

## 解説

## 仮想環境社会の展望

## 3. 仮想環境を実現するヒューマンインタフェース

Research Issues on Human Interface Technologies for Cyber-space Communications by Shigeru AKAMATSU (ATR Human Information Processing Res. Labs.).

赤 松 茂<sup>1</sup>

<sup>1</sup> ATR 人間情報通信研究所第2研究室

## 1. はじめに

映像や音響を中心としたマルチメディア情報を駆使することで臨場感のある仮想的環境を生成し、人間にあたかもその環境の中にいるような感覚を味わせるバーチャルリアリティ技術<sup>1)</sup>の進展は著しい。人間とのインタラクションを可能とする画像認識や音声認識などの AI 技術あるいは高度なネットワーク技術とあいまって、このような仮想環境は近い将来、人與人、および、人とコンピュータとのコミュニケーションの新しい舞台として発展し、これによって仮想環境社会の到来をもたらすものと期待されている。

仮想環境を介したコミュニケーションが人間にとって使いやすい自然なコミュニケーション手段として受け入れられるためには、実世界の人間とコンピュータ内に構築される仮想環境との間の接点を担うヒューマンインタフェース技術の果たす役割は大きい。本稿では、その将来展望として、ユーザである人間の情報処理メカニズムや情報に対する受容特性にフィットした技術を確立することの重要性を強調する。また、視覚の高次認識機能について、このような視点にたって、人間の情報生成・認知機構の解明とその工学的モデルの提案をめざして進めている研究の具体例を紹介する。

## 2. 仮想環境コミュニケーションにおけるヒューマンインタフェースの課題

図-1 は、人間と人間のコミュニケーション形態の変遷を、人間、機械、環境のかかわりという観点で比較したものである。まず、電気通信のようなコミュニケーション技術が登場する以前に

は、手紙や書物による文字を媒介とした情報の交換を除けば、人間と人間との意思の疎通はもっぱら対面対話、いわゆる face-to-face でのコミュニケーションによっていた。そこでは自らが存在している時間・空間的に限定された環境の範囲内という制約はあるものの、図-1(a)のように五感を自在に使った情報交換が行われていたといえよう。電気通信技術の進歩は、図-1(b)に示されるようにネットワークや計算機といった機械を媒介とすることで、時間・空間的な距離を克服することを可能とした。しかし、電話は音声しか伝えられない、というような「機械」の側の技術的な制約によって、人間と人間の間で交換される情報は、質と量の両面で、face-to-face による対話には遠く及ばないという状態が続いてきた。マルチメディア通信は、このような従来の電気通信の

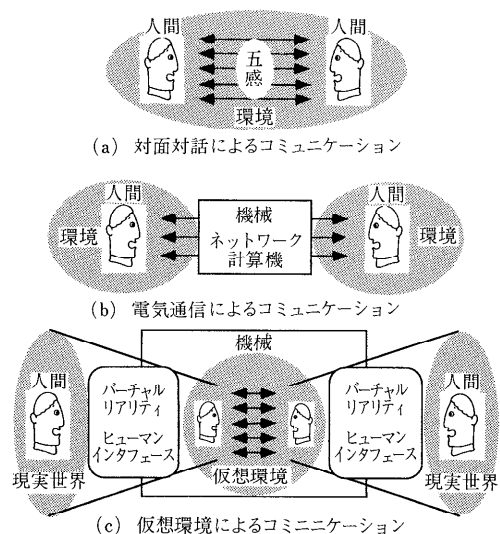


図-1 コミュニケーションにおける人間・機械・環境

技術的制約を取り除き、人間にとってより自由で豊富な情報の交換を可能にするものと期待されているといえよう。

一方、近年、臨場感のある仮想的な環境を計算機によって作り出し、人間にあたかもそのような人工的環境の中にいるような感覚を与えるというバーチャルリアリティ技術の急速な発展を背景として、こうした計算機上に構築された仮想環境を舞台とした新しいヒューマンコミュニケーションの実現形態を仮想環境コミュニケーションと呼び、マルチメディア通信の将来イメージの1つとして注目を集めるようになってきた。これは、時間・空間的に離れた異なる「現実世界」の環境に生きている人間同士を結びつけるネットワークが提供する情報チャンネルの量的拡大を意味するだけではない。図-1(c)に示すように、計算機という機械の中に構築された「仮想環境」に人間の感覚や意識を入り込ませ、そのような仮想環境を互いに共有させることで、ユーザが所属している現実世界の制約に必ずしもとらわれない質的に新しいコミュニケーションを可能にするものといえる。したがって、現実世界での face-to-face の対面対話におけるリアリティを仮想環境においても実現することだけに留まらず、アートの視点にたって人間の感覚・感性を刺激することにより、人間同士のイメージレベルでのコミュニケーションの活性化を目指したハイパリアリティ追求の試み<sup>2)</sup>も盛んになってきている。

このような仮想環境コミュニケーションの場を提供する計算機には、現実世界の人間と人間、人間と環境との間を、計算機の内部に構築された仮想環境を媒介として橋渡しをするメディアとしての役割が期待されている<sup>3)</sup>。そのためには、人間に負担をかけず疲れさせない、自然な使用感を与える、といった「人に優しいインタフェース」、そして、人間がコミュニケーションを通じて伝えたい知・情・意の情報を上手に汲みとることができる「人の欲求を満たすインタフェース」を計算機によって実現する技術の確立が望まれているといえよう。

このようなインタフェースの実現にあたって、現実世界と仮想環境とを橋渡しする情報メディアとしての計算機に要求される機能としては、まず、人間の感覚に効果的に訴えてその意識を仮想環境

の中に引き込ませるため、仮想環境に構築された情報をできるだけ高い臨場感を与え、より自然な形で人間に提示する優れた情報表出の能力をもつことがあげられる。さらには、現実世界の環境や人間が発するコミュニケーションに有用な情報を仮想環境に効果的に取り込むとともに、必要に応じて現実世界の投影像あるいは人物の分身を仮想環境内に作り出せるように、現実世界を認識する能力を計算機がもつことも必要であろう。ここで、前者については提示された情報を人間がどのように感じているのかという点で、また後者については人間にとってどのような情報がコミュニケーションにおいて重要かという点で、どちらの機能に関しても、情報に対する人間の受容特性が重要な要因となっているといえよう。

図-1(a)の face-to-face でのコミュニケーションに立ち返ると、人間はその五感をフルに使うことによって、感覚や論理あるいは感性といったさまざまなレベルの情報を互いにやりとりしている。五感にかかわる情報は目や耳といった末梢器官を通じて入出力されるものであるが、情報の発信、受信をつかさどるコミュニケーションの主体は、人間の「脳」にほかならない。したがって、仮想環境のコミュニケーションにおいて、人に優しい、人の欲求を満たすインタフェースを実現するためには、情報の送り手であり受け手でもある人間の「脳における情報の認知・生成」のメカニズムや特性を、人と人や環境とを橋渡しするメディアとしての計算機の情報処理に反映させていくことが必要と考えられる。

人間は、外界から獲得している情報の3分の2以上を視覚から得ているといわれる。したがって、ヒューマンインタフェースを左右する要因として視覚の果たす役割は五感の中でもとくに大きい。実際、人間の意識を仮想環境の中に引き込ませる効果のある「臨場感」に、主として視覚によって感じとられる立体感が大きく寄与することはよく知られているところである。そして立体感を与えるための視覚情報の表示技術に関しては、特殊な表示装置が人間の体に不自然な負担を強いたり、立体表示原理が人間の視覚メカニズムと完全には一致していないことによって人を疲れさせやすいなどの従来技術の問題点を解決し、より自然な立体視の実現を目指すさまざまな研究が行われて

いる<sup>4)</sup>。

一方、現実世界の環境のもつ情報や人間が発信するコミュニケーションに有用な情報を仮想環境に取り込むことによって人の欲求を満たすインタフェースの実現をはかるという観点からは、目で捕えた視覚情報から対象を認識するという人間のより高次の情報処理の特性を明らかにすることが必要である。

人間による高次の情報生成・認知メカニズムには次のような特徴がある。まず第1に、コミュニケーションにおいては感性的なイメージの伝達が重要な役割を果たしていることである。これは、会話によるコミュニケーションにおいて、言葉自体が言語的に説明する論理的な情報もさることながら、口調の中に含まれている感情情報が重要なメッセージとなっていることが多いことから明らかであろう。また第2点として、人間は五感を駆使して複数の感覚を同時に使ったマルチモーダルなコミュニケーションを行っていることである。雑音が多くて相手の声が聞きとりにくい音環境では、人は唇の動きをみることで話しを聞きとろうとするが、これなどは人間のコミュニケーションがもつマルチモーダルな本質の現れといえよう。また、注目すべき第3の性質として、人間による高次の情報獲得と認知過程は、けっして情報の入力を受動的に待つという姿勢にあるのではなく、注意の集中などによって必要な情報をより積極的に獲得しようとする能動的な性質をもっているという点である。さまざまな音が充満している雑踏の中でも、関心のある人物の声や話題については、注意を集中すれば話しを聞きとることができるというカクテルパーティ効果は、このような人間の情報認知における能動的な性質を如実に示す例といえよう。

仮想環境におけるコミュニケーションを支える人に優しい、人の欲求を満たすインタフェースの実現を展望するとき、人間による高次の情報生成・認知の過程について、以上に示すような視点にたって、その特性やメカニズムを探る基礎研究が必要と考えられる。

本稿では、現実世界と仮想環境との橋渡しする役割を担う計算機の働きのうち、とくに視覚によって現実世界をとらえることで構築される知的ヒューマンインタフェース、そのうちとくに人間の

顔を媒介としたヒューマンインタフェースを題材として、以上の3つの視点にたった研究の具体例を紹介することにしたい。

### 3. 顔を媒介としたインタフェース

現実世界と仮想環境とを橋渡しするヒューマンインタフェースの機能として、現実世界の人間がコミュニケーションのために発する情報を計算機で読みとれることが重要であることはすでに述べた。そのような認識機能に関してはこれまで主に音声認識や文字認識などの言語情報を対象とした研究が進められてきた。しかし、アメリカの心理学者マレービアンの実験によれば相手に好意(好感)を伝える力は、言葉が7%、声の調子が38%に対して、顔の表情は55%に達するという結果が得られている。視覚によって顔は人物の個性、社会的帰属、印象、感情、意図、関心などさまざまな主に非言語情報を伝えており<sup>5)</sup>、face-to-faceのヒューマンコミュニケーションではこのような顔が表す情報が文字どおり大きな役割を果たしているといえる。したがって顔が伝える情報を計算機で認識・生成できるようになれば、人間と仮想環境とを橋渡しするヒューマンインタフェースはより高度なものになると期待される。実際、仮想環境コミュニケーションの将来イメージの1つとして、現実世界の人間が計算機上に構築された仮想環境に自分自身を投影して分身としてのエージェントを作り出し、そのようなエージェントと現実世界の人物とがface-to-faceのコミュニケーションを行うことが考えられる。そのような仮想環境コミュニケーションを実現する上で、顔画像の計算機による認識・生成技術は不可欠である。

実際、90年代の初頭から、顔画像による個人識別<sup>6)</sup>や表情認識<sup>7)</sup>などのように顔が表す情報を計算機に読みとらせる試みや、表情や発話に応じて変化する顔画像を生成する試み<sup>8)</sup>など、顔を媒介とした計算機インタフェースの研究<sup>9)</sup>がさまざまな研究機関で活発に行われるようになった。それらの研究の多くは、顔という対象のもつ複雑で微妙な形、色合い、動きの変化を、いかに正確に、より安定に、より速く認識あるいは生成するかという画像処理、パターン認識、画像生成に関する工学的な基本課題の解決をはかることを主眼とし

て取り組まれてきたといえよう。

現実世界での人間同士の face-to-face コミュニケーションを仮想環境にまで自然な形で拡張できるようにするためには、これらの工学的課題の解決が不可欠であることはいうまでもない。しかし、顔を媒介として伝えられる情報は人間に認知されることによって初めて具体化するものであることを考慮すれば、顔が伝える情報の計算機による認識・生成においては人間の判断や好みに忠実であることが求められる。このためには人間が顔という対象のどのような視覚的特徴に着目し、どのような特徴表現に基づいて個性や表情などの認知に関する高次の判断を行っているのかのメカニズムを探り、そのような認知特性を計算機による情報処理に反映させようとするアプローチが必要と考えられる。

人間による高次の情報生成・認知メカニズムの特徴として先に前章では3つのポイントをあげた。ここではそれらに対応する形で、顔を媒介とした仮想環境とのインタフェースの実現に向けて取り組むべき基礎研究課題として、人間による顔の情報処理のメカニズムや特性を、以下の3つの視点に基づくアプローチによって明らかにすることの重要性を指摘しておきたい。

#### (1) 感性的イメージの情報源として顔をとらえる

非言語コミュニケーションにおいては感性的なイメージの伝達が重要である。こうした観点から、顔によって伝えられるさまざまな感性的イメージが人間によってどのように認知されるかを探ることにより、顔画像の物理的パターン情報から人間の認知するイメージを予測できるようにする計算モデルを明らかにしていく必要がある。

#### (2) マルチモーダルな情報源として顔をとらえる

人間のコミュニケーションは複数の感覚を駆使したマルチモーダルな情報処理が行われている。こうした観点から、face-to-face の対話において顔のもつ視覚・聴覚の異種感覚情報がどのような相互作用をもたらししているかを明らかにしていく必要がある。

#### (3) 3次元物体として顔をとらえる

現実世界の顔は3次元空間での位置・姿勢や周囲の環境に応じてその見え方がさまざまに変化する。

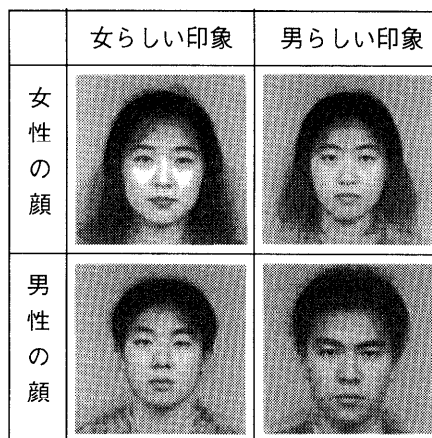


図-2 顔の印象に寄与する2次元的特徴の比較  
「男らしい顔」と「女らしい顔」の例

る。このような変化に対処するためには、必要な情報を能動的な働きかけによって積極的に獲得しようとする人間の方策に学んだアプローチについて研究する必要がある。

筆者らのグループでは、人間による顔の認知という高次の視覚情報処理の特性とそのメカニズムを上記3つの視点にたって探るとともに、人間のコミュニケーションにおいて顔が伝えている情報を計算機によって認識・生成することを目指して研究を進めている。仮想環境を実現するヒューマンインタフェース研究の将来展望を考える一助としてもらうため、これらの研究の具体例を以下に簡単に紹介する。いずれも緒についたばかりの基礎検討レベルであるが、それらの成果は将来、現実世界の人間と仮想環境との間を「顔」を媒介として橋渡しするインタフェース実現に何らかの形で寄与するものと信じている。

#### 4. 顔の感性的イメージを探る

人の顔をみて我々は、たとえば「優しい顔」とか「男らしい顔」といった印象に対応した感性的なカテゴリーへの分類判断を行うことができる。このような顔の感性的イメージが視覚パターンのどのような物理的な特徴に起因して形成されるのかを探ろうとする研究を進めている<sup>10)</sup>。

顔の印象というような感性的な概念は、個人差も大きく曖昧であり時間とともにうつろいやすいなどの理由から客観的にその基準を求めることはこれまで困難とされてきた。しかし感性的判断に

よる顔の分類は人間が顔の集合に存在する視覚的な規則性を暗黙的に学習した結果ととらえることができる。そこで多数の顔画像データベースからあるイメージにもっとも合致する顔を検索するという課題を繰り返し、選ばれたパターンの特徴値のばらつきを解析することによって、分類判断の「規則」として用いられたパターンの特徴の存在を浮かびあがらせるという新しい心理実験手法を導入した。そして顔画像上に目視で抽出した特徴点に基づいて、目や眉といった顔の造作の形態や、目と口の間の距離といった配置を表す特徴値を求め、人間が顔の印象を分類する際にそのような顔の2次元的特徴のレベルで重視している要因があるかどうかを調べた。

一例として、20代の男女各50名の顔を対象として、男らしさ・女らしさという性別の印象の決定に顕著な役割を果たしている特徴を分析した。図-2は、印象の認知に大きく寄与していると判定された特徴をもつ顔を計算機によって平均化し、各印象の典型とみなせる顔を対比したものである。男性の顔の男らしさの評定には、顔の造作の全体的な配置が関連しているのに対して、女性の顔での女らしさの評定には眉の形や唇の厚さといった顔の部分的形態の特徴が関連している。この結果から、顔の男らしさ・女らしさは、男女で異なる顔の特徴から判断されるらしいことが示された。

一方、子供・大人という年齢の印象に寄与する顔の特徴要因についても調べた実験の結果から、男性の顔については、性別と年齢という異なる次元の感性的判断が、顔の全体的な配置に関する共通の物理的特徴に基づいて認知される傾向にあることも明らかにされている。

このように本研究は、年齢や性別のような生物学的な属性の違いや、日本人と外国人の文化的・言語的な背景の違いによって、顔が伝える情報がどのように影響されるかというヒトの感性の本質に迫る展開が期待されている。と同時に、顔の印象を左右する物理的要因として、2次元の特徴から3次元形状、色合い、テクスチャと分析対象をさらに広げることで、ヒトの感性をなぞるような顔の情報処理を計算機によって実現することへの貢献も期待される。仮想環境コミュニケーションにおけるインタフェースという観点からは、現実

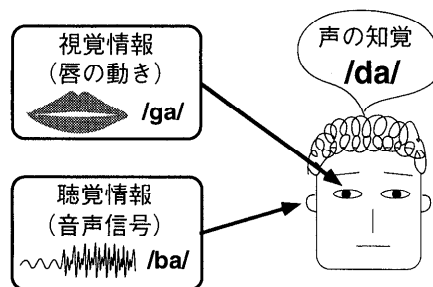


図-3 マガーク効果にみる視聴覚情報の相互作用

世界の人間が顔によって表現しているさまざまな感性的情報を、どのようにして忠実に、あるいはコミュニケーションの状況に応じてはどのようにデフォルメした形で仮想環境に取り込み、分身としてのエージェントを表現するかという問題ともかわる研究といえよう。

##### 5. 顔が伝える視聴覚情報の相互作用を探る

人のコミュニケーションが複数感覚を駆使して行われていることを示す例として、相手の顔を観察することによって騒音下での発話内容の聞き取りが容易になることをとりあげた。しかし視聴覚情報は単に互いに相補的であるというだけではなく、人間は両者を統合的に処理していることが知られている。これを示す代表的な現象が、聴覚による音声知覚が視覚によって影響されて生じるマガーク効果である。これはたとえば、図-3に示すように、/ga/を発話しているときの顔の映像をみながら/ba/という音声を聴くと、多くの場合、人の耳には/da/と知覚されるという現象である。マガーク効果に代表されるマルチモーダルな知覚過程<sup>11)</sup>は多くの研究者の関心を集めてきており、どのような情報が統合されるかに関して、さまざまな観点から比較した実験結果の報告がある。たとえば、言語の違いや母国語と外国語による違い、話者の顔と声の性別の違い、視聴覚情報の時間的同期の違い、倒立提示のような顔の向きの違い、などによる比較である。しかし、これらの知見をカバーし、顔のもつどのような視覚情報がマガーク効果を引き起こしているのかを統一的に説明するモデルはまだ確立されてはおらず、引き続き多くの実験的検討が期待されている。

顔と声の情報は、人間のコミュニケーションにおいて相手が誰であるかを知る手がかりとしても

重要である。マガー効果でみられた視聴覚情報の相互作用が、こうした人物同定に関する高次の記憶レベルでも起こるだろうかという疑問は自然であろう。そこで、顔と声を対にして学習させた場合に、聴覚および視覚の文脈刺激によって顔と声の記憶がどのように影響されるのかを明らかにすることも重要な検討課題である<sup>12)</sup>。

このような人間の情報処理のマルチモダリティに関する基礎的研究は、映像や音声といった複数モードを駆使した仮想環境コミュニケーションのためのヒューマンインタフェース設計に寄与する重要な科学的指針を提供するものと期待される。

### 6. 3次元物体としての顔の見え方の変化を克服する

3次元物体である顔は、みる角度によって見え方が大きく変化するが、多くの場合、人間はそのような変化にもかかわらず相手の顔を正しく認識することができる。このような人間の優れた3次元物体認識能力がパターンの見え方にどのように影響されるのかについて基礎的な研究が進められている<sup>13)</sup>。

たとえば、ある人物の顔を1枚の顔画像で学習した後、与えられた顔画像が同じ人物かどうかを同定する課題に着目した研究を紹介しよう。図-4に示すように、学習した顔画像とテストに用いた顔画像との間で、顔をみる視点を変化させた場合、同一人物と同定する能力がどのように変化するかを心理実験によって測定した。その結果、未知の人物について斜め顔を学習した場合には、正面顔や横顔を学習した場合に比べて、学習していない新規の見え方の顔画像に対しても高い同定能力が得られるという、人間による顔認識の視点依存性に関する基本的性質も明らかになっている。このような研究で得られる人間の顔認識メカニズムに関する知見は、計算機で顔を認識する場合に顔をどのように表現することが望ましいかという点で大いに参考になるとともに、顔画像検索システムのように人間による記憶を前提としたシステムのヒューマンインタフェースの設計という面からも有用であろう。

また現実世界の3次元環境中を自在に動き回る人間を相手に仮想環境とのインタフェースを実現する方策を探るという観点からは、環境中での位

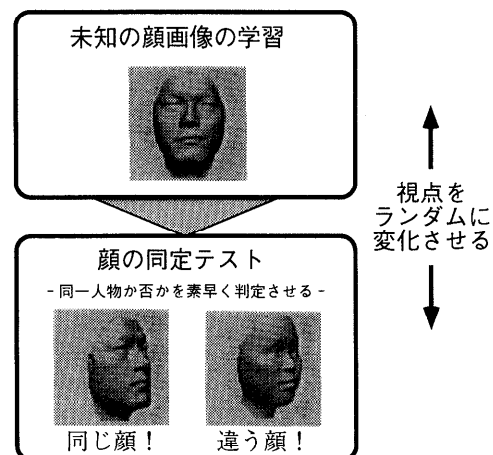


図4 人間の顔認識能力の視点依存性を調べる心理実験

置の変化や遮蔽物の存在による見え方の変化も考慮する必要がある。計算機による顔の認識に関するこれまでの研究はもっぱら正面顔画像が与えられるものという前提で進められてきた<sup>6)</sup>。今後は、必要な情報を積極的に獲得しようとする能動的な性質をもつ人間の高次の情報獲得と認知過程に学ぶことにより、“必要なところだけみやすいようにしてみる”<sup>14)</sup>というアクティブビジョン(能動視覚)のアプローチについての研究が必要であろう。

### 7. む す び

仮想環境を介して自然なヒューマンコミュニケーションを実現するために、現実世界と仮想環境を橋渡しするヒューマンインタフェースを担う計算機には、現実世界の多様な環境の中を自在に振る舞う人間が発する知・情・意の信号を五感を通じて認識するという人間を見守るエージェントとしての働きが求められよう。このためのインタフェース技術の将来展望としては、まず人間における情報の認知・生成のメカニズムと特性を明らかにし、それらを計算機による情報処理に反映させていくというアプローチが大切であることを述べた。とりわけ視覚を通じて人間の顔が発信するさまざまな情報は、コミュニケーションにおいて重要な役割を演じている。これらの情報を計算機で取り扱うことにより顔を媒介としたインタフェースを実現するためには、感性の把握、他感覚との融合、能動的観測という人間の高度視覚機能の特

質に着目する視点にたった研究が重要と考えられる。

### 参 考 文 献

- 1) 廣瀬通孝：バーチャル・リアリティ，産業図書 (1993)。
- 2) 中津良平：アーティスティックな要素を取り入れたバーチャルリアリティ技術—アートと工学の融合をめざして—，日本 VR 学会論文誌 JVR SJ, Vol.1, No.1, pp.1-9 (Sep. 1996)。
- 3) 田村秀行，池田克夫：知能情報メディア，総研出版 (1995)。
- 4) 大頭：二眼式立体視技術と視覚機能，精密工学会誌，Vol.54, pp.5-11。
- 5) ブルース，V.：顔の認知と情報処理，吉川左紀子訳，サイエンス社 (1990)。
- 6) 赤松 茂：コンピュータによる顔の認識の研究動向，電子情報通信学会誌，Vol.80, No.3 (Mar. 1997)。
- 7) 松野勝弘，李 七雨，辻 三郎：ポテンシャルネットと KL 展開を用いた顔表情の認識，信学論 (D)，Vol.J77-D-II, No.8, pp.1591-1600 (Aug. 1994)。
- 8) 森島繁生，岡田信一，原島 博：知的インタフェースのための顔の表情合成法の一検討，信学論 (D)，Vol.J73-D-II, No.3, pp.351-359 (Mar. 1990)。
- 9) 末永康仁，間瀬健二，福本雅朗，渡部保日児：Human Reader：人物像と音声による知的インタフェース，信学論 (D)，Vol.J75-D-II, No.2, pp.190-202 (Feb. 1992)。
- 10) 山口真美，加藤 隆，赤松 茂：顔の感性情報と物理的特徴との関連について—年齢／性の情報を中心に—，信学論 (A)，Vol.J79-A, No.2, pp.279-287 (Feb. 1996)。
- 11) 近藤公久：マルチモーダルな知覚過程，電子情報通信学会誌，Vol.78, No.12 (Dec. 1995)。
- 12) 加藤 隆，神崎利佳，東倉洋一，赤松 茂：顔と声の再認における文脈情報の効果，東北心理学研究，No.45, 東北心理学会第 49 回大会研究発表抄録，p.52 (July 1996)。
- 13) 行場次朗，柳田多聞，赤松 茂：人間の物体認知における視点依存性と非依存性，信学論 (A)，Vol.J79-A, No.2, pp.158-165 (Feb.1996)。
- 14) 久野義徳：アクティブビジョン—歴史と展望—，人工知能学会誌，Vol.10, No.4 (July 1995)。  
(平成 9 年 1 月 14 日受付)



赤松 茂 (正会員)

1952 年生。1975 年東京大学工学部計数工学科卒業。1977 年同大学院修士 (計数工学) 課程修了。同年日本電信電話公社 (現 NTT) 入社。以来，文字認識，画像処理の研究・実用化に従事。1985～86 年米国カリフォルニア大学 Irvine 校客員研究員。1992 年より (株) ATR 人間情報通信研究所第二研究室長，現在に至る。工学博士。パターン認識のヒューマンインタフェースへの応用，感性情報処理，とくに人間とコンピュータにおける顔情報処理が主な研究テーマ。IEEE, ACM, 電子情報通信学会，日本バーチャルリアリティ学会各会員。