

# 5

## 生命的エージェントによる インタフェース／メディア

石塚 満 東京大学 情報理工学系研究科

### 顔と姿を持つエージェント

顔と姿を持ち音声機能を有する生命的エージェント (lifelike agent) を用いるマルチモーダルインタフェースやメディアコンテンツが出現し始め、複雑化が進む情報化社会の中で自然で理解しやすく親しみやすいマルチモーダルメディアとして発展が期待されている<sup>1), 2)</sup>。これらのエージェントは情報空間でガイド、チュータ、執事、パートナーなどの役割を果たすことになる。先行的研究によりそれらの効果も実証されるようになってきている。複数の技術要素の集積が必要なため以前は実現が難しかったが、最近ではツールとして提供される要素技術が利用できるようになっており、試行的使用は可能になりつつある (生命的エージェントシステムの構成要素についてはコラム参照)。Web 等で見かけることもあるようになってきており、図-1 はそのような例を示している (Web では音声出力はあるがユーザ入力はいずれもマウスによる選択やテキスト入力であり、音声入力はまだ一般的でない)。最近では、ヒューマノイドロボットもこのようなエージェントの一種として用いる研究も行われている。

しかし一方で、社会への浸透は期待されたほどには進んでいないという現実もあり、広く一般の人々が感性的で魅力的なインタフェースやコンテンツを容易に作成できるようになるには、課題も多く残されている。本稿では、このような技術の経緯と現状を紹介し、広い普及を可能とするためのいくつかの課題について記す。

### 経緯と現状

生命的エージェントは、一定程度の自律性を有する機能モジュールという点では他のエージェント技術と

共通性を持つが、目的や外見として身体を持つこと (embodiment) を始めとして違いも多く、半ば独自領域として発展を遂げてきたともいえる。1987年にApple社が将来のインタフェースのコンセプト・ビデオとして Knowledge Navigator を示したのが1つの契機となり、1990年に代入すると欧米で具体的に実現する技術の研究開発が進んだ。初期段階で我々も含め日本では、テクスチャマッピングによる自然感の高い顔を持つ擬人化エージェント (anthropomorphic agent) の具現化の点で特徴があったといえる (しかし、自然感の高さは人間と同等の能力を有するといった過度の期待を抱かせてしまう点で、必ずしも好ましいとはされなくなっている)。また、東芝で開発された TOSBURG II はエコーキャンセラを備えるなど、音声機能に特徴があった。

Microsoft社が1997年に Microsoft Agent キャラクタとシステムを無償公開したことは、この分野の開発を促進し、またシステム構築上のインタフェース等についての1つの参照ベースを与えることになった。Microsoft社はこれを用いたアシスタント・ツールを Office 2000 に標準装備したが (代表的にはイルカのキャラクタ)、煩わしいと感じるケースが多く、必ずしも成功しなかった。教訓としては、アシスタントは不用意に出しゃばるような形態はよくなく、控え目であるべきということになる。1990年代後半には、直接操作型 GUI 対エージェントインタフェースの論争とまではいかないが、議論があった<sup>3)</sup>。このエージェントインタフェース主張者の P. Maes (MIT メディアラボ) は、身体を持つ生命的エージェントに特に限定されず広くソフトウェアエージェントによるインタフェースを対象にしたのであるが、議論の内容はほぼ生命的エージェントについて当てはまる。複雑化する情報社会の中で未知な環境や未知で多様な機能に遭遇する機会は増えてきている。物理的世界ではこのような場合は知識を持つ人に尋ねるのが最もよく、特別

な操作法を知る必要がないという意味で汎用的な解決法である。直接操作型 GUI の優位性を主張する B. Shneiderman (メリーランド大) の賛同は得られないかもしれないが、このような場合には、物理的世界の人対人の汎用的な対話やインタフェースを模擬するエージェントインタフェースは、特別な操作知識を要しない自然なインタフェースということになるのではな

かろうか。日常的に使用する機能については、エージェントを介して問い合わせたり動作させるのは煩わしいということになる。教育システムにおける教師エージェントなどは、単にインタフェースとしてだけではない役割を果たすことになり、有効性が高い。

2000 年代になると、米国では海外に展開する兵士が会話も含む異文化環境へ適応するためのマルチモーダル訓練システム開発に向け陸軍から資金が投じられ、南カリフォルニア大に ICT (Inst. for Creative Tech.) が設立され、この方面への応用が進展している (当初の対象地域はボスニアだったが、その後アフガニスタン、イラクと広がった)。一方 EU では、IST (Info. Society Tech.) プログラムの下でマルチモーダルインタフェースと HUMAINE と称される感情に関するシステム (emotion-oriented systems) がテーマに設定されており、多数の大学が参画しており、多くの研究が進んでいる。

生命的エージェントによるメディアコンテンツを単なる技術としてだけでなく、芸術的面も評価し、総合的に強化を図ろうとする動きも始まっており、2005 年からコンテスト形式の GALA (Gathering of Animated Lifelike Agents; <http://hmi.ewi.utwente.nl/gala/>) が行われている。この作品 (動画) を見るとこの分野のレベルを実感することができ、総作品として評価するこのような動きは、メディアクリエイターも巻き込む新段階へ進ませる動力となる。

インタフェースやメディアにおける生命的エージェントの使用に関し、その認知的背景として把握しておくべき事項を以下に 3 点挙げておく。

- A. Mehrabian による 1960 年代のノンバーバルコミュニケーションの役割と重要性の指摘 (日常のコミュニケーションにおいて身振り、表情、声調やイントネーションによる情報伝達の部分がきわめて大きい (90% 以上である) ことを示した)。
- メディアの等式 (Media Equation) : B. Reeves と C.



図-1 Webに現れる生命的エージェントの例。(左は上半身のみのキャラクター)

Nass は 1996 年発刊の著書において "Media = Real Life", すなわち人間はインタラクティブする人工物であるメディア対象物を自然に、人間のように人格を持つ対象として扱う傾向があることを指摘した<sup>4)</sup> (これは、人間は主として人間社会の中でのインタラクティブにより進化してきたことに起因する)。

- ペルソナ効果 (Persona Effect) : 生命的エージェントの存在は、たとえば教育用メディア等において、必ずしも直接的に作業や学習の効率に影響を与えるものではないが、モチベーションを高めたりする効果を持ち (学習などにおける) 経験の認知性を高める効果を持つ<sup>5)</sup>。

## 記述言語

共通的な HTML により多くの人々が容易に Web コンテンツを記述・作成し発信できることで Web が急速に拡大、普及したように、生命的エージェントによるコンテンツの普及に関しても共通的あるいは標準のベースとなる記述言語の設定が重要となる。この考えの下で我々は MPML (Multimodal Presentation Markup Language)<sup>6), 7)</sup> を開発し、国際的標準化に向けての議論にも努力してきたのであるが、2005 年までは大きな進展はなかった。詳細は省くが、世界各国で開発された記述言語には、CML/AML (英インペリアルカレッジ)、APML (伊パリ大、仏パリ大など)、RRL-NECA (英ブライトン大など)、STEP (蘭 Vrije 大アムステルダム)、VHML (豪 Curtin 大)、BEAT (米 MIT)、PAR (米ペンシルベニア大) などがある<sup>2)</sup>。Web との整合性を考慮し、MPML を含め半数以上は XML 準拠のマークアップ型記述言語になっている。

標準化の議論が進展しなかった要因として、次の 2 つの異なる指向性があり、この間での調整が難しかったことがある。第 1 は、多くの人々が (HTML での Web コ

ンテンツ記述のように) 無理なく記述できるようにすることを重視する立場であり, 第2はプロのクリエイターによる使用にも耐えるように細部制御も含む高い表現力を追求する立場である. これらの違いは用いるキャラクターエージェント機能にも反映する. 第1の立場のアプローチではエージェントへの動作指示は中レベルの抽象度であり, 多くの場合, キャラクターエージェントはあらかじめ定義された30~50種程度のジェスチャパターン(これを最近ではgesticonとも称する)を持ち, これらから選択して実行するかたちをとる. 第2の立場のアプローチではエージェントの動作指示はパラメータも含むような低レベルの制御も可能とすることから, キャラクターエージェントも動作に高い自由度を持つものが用いられる. 我々のMPMLは第1の立場に立脚している. 以上の2つのアプローチは将来的には統合されるべきものであるが, どちらを起点にして他方を取り込んでいくかが明確にならなかった.

2005年からEUの研究者を中心に記述言語の標準化についての議論が進められ, 方向性が打ち出されるようになってきた<sup>8)</sup>. この構成法では従来は混在していることもあった, マルチモーダルなコミュニケーションのプラン生成部と, エージェントの具体的な動作生成部とを分離する. そして, 後者への入力となる記述言語としてBML (Behavior Markup Language) を定めている. このBMLはまだ骨格部分だけで, 具体的に動作するシステムとなっているわけではないが, 上記の第1の立場を基礎にして第2の立場への拡張を考慮するアプローチをとっている. 今後の展開は不透明な部分もあるが, 共通のあるいは標準としての記述言語設定への促進材となることを期待したい.

話を我々のMPMLに戻すと, MPMLの記述の主対象はキャラクターエージェントを用いるマルチモーダルWebコンテンツであるが(プレゼンテーションを主とするが質問を受け付けたりの対話機能もサポートしている), 他の情報環境へ対応するいくつかのバージョンが開発されている. MPML-Mobileは携帯電話向けのバージョンである. この開発動機は, (1)通常のWebコンテンツで有料のものを広めるのは非常に難しいが, 2001~2004年当時の携帯コンテンツは月額105円, 210円といったものが伸びており, コンテンツビジネスとして見込みがあったことと, (2)我が国が先行している携帯用Webコンテンツにおいて実用システムとしての地位



図-2 GALA2006受賞作品

を確立できれば, MPMLのデファクト標準化にも寄与するであろうということであった. このMPML-Mobileを用いてHottolink社と共同し, 2004年にはau携帯の有料公式サービスを開始した(しかし残念ながら, 十分な顧客の獲得に至らず, サービスは停止された).

MPMLベーシックバージョンはディスプレイ上の2次元空間のコンテンツを記述するものであるが, MPML-VR (VR: virtual reality) は3次元VRML空間でのエージェントを用いるコンテンツ記述用のバージョンである. 図-2は同様な3次元空間用で, かつ利用者の視線を検知して適応的にコンテンツを切り換える機能を付加した記述言語によるプレゼンテーションコンテンツの画面であり, 前述のGALAコンテストで2006年受賞作品となったものである(制作ディレクターはH. Prendinger (NII)でE. Andre (独アウグスブルグ大)と筆者が協力し, 3名のドイツ人大学院生が制作した).

MPML-HR (HR: humanoid robot) はエージェントとして物理的なヒューマノイドロボットを用いるコンテンツを記述するバージョンである. 現在, 筆者の研究室とホンダ・リサーチ・インスティテュートが共同で開発を進めており, 図-3のようなAsimoを用いるプレゼンテーションコンテンツが作成されている. ヒューマノイドロボットの使用できる環境は限定されており, ソフトウェアによるキャラクターエージェントのようにディスプレイがあればどこでも使用できるというわけではないが, キャラクターエージェントとは異なるより認知度の高い強い印象を与える. ヒューマノイドロボットに物の把持などの作業を行わせるには詳細な制御を必要とするが, 非専門家でも記述可能なメディアとしてのヒューマノイドロボット用の記述言語として, MPMLのようなレベルの言語の必要性はあると考えられる. 二足歩行ヒューマノイドロボットは我が国が現在最も進んでおり, 他国には

まだ十分なものはないが(頭部だけのものはある), 関心は高くなっている。

## エージェントの感情

感情表現はエージェントに生命感, 信憑感 (believability) を与えるために, そしてまた共感するエージェント (empathic agent) を実現するために重要な要素である。プレゼンテーションにおいてもエージェントによるものは人間に比べると平板になりがちであり, 心の動きを感じないエージェントによるものは少し長く (4 ~ 5分以上) になると退屈に感じられるようになってしまう。そこで, MPML を始めいくつかの記述言語はエージェントの感情を定める機能を具備しており, 今日でもエージェントの感情に関する研究は多く行われている。

よく知られた基本 5 感情は, 「怒り, 喜び, 驚き, 恐れ, 嫌悪」であるが, これ以外にもさまざまな感情モデルが存在する。最も包括的といえるのが, OCC 感情モデル (OCC はこのモデルを提唱した Ortony, Clore, Collins 3 名の頭文字) であり, 22 感情を含んでいる (たとえば, 喜びでも自分自身に関する事象に対する感情は "joy", 自分以外の相手に関する事象に対しては "happy-for" と分けている)。幾分過剰なのであるが, MPML ではこの OCC 感情モデルでエージェントの感情を指定して記述すると, 音声の質が変化し<sup>☆1</sup>, 発声時の前後を中心にして感情に対応するジェスチャを生成するようにしている。

エージェントの感情を必要とされるたびに毎回記述するのは煩雑であるので, 人工感情機能も開発されている。この簡単なものはペットロボット等でも実現されているが, 我々の開発した SCREAM<sup>9)</sup> は前述の OCC 感情モデルに基づくものであり, かつ, 社会的関係により感情の表出を制御するソーシャルフィルタリング機能 (上司や顧客に対しては怒りの感情は抑制して対応するなど) なども備えている。この SCREAM を外部モジュールとして MPML に結合すると, エージェントの感情が生成され, 動作や表情に反映されることになる。

感情は数秒から数十秒間といった短期間の心の動きであるが, ムードはより長い数分から数十分の期間の心の状態を表し, さらにパーソナリティは半永久的な性向を表す。感情と同様にパーソナリティについても種々のモデルが存在するが, よく知られているのが Big Five あるいは 5 因子モデルと呼ばれる以下の因子である。

☆1 現状は TTS (Text-to-Speech) エンジンの制御可能なパラメータ (スピード, 平均ピッチ, ピッチ変化幅, 強度など) を変化させることで対応している。音声分野では感情的音声合成の研究が進んでいる。

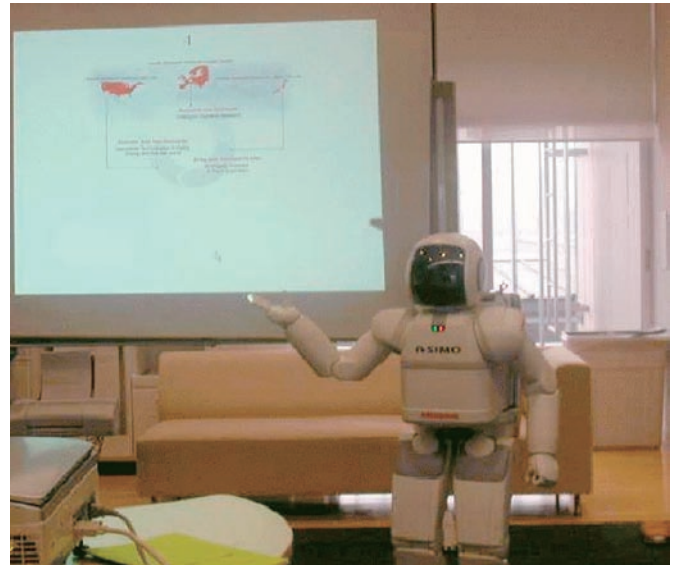


図-3 MPML-HR 記述に基づく Asimo によるプレゼンテーション (ホンダ・リサーチ・インスティテュートと東大の共同研究による)

- 神経質 (Neuroticism) <--> 安定的 (Emotional stability)
- 内向的 (Introversion) <--> 外向的 (Extraversion)
- 経験に対して非開放的 (Closeness to experience)
  - <--> 経験に対して開放的 (Openness to experience)
- 調和的 (Agreeableness) <--> 敵対的 (Disagreeableness)
- 生真面目 (Conscientiousness)
  - <--> 無頓着 (Lack of conscientiousness)

個性的なエージェントを実現するために, このようなパーソナリティの付与も行われるようになってきている。

感情的 (affective) あるいは共感的 (empathic) インタラクションを達成するためには, 相手の心的状態 (感情状態) を推測する機能が必要となる。上記のエージェント人工感情モジュールとしても, 相手方の感情の情報は入力として重要である。表情や発話音声から感情認識を行う研究も存在するが, 必ずしも信頼性は高くはないことから, 最近では生体センサがよく用いられるようになってきている。皮膚導電率 (SC) は感情の高揚度, 興奮度 (arousal) に対応し, 心拍数 (BVP) は感情のポジティブ/ネガティブ度 (valance) に対応することになるので, Lang の 2 次元感情モデルを用いるとこれらの計測値を感情に対応させることができる。

生体センサとは別に, 視線情報も相手方の関心や心の動きを知る上で大変重要である。視線追跡センサを用いてこのような情報を推定し, 提示するコンテンツやエージェントの対応の仕方を適応的に変化させることも行われるようになってきている。図-2 のエージェントによるプレゼンテーションにも前述のように, この機能が組み込まれている。

生命的エージェントとのインタラクションにおいて以

## 生命的エージェントシステムの構成要素

### ◇エージェントキャラクタ

全身のもの、上半身のもの、頭部だけのもの（トーキングヘッドと呼ばれ、普通、豊かな顔表情表現を可能にしている）がある。2次元と3次元のキャラクタがあり、またあらかじめ定められた30～50種のジェスチャを行うものと関節角制御等も可能な自由度の高い動作を行えるものがある。発話に同期した唇の動き（lip-sync）機能は必須である。一定期間動作指令がないとアイドル動作を行うような簡単な自律性は与えられているが、より高い自律性は今後の課題である。

### ◇TTS(Text-to-Speech)エンジン(+音声認識ソフト)

エージェントの発話用にTTSエンジンを組み込む。TTSはインタフェースとして標準のSAPIを持つので使いやすくなっている。音声の性質を変え感情的な音声を生成できるものが好ましい。音声入力も行う場合には音声認識ソフトの組み込みも必要になる。

### ◇コンテンツ記述言語と関連ツール

通常のプログラミング言語（Webコンテンツの場合はJavaScriptなど）から上記のような機能モジュールを呼び出すのがベースとなるが、これではその言語でプログラミングしなければならず、一般人には敷居が高い。そこで、容易にキャラクタを含むマルチモーダルコン

テンツを記述する言語の開発が行われているのだが、残念ながらまだデファクトも含めて標準はない（本文参照）。この記述言語は背景画面等を変えるなどの場面制御や、さらに対話管理機能をサポートする必要がある。その他、魅力的なコンテンツの容易な作成を可能にする知的支援機能が必要とされている（本文参照）。

### ◇その他の付加I/O機能

ユーザ状態を観測して感性的あるいは共感的対応を可能とするための感情センサ入力、カメラ画像入力等も使うことがある。また、知識・情報ベースとしてのWebへのアクセス機能等も必要になることもある。

上のようなセンサの使用はどの程度普及するかは不明確なところもあるが、ゲームでは可能性が高く、またAffective Computingの研究とも関連し手軽なセンサが普及するようになれば、他の一般の場面でも浸透することが考えられる。

## エージェントの自律性の向上

エージェントの発話や動作は現状では人がスクリプティングするのが最も実用性が高いのであるが、シナリオとともにエージェントの（感性的な）マルチモーダルな行動をスクリプティングするのは、コンテンツ作成のコストが大きい。エージェントの自律性を次第に高めていき、このコストを低減する必要がある。プレゼンテーションの場合では、理想的にはTVアナウンサーに原稿あるいは伝達内容の情報をプレゼンテーション用資料（図や表）とともに渡せば、感情や強調も含む適切な様式で伝えてくれる、といったようにエージェントもなってほしい。これまでのエージェントもアイドル動作（一定期間動作コマンドが与えられないときに自発的に起動する動作）といった単純な自律性を備えていたが、自律性の幅と深みを増していくことが必要となる。ここではそのような方向へ向けての研究を紹介する。

発言テキストに対応した適切なジェスチャを付与することは負担が大になることから、ジェスチャの自動生成に向けた研究が進んでいる。CassellらのBEATでは、発話の特定の動詞に関連付けて表象的（iconic）ジェスチャ

を生成する機能を実現している。最近では、事例ベースのアプローチが取り入れられるようになってきている。KippはTVトークショーの発話とジェスチャを解析して、両者の要素（lexiconsとgesticon）間の関係を見出し、これによる発話テキストからのジェスチャの自動生成を行った。Stoneらは人間の発話に伴う手の動きを記録し、この解析に基づき発話から手の動作を生成している。SmidらはTVニュースキャスターの発話と頭部の動きの関連付けの解析に基づき、発話テキストから頭部の動きと表情を生成している。

前章のエージェントの感情についても、テキストから抽出する研究が進んでいる。現状で評価が高いのはHugo Liu（MIT）らによる研究<sup>10)</sup>であり、これは多く見られる感情関連の単語をキーとする単語レベルの解析によるのではなく、常識ベース（集合知として収集した事実を表すセンテンスのコーパスであるOpen Mind Commonsense）を利用してテキストからの感情抽出を行っている。我々のテキストからの感情抽出研究もあり<sup>11), 12)</sup>、最近のものではConceptNetやWordNetの大規模知識ベースと自然言語解析を利用するようになってきている。これにより、前述のHugo Liuらのシステムでは基本5感情の抽出であるのに対し、より多種の感情抽出を可能にしている。発話テキストから感情が抽出されると、発話音声を感情により制御するのに加えて、付加的にジェスチャも生成できることから、ジェスチャの自動生成にもつながる。

音声対話の自由度向上も大きな課題である。現状で

は、状態遷移モデルに基づき各状態に応答パターンルール(ワイルドカードや変数も含む)を用意する方法が最もよく用いられている。音声対話の研究領域で長年研究が行われてきており、他の高度な方法として、フレームベースのモデル、プラン生成のモデルなどがある。しかし、いずれも想定していた事態には対応できるものの、想定外の問いかけには十分に対応できないといった自由度、柔軟性の幅の狭さがあり、対話の自然さを損なうことになっていた。

これらに対し、J.Weizenbaum の ELIZA (1966 年) に端を発するチャットボット (chatbot あるいは chatterbot) は、1990 年代からの Loebner コンテストを中心に進歩を続けている。これらは数万の応答パターンルールを持つことにより (学習によって獲得するものもあり)、自由度の高い話題に対して対話を自然なかたちで続けられるようになってきている。ただし、その対話内容は表面的、反動的なものであり、目的を持つ意味を伝えるようなものではない(人工無能とも称される)。

我々はエージェント対話の自由度を向上させるのにこのチャットボットは有効であると考え、我々の MPML は基本的に状態遷移モデルによる対話をサポートしているが、これにチャットボットを組み合わせた<sup>13)</sup>。すなわち、想定している話題に関しては MPML 中にスクリプトとして記述された応答パターンにより対応し、それ以外の話題の場合にはチャットボットに対応を委ね、この間の移行をスムーズに行えるようにした。チャットボットによる自由対話中に本来の意味ある話題に戻るきっかけを与えるようにしている。ここで使用したチャットボットは ALICE (Loebner コンテストに 3 度優勝している) であり、音声対話は残念ながら認識性能の点で実現できておらず、テキスト入力による対話形式をとっている。チャットボット技術は今後、エージェント対話の自由度、柔軟性の向上に大きな役割を果たすと思われる。

## マルチモーダルコンテンツの自動生成

キャラクターエージェントを含むマルチモーダルコンテンツ量はまだ決して多くはなく、かつその制作もこれまでに挙げてきたような要因により順調に進んでいるわけではない。一方で、Web を中心として利活用できる情報は膨大になってきている。

そこで考えられるのが、利用できる Web 情報からマルチモーダルコンテンツを自動生成できないかというアプローチである。当初の品質は十分でないかもしれないが、効用が認められれば、品質は向上していくと考えられる。オンラインニュースのアナウンサーエージェントに

よる読み上げなどはこの直接的な適用であり、テキストからのジェスチャ、頭部動作や顔表情の自動付与などが試みられている。

我々は複数テキスト文書要約の研究も行っていることから、この研究も背景として開発したのが Auto-Presentation システム<sup>14)</sup> である。これは検索語に対する関連情報を Web 検索エンジンにより収集して要約を作成し、キャラクターエージェントによるマルチモーダルな説明プレゼンテーションを自動生成する。ここでの要約はテンプレートベースのデータマイニングに基づく方法をとっており、以下のテンプレートを用いている。

- i) トピックの What/Who
  - ii) Whereabouts (情報源の場所、コンタクト先など)
  - iii) Why
  - iv) How
  - v) 強調された項目 (<h1>,...,<h4>,<b>,<strong>,<br><big>,<em>,<dt> のタグ等で判別)の短い説明文
- Wikidedia にある項目であると比較的うまく生成できるのだが、Google, Yahoo! から収集した複数 Web ページ情報の要約に関してはまだ課題が多い状態である (一方で、Wikipedia 項目の成長、増加により、この問題は次第に軽減されている面もある)。

図-3は検索語 "Big Bang" に対する Auto-Presentation システムにより自動生成された、2 体のキャラクター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画面を示している。この例に示されているように、関係する画像ファイルも抽出し、要約テキスト文とともに画面に配置するようになっている。

この種のコンテンツ自動構成に関係深い技術として、Storytelling の研究も進んでいる。

## ブレイクするのはいつか?

自然で親しみやすいインタフェースやマルチモーダルコンテンツで重要な役割を果たすとして期待されている、顔と姿を持つ生命的エージェントに関する現状と研究について記した。技術としては実用レベルになっているものの、期待したほどにはポピュラリティは上がっていないという事実もある。そこで、その飛躍のために必要な課題である、(1) 標準プラットフォーム(特に標準記述言語)、(2) エージェントの感情や共感機能、(3) エージェントの自律性向上、(4) コンテンツの自動生成、に関する研究を中心に紹介した。

非専門家である普通の人が Blog 等で情報発信するように、容易に生命的エージェントを含むマルチモーダル情報コンテンツを作成し、発信するようになると思うが、

トル

図-4 ?

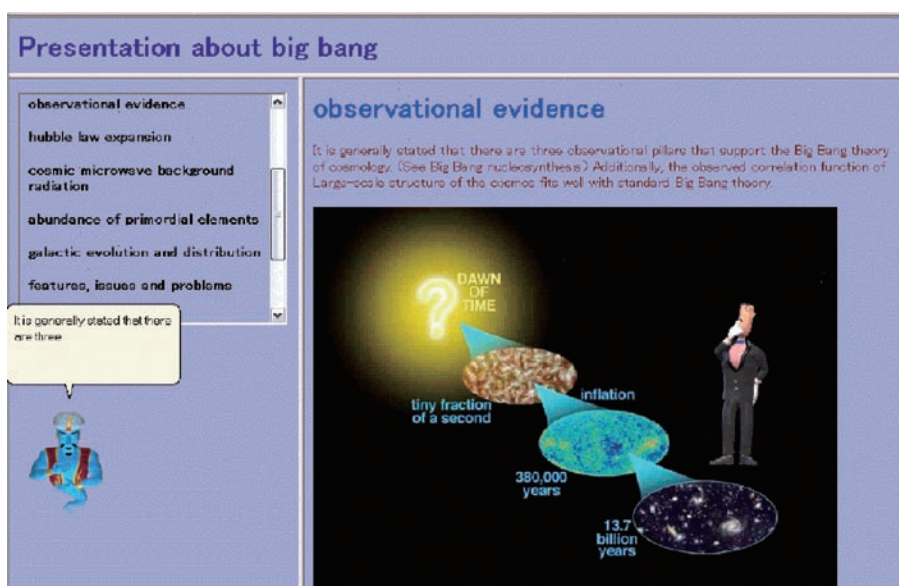


図-4 Auto-PresentationによるWeb情報からのマルチモーダルコンテンツの自動生成画面（ここのトピックは“Big Bang”）

ブレイクするのはいつの時点となるであろうか。本稿で述べたような機能を一定程度可能にした記述言語を中心とするプラットフォームが標準として現れ、誰でもが多彩で魅力的なエージェントキャラクタとコンテンツの制作ができるようになり、これらのツールが無料に近いかたちで提供されるときだと思うが、どうであろうか。

**参考文献**

- 1) Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. (eds.): Embodied Conversational Agents, The MIT Press (2000).
- 2) Prendinger, H. and Ishizuka, M. (eds.): Life-like Characters - Tools, Affective Functions and Applications, Cognitive Tech. Series, Springer-Verlag (2004).
- 3) Shneiderman, B. and Maes, P.: Direct Manipulation vs. Interface Agents, Interactions, Vol.4, No.6, pp.42-61 (1997).
- 4) Reeves, B. and Nass, C.: Media Equation: How People Treat Computers, Television and New Media like Real People and Place, Univ. of Chicago Press (1996).; 細野 (訳): 人はなぜコンピューターを人間として扱うか「メディアの等式」の心理学, 翔泳社 (2001).
- 5) Lester, J. et al.: The Persona Effect: Affective Impact of Animated Pedagogical Agents, Proc. CHI-97, pp.359-366, Atlanta, Georgia (1997).
- 6) Prendinger, H., Descamps, S. and Ishizuka, M.: MPML: A Markup Language for Controlling the Behavior of Life-like Characters, Jour. of Visual Languages and Computing, Vol.15, No.2, pp.183-203 (2004).
- 7) Ishizuka, M. and Prendinger, H.: Describing and Generating Multimodal Contents Featuring Affective Lifelike Agents with MPML (invited paper), New Generation Computing, Vol.24, pp.97-128 (2006).
- 8) Kopp, S. et al.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language, in Intelligent Virtual Agents 2006 (J. Gratch et al. (eds.)), LNCS 4133, pp.205-217, Springer (2006).
- 9) Prendinger, H. and Ishizuka, M.: SCREAM: Scripting Emotion-based Agent Minds, Proc. 1st Int'l Joint Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS-02), pp.350-351, Bologna, Italy (2002).
- 10) Liu, H., Lieberman, H. and Selker, T.: A Model of Textual Affect

Sensing using Real-World Knowledge, Proc. Int'l Conf. on Intelligent User Interfaces (IUI2003), pp.125-132, Miami, Florida (2003).

- 11) Ma, C., Prendinger, H. and Ishizuka, M.: Emotion Estimation and Reasoning Based on Affective Textual Interaction, in Affective Computing and Intelligent Interaction, First Int'l Conf. ACII 2005 (Tao, J., Tan, T. and Picard, R. W. (eds.)), Springer LNCS 3784, pp.622-628, Beijing, China (2005).
- 12) Masum, S. M. A., Islam, Md. T. and Ishizuka, M.: ASNA: An Intelligent Agent for Retrieving and Classifying News on the Basis of Emotion-Affinity, Proc. Int'l Conf. on Intelligent Agent, Web Technologies and Internet e-Commerce (IAWTIC'2006), pp.133-138, Sydney (2006).
- 13) Mori, K., Jatowt, A. and Ishizuka, M.: Enhancing Conversational Flexibility in Multimodal Interactions with Embodied Lifelike Agents, Proc. Int'l Conf. on Intelligent User Interfaces (IUI2003), pp.270-272, Miami, Florida (2003).
- 14) Masum, S. M. A., Ishizuka, M. and Islam, Md. T.: 'Auto-Presentation': A Multi-Agent System for Building Automatic Multi-Modal Presentation of a Topic from World Wide Web Information, Proc. 2005 IEEE/WIC/ACM Int'l Conf. on Intelligent Agent Technology (WI/IAT2005), pp.246-249, Compiègne, France (2005).

(平成 19 年 2 月 7 日受付)

**石塚 満 (正会員)**  
ishizuka@i.u-tokyo.ac.jp

1971 年東大・工・電子卒業, 1976 年同大学院博士課程修了。工学博士。同年 NTT 入社, 横須賀研究所勤務。1978 年東大・生産技術研究所・助教授, 1992 年東大・工・電子情報・教授, 2001 年情報理工学系研究科・電子情報学専攻, 2005 年同創造情報学専攻 (電子情報学専攻兼任)。研究分野は人工知能, Web インテリジェンス, 次世代 Web 情報基盤, 生命的エージェントによるマルチモーダルメディア。IEEE, AAI, 人工知能学会 (前会長), 電子情報通信学会, 映像情報メディア学会, 画像電子学会等各会員。