しからう) を表現できているし、モーションキャプチャのデータは人間の動きをとったのであるから自然であることはいうまでもない。これらのデータ素材をカタログ化しておいてその動きを接続合成することで連続した動きが自由にデザインできる可能性がある。

本文ではこれら両極端の手法ではなく、アニメータが使う経験的知識や物理法則に従った知識をとりこんで、インタラクティブに動きをデザインする方法を紹介する。これもまだインタラクティブな手法ではあるが、リアルな、あるいはそれらしい動きをデザインするのに専門のアニメータのような知識や経験が必要ないため、自動化したエージェントの動き制御に適用することが可能である。

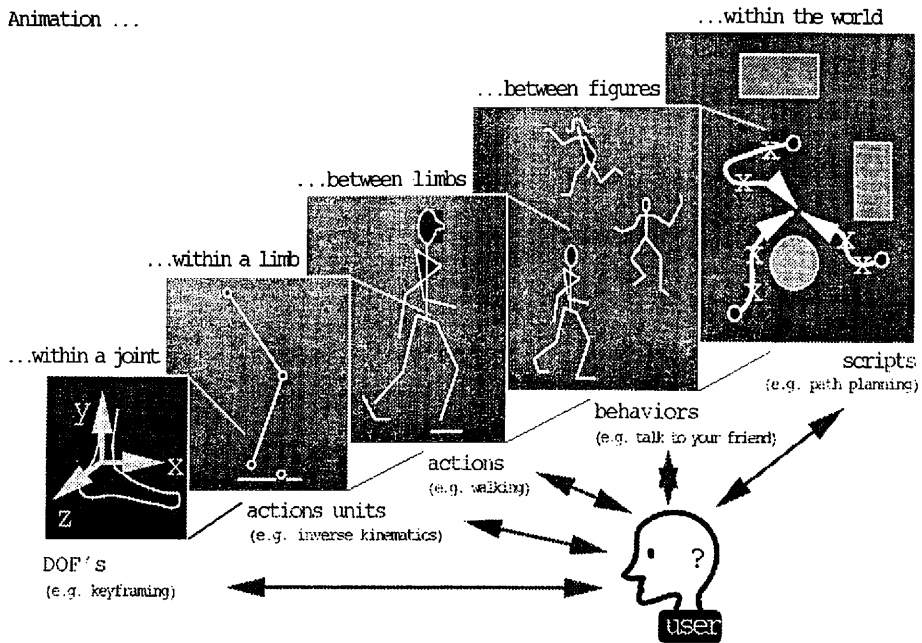


図 2: 階層型の擬人化エージェントキャラクタデザイン

3.1 エージェントキャラクタの自動動作生成

キャラクタのアニメーションの制御・生成は本質的に階層的(図2)になっていると考えられる[4]。上位から説明すると台本(script)にしたがって一人一人のキャラクタの大まかな演技が決められ、その中でキャラクタ相互の振る舞い(behavior)が決まり、歩いたり笑うという行動(action)があって、その下には各部品動きの調和をとるような行動単位(action unit)が規定され、最後には関節の角度の自由度(degree of freedom)を決定する。ところが、それぞれの階層での微調整が他階層での制約を満足させるようにするのは非常に困難な作業である。例えば扉まで歩くようなアニメーションを作ったときに歩き方に個性をつけようと歩幅を変えた途端に、扉のところで足がそろわなくなってしまうようなことが生ずる。実際のCGアニメーションの制作現場ではこれらのステップが分業化され、アニメータの感性や密接なコミュニケーションや作業のやり直しなどで実現されている。階層ごとの制約を満足しながらパラメータの変更が他の階層に波及するシステムが必要となる。そして、アニメータの独創性や個性を発揮できるようにするためにも、全部を自動化するのではなく、例えば全体を自動化しておいて、微調整も可能にすることがデザインツールとしては非常に重要である。

図3は歩行と走行のみを対象にして、そのような調整を簡単にし、さらに自然な動作を生成できるようにしたデザインツールである[5, 6]。図2の行動と行動単位の2階層を自由に行きつ戻りつしてアニメーションを会話的に生成することができる。例えば、行動レベルでは

歩く速度や歩幅の関係が、沢山の人の歩くデータから導いた関係式で表現されており、速度をかえるとその速度にあった自然な歩幅とピッチに変更される。その際に、行動単位レベルのパラメータ(足の振り上げや腰の曲げぐあいなど)にも微妙に影響をおよぼし自然な歩行の動作が生成できるようになっている。さらに行動単位レベルのパラメータも自由に調整できるようになっている。このようなインタラクティブなシステムを作ってアニメータに使ってもらい、いろいろな動きのカatalogを作り、エージェントの歩行や走行のデータベースとして登録できるようにすることが可能である。そうすれば、本ツールを用いてデザインした振る舞いの個性にしたがって、パラメータのリストが入手できるため、それキーフレーム的に順次変更するようなbehavior/actionモデルを作ることによって自然に歩行するキャラクタをインタフェース・エージェントとして登場させることができるようになる。

図4はこのツールを使って歩行と走行をデザインしたアニメーション化したフィルム“Runs”の1シーンである。比較的容易に、スムーズでリアルな歩行や走行を生成できている。ちょっとジョークの効いた短編の中で使われている歩行と走行のアニメーション生成技術の有効性が良く解る。(ちなみに、“run”には非常に多くの意味がある。辞書で引くと:「走る」の他に[the ~s; 単数または複数扱い(口語)下痢]となっている。)

3.2 擬人化による認知性や身体性

エージェントに擬人性をもたせることは、エージェントの作業空間を可視化したり、身体性をあたえることにつながる。

3.2.1 エージェントによる作業空間の可視化

エージェントが有するあるいは作業している各種リソース（データ、知識ベース、対象物など）をパターン化して表示したり、またエージェントの活動状況（煩雑度、困難度、結果の評価など）を表示することは、人間がエージェントを理解するのに手助けとなる。その際にインタフェースエージェントのジェスチャや表情で表すことができる対象として、忙しくデータを検索しているエージェントが走り回ったり、目がまわる表情をつけることなどが考えらる。我々が物事を記憶したり考えたりするとき自分の体の全体や一部をリファレンスにすることがあるように、エージェントを擬人化することはその理解の補助にもなると考えられる。例えばガイドがが順路を示すときにいくつか方法があつて、(i) 文章で示したり、(ii) 地図を示して現在地、目的地、順路を示したり、(iii) 位置関係を身ぶりですす、などがある。このとき(iii)のように動作で示された順路は、自分の身体を手がかりとできるため記憶が容易になると考えられるのではないだろうか。また、人間のジェスチャについては、音声と共に(1)形を直接表現するジェスチャ、(2)比喩的に位置関係を表現するジェスチャ、(3)タイミングをとるジェスチャ、(4)指示やポインティングをするジェスチャに分類することができ[3]、これらの情報がマルチモーダルなコミュニケーションを成立させていると考えられる。エージェントにもこのようなジェスチャをつけることで、分かりやすく、気軽に、少々いい加減でもいいインタフェース[7, 8, 9]を作ることができる。

3.2.2 リアリティ、ビリーバビリティ、擬人性、生きいき性?

最近エージェントについて、believabilityやlife-likenessの議論が盛んになっている。適当な日本語がみあたらないので、あえて、それぞれ「それらしさ」と「生きいき性」と訳しておく。はない。3節で書いたようにアニメーションのキャラクターがいかにそれらしく動いているというときにbelievableなキャラクターだと言える。さて、そういった「それらしさ」や「生きいき性」を表現するポイントは何だろうか。アニメーションの分野ではこれについて良く研究がされており、「誇張」、「予測」、「フォロースルー」などが鍵とされる。アニメーターたちの経験的な知識を積み重ねて、これらをどのように表現するかがある程度整理されている。図1をみてわかるように、これまでの研究は擬人化度をあげるためにリアルな顔や着衣の表示やその動きの表現研究がおこなわれてきたが、リアルさを追求するほど現実とのギャップが目立つようになってきている。アニメーションに現



図4: コンピュータで生成した歩行するエージェントの例(歩行動作の自動アニメーションツールを用いたアニメーション「Runs」の1シーン)

れるマンガ的なキャラクターでも充分にその世界では生きいきしていると感じることができるのはなぜだろうか。インタフェース・エージェントのキャラクターも現実をまねる必要はなく、それらしさや生きいき性のための要件を検討し、キャラクターにくみ込むことが必要である[10]。

4 エージェントの振る舞いの生成

インタフェース・エージェントとのやりとりのデザインは、エージェントの振舞いや行動のデザインに関わる。エージェントに自律性をもたせるためにはその行動を生成するメカニズムが必要で、エージェントの振舞いや行動はゴールを与える(goal-based)か、台本を与える(script-based)かの2種類が考えられる。goal-basedなシステムは途中の振舞いや外部刺激に対する反応をエージェントに任してしまうため直接に操作している感覚が低いことが欠点といえば欠点であるが、自律的な振舞いのなかから当初予想しなかった結果をもたらしてくれる可能性がある。一方script-basedなシステムではユーザの台本が、エージェントの活動を支配するためエージェントの自由度は低くなり、ユーザもスクリプトの範囲で振る舞うことが要求される。ちょうど、ユーザの立場は、前者は課題だけ与えてあとは部下に任せてしまう上司であり、後者はあれこれと細かく指示をして作業を指導する上司である。

Goal-basedなエージェントの振舞いモデルを用いたアーキテクチャとし非同期階層型エージェントアーキテクチャ(Asynchronous-Hierarchical Agent Architecture, A-HA アーキテクチャ)が考えられる[11]。それはちょうどキャラクターの動きをデザインする階層とおなじように、図5のように各エージェントはボトムアップの認識系とトップダウンの行動系からなる。それぞれの系は上から動機(motivation)、計画あるいはポリシー(plan or policy)、振舞い(behavior)、行動(action)、行動単位(action unit)というぐあいに階層的になっている。例えばトップダウンの行動系では、動機またはゴールが与えられ、ゴールへ到達するための全体の計画をきめ、計画

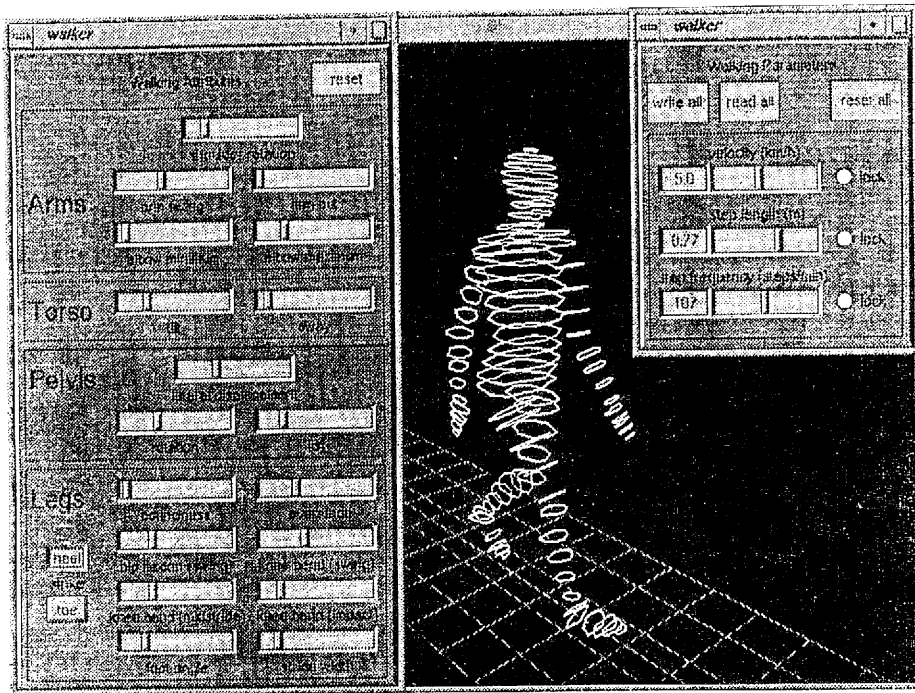


図 3: エージェントの歩行アニメーションデザインツールの画面

の各段階での振舞いを決定し、振舞いにしたがって行動する、というように詳細化されていく。

5 エージェントのタスクと応用

インタフェースとして考えるときに、タスクを切り放して設計することは困難であるし、良いものを作ることにはできない。具体的なタスクにおいて、インタフェース・エージェントがどのように振る舞うことが期待されるかを考えると、そのときの擬人化程度を設計することができるだろう。具体例をあげて示す。

5.1 メタミュージアムにおけるエージェント

メタ・ミュージアムとは、博物館などの実体の展示物がもつ感性的な豊かさを窓や道標にして、外在するサイバースペースを自由に探検したり経験できるようにすることにより、展示者と来館者あるいは専門家と非専門家のコミュニケーションを支援する環境を提供することを目的とした、我々が提案している博物館の将来像 [12, 13] の名称である。

美術館や博物館は、歴史的遺産や自然現象や芸術作品に関する「もの」や知識の大規模な集積場として、来館者に情報を提供している。それは単に事実をみせる場であるだけでなく、製作者や展示者の知識や考えなどの展示物が備えている情報を共有するコミュニケーション環境と、とらえることができる。そしてその情報はコンピュータとネットワークの普及によりデジタル化され量的に増大し、分散されていたものが結合されて複雑化

しつつある。そのような膨大なデータベースを探索するのにインタフェース・エージェントが力を発揮できると考えられる。

メタ・ミュージアムのガイドエージェントは拡張現実空間にあらわれて、ユーザの興味、スケジュール、位置、館内の状況を認識して次の展示への移動スケジュールをたててユーザへ提示する。その提示方法は、例えば、来館者をせき立てたり、鑑賞するのをじっと待ったり、あるいはユーザの足下への注意を促すなどの動作を使うことが考えられる。

また、歴史的展示などでは、擬人化されたエージェントは、当時の人に変身して当時の様子をドラマチックに演技して説明することも要求される。あるいは自分の分身としてエージェントに演技してもらいエージェントを通じた擬似体験ができるようにすることも必要である (図 6)。

5.2 議論活性化エージェント

議論を活性化するエージェントには、人間同士の会話に参加してタイミングをとって適切な意見を提示するという、もっぱら自律的で知的な活動が期待される。したがって擬人化はあまり重要ではなくなり、とくに身体を全部表示する必要はなく、顔と手があれば充分であると考えられる。むしろ、まずエージェントの作業している空間 (例えば、エージェントが整理した議論の内容) を見せることが重要と考えられる。また、エージェントの状態 (例えば、新しい意見を発せられる状態にあるかど

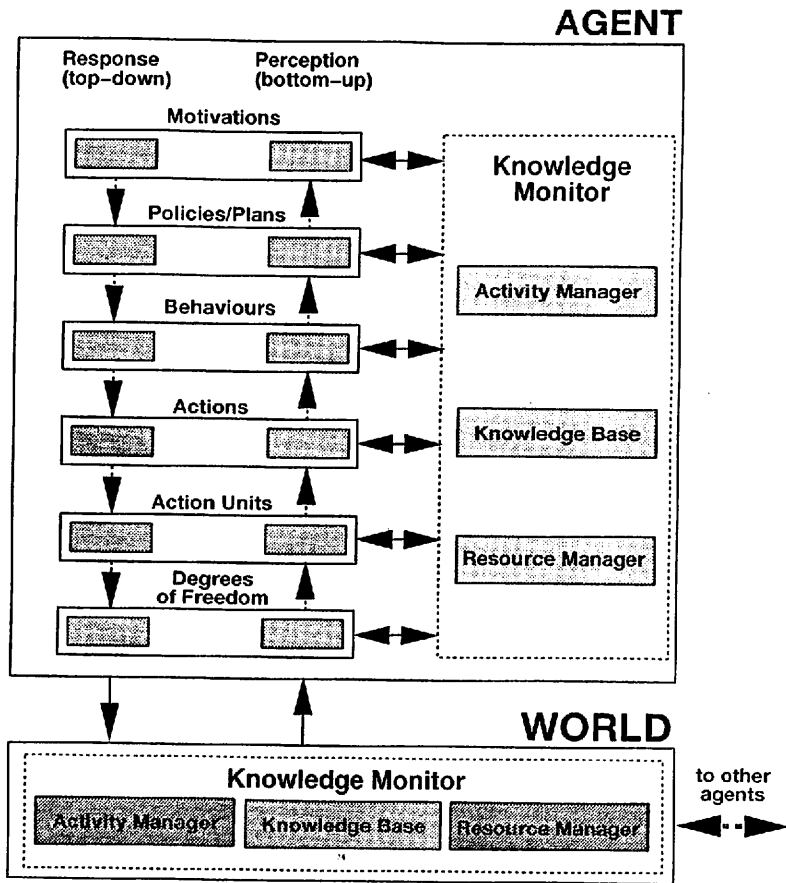


図 5: 非同期階層型エージェントアーキテクチャ

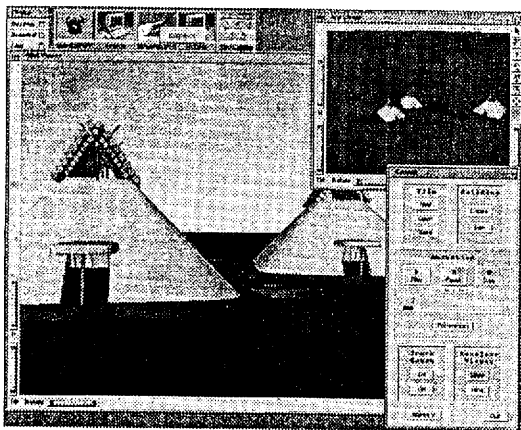


図 6: 弥生時代の集落の再現(家屋建立年代変遷の仮説検証ツール)

うかを表情で、さも一言ありそうにする)を可視化したり、エージェントのしている世界がどこであるかを明確にすることなどが重要ではないか。図7はAIDEと呼ぶオンライン型の非同期議論活性化環境の利用の様子である[14]。議論支援エージェントが議論の膠着などを検出すると、これまでの議論の展開を切らないで、かつ新しい話題に展開するように自分の知識データベースからコメントを提供する。

5.3 MIC インタフェース

人間とエージェント間のインタフェースにおいて3つの要素を見逃すことはできない。すなわち、マルチモダリティ(M)、親密(intimate)で賢い(intelligent)インタラクティビティ(I)およびクリエイティビティ(C)である(図8)。自律的に動くエージェントにとって擬人性を備えることは、これらの要素を備えるひとつの方向であり、よりよいインタフェースの実現に一歩近づくと考えられる。

6 おわりに

擬人化したキャラクターを用いたインタフェースやマルチモーダルインタフェースについてはこれまでいろいろと研究がされてきているが、インタフェース・エージェントの研究はまだ始まったばかりである。このようなインタフェースをただの自動プログラムではなく、なぜエージェントと呼ぶかについては多くの議論があろう。エージェントとよぶことによって、自律性や擬人性に考えが及び、例えばA-HA アーキテクチャのように、知的なインタラクションが可能となるインタフェースや機能あるいはプログラム構成のデザインをどう考えたら良いかについてある指針を与えてくれ、具体的に考える場を提供してくれるのではないだろうか。

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”, 情処研資 HI, 情報処理学会, November 1996.
- [2] 石田 享, “エージェントを考える”, 人工知能学会誌, 10, 5, pp. 663-667, Sept. 1995.
- [3] Justine Cassell et al., “Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents”, In *SIGGRAPH94 Proceedings*, pp. 413-420, Orlando, FL, 1994.
- [4] Armin Bruderlin, “The creative process of animation human movement”, *Knowledge-Based Systems*, 1996.
- [5] Armin Bruderlin and Tom Calvert, “Knowledge-driven, interactive animation of human running”, *Graphics Interface '96*, pp. 213-221, May 1996.
- [6] Armin Bruderlin and Kenji Mase, “Interactive, real-time animation of human locomotion”, *Lifelike Computer Characters '96*, pp. 73-74, 1996.
- [7] Richard A. Bolt, “The integrated multi-modal interface”, 信学論, J70-D, 11, pp. 2017-2025, Nov. 1987.
- [8] 間瀬 健二, “動画像処理を用いた新しいマンマシンインタフェースの研究”, PhD thesis, 名古屋大学学位論文, Mar. 1992.
- [9] Alex P. Pentland, “Smart rooms”, *Scientific American*, pp. 54-62, April 1996.
- [10] Ryohei Nakatsu and Kenji Mase, “Tutorial Course on Life-like Believable Communication Agents”, *SIGGRAPH96*, New Orleans, Aug. 1996.
- [11] Sidney Fels et. al., “Agent 99: Implementing a simple card game using agents”, 知能情報メディアシンポジウム, Dec. 1996.
- [12] Kenji Mase, Rieko Kadobayashi, and Ryohei Nakatsu, “Meta-museum: A supportive augmented reality environment for knowledge sharing”, *Intn'l Conf on Virtual Systems and Multimedia '96*, pp. 107-110, Sept. 1996.
- [13] 門林 理恵子, 間瀬 健二, “新しいコミュニケーション環境としてのmetamuseum”, マルチメディア通信と分散処理ワークショップ論文集, pp. 71-78, 情処, Oct. 1995.
- [14] 西本 一志, 角 康之, 間瀬 健二, “一参加者として対話に加わる対話活性化エージェント”, 信学技報 TL96-7, 電子情報通信学会, 1996.

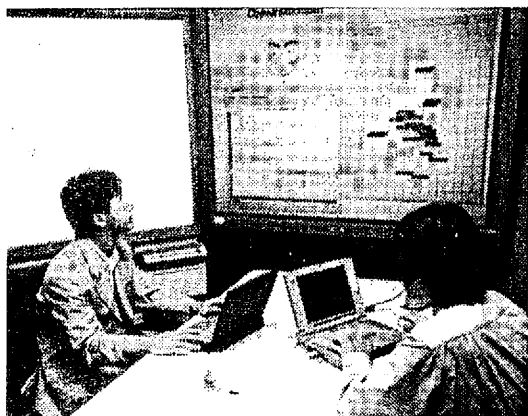


図 7: 議論活性化環境 AIDE の利用の様子: 画面右側がエージェントが整理している議論空間マップ

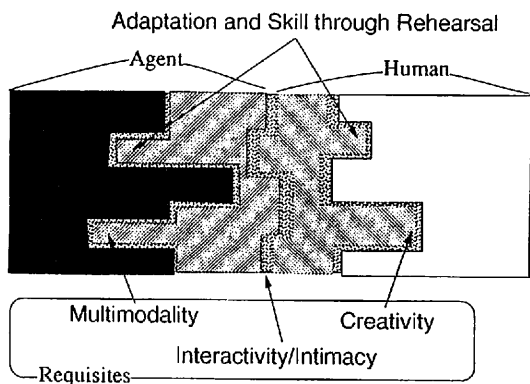


図 8: MIC インタフェース