

## 解 説

## ネットワーク社会を支援する新しい知能メディア技術

## 5. 放送メディアにおける知能情報処理技術

Intelligent Information Processing Technology in Broadcasting by Toshiro YOSHIMURA and Yeun-Bae KIM (NHK Science and Technical Research Laboratories).

吉 村 俊 郎<sup>1</sup> 金 淵 培<sup>1</sup>

<sup>1</sup> NHK 放送技術研究所

## 1. はじめに

放送は広く一般視聴者に向けて、社会におけるさまざまなできごとを伝えるとともに、芸能・教育・娯楽などを提供しているメディアである。この放送メディアは、今、デジタル技術、マルチメディア技術の発展にともない、多チャンネル化など新しいサービスの可能性が広がってきている。

このような中で、たとえば、放送を受信利用する一般視聴者の情報処理活動を支援し、拡張する知能情報処理技術が期待されている。その1つとしてエージェントがあげられる。

本稿では、まずマルチメディア時代における放送サービスの方向性を紹介し、知能情報処理技術と関連が深いと考えられるサービスイメージについて説明する。次に、放送における知的エージェントの枠組みと実現するための技術について解説する。

## 2. マルチメディア時代における放送サービス

コンピュータ、通信、パッケージなどの分野でマルチメディアが進展するにともない、放送をはじめとするオーディオビジュアル情報メディアの利用者の情報消費活動も性格が変わってきている。すなわち、さまざまな伝送媒体を利用して多様な情報を手に入れることが可能であることから、利用者の側が自分の欲しい情報を手に入れる際に自分の条件に適した伝送媒体を選びとるようになる。また、「ちょっと見」や曖昧指向の情報選択または情報サーフィンのようなタイプの情報消費活動も一般的になってきている。

一方、マルチメディアのインフラであるディジ

タル技術のメリットを放送にも生かすことにより、伝送路の高品質化とともに多チャンネル化、多情報化が可能となる。番組を構成する映像、音声などの各種表現メディアはデジタル信号として統一して処理できるようになる。また、高速処理機能と大容量メモリを有する高機能な放送受信端末が今後より容易に実現しやすくなっていくであろう。放送サービス全体も、デジタル技術をベースに、高品質化とともに多情報化、多機能化の方向で発展していくであろう。

このような放送サービスの変化の流れは国際的なものであるが、NHKではデジタル化のメリットを生かした統合デジタル放送 ISDB (Integrated Services Digital Broadcasting) の研究を進めている。ISDBは放送番組の制作、伝送、受信までの範囲すべてと関係する。そのシステムイメージを図-1に示す<sup>1)</sup>。21世紀の放送サービスとして、無線 ISDB を基に、有線ネットワークも利用する新しい形態の放送サービスを想定し、サービスイメージの検討を進めている。以下では、今後の放送サービスのイメージについて、知能情報処理技術の応用の観点から説明する。

(1) イントロダクション、キャプション、インフォメーション<sup>2)</sup>

デジタル放送システムでは、マルチチャンネル放送の番組案内(イントロダクション)に関するデータを放送して受信者の番組選択を支援するサービスや、受信側で字幕をオンオフ制御できるクローズドキャプションサービス、受信者側で選択可能な複数言語の字幕放送サービス、ドラマのあらすじや、スポーツの選手データなど番組の内容に関連する付加情報(インフォメーション)を並行して放送しその情報をいつでも呼び出して利

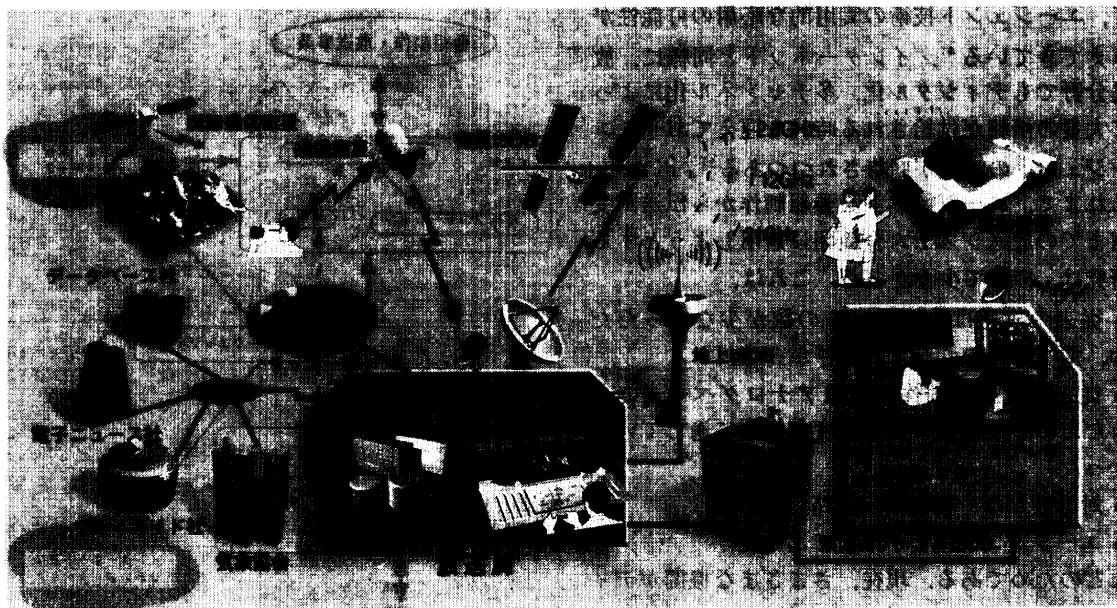


図-1

用できるようにするサービスが考えられる。このようなサービスを受信者に相応しいスタイルで容易に利用できるようにする支援技術が重要である。

### (2) マルチチャンネル

デジタル放送では、1つの放送チャンネルの番組で複数映像の伝送機能を利用することも考えられる。たとえば、スポーツ中継番組で、1つの競技を複数カメラでとらえ、これらの複数の素材画像を放送し受信側で視聴者の好みなどに合わせて自由に選択可能とする番組も考えられる。このような形態の番組においては、視聴者の状況に応じて複数の映像を組み合わせて視聴するための案内方法や自動提示制御の技術が必要とされる。

### (3) インタラクティブ

インタラクティブなマルチメディアサービスが、今後放送においても実現されるようになるであろう。たとえば、ニュース、生活情報などの番組をマルチメディアコンテンツとして制作放送することが考えられる。この場合、番組の中の各項目間あるいは他番組を含めたマルチメディア情報の情報検索・選択機能が必要とされる。また、教育娯楽番組では、視聴者と対話的に提示内容を変化させるプレゼンテーション制御が必要とされる。

以上例示したような高機能かつ多機能なサービ

スを幅広い視聴者層が容易に楽しむことができる受信端末機能が必要とされる。受信端末をインテリジェント化し、ヒューマンインタフェースを向上させて使い勝手をよくすることが重要である。

受信機と視聴者間のインタラクションに関するヒューマンインタフェースとして、GUI (Graphical User Interface)の応用が考えられる<sup>3)</sup>。また、家庭における放送受信に適した指示入力については、種々のポインティング手段を統合的、効果的に利用することが考えられる。たとえば、トラックボールなどの物理デバイスと音声認識(単語、自然言語)などの組合せにより、自然なインタフェースを実現する。受信機の応答および指示の案内情報については、音声合成による応答出力などの手法も考えられる。

このような処理制御を自律的に行い、放送を受信利用する一般視聴者の情報処理活動を高度に支援する手段として以下で述べるエージェントが期待されている。

## 3. 放送におけるエージェントの枠組み

エージェント技術は、ロボティクスや分散AIなどの人工知能分野で理論的な側面を中心に発展してきた。最近インターネットのような膨大な情報を提供できるダイナミックな環境を処理対象とするソフトウェアエージェントの登場によっ

て、エージェント技術の実用的な応用の可能性がみえてきている<sup>4)</sup>。インターネットと同様に、放送分野でもデジタル化、多チャンネル化によって大量の情報が供給される時代を迎えており、エージェントの活躍が期待されつつある<sup>5)</sup>。しかし、これまで、エージェントを番組制作から伝送、受信までの放送システム全範囲に適用しようとする研究は、極めて少なかった。これは、

- 従来の放送はチャンネルの数が少なかったため、情報管理のニーズが小さかった
- 従来の放送や制作機器はアナログベースであったため、計算機による高度な処理に適していなかった
- 対話性がなく単方向伝送路であった
- エージェント技術が未熟であった

などのためである。現在、さまざまな機関がデジタル化、多チャンネル化を検討し、PerfectTVのように通信衛星を利用してすでに70チャンネル以上のデジタル放送が実施されている。また、前章で述べたようにISDBで期待されるサービスを実現するための技術が求められている。このような意味で現在は、エージェント技術の利用を検討する上で格好の時期といえよう。

放送分野への応用を目的とするエージェントシステムを構築するために、考慮すべき点は、

- 放送側と受信側の2種類のユーザが存在すること
  - Hardware エージェントなどを含むさまざまな種類のエージェントが要求されること
- である。エージェントを利用した放送の枠組みを図-2に示す。

以下、放送側と受信側の2つのカテゴリに分けて放送におけるエージェントの応用について述べる。

### 3.1 放送側でのエージェントの応用

放送側でのエージェントの応用分野は広く、これらは大別して番組制作と事業分野に分けることができる。

#### • 番組制作におけるエージェント

番組制作の効率を上げるための支援技術としてのエージェントの利用が第一目的である。そのために、まず番組制作の全体プロセスを検討し、どのような形態でエージェントを制作プロセスに取り込むべきかを考える必要がある。番組制作は

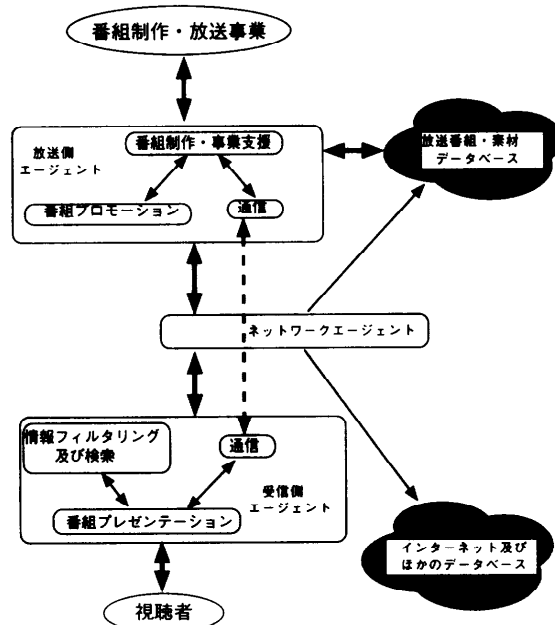


図-2

(1) 発想, (2) 企画・台本, (3) 取材・素材収録, (4) 素材加工, (5) 編集の5つのプロセスから成り立っている。これらのプロセスは、制作者の感性や意図に大きく依存するので、番組制作側でのエージェントの利用は、全自動番組制作システムでなく、制作者のニーズに合ったさまざまな情報を多角的に提供することで番組制作時間を短縮し、かつ品質の高い番組の制作支援するのが現実的であろう。

たとえば、エージェントはインターネットやほかのデータベースを能動的に検索、またはインターネットの専用エージェントと通信しながら、常時適切な情報を制作者に提供することで番組の発想や企画を支援することが可能である。また、制作者の簡単な指示に従って、エージェントがTVカメラを操作しプロのカメラマン並みの撮影を行うことも考えられる。さらに制作者の指示によって映像素材を検索したり、編集機器をコントロールして編集作業を支援することで番組制作の効率をあげることもある。

番組制作における実験的な応用例としてPinhanezらが提案したインテリジェント・スタジオ・プロジェクトがある<sup>6)</sup>。インテリジェント・スタジオでは、スマートキャムス(SmartCams)と呼ばれるカメラがあり、ディレ

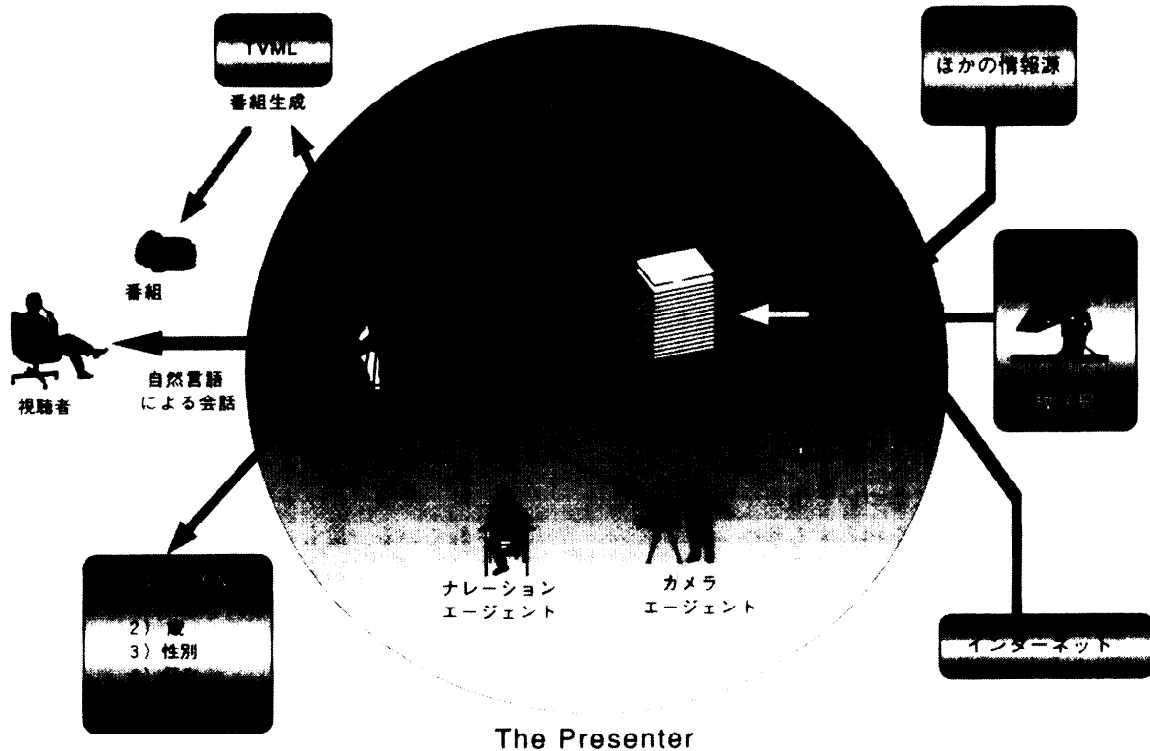


図-3

クタの言葉による指示にしたがってスタジオの中の人間や物体を能動的に撮影することを目的としている。これらのスマートカムスは、カメラレンズから得た視覚データをそのシーンや物体に対する記号で表現された知識と関連させることで、物体の撮影を行う。また、複数のスマートカムスが共通の知識を参照しながら協調することで、多様な角度からの撮影が可能である。

このような技術は、制作プロセスの単純化をはかることができ、制作コストの負荷を軽減することができる。一方、人工生命のような自律エージェントや擬人化されたエージェントは、番組の仮想キャラクターとして、視聴者と仮想現実の世界の中で対話的にやり取りできる参加シアタ (Participatory Theater) のような番組への利用が期待できる<sup>7)</sup>。

一方、放送の事業分野を支援する番組プロモーションエージェントや視聴者の意見を集めるサービスエージェントなどのさまざまな種類のサービスの実現が期待される<sup>5)</sup>。

### 3.2 受信側でのエージェントの応用

受信側のエージェントは、数多くのチャンネル

から送られる番組やほかの情報源から送られてくる情報をフィルタリングし、視聴者のみやすい形態で番組を編成して情報をプレゼンテーションすることを第一の目的とする。このような機能を果たすためには、受信側のエージェントは、放送されてくる番組の内容と各視聴者の好みや趣味に関するさまざまな情報や知識を知る必要がある<sup>8)</sup>。

とくに番組の内容に関する情報は、たとえば、その番組のハイライトを作成して視聴者に提供するために、不可欠である。このような番組の内容に関する情報を記録したものをプログラムインデックスと呼び、視聴者に関してエージェントがもっている知識を記録したものをユーザプロフィールと呼ぶ。

このようなインデックスとプロフィールを用いた受信端末の応用例として、NHKで開発を目指している実験システム The Presenter がある。The Presenter は、ある家族を想定し (実験では3人家族)、番組を視聴する各人の好みや趣味にあった番組をプレゼンテーションするのが目的である。このシステムの構成を図-3に示す。

The Presenter の中心要素であるディレクタエ

エージェントは、家族の各人を識別し、その人のプロフィールを参照したり、視聴者との自然言語による会話を介して番組や情報を検索・フィルタリングまたは再編成して視聴者に提供する。また、インターネットやほかの情報源からデータを集めてまったくユニークな番組をその場で生成して提供することもある。このようなリアルタイム番組生成のためには、TVML (TV program Making Language) のようなマルチメディア・プレゼンテーション言語またはオーサリングツールを駆使する<sup>9)</sup>。このような番組生成には、番組制作のプランに相当するあらかじめ用意された番組シナリオテンプレートを用いる。ディレクターエージェントは、インデックスに基づいて情報を集める情報エージェント、効果音を担当するサウンドエージェント、CGを駆使してスタジオセットを生成するスタジオエージェント、CGのスタジオセットをズーム、パン、チルトなどができるカメラエージェント、字幕を生成するナレーションエージェントを用いて視聴者の好みに合った簡単な番組を生成するものである。

#### 4. 放送エージェントのための基盤技術 —映像インデキシング—

人間は環境を作り、環境は人間を作る。このことは、エージェントについても同じことがいえる。すなわちエージェントは環境とともに存在するものである。このようにエージェントにとって情報の変化や既存のソフトやハードまたはツールとの接続やデータの認識技術は、エージェント間の通信、人とエージェント間のコミュニケーション技術とともにきわめて重要である。とくに放送のように映像、文字、音などマルチメディアを駆使してさまざまな情報の発信を行う分野では、エージェントは自分が扱っているメディアの内容を正確に知る必要がある。そのため、映像認識、音声認識、言語処理技術を用いて、番組や情報の内容を把握することが必要である。しかし、各メディアに対する自動認識率は数年前と比べてかなり高くなっているものの、実用システムへの利用にはまだ十分ではない。このような認識の負担を軽減できる技術の1つが「内容に基づくインデキシング」手法である。ヨーロッパの標準化機構である ETS (European Telecommunication Standard) やアメ

リカの ATSC (Advanced Television Systems Committee) はデジタルビデオ/TV放送システムのためのサービス情報の仕様を標準化しているが、その仕様では、番組のジャンルとサブジャンルに関する記述程度しか規定していない。したがって、この情報を基に、VOD (Video-On-Demand) のようなサービスには利用できるが、エージェントによる番組のダイジェスト版の生成や番組の中からある特定シーンを検索するなどのより高度なサービス機能の実現のためには、より細かい情報の提供できる映像インデキシングのような技術が要求される<sup>10)</sup>。

映像インデキシングによって作成されたインデックスには、映像(カット単位)の意味的な内容が記述されている。最も一般的な映像インデキシング手法として、人間があらかじめ与えたアノテーション(注釈)や番組についているキャプションから重要と思われるキーワードを抽出し、そのキーワードを基に認識や検索を支援する「キーワードに基づくインデキシング」がある<sup>11)</sup>。この手法の最も優れている点は、検索に対するロバストネス(頑健性)であるが、キーワードによるインデックスは、映像の中の被写体同士、あるいは、被写体とその動作・状態との間の意味的な関係を表現できない弱点がある。一方、自然言語処理技術を用いてアノテーションやキャプションを処理し、構文木を基にインデックス生成することで、キーワードの弱点を補う手法も提案されている<sup>12)</sup>。そのほか、Informedia プロジェクトのように画像処理、音声処理、キーワードを組み合わせてインデックスを生成する手法<sup>13)</sup>などさまざまな技術が提案されている。

#### 5. む す び

ここでは、マルチメディア時代における放送サービスの方向性を紹介し、知能情報処理技術と関連が深いと考えられるサービスイメージについて説明した。また、放送における知的エージェントの枠組みと実現するための技術について解説した。

エージェント技術の最も重要な性質の1つは、分散指向である。複数のエージェントがお互いに協力または競争しながらあるタスクの達成を果たすことである。このような分散指向の計算環境が

成功するためには、エージェント、オブジェクト、既存のソフトやハード間のインタオペラビリティ（操作の共通性）が本質的な条件となる。このようなインタオペラビリティを確保するための1つの手段が共通のインタフェースやプロトコルを標準化することである。エージェント技術の標準化を目指す国際組織フィパ(FIPA -Foundation for Intelligent Physical Agent)が1996年9月に設立され<sup>14)</sup>、活動を開始している。当面の標準化の対象は次の3つのカテゴリである。

- エージェントの管理技術
- エージェントと人、エージェント同士の通信技術
- エージェントと既存のソフトとの通信インタフェースやプロトコル

エージェント技術のようなまだ未熟である技術を標準化の対象にすることは、時期尚早という声もあるが、すでに実用化の試みも始まっており、これらが個々別々に行われたのではエージェントの特徴である分散指向は十分に発揮されなくなってしまう。そこで、今の段階から標準化の努力を開始することは意味のあることといえる。

#### 参 考 文 献

- 1) 柳町：ISDBによる放送サービスの高度化とマルチメディア、NHK技研R & D, No.33 (Sep. 1994).
- 2) 白井他：「ISDBでのテレビジョン高機能化」テレビ学技報, BCS94-10 (Mar. 1994).
- 3) 柳町他：放送とヒューマンインターフェース、テレビ誌, Vol. 48, No. 8 (Aug. 1994).
- 4) Etzioni, O. and Weld, D. : A Softbot-Based Interface to the Internet, Communications of the ACM, Vol.37, No.7 (July 1994).
- 5) 金 淵培, 柴田正啓, 林 正樹：エージェントによる新放送サービス, 電子情報通信学会技術研究報告, Vol. 95, No. 265 (1995).
- 6) Pinhanez, C. and Bobick, A. : Intelligent Studios : Using Computer Vision to Control TV Cameras, IJCAI'95 Workshop on Entertainment and AI/Alife (Apr.1995).
- 7) Maes, P et al. : The ALIVE System : Wireless, Full-body Interaction with Autonomous Agents, MIT

Media Laboratory Perceptual Computing Technical Report No. 257.

- 8) Kim, Y-B. et al. : Agent-based Broadcasting with Video Indexing, IEEE Transactions on Broadcasting, Vol. 42, No.3 (Sep. 1996).
- 9) 林 正樹：テキスト台本からの自動番組制作～TVMLの提案～, テレビジョン学会年次大会講演予稿集 (1996).
- 10) 柴田正啓：映像の内容記述モデルとその映像構造化への応用, 信学論(D-II), Vol.J78-D-II, No.5 (1995).
- 11) Smoliar, S.W. et al. : Content-Based Video Indexing and Retrieval, IEEE Multimedia, 1(2) (1994).
- 12) Kim, Y-B. and Shibata, M. : A Video Indexing Method using Natural Language Memo for TV Program Production, 12th European Conference on AI, ECAI-96 (Aug. 1996).
- 13) Hauptmann, A. and Smith, M. : Text, Speech, and Vision for Video Segmentation : The Informedia Project, AAAI Fall 1995 Symposium on Computational Models for Integrating Language and Vision (1995).
- 14) <http://drogo.cselst.stet.it/fipa/>  
(平成8年11月5日受付)



吉村 俊郎

1977年東京大学工学部電子工学科卒業。同年NHKに入局。京都局、技術本部を経て1987年より同放送技術研究所に勤務。以来、衛星データ放送、ISDB利用技術の研究に従事。電子情報通信学会、テレビジョン学会各会員。



金 淵培 (正会員)

1983年サンパウロ大学工学部化学工学科卒業。1984年(株)ラビリス・インターナショナル入社。1991年NHK入局。現在、放送技術研究所・先端制作技術研究部において、ニュース用機械翻訳、自然言語処理、マルチメディア情報処理、放送におけるエージェントの研究に従事。言語処理学会、テレビジョン学会、AAAI各会員。