

ATRの最近の研究

— 知能映像通信研究所におけるインタフェース・エージェントの研究 —

(株) ATR知能映像通信研究所
第二研究室 間瀬 健二

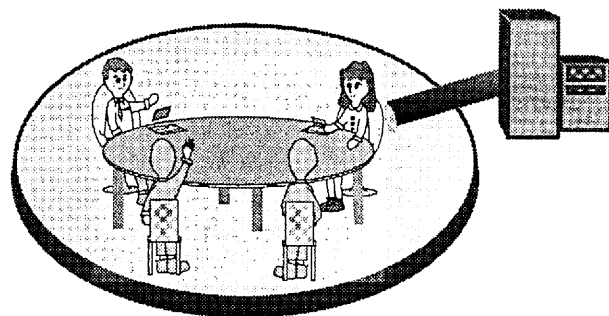


図1: コンピュータ・エージェントが会議に参加する

1 はじめに

ATR知能映像通信研究所は、映像を中心としたマルチメディア情報を駆使することにより、21世紀の新しいコミュニケーション文化の創出に貢献するコミュニケーション技術の研究開発を目標にかかげています。そこでは、距離と時間を越えながらいかににも実際に会話をしているようなリアルな感覚を与えてくれるコミュニケーションと、現実にはありえないが人の感性を刺激して活性化する超リアルなコミュニケーションを実現するための技術が研究課題となっています。現在、研究所ではテーマを、(1)コミュニケーション環境生成技術、(2)エージェントインタフェース技術、(3)イメージコミュニケーション技術、(4)人のコミュニケーション過程の分析、(5)バーチャルリアリティ技術、(6)アート&テクノロジーの融合、などのサブテーマに分けて研究をすすめています。本文では、エージェントインタフェース技術の研究を中心にプロジェクトの課題や関連する技術の成果を紹介します。

2 コミュニケーションの目的とその支援

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ります。我々は、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するために、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したいと願っています。そして、そのアプローチとして、使われるツールやインタフェースを独立に考えるのではなく、インタフェースにタスクを取り込む形で、そこに自律性(外部からのコントロールではなく、自分の立てた規範にしたがって行動すること) ~~立てた~~や擬人性(人のように見えても、実際はそうでないコンピュータなどを人に見立てること)を持たせることによって、ただの自動プログラムではなく、エージェントとして捉えることにしています。これをインタフェース・エージェントとよんでいます。

エージェントということばは、旅行代理店のエージェントということばがあるように、自分の代理人にある仕事を任せることを示唆します。最近ではいろいろなコンピュータプログラムをあたりかまわずエージェントと命名する動きがあり、非常に広い意味で用いられています。ここでいうインタフェース・エージェントもマクロな意味で捉えています。エージェントと呼ぶことの功罪やエージェント的システムそのものの功罪(自律的に作業するコンピュータ

プログラムの結果など信用できないなどの反論が典型的)はいろいろありますが、エージェント性を持ち込むことで、人間の認知性、社会性、身体性などをコンピュータとのやりとりの枠組みでも考えることができ、考え方に幅が出てくるのではないかと考えています。

インタフェース・エージェントの例として、我々は、オンラインのミーティング(例えば、実務的なものとしてはデスクトップTV会議と共同作業空間を提供するシステムがあり、遊び的なものにはオンライン・おしゃべりシステムがあります)に介在し発想支援をするエージェント(図1)や仮想空間に現れるサイバースペースのガイドエージェントを考えています。そのための基本技術として次のようなテーマを設定して研究を進めています。すなわち、(i)擬人化したエージェントの動き生成の研究、(ii)エージェントの振る舞いの生成制御の研究、(iii)エージェントのタスクと応用の検討、などです。以下、簡単に個々の研究を紹介します。

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わせて動作させる必要があります。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りました。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りましたが、必要な動きがあればその都度、実際の人間に動いてもらわないといけません。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困難です。

例えば、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ということです。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ります。

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ました。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す。それでもフレーム数は、10分間で18000枚、1時間なら108000枚にもなります。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のですが、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ります。また最近CGで作られたアイドル歌手がデビューしたということですが、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のです。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す。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ことでしょう。このモーションキャプチャの問題も重要で、ここでは詳しく紹介しませんが、ATRでは非接触の赤外線センサやTVカメラを使ってジェスチャや表情の抽出や認識をする研究もすすめています。

3.1 エージェントキャラクタの自動動作生成

我々は、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ます。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、その下には各部品の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）を決定するわけです。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す。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じます。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わけです。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ります。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す。

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、さらに自然な動作を生成できるデザインツールです。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ます。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に変更されま

す。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ます。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ます。我々は、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したいと考えています。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、パラメータを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ります。

なお、本ツールは現在ニューヨーク・クイーンズ市の動画博物館（Museum of Moving Image）で行われている映画の裏舞台（Behind the Screen）という特別展で展示され、最新のコンピュータ技術が映画制作でどのように使われているかを紹介しています。このツールは来訪した中学生や高校生がいろいろな歩き方を自分でデザインでき、好評だということです。また、図4はこのツールを使って歩行と走行をデザインしアニメーション化したフィルム“Runs”の1シーンです。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ます。国内外のCGフィルムショーでも紹介されつつあり、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を見て頂く機会があるのではないかと思います。（ちなみに、“run”には非常に多くの意味があります。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] というものもあります。）

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ります。これはどういうことでしょうか。

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ります。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れます。我々が物事を記憶したり考えたりするときに自分の体の全体や一部をリファレンスにすることがありますが、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ます。例えばガイドがが順路を示すときにいくつか方法があって、(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、などがあります。このとき (iii) のように

Animation ...

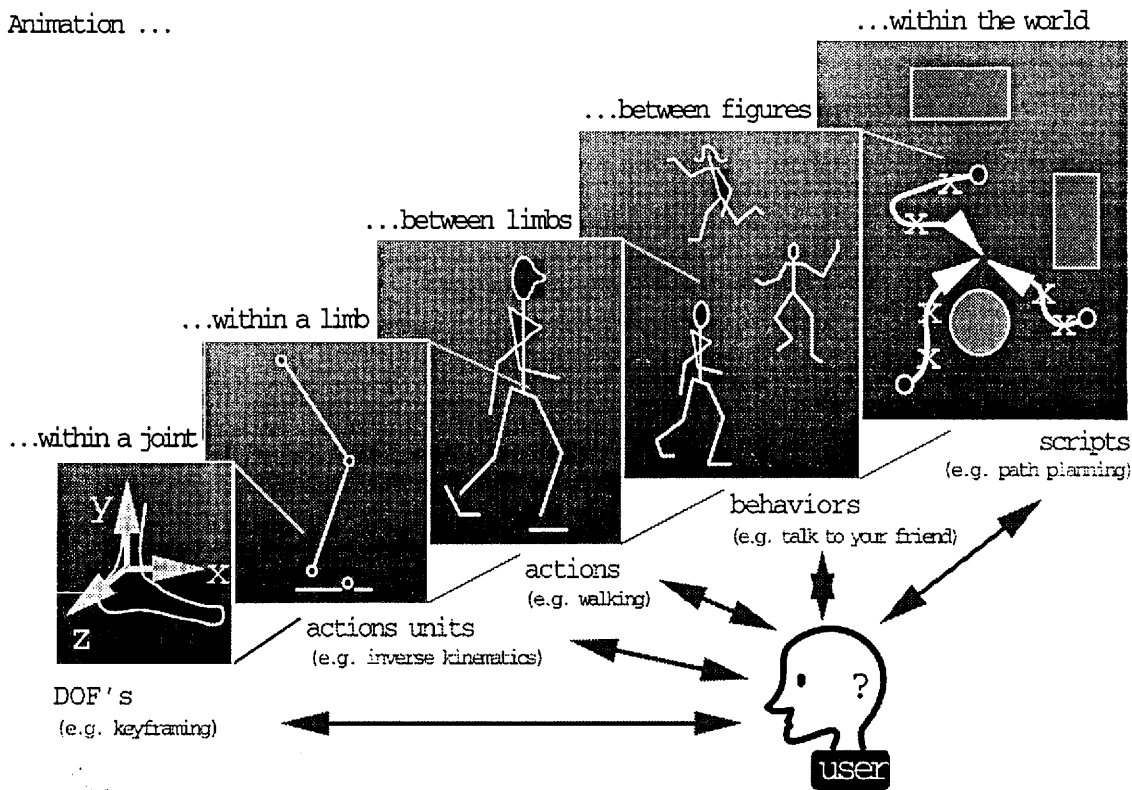


図 2: 階層型の擬人化エージェントキャラクターデザイン

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のでしょうか。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、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ます。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、分りやすく、気軽に、少々いい加減でもいいインタフェースを作ることができるのではないのでしょうか？

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ます。適当な日本語がありませんが、あえて、それぞれ「それらしさ」と「生きいき性」と訳しておきます。believe はここでは「信頼する」という意味ではないので、believable agent は「信頼できる秘書」とかという意味ではありません。アニメーションのキャラクターがいかにも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ます。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でしょうか。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」、「フォロールー」などが鍵のようです。これまでの研究は擬人化程度をあげるためにリアルな顔や着衣やその動きの研究がお

こなわれてきましたが、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ます。アニメーションに現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でしょうか。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だと考えています。現在 ATR にはアーティストが同僚として活動しています。まさしくアーティストの創造するキャラクターはエンジニアが考えることのできない、生きいき性をもっています。その秘訣をアーティストから引き出すことができれば、それらしく相手をしてくれるエージェントを設計できるのではないかと思います。

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、エージェントの振る舞いや行動のデザインに関わります。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、エージェントの振る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 種類が考えられます。goal-based なシステムは途中の振る舞いや外部刺激に対する反応をエージェントに任せ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れば欠点ですが、自律的な振る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ります。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

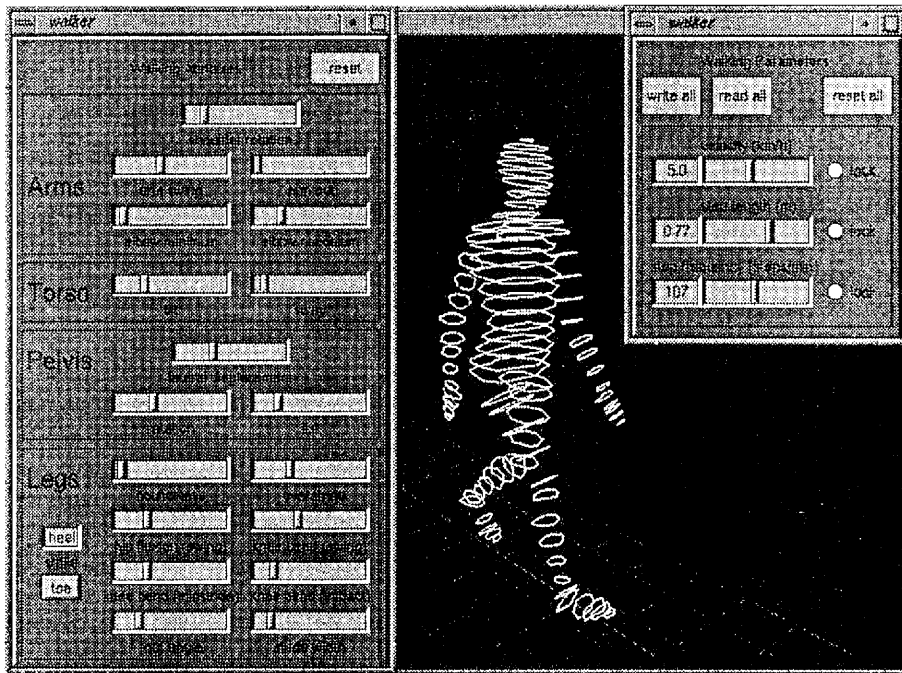


図 3: エージェントの歩行アニメーションデザインツールの画面

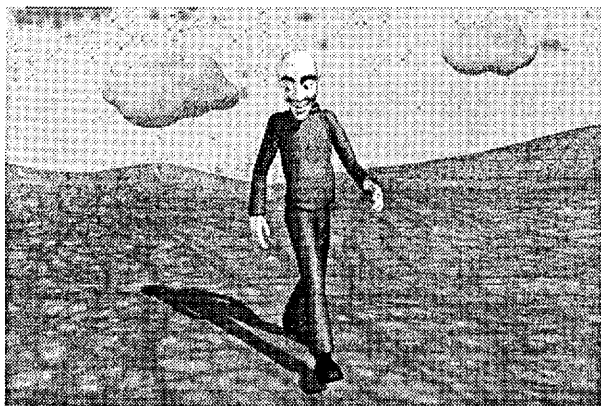


図 4: コンピュータで生成した歩行するエージェントの例 (歩行動作の自動アニメーションツールを用いたアニメーション「Runs」の1シーン)

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ます。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のどちらかになることとなります。エージェントが何もかもまかせられる部下にまで成長しているかどうか、その判断の分かれ目です。エージェントの作業の信頼性はそのタスクの困難性にもよるので、一概にどちらが良いとは言えません。どちらも使えるようにしておくのが理想でしょう。現在は両方のアプローチからエージェントの振る舞いを生成するための手法を検討しています。

まず、Goal-based なエージェントの振る舞いモデルを用いたアーキテクチャとして我々は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: アッハーと呼びます)

を構築中です。それは図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り、認識結果が行動系にトリガーをだしてドライブするようなシステムを考えています。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ます。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きます。

一方、script-based な方法として、現在はまずマルチモーダルなインタフェース環境を作るために、ビデオカメラなどによる動作認識などの現実世界知覚モジュールと、エージェントの振る舞いや音生成をする生成モジュール群のイベントスケジューリングを区間算法を用いて高速に行う手法の研究を進めています。この手法によれば、インタラクションの台本をプログラムと独立に記述することができ、将来スクリプト作成の効率化をはかることができると考えられます。

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、タスクを切り放しで設計することは困難であるし、良いものを作ることはできないでしょう。まえがきで述べたように、私たちは発想支援や相互理解などについて具体的なタスクをあげ、そのときのインタフェース・エージェントがどのように振る舞うべきかを検討しています。

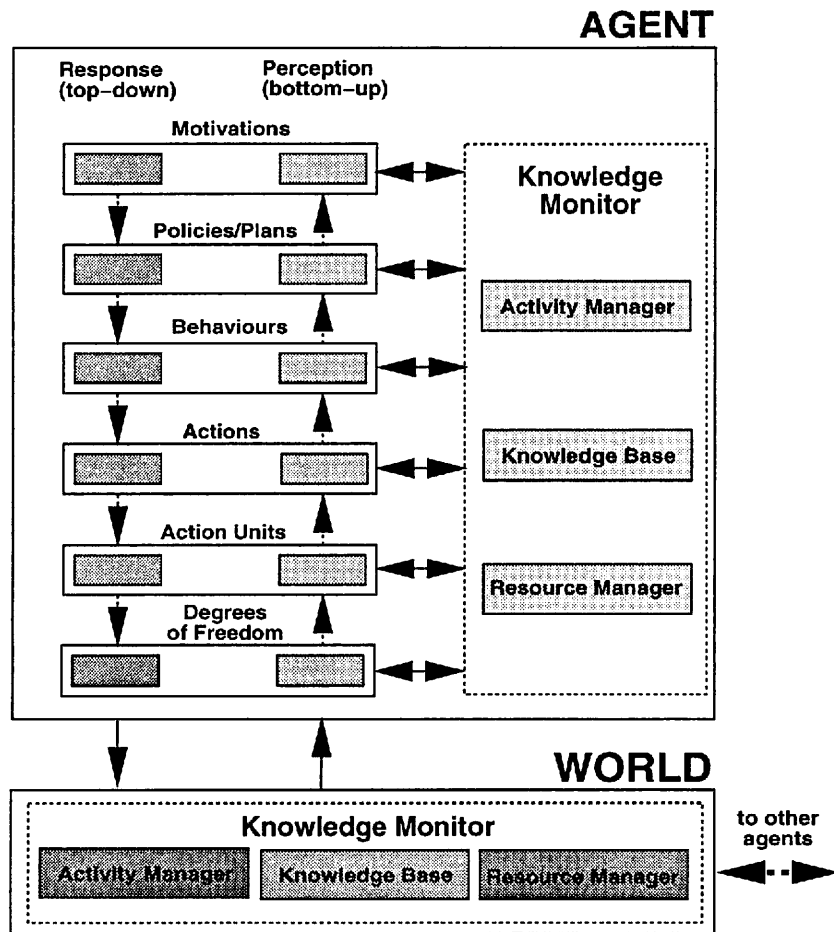


図 5: 非同期階層型エージェントアーキテクチャ

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を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の名称です。

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ます。それは単に事実をみせる場であるだけでなく、製作者や展示者ら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ます。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ります。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ます。

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立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

られます。

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可能となります。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することも可能となります。

また、研究ツール共有の具体的効果を示すために、考古学的にも新しいテーマである、弥生時代の集落の遺跡発掘データから家屋建築の変遷の様子をシミュレーションするツール VisTA (図 6) を作成中です。これは家屋の存続年代について仮説を立てると集落の変遷をシミュレーションシグラフィックスで可視化するものです。仮説にもとづく集落の様子が3次元的に再現され、隣同士が近すぎるとか互いに内部が見えてしまうなどの物理的な制約条件を容易に認識して仮説を検証することが可能になります。また、HTMLを用いて発掘データをデータベース化し、シミュレーション空間から発掘データへのハイパーリンクを提供しています。このようにして分散化されたデータベースへのアクセスが可能となります。

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

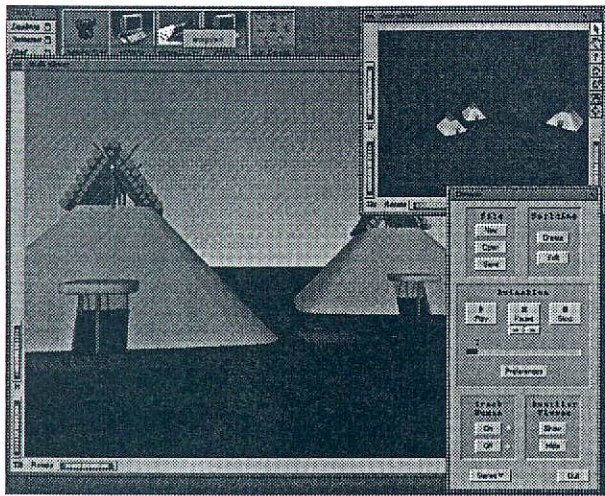


図 6: 弥生時代の集落の再現 (家屋建立年代変遷の仮説検証ツール)

う、もっぱら自律的で知的な活動を期待します。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ます。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ます。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、エージェントの見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でしょうか。

さて、前述のガイドエージェントは、展示者や研究者とのコミュニケーションを支援できなければなりません。ところが研究者のノートや研究者が使うツールを見せられても一般人にはちんぷんかんぷんです。ここで、研究者のノートやツールを理解するには、専門用語のマッピングをする必要があります。エージェントの仕事はこのような用語がお互に通じるように、来館者と展示者の概念空間を擦り合わせることから始まります。

また、興味 の推定や相互理解のための概念空間のブラウジングには、思考空間の可視化などの技術が役に立つと考えています。電子化された個人のメモや発言に含まれるキーワードから、統計的処理により思考空間を自動的に形成し可視化することが可能となりました。可視化されたパターンから、いろいろな話題について個人がどんな視点でみているかを知ることができます。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す。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します。

6 おわりに

ATR で進めている最新の研究テーマのいくつかを第二研究室の活動を中心に紹介しました。最初にも述べた通

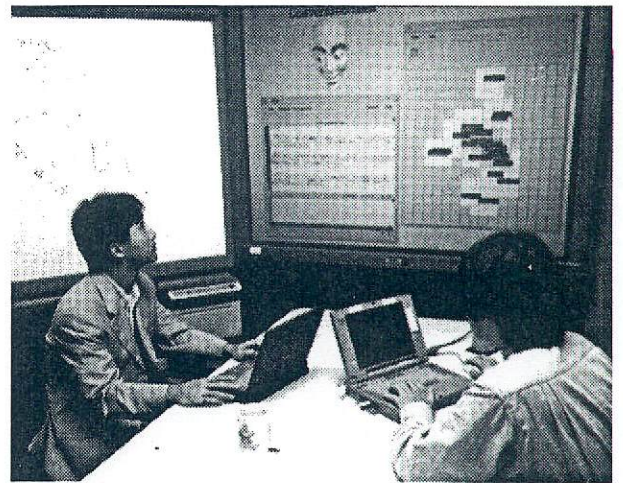


図 7: 議論活性化環境 AIDE の利用の様子: 画面右側がエージェントが整理している議論空間マップ

り、ほかにも新しいコミュニケーション形態の創出にむけていろいろな方向から研究テーマが設定され活発に研究が進んでおり、ここで紹介したのはそのほんの一部です。

擬人化したキャラクタ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ますが、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という感じです。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かと思えます。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かと思います。研究を進めています。

最後に、日頃ご指導頂く中津良平社長ならびに、ここで紹介した研究に携わる第二研究室の皆様へ感謝します。