

# 1. 知的画像符号化と知的通信

正会員 原 島 博<sup>†</sup>

## 1. ま え が き

画像通信は今後どのような方向へ向かうのであろうか。本稿では、主として画像符号化の立場から、その将来像を探ってみたい。

我々人間は、情報の大部分を視覚を通じて受け取っているといわれ、将来の通信メディアとしての画像通信の重要性が古くから指摘されながら、その歩みは必ずしも順調ではなかった。たとえば、1960～70年代には、アナログテレビ電話の商用サービスが試みられ、またデジタルテレビ電話を目的とした高能率符号化の研究も行われた。しかし、これらは結果的には成功しなかった。

画像通信が本格的に実用に至ったのは80年代になってからであろう。すなわち、1980年に標準化されたデジタルファクシミリがビジネス分野を中心として広く普及し、また、同じくビジネスの利用を目的としたテレビ会議システムも実用化された。ビデオテックスなどの画像データベースサービスも開始されている。

さらに最近、画像通信あるいは画像符号化の分野において、いくつかの新しい動きが見られるようになった。

そのひとつは標準化へ向けた動きで、たとえば、テレビ会議を目的とした1.5 Mb/s/384 kb/sの符号器の標準化が急ピッチで進められている。また、デジタル電話回線で伝送可能なテレビ電話用64 kb/s符号器が相次いで発表され、その標準化作業も始まりつつある。さらに、静止画像の符号化についても、ISOとCCITTの協同作業で、その標準化の方向が固まりつつある。

いまひとつは画像情報ネットワークへ向けた動きで、光ファイバを用いた広帯域LAN、あるいは次世

代のISDNとしての広帯域ISDN(B-ISDN)の検討が盛んである。さらに、非同期転送モード(ATM)の映像パケット伝送技術も注目されている。

放送分野でもHDTVやEDTVなどの高品質放送が最近の話題であり、そのための画像符号化の研究が精力的に進められている。さらには、各種画像情報検索サービスあるいはCD-ROMなどのパッケージ系メディアも注目されている。

このように多様化する画像通信技術あるいは画像符号化技術は、今後どのように展開するのであろうか。

本稿では、将来の画像通信の一形態として期待されている「知的通信」の立場から、その基盤技術としての「知的画像符号化」の構想と研究の動向を解説し、あわせて画像を中心とする次世代の通信のありかたを探ってみたい。

## 2. 知識を用いた通信と符号化

「知的」という言葉は様々な意味で用いられている。また、時代とともに変化している。現状よりも少し人間に近いと思われる機能を持っているとき、そのシステムに「知的」なる接頭語がつくことが多い。その意味では、「知的」なる用語は現在の状態ではなく、その目指している方向を示すものであると解釈した方がよい。本稿で述べる「知的通信」あるいは「知的符号化」も同様であって、現状よりも少し人間に近い機能を持った通信あるいは符号化を志向した概念である。しかし、これではあまりにも曖昧であるので、本稿ではとりあえず次のように「知的通信」を定義して議論を進めることにしよう。

「知的通信」(Knowledge-based Communication または Intelligent Communication)とは、送信端と受信端あるいはネットワークが互いに知識を共有し、通信路においては真に本質的な情報のみのやりとりを行い、利用者に対してはその情報処理様式に最も整合したインタフェース環境を提供する通信である。

もともと通信とは、その原点に戻って考えると、送

<sup>†</sup> 東京大学 工学部

“Intelligent Image Coding and Communication” by Hiroshi Harashima (Faculty of Engineering, The University of Tokyo, Tokyo)

り手と受け手が、それぞれの知識を共有する営みである。もし、送り手と受け手において、あらかじめ共有している知識がほとんどなければ、多くの情報=知識を受け手へ伝えなければならない。逆に、共有している知識が豊富であれば、実際に伝えている情報の量はわずかであっても十分に通信の目的は達せられよう。

これまでの通信は、信号を忠実に伝えることのみを目的としており、送るべき情報に関連した知識を積極的に利用することはなかった。知識を蓄えておくには膨大なメモリー量が必要とされる。また、通信において知識をいかに利用するかというソフトウェア的な問題もある。知識を利用する知的通信は、文字通り夢の通信であったのである。

しかし、最近における情報通信ならびに情報処理技術の進歩は、知的通信が単なる夢でなく、近い将来に現実になるのではないかと期待も少しずつ抱かせてくれるようになった。たとえばメモリーに関しては、近い将来、それこそ湯水のごとく使える日が来るであろう。知識を利用するソフトウェアについても、知識工学あるいは人工知能の分野で精力的に研究が進められている。また、信号から知識を抽出する広い意味でのパターン認識・理解技術も着実な進歩を見せている。

このような知的通信において、知識を利用しつつ、情報信号をその本質的な部分に着目して、記述あるいは表現する操作が「知的符号化」(Knowledge-based Coding または Intelligent Coding) である。知的符号化において知識を利用するためには、与えられた情報あるいは信号の内容を抽出・理解することが必要であ

る。この観点から、知的符号化を「情報の内容にまで立ち入った符号化」と特徴づけることもできよう。

知的符号化によって、情報信号がその意味内容にまで立ち入って記述されていれば、意志の伝達に本質的な情報のみを実際に伝送することにより、真に効率的な通信が実現される。また、推論操作あるいは必要に応じてメディア変換を施すことにより、利用者に最も整合した「知的インタフェース」を提供することも可能になろう。その意味では、知的符号化は知的通信を実現するための基盤技術である。

以下、このような知的通信ならびに知的符号化の目指すべき方向を、主として画像通信の立場から考えてみたい。

### 3. 画像符号化の世代区分と階層モデル

まず、知的通信の期盤技術としての知的符号化の概念を明確にすることから始めよう。知的符号化には、情報のいかなるレベルまで立ち入るかによって様々な発展段階がある。筆者は、画像符号化に関して、表1に示す世代区分を提唱した<sup>1)</sup>。この世代区分そのものはあくまで便宜的なものであるが、現在および将来の画像符号化を分類し、その発展方向を占うひとつの試みとしての価値はあるであろう。

まず第0世代は、単純な「直接符号化」であって、PCMのように画素単位の明るさのみを符号化する。

第1世代は、「統計的冗長圧縮符号化」であって、近傍画素間の統計的な相関を積極的に利用して画像の冗長な部分を取り除くことにより、高能率な情報圧縮符号化を行う。現在実用化されている各種の予測符号化

表 1 画像符号化の世代区分

世代区分	符号化方式	符号化に用いる知識	符号化レート (b/s)	符号化方式の例
第0世代	直接符号化	振幅分布	$10^7 \sim 10^8$	標本値ごとのハフマン符号化など
第1世代	統計的冗長圧縮符号化	画素間相関などの統計的性質	$10^5 \sim 10^7$	予測符号化 変換符号化 ベクトル量子化など
第2世代	構造抽出符号化 特徴抽出符号化	動き情報・輪郭情報などの特徴情報	$10^4 \sim 10^5$	輪郭符号化 構造・特徴情報に基づく適応符号化
第3世代	分析合成符号化 (モデルによる符号化) (パラメータ符号化)	画像のモデル 図形分離 特徴抽出	$10^3 \sim 10^4$	構造モデルに基づく顔画像の分析符号化
第4世代	認識合成符号化 (コマンド符号化)	画像の認識と生成に必要な知識と規則	$10^2 \sim 10^3$	
第5世代	知的符号化	画像の概念モデル 推論・思考 異種情報の統合	$10 \sim 10^2$	

方式および変換符号化方式などは、ほとんどすべてがこの第1世代の統計的冗長圧縮符号化である。

**第2世代**は、「構造抽出符号化」あるいは「特徴抽出符号化」と呼ぶべき符号化方式である。第2世代では、画像を確率的な信号ではなく、構造を持つ物体像の集まりであると考え、すなわち、動き情報や輪郭情報によって画像の構造情報を抽出し、これらの情報を利用して高効率な符号化を実現する。動き情報や輪郭情報などの分析情報は、必要に応じて補助情報として受信側へ伝送される。

**第3世代**の「分析合成符号化」では、むしろ分析情報そのものが符号化の対象になる。そのひとつとして、画像をモデルに基づいて分析し、そのパラメータを数値情報の形で伝送する「モデルに基づく符号化方式」が考えられる。受信側では、伝送されたパラメータによってモデルを変形して画像を再合成する。画像の場合に、汎用的なモデルを想定することは難しいが、4章で述べるように、符号化対象がたとえば人物像などに限定されている場合は、その3次元モデル化が可能になろう。なお、この分析合成符号化は、伝送すべき情報がモデルのパラメータであるので、「パラメータ符号化」と呼ばれることもある。

**第4世代**の「認識合成符号化」では、この合成符号化の考え方をさらに一歩進めて、画像の構造をその要素に分けて認識し、その認識結果を記号情報の形で受信側へ伝送する。この認識合成符号化では、記号的な認識情報のみを符号化すればよいので、より超低レートでの画像通信が可能になろう。また、認識合成符号化は、合成の立場からは画像合成に必要な「コマンド」の符号化伝送となる。すなわち、受信側では伝送されてきた図形コマンドに基づいて、一定の規則(ルール)に従って画像信号が再合成される。

そして、我々の最終目標が、**第5世代**の画像符号化である「知的符号化」である。知的画像符号化では、画像の概念、意味あるいは画像が発生された意図や動機などが解釈・理解されて伝送され、受信側では必要に応じて推論を加えながら信号が再合成される。この知的符号化では、送受信端で推論を行うための知識ベースや推論ベースが必要になる。また、この世代では、画像が概念レベルで扱われているので、音声などの異種情報と統合して処理することが可能となろう。あるいは、受信者にとって真に必要な情報のみを符号化するために、図1に示すように受信者から符号器へのフィードバック路が必要になるかも知れない。

第2世代までの符号化は、受信側で忠実な原画像を再現することを目的とする「波形符号化(Waveform

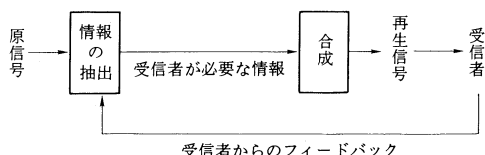


図1 受信者からのフィードバックのある符号化

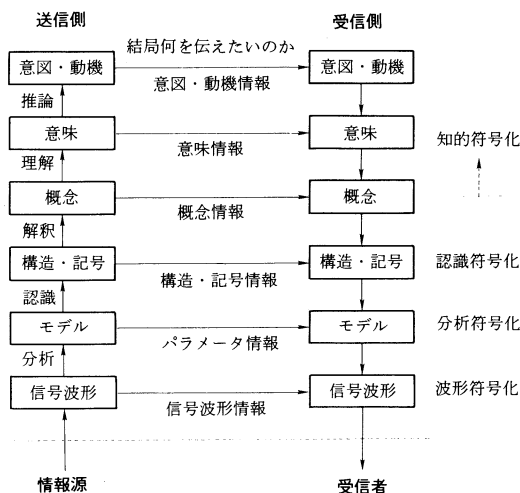


図2 符号化の階層モデル

Coding)」であった。これに対して、第3世代以降の画像符号化は、「合成符号化(Synthetic Coding)」となる。合成符号化では、波形そのものの忠実な再現は重要でなく、そこに含まれている「より本質的な情報」のみが符号化の対象である。この「より本質的な情報」には、通信の目的によって様々なレベルがありうるが、極端な場合は送信者の通信意図だけが伝わればよいこともあろう。これは究極の通信形態であるといえる。

このように符号化を「伝送すべき情報を抽出するプロセス」とみなすと、この立場から、画像符号化における各世代を図2のような「符号化の階層モデル」で表現することができよう\*。ただし、このモデルでは、第0世代から第2世代までを一括して「波形符号化」として位置づけ、その上に第3世代の「分析符号化」と第4世代の「認識符号化」を位置させている。また、第5世代以降の「知的符号化」は、そこで抽出されかつ伝達される情報のレベルに応じて、さらに細かく分類されている。

\* このモデルの発想は森島繁生氏(現成蹊大学)による。

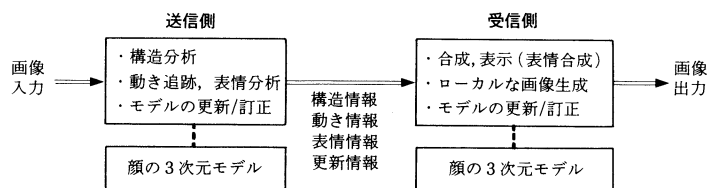


図3 分析合成符号化方式の構成

#### 4. 知的画像符号化へ向けて ——顔画像の分析合成符号化の試み——

前節で述べた世代区分に従えば、画像符号化の現状は第1世代後期あるいは第2世代の初期の段階に位置していることになる。現在の画像符号化方式は画像の動きや情報や構造情報などを充分に利用しているとはいえず、今後はこの方面の研究が精力的に進められることとなる。

さらに、第3世代の分析合成符号化の研究も始められている<sup>2)</sup>。筆者の知る限り、我が国において少なくとも10の研究機関が関連の研究を進めている<sup>3)</sup>。

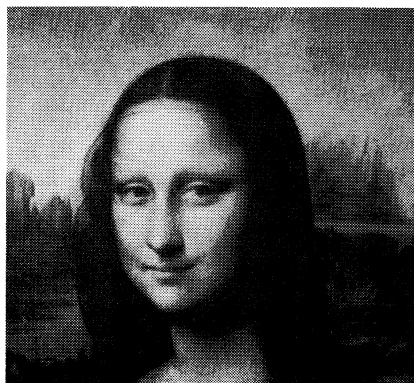
分析合成による画像符号化の一般的な構成を図3に示す。この方式の特徴は、送信側と受信側で対象画像に関する知識を共有し、符号化および復号に際してその知識を積極的に利用することである。画像が人物肩上像に限定できる場合は、顔および頭部についての3次元構造モデルが知識の中心となる。筆者らは、構造モデルとして3次元のワイヤフレームモデルを用い、これに符号化対象人物の顔写真を張り付けて個人のモデルとすることを提案した<sup>2)</sup>。

送信側では、このような知識（モデル）に基づいて、被写体画像の構造の分析、動きの分析、表情等の分析を行い、その分析結果を構造情報、動き情報、表情情報として受信側へ伝送する。また、必要に応じて、通信の開始時には不完全であった知識を更新する情報も伝送する。

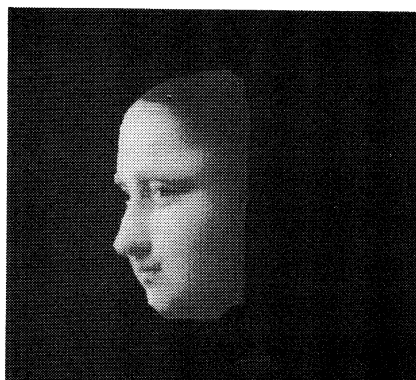
受信側では、伝送されてきたパラメータをモデルにあてはめて顔画像を合成し、動きと表情を与えて表示する。この際、モデルに含まれていない部分（たとえば背景）は、受信側のローカルな画像生成により補う。

このような分析合成符号化では、送信側において画像の構造・動きなどを同定して、そのパラメータを抽出する技術が必要となる。一方受信側では、伝送情報と知識（モデル）に基づいて画像を再合成する技術の開発が必要である。

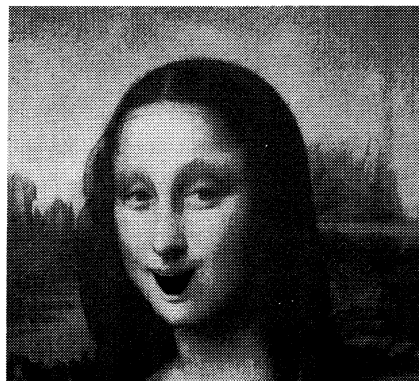
前者すなわち送信端における画像分析に際しては、



(a) 「モナリザ」



(b) 横顔の合成



(c) 表情の合成

写真1 「モナリザ」の横顔と表情の合成例

多くの研究者によってその基盤が整備されつつある画像処理技術やパターン認識／理解技術が応用されよう。一方、後者すなわち画像の再合成に際しては、コンピュータグラフィックス技術やコンピュータアニメーション技術が駆使されることになる。さらに最終的には、送受信端で共有されている知識の処理技術、および知識を操作する人工知能技術も導入が必要になる。分析合成による画像符号化は、このように従来比較的独立して発展してきた「画像処理、パターン認識／理解技術」と「コンピュータグラフィックス／アニメーション技術」、さらには、「人工知能／知識処理技術」を通信固有の「符号化技術」を通じて統合しようとする試みである。

写真1(a)のモノリザ像を対象として、分析合成符号化の手法によって作成した横顔を写真1(b)に示す。また表情付けを行った例を写真1(c)に示す。

## 5. 知的符号化による知的通信の展開

知的符号化の対象は必ずしも画像には限られない。たとえば音声の知的符号化も考えられる。音声の場合は、その生成モデルのパラメータを伝送する分析合成符号化方式が、PARCOR やマルチパルス方式としてすでに実用化されている。将来は、音声認識あるいは言語認識をとまなう符号化も実現されよう。

それではこのような知的符号化は、次世代の通信技術という観点からいかなる意義があるのであろうか。筆者自身は、知的符号化は情報通信のみならず情報処理も含めてその基盤技術であると考えている。そのすべてを述べるには紙面の余裕がないので、ここではとりあえず通信の立場から、知的符号化のいくつかの側面を述べておこう。

### 5.1 超低レート情報圧縮技術としての知的符号化

知的符号化によって大幅な情報圧縮が期待される。それでは、知識を利用することにより伝送すべき情報量はどこまで圧縮できるのであろうか。もちろん目的によって異なるであろうが、たとえば、画像や音声の通信においては、人間が最終の受信者であることを考えると、情報圧縮の究極の限界は、人間の情報受容能力あるいは情報処理能力で与えられることになる。これについては、主として心理学の分野で数多くの測定例がある。

たとえば、人間の視覚から受容される情報量は、視覚や網膜の分解能を考慮することによって、(読書に適当な明るさの下で) $10^6 \sim 10^8 \text{ b/s}$  であろうと推測されている。また、聴覚から受容される情報量は  $10^3 \sim 10^5 \text{ b/s}$  程度とされている。これらは、テレビ信号あるい

は電話信号を直接 PCM 符号化したときの情報量にほぼ等しい。

しかし我々人間は、このような多量の情報量をすべて脳の中核において意識し、処理しているわけではない。各種の心理学実験によれば、人間が意識して処理している情報量はたかだか1秒間あたり数十ビットに過ぎない<sup>1)</sup>。このことは、感覚レベルから意識レベルに達するまでに、人間にとって真に必要な情報の選択操作が行われていることを意味している。

人間が通信の最終受信者である場合は、人間の最終的な情報処理能力に見合うだけの情報が送信側から送られていればよい。この立場からは、画像あるいは音声の通信において真に必要な符号化情報量は、1秒間あたり数十ビットでよいことになる。もちろん、これは理想的な究極の符号化の場合であって、実際にはこの数字を実現することは至難であろう。しかし、少なくとも我々の努力目標は与えてくれる。

表1に、先に述べた画像符号化の各世代において、目標とされている符号化レートを示しておいた。この表によれば、究極の画像符号化である知的符号化では、 $10 \sim 100 \text{ b/s}$  の情報量で画像を伝送することが目標となる。これに対して、認識符号化では  $100 \sim 1 \text{ kb}$ 、分析符号化では  $1 \text{ k} \sim 10 \text{ kb}$  程度の伝送情報量となることが期待される。もし  $10 \text{ kb}$  以下で動画像が伝送できれば、 $1200 \sim 9600 \text{ b/s}$  のアナログ電話回線が使用できることになり、社会的なインパクトも大きいであろう。

### 5.2 情報の構造化技術としての知的符号化

以上述べたように、知的符号化により従来にない超低レートの情報通信が実現されよう。しかし、筆者の私見では、情報圧縮は必ずしも知的符号化の本質ではない。

知的符号化の本質は、むしろ「画像あるいは音声などの情報信号を意味内容にまで立ち入って解析して、

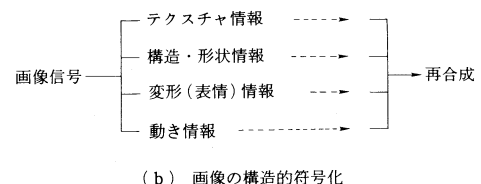
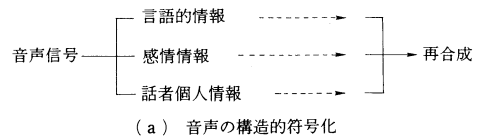


図4 音声と画像の構造的符号化

構造的に記述すること」である。信号が構造的に記述されていれば、通信に本質的な部分のみを伝送することによって情報圧縮が可能になるし、あるいは、必要に応じて編集・加工やメディア変換を行い、いわゆる付加価値通信を実現することもできよう。

たとえば、音声信号の場合、図4(a)に示すように言語情報、感情情報、話者情報などに分解できれば、必要に応じて言語情報のみを送ることにより大幅な情報圧縮が可能になろう。さらに、言語情報から概念情報、意味情報などを抽出しておけば、その概念・意味情報に基づいて、異なる言語間の自動翻訳通信を実現することもできよう。

画像の場合も、図4(b)に示すように画像信号をその構造情報、テクスチャ情報、動き情報、各種の変形情報（たとえば顔における表情情報）などに分解して構造的に記述しておくことにより、様々な編集・加工あるいはメディア変換が可能になろう。あるいは、画像において人間が注目している部分（例えば顔画像では目や口）を中心に符号化することによって、主観的な品質を落とさずに大幅な情報圧縮が期待できよう。

知的符号化のひとつの応用として、データベースサービスがある。データベースには、一般に極めて多量の情報を蓄積しておく必要があり、超低ビットに情報圧縮できればその効果は大きい。また、データベースサービスでは、検索要求に応じて情報の編集加工が必要になることも多い。その際、情報が抽象的な記号情報の形に符号化されていれば、その記号操作によって自由な情報の編集加工ができよう。また、必要に応じてメディア変換を行うことによって、真にマルチメデ

ィアのデータベースサービスを提供することが可能になろう。

### 5.3 知的インタフェース技術としての知的符号化

最近、情報システムとその利用者の間のインタフェース環境を改善するために、「知的インタフェース」と呼ばれる技術が注目されている。知的符号化は、通信システム端末あるいは計算機システム端末における知的インタフェース技術としての役割も期待されている。

利用者から与えられた画像信号や音声信号をその意味内容にまで立ち入って構造的に記述し、記述された内容进行操作し、必要に応じて信号を再合成して利用者に提示することは、知的インタフェースを実現するための必須技術である。また、利用者に最も適したメディアに情報を変換して提示することによって、たとえば視覚障害者には音声で、聴覚障害者には画像で情報を提示し、両者が自由に会話することも可能になろう。

知的インタフェースによって、情報システムと利用者との間の真の整合を図るためには、利用者である人間自身の情報処理機能を解明し、これを巧妙にシステムに組み込む必要がある。知的符号化は、図2の符号化の階層モデルからも明らかなように、人間における情報処理機能をそのまま模倣した形をとっている。したがって、知的符号化におけるそれぞれのレベルにおける処理は、図5に示すように、情報源あるいは受信者である人間の情報生成プロセスと情報処理プロセスに対応させることができよう。

図5は、知的符号化を用いたヒューマンコミュニケーション

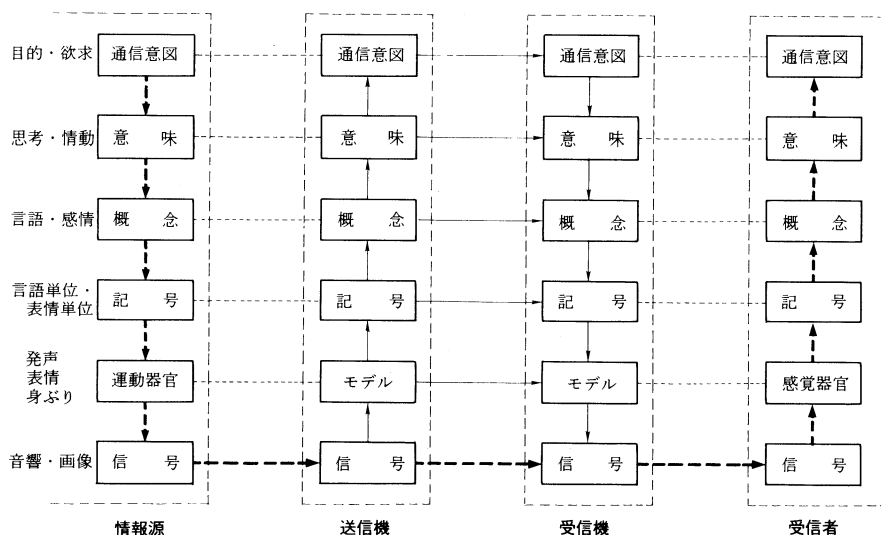


図5 ヒューマンコミュニケーションのモデル

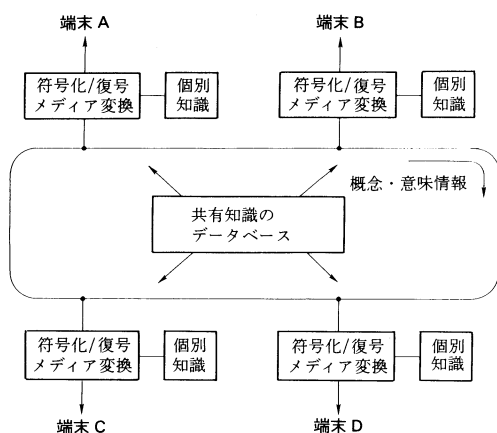


図6 知的情報通信ネットワーク（ネットワークの形態はリング状でなくてもよい）

ーションのモデルとみなすことができる。このモデルの垂直方向は情報処理あるいは情報生成の軸であり、水平方向は情報通信の軸である。このようなヒューマンコミュニケーションモデルにおける情報の階層構造を明らかにすること、および、それぞれのレベルにおける情報の伝達様式を解明することが今後の課題である。

#### 5.4 知的符号化による知的情報通信ネットワークの実現

知的符号化による知的通信の最終形態は、上で述べた情報圧縮技術、メディア変換技術、知的データベース技術、知的インタフェース技術などを統合した「知的情報通信ネットワーク」（図6）であろう。このネットワークには、利用者（端末）毎の個別知識のデータベースとシステム全体で共有されている知識データベースが分散配置されており、利用者は必要に応じてこれらのデータベース間で知識のやりとりを行う。また利用者は、最も適した情報表現メディア（言語、音声、画像など）に情報を変換・合成することにより、最終的に必要とする情報を受け取っている。このように、ネットワーク上は真に重要な情報のみをやりとりし、利用者に対しては最も適したインターフェースによ

て情報を提供することが、知的情報ネットワークの目的である。

この知的情報ネットワークを実現するための基盤技術は、上述の知的符号化技術であるが、さらにこれに加えて、最も効率的な知識の分散配置問題、プロトコル、メディア変換、情報セキュリティなど、解決すべき課題は山積みされている。

## 6. む す び

本稿で述べた知的通信および知的画像符号化は、次世代の画像通信ではなく、次の次、あるいはさらにその先の通信形態かも知れない。このような通信がいつ頃実現するかという問いに対しては、残念ながら筆者は回答を持ち合わせていない。しかし、少なくとも将来進むべきひとつの方向として存在することは確かであろう。かつて、1940～50年代に行われたデジタル通信の先駆的研究が今日のISDNの時代の礎になったように、来たるべき新しい世紀へ向けた研究が、いまこそ要請されているのではないだろうか。通信は決して成熟した技術ではなく、これから面白くなる分野であるし、またそうでなければならない。

（昭和63年4月30日受付）

## 〔参 考 文 献〕

- 1) 原島 博, 相沢清晴, 斎藤隆弘: “次世代画像符号化の構想—分析符号化から知的符号化へ—”, 信学技報, IE87-1 (Apr., 1987)
- 2) K. Aizawa, H. Harashima and T. Saito: “Model-Based Synthesis Image Coding System—Modeling a Person's Face and Synthesis of Facial Expressions—”, IEEE/IEICE, Global Telecommunications Conference, 2.3 (Nov., 1987)
- 3) 原島 博: “顔画像の分析合成符号化方式の動向”, テレビ学技報, ICS88-17, IPA88-11 (Feb., 1988)



はらしま ひろし  
原島 博 昭和48年、東京大学大学院博士課程修了。同年、同大学講師、50年、同助教授となり、現在に至る。現在、同大学工学部総合試験所に勤務。通信理論、通信方式、画像の符号化と処理、デジタル信号処理の研究に従事。工学博士。正会員。