

## マルチメディア技術とマン・マシンコミュニケーション

谷口, 倫一郎  
九州大学システム情報科学研究所知能システム学部門

<https://hdl.handle.net/2324/5767>

---

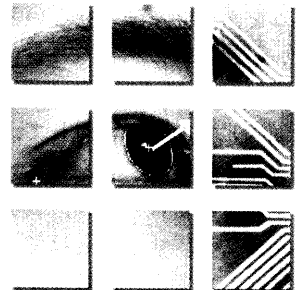
出版情報 : 電子情報通信学会誌. 88 (10), pp.799-803, 2005-10-01. 電子情報通信学会  
バージョン :  
権利関係 :





# マルチメディア技術と マン・マシンコミュニケーション

## Multimedia Technology and Its Application to Man-Machine Communication



谷口 倫一郎

### 1. マルチメディア全盛期

情報処理やコンピュータ、インターネットに関連してマルチメディアという言葉がよく使われるようになって久しい。マルチメディアの持つ本来の意味は、情報を伝達する媒体（メディア）が複数（マルチ）存在することであり、原理的には様々なメディアの組合せを考慮することができる。しかし、情報処理技術の世界では、多くの場合、文字情報に加えて、音声、音楽などの聴覚情報、画像やビデオ映像などの視覚情報をミックスしたものをマルチメディアと呼ぶことが多い。これは、主として以下の点によるものと考えられる。

- ① 人間の情報獲得の多くが聴覚と視覚の二つのメディアに依存していること。
- ② テレビの普及で視覚情報と聴覚情報だけで情報を取得することに慣れていること。
- ③ 視覚、聴覚情報を入出力するための装置が安価に入手できること。

これまで、マルチメディア技術は主に映像や音楽などのコンテンツを伝送・蓄積し、視聴するための技術として開発が進められてきた。代表的な技術としては映像や音楽の伝送・蓄積を効率的に行う情報圧縮技術<sup>(1)</sup>や蓄積された映像情報を検索する技術<sup>(2)</sup>などである。しかし、これからは、マルチメディア技術は人間が複雑なシステムや機械を容易に使うための技術、つまり、人間からシステムが情報を獲得し、そしてシステムが人間に情報を提供するというマン・マシンコミュニケーションのため

の技術として利用される。ここでは、そのような観点からのメディア技術について簡単に紹介したい。

### 2. マン・マシンコミュニケーションのためのメディア技術

システムや機械のほとんどは人間が制御するものであり、システムが複雑になればなるほど、マン・マシンコミュニケーションが重要になってくる。最近のPCの性能向上のかなりの部分は、ユーザインタフェースの向上に利用されているともいわれているほどであり、使いやすいマン・マシンコミュニケーションへの要求が高まっている。マン・マシンコミュニケーションにとって最も基本的な点は、人間の意図が的確にシステム側に伝わること、そしてシステムの反応が人間に分かりやすい形で提示されることである。従来、情報処理システムにおけるマン・マシンコミュニケーションには主としてキーボードやマウスなどの道具が用いられてきたが、これらは必ずしもコミュニケーションに最適な道具ではない。人間同士のコミュニケーションでは、言葉による言語情報と身振りや表情などの非言語情報が適切に組み合わせられて用いられる。したがって、システムとのコミュニケーションやインタラクションにも言語情報と非言語情報を上手に組み合わせる利用することが有用になってくると考えられる。特に、非言語情報を効果的に利用することが円滑で効率的なマン・マシンコミュニケーションを実現するために極めて重要である。

#### 2.1 非言語情報の利用法——非言語情報の獲得——

非言語情報の利用法についても、人間からの情報獲得と人間への情報提示の二つの側面を考える必要がある。まず、人間から非言語情報をどのように獲得するかという点についてであるが、様々なアプローチで研究が進められている。重要なアプローチの一つは、コンピュータ

谷口倫一郎 正員 九州大学大学院システム情報科学研究所知能システム学部門  
E-mail rin@computer.org  
Rin-ichiro TANIGUCHI, Member (Faculty of Information Science and Electrical Engineering, Kyushu University, Kasuga-shi, 816-8580 Japan).  
電子情報通信学会誌 Vol.88 No.10 pp.799-803 2005年10月

ビジョン（画像認識）の技術を基にした映像情報から身振りや表情などの非言語情報の獲得である。コンピュータビジョンによるアプローチの優れた点は、入力装置としてカメラを用いるだけなので、対象となる人間に装置やマーカーなどを装着せずに非言語情報を獲得することができる（非接触で観測可能な）点であり、人間にシステムをあまり意識させないという点で自然な手法であるといえる。

コンピュータビジョンによるアプローチでは、映像が二次元であるのに対して、観測対象である人間は本来三次元世界の対象であるということに注意しなければならない。厳密に言えば、二次元の情報から三次元の世界を完全に獲得することはできない問題（不良設定問題と呼ばれる）であり、幾つかの付加的な制約条件を設けることによって解を求めることになる。例えば、複数のカメラを用いて別の角度から撮影した映像を用いたり、人間の身体の高さや構造をあらかじめ想定するモデルベースのアプローチを用いたりすることになる。

例えば、図1に示す例では、以下のような手順で人間の全身の姿勢を実時間で推定している<sup>(3)</sup>。

- ① カメラで撮影した身体の画像から、手先、足先、顔といった画像から抽出しやすい特徴を検出する<sup>(注1)</sup>。
- ② 複数のカメラで上記の特徴を検出した上で、それらの三次元位置を三角測量の原理で計測する。
- ③ 身体の形状モデルに基づき、得られた特徴の三次元位置に最もよく整合するような身体姿勢を推定する。

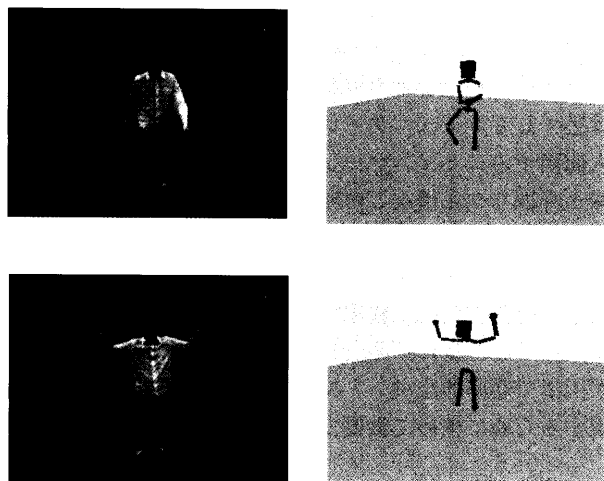


図1 コンピュータビジョン技術による身体姿勢の推定 複数のカメラから得られた画像から画像特徴を抽出し、それらから身体の三次元的な姿勢を推定している。複数のPCを並列処理することにより、実時間処理が実現されている。

(注1) 本稿で紹介したシステムでは、肌色などの特徴的な色領域（ブロップと呼ぶ）を検出することで、身体の特徴を抽出している。

そして、得られた身体姿勢の推定結果に基づいて人間の動作を認識し、仮想空間のナビゲーションや仮想物体の操作などが実現されている。

また、身体動作以外の典型的な非言語情報の獲得としては、視線検出が挙げられる。視線は人間の意図を表す重要な非言語情報であることが知られており、視線検出は、人間の意図推定には極めて重要な技術である。図2に示した例では、顔を撮影した画像から以下の手順で視線検出を実時間で行っている<sup>(4)</sup>。

- ① 虹彩付近の領域を切り出し、画像の明るさに基づいて2値化し、虹彩候補を抽出する。
- ② 抽出した虹彩候補の境界を検出し、楕円の当てはめを行う。
- ③ 得られた楕円の形状から虹彩（円）の三次元空間内での向きが計算できる<sup>(注2)</sup>。この向きが視線方向である。

視線を観測することにより、人間の意図する操作を事

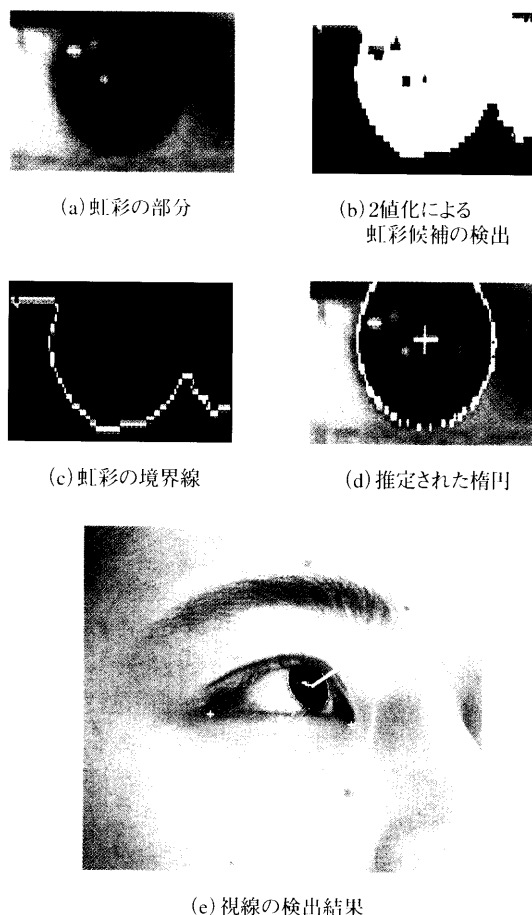


図2 コンピュータビジョン技術による視線の検出 画像から虹彩を楕円として抽出し、楕円の形状（長軸と短軸の長さ、軸の向き）から虹彩の向きを推定し、視線を検出する。

(注2) 本来、虹彩は円であるのに、楕円として画像に写るのは、その円が三次元空間内で傾いているからである。

前に予測したり、迷いといった人間の心理状態を推定したりできるので、システム側から人間の操作を支援するための補助操作や情報提示を行うことができる<sup>(5)</sup>。ここで述べた人間の身体動作や視線の獲得以外にも、手指の動作、顔の表情などを実時間で計測・認識するシステムの開発・研究が精力的に行われており<sup>(6),(7)</sup>、様々なマン・マシンコミュニケーションに利用され始めている。

## 2.2 非言語情報の利用法——非言語情報の提示——

一方、人間への情報提示に関しては、非言語情報を提示することによって、システムと自然なコミュニケーションを実現する方式の研究が進んでいる。例えば、ある情報を言葉でユーザに提示する場合、単にテキストによる言語情報だけで情報を伝えるより、コンピュータグラフィックスで作成した擬人化エージェント（キャラクターやアバタと呼ばれることもある）をディスプレイに表示し、音声情報に合わせて身振りや表情を付加した方が、その内容をよりの確に伝えることができるといわれている。例えば、「大きな」という語を発話するときに、両手を広げて大きい様子を表すような動作を提示することで、「大きい」ということを強調して伝えることができる。

言語情報に非言語情報を自動的に付加するための方式としては、以下のような方式が考えられている<sup>(8)</sup>。

- ① 発話すべきテキストを言語解析し、文節単位に分割する。
- ② 取り出された各文節の役割を識別し、各文節の役割に応じて、適当な非言語情報の表現法を知識ベースから検索する。
- ③ 得られた非言語情報を CG 技術でアニメーションとして生成する。

このような方式で、自然なアニメーションを生成するためには、音声合成による発話のタイミングと非言語情報を表す身体動作の動きが同期するような仕組みも重要であることに注意して欲しい。

非言語情報の提示が有効なのは、単にシステム側から人間に情報を提示する場合だけに限らない。人間が音声言語で情報をシステムに提示する場合も、人間の相手をする擬人化エージェントを提示することが有効であると考えられている。人間の発話に応じて、擬人化エージェントがうなずく、視線を合わせるというような非言語情報を適切に提示すると、人間側は安心して発話を行うことができ、システムと自然なコミュニケーションが可能になる。どのようなタイミングで擬人化エージェントにうなずかせるかというような問題も興味深い問題である<sup>(9)</sup>。

## 2.3 言語情報の利用法

言語情報の利用に関しても、音声言語の認識に加えて、音声から話し手の感情を抽出し利用する試みも進められており<sup>(10)</sup>、既に、ユーザインタフェースに利用することを前提とした感情認識システムも商品化されている。それらの基本的な考え方は、発話するときの感情によって、音声の強弱や高低、発話の速度やリズムなどが異なることを利用するものである。具体的には、音声情報に周波数分析などを施して特徴を抽出し、得られた特徴量から適当なパターン識別器を用いて発話者の感情を分類するといったことが考えられる。話者の感情が推定できれば、それに応じて、システム側も適切な反応をすることができ、より自然で、効率的なコミュニケーションが可能になると期待されている。ただ、感情表現や表情などは音声に限らず個人による違いが大きいため、それらを完全に認識することはなかなか難しい問題である。

また、音声と映像の二つのメディアから得られる情報を統合的に解析するという研究も行われている。音声と映像は相補的な情報を提供する部分があるので、両者の情報を用いることによって、より正確にコミュニケーションの内容を抽出できるという考え方である。例えば、音声認識において、口唇を撮影した画像から口唇の形状や動きを画像処理で検出し、音声情報から得られる特徴とともに、HMM (Hidden Markov Model: 隠れマルコフモデル) を用いて認識するような方式が考えられている<sup>(11)</sup>。雑音のある環境下で、雑音のない音声信号が得られないような状況では特に有効であると考えられている。感情の認識に関しても、音声情報と映像情報を融合して認識するような研究が行われている。この場合は、顔画像から目や口などの主要器官を抽出し、その位置関係を画像からの特徴として用いており、音声から得られた特徴とともにパターン認識器で認識する<sup>(12)</sup>。

## 3. マン・マシンコミュニケーションの関連技術

これまで述べてきたマン・マシンコミュニケーションに関する技術は、情報の獲得と提示を組み合わせることで、システムを媒介とした人間同士のコミュニケーションにも適用することができる (図3 (a))。時間的、地理的に離れた人間同士のコミュニケーションを支援する有効な道具として期待されている<sup>(13)</sup>。また、同様の考え方で、ハンディキャップを持った人を支援する技術としても利用することができる。マン・マシンコミュニケーションの技術を応用すれば、情報の伝達手段を変換することも可能になるので (図3 (b))、ある情報伝達のメディアを利用するのが困難な人との間にコミュニケーションチャネルを確立することができる。典型的な例は、手話の翻訳である。手話によるコミュニケーションでは音声というメディアが利用できないが、身体動作 (特に、

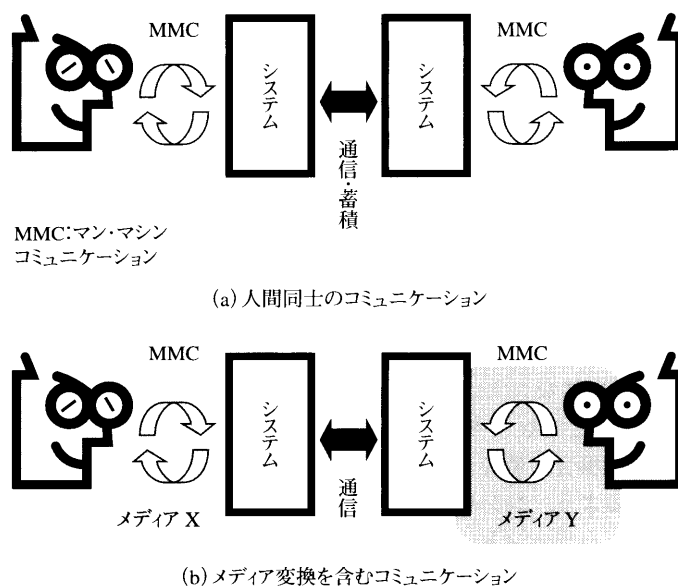


図3 マン・マシンコミュニケーションを基にした人間同士のコミュニケーション (a) マン・マシンコミュニケーションを組み合わせることにより、人間同士のコミュニケーションの道具として利用することができる。これにより、コミュニケーションの地理的、時間的な制約を緩和することが可能になる。(b) 異なるメディアによるマン・マシンコミュニケーションを組み合わせると、メディア変換が実現でき、ハンディキャップを持った人の支援技術として利用することができる。

手指の形状や動き)の解析技術を利用することによって、手話の内容を認識することができる<sup>(14)</sup>。認識した内容を、情報を受け取る側にとって便利なメディア、例えば健常者にとっては音声言語を用いて提示することにより、手話翻訳が構成されることになる。

マン・マシンコミュニケーションのためのメディア技術に共通する課題の一つとして、人間とのコミュニケーションを円滑に行うために実時間処理が必須であるという点が挙げられる。画像や音声などの信号は情報量が多いため、高度な処理をいかに少ない計算コストで実現するかも現実的には重要な課題となってくる。アルゴリズムの開発だけでなく、高速化を実現できるマイクロプロセッサや DSP (Digital Signal Processor: 信号処理用プロセッサ) の開発、それらを組み合わせた並列処理手法の開発などもシステム開発という観点からは重要なテーマである<sup>(15)</sup>。また、動作の予測をうまく利用して、見かけ上の遅延を小さくし、円滑なコミュニケーションができるような仕組みの研究も進められている<sup>(16)</sup>。

#### 4. 終わりに

ここで述べたように、マン・マシンコミュニケーションを円滑に行うために様々なマルチメディア技術の研究開発が進められている。身振りや表情など非言語情報の基本的なものはかなり認識できるようになってきたものの、まだ、どのような状況でも高精度で認識できるような頑健なシステムは完成していない。特に、細かいニュアンスや仕草といったものは、まだ十分に獲得できるわ

けではない。いわゆる「気の利いた」システムが実現され、人間とシステムのコミュニケーションが、人間同士のコミュニケーションと同じようになるには、様々な非言語情報を認識し、総合的に意図を理解しなくてはならず、解決すべき問題が多く残っている。このような知的な情報を獲得するためにメディア処理技術の果たす役割は極めて大きい。

今後、ユーザからは、より実在感のあるコミュニケーションが要求されるようになると思われる。そのためには、視覚、聴覚以外のメディアを有効に利用する技術も必要である。既に、触覚情報や味覚・嗅覚情報はデジタル化できるようになっている<sup>(17)</sup>。これらの装置が手軽に使えるようになれば、より臨場感、実在感のあるコミュニケーションが可能になると期待できる。また、ロボットのような物理的な実体のあるものを利用したコミュニケーションも、実在感のあるコミュニケーションという意味で非常に興味深い問題であり<sup>(18),(19)</sup>、今後の発展が期待される。なお、本稿では誌面の都合上、個々の詳細な技術については省略させて頂いた。興味のある方は文献を御参照頂ければ幸いである。

#### 文 献

- (1) 酒井善則, 吉田俊之, 映像情報符号化 (ヒューマンコミュニケーション工学シリーズ), オーム社, 2001.
- (2) 国枝孝之, 脇田由喜, 高橋 望, MPEG-7と映像検索, CQ 出版社, 2004.
- (3) 伊達直人, 吉松寿人, 有田大作, 谷口倫一郎, “多視点動画画像解析による身体の実時間姿勢推定,” 画像の認識・理解シンポジウム (MIRU2004) 講演論文集, vol.I, pp.678-683, 2004.

- (4) 辻 徳生, 柴田真吾, 長谷川 勉, 倉爪 亮, “視線計測のための LMedS を用いた虹彩検出法,” 画像の認識・理解シンポジウム (MIRU2004) 講演論文集, vol.I, pp.684-689, 2004.
- (5) 崎田健二, 小川原光一, 木村 浩, 池内克史, “視線を用いた人間の意図推定に基づく人間とロボットの柔軟な協調作業,” 画像の認識・理解シンポジウム (MIRU2004) 講演論文集, vol.I, pp.475-480, 2004.
- (6) 浜田康志, 島田伸敬, 白井良明, “遷移ネットワークに基づく多視点画像時系列からの手指形状推定,” 信学論 (D-II), vol.J85-D-II, no.8, pp.1291-1299, Aug. 2002.
- (7) 赤松 茂, “人間とコンピュータによる顔表情の認識 [I] — コミュニケーションにおける表情とコンピュータによるその自動解析 —,” 信学誌, vol.85, no.9, pp.680-685, Sept. 2002.
- (8) 岡本和憲, 中野有紀子, 西田豊明, “台本に基づく会話エージェントのジェスチャ自動生成,” 人工知能学会全国大会講演論文集, no.1E1-02, 2004.
- (9) 角所 考, 伊藤淳子, 美濃導彦, “会話エージェントのためのノンバーバル表現間の相互依存性のモデル化,” 合同エージェントワークショップ & シンポジウム (JAWS2003) 講演論文集, pp.366-367, 2003.
- (10) 伊藤亮介, 駒谷和範, 河原達也, 奥乃 博, “ロボットとの音声対話におけるユーザの心的状況の分析,” 情処学音声言語情報処理研報, no.SLP-45-18, pp.107-112, 2003.
- (11) 吉永智明, 田村哲嗣, 岩野公司, 古井貞熙, “横顔の動画像情報を用いたマルチモーダル音声認識,” 情処学音声言語情報処理研報, no.SLP-46-11, pp.61-66, 2003.
- (12) 松本祥平, 山口 健, 駒谷和範, 尾形哲也, 奥乃 博, “ロボットでの利用を目的とした顔画像情報と音声情報の統合による感情認識,” 日本ロボット学会第 22 回大会, no.3D14, 2004.
- (13) 西田豊明, “人間同士の自然なコミュニケーションを支援する知能メディア技術,” FIT2002, 学術系・企業系予稿集, no.MK-3, pp.24-25, Sept. 2002.
- (14) 松尾英明, 呂 山, 猪木誠二, 今川和幸, 高田雄二, 長嶋祐二, “動作構成要素に基づく非接触手話動作認識方式,” ヒューマンインタフェース学会論文誌, vol.3, no.3, pp.135-144, 2001.
- (15) 有田大作, 花田武彦, 谷口倫一郎, “分散並列計算機による実時間ビジョン,” 情処学論, vol.43, no.SIG11 (CVIM 5), pp.1-10, 2002.
- (16) 内田誠一, 森 明慧, 倉爪 亮, 谷口倫一郎, 長谷川 勉, 追江博昭, “動作の早期認識およびその予測への応用に関する検討,” 信学技報, PRMU2004-94, pp.7-12, Nov. 2004.
- (17) 都甲 潔, 感性バイオセンサー味覚と嗅覚の科学, 朝倉書店, 2001.
- (18) 倉爪 亮, 内田誠一, 谷口倫一郎, 長谷川 勉, “プロアクティブヒューマンインタフェースの研究—第 1 報 人間型アクティブインタフェースの開発—,” 日本機械学会ロボティクス・メカトロニクス講演会 '04 講演論文集, no.1A1-H-76, 2004.
- (19) 小野哲雄, 今井倫太, 石黒 浩, 中津良平, “身体表現を用いた人とロボットの共創対話,” 情処学論, vol.42, no.6, pp.1348-1358, 2001.



谷口 倫一郎 (正員)

昭 53 九大・工・情報卒, 昭 55 同大学院修士課程了. 同年同大学助手, 平 8 同大学院システム情報科学研究科 (現研究院) 教授. 画像処理, コンピュータビジョン, 並列処理等の研究に従事. 昭 62 年度本会篠原記念学術奨励賞, 平 6 年度情報処理学会坂井記念特別賞等を受賞.