シンビオティック・システムの実現に向けて
―人，社会，環境，情報システムの協調系―

7 人間と共生する情報システムの実現を目指して

松山隆司 * tm@i.kyoto-u.ac.jp
川嶋宏彰 * kawashima@i.kyoto-u.ac.jp
鷲見和彦 *,** Sumi.Kazuhiko@ds.MitsubishiElectric.co.jp
* 京都大学情報学研究科 ** 現在，三菱電機（株）先端技術総合研究所

「人間と共生する情報システム」という概念は，平成 13 ～ 17 年度の 5 年間に渡って実施された文部科学省科学研究費補助金特定領域研究「ITの深化の基盤を拓く情報学研究」（領域代表者：安西祐一郎（慶應義塾大学塾長））における A03 柱「人間の情報処理の理解とその応用に関する研究」での研究目標として掲げてきたものである．A03 柱では，「人間と共生する情報システム」の実現を目指して，神経生理学，心理学，マルチメディア情報処理，ロボティクスといった多様な分野の研究者に集まっていただき，学際的な研究を推進してきた．本稿では，研究の背景とねらい，A03 柱の中で筆者の研究室で行われた研究の成果概要を述べる☆1．なお，A03 柱全体における具体的な研究内容や成果については，Web ページ（http://vision.kuee.kyoto-u.ac.jp/Informatics-A03/）を参照していただきたい．

人間と共生する情報システムとは

21 世紀社会は，政治，経済，文化，教育，娯楽などあらゆる個人的，社会的活動の分野において，ネットワーク上でのマルチメディア情報処理システムがその活動基盤を支えるようになると言われている．このことは，我々人類は，生身の人間としてこれまで暮らしてきた「実世界」に加え，情報ネットワークシステム上に作られた「情報ネットワーク社会」においても同時に暮らすことを意味する．すなわち，21 世紀社会における大きな課題は，「我々人類が実世界と情報ネットワーク社会という特性が異なる 2 つの社会において，いかにうまく暮らしていくことができるか」ということになり，現在はまさにその課題解決の路を模索している時代であるといえる．

人間と情報ネットワークシステムとのかかわり合い方としては，以下の 2 つのタイプがある．

環境型： 人間が情報ネットワーク社会の中にメンバとして入り，情報ネットワークシステムが環境として人間の活動支援を行う．ユビキタス環境では，RF タグやセンサネットワークなどによって実世界の中に情報ネットワークシステムが埋め込まれ，実世界での活動がそのまま情報ネットワーク社会での活動となることを目指しているといえる．

パートナー型： 擬人化された情報システムが人間のパートナーとして共に活動を行う．ネットワークロボットは，背後に持つネットワーク上の膨大な情報資源を活用し人間とのコミュニケーションによってその活動を支援する．

我々の提唱している「人間と共生する情報システム」という考え方は，後者の立場に立ったものであり，以下のような議論に基づいている．

図 -1 は，ヒト（生物としての人間）と「もの」（道具や機械）とのかかわり合い方を歴史的経緯に沿って左から右に向けて整理したものである．

まず，「ヒトは道具を使う動物である」という言い方をするように，人間と「もの」とのかかわり合いを考える上で，道具という「もの」は特別な意味を持つと考えられる．

道具は，古来より人間にとって生活を営んでいく上で必要不可欠なものであり，道具によって我々人間は現在

☆1 本稿は，本特集の趣旨を踏まえ，文献 1）を改編したものである．

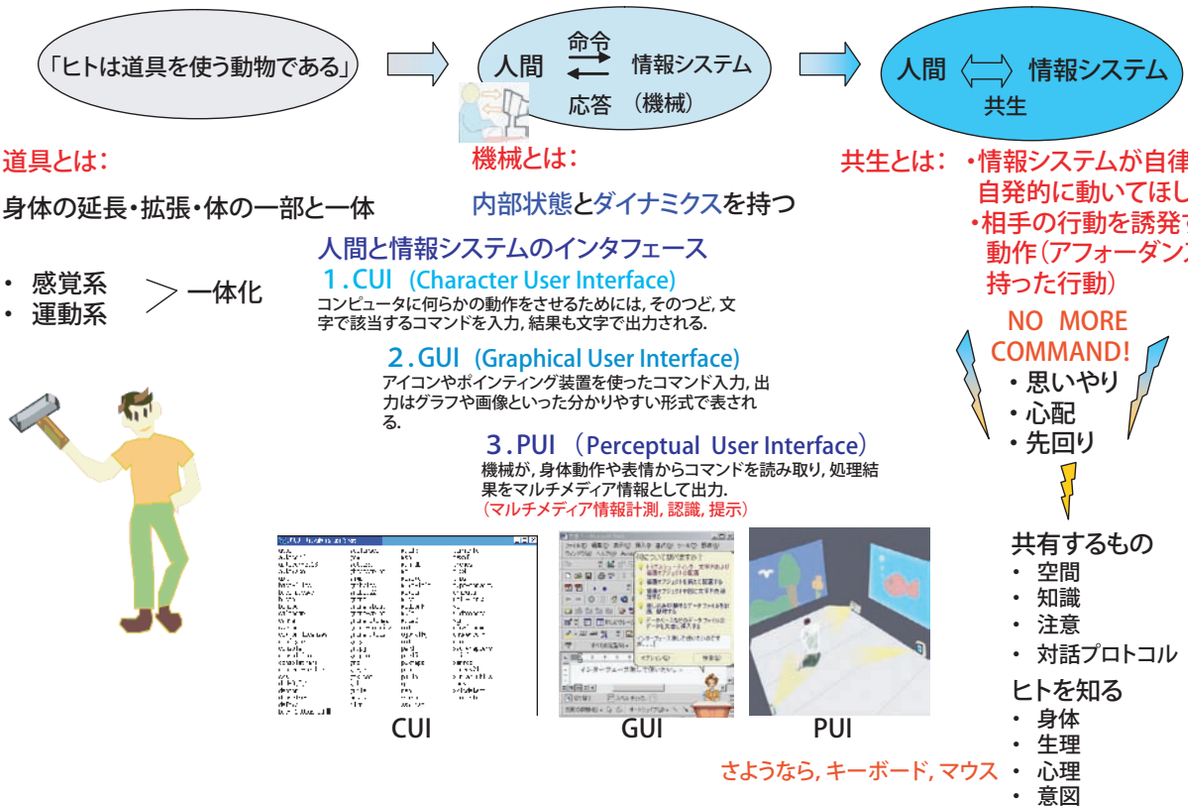

図-1　人間と共生する情報システムの意義・位置づけ

の繁栄を築くことができたといっても過言ではない.

では，ヒトにとって道具とは何か，あるいはヒトは道具とどのようにかかわっているのであろうか．我々は，道具は身体の延長，拡張，体の一部であると考える．すなわち，「もの」である道具がヒトの身体と同化されるというのが，道具を使いこなすということの具体的な姿であるといえる．このことは，ヒトが道具とかかわる際には，特別な仕組みは必要なく，環境の整備と訓練を通じて身体との一体化を図ればよいことを意味する.

次に，人間と「もの」との関係を大きく変化させた出来事として産業革命による自動機械の発明がある．産業革命後飛躍的な進歩を遂げた自動機械は，主体である人間に対して従者であると位置づけられ，人間の命令を受けその活動を支援するためのものとして捉えられてきた.

道具と異なり，自動機械は

- それ自身の固有の内部状態（記憶）
- その動作を定めるダイナミクス

を持ち，それらに基づいて主体からの命令に反応・応答する．このため，人間が自動機械とうまくかかわり合う――「命令と反応・応答」をスムーズに行う――には，

- 現時点での機械の内部状態を知る（同じ命令を与えても状態が異なっていれば異なった反応・応答が返される）
- 機械のダイナミクスを知る（命令を与えると機械が人間とは独立に動くため，機械の将来における動作を予測して命令を与える必要がある）

ことが必要となる．このため，人間と自動機械とのかかわり合い方をスムーズに行うことを目指したヒューマン・マシン・インタフェースの開発が必要となり，さまざまな自動機械を対象としたインタフェース装置――たとえば，スピードメータ，電圧計，流量計などの計器類および，弁やスイッチ，アクセルペダル，ハンドルなどの操作器具類――が作られてきた.

20世紀後半に発明されたコンピュータは，

- それまでの自動機械と比べ，圧倒的に複雑な内部状態を持つため，その状態を把握，理解することが難しい
- プログラムによってその動作特性が動的に変化するため，その動作を予測することが難しい

といった特性があり，人間がコンピュータを使いこなすには，高機能かつ人間にとって分かりやすいインタフェースの開発が不可欠になった.

コンピュータに対するインタフェースの研究は，図-1中央に示したように，キーボード・文字ディスプレイを使ったCUI（Character User Interface）から始まり，

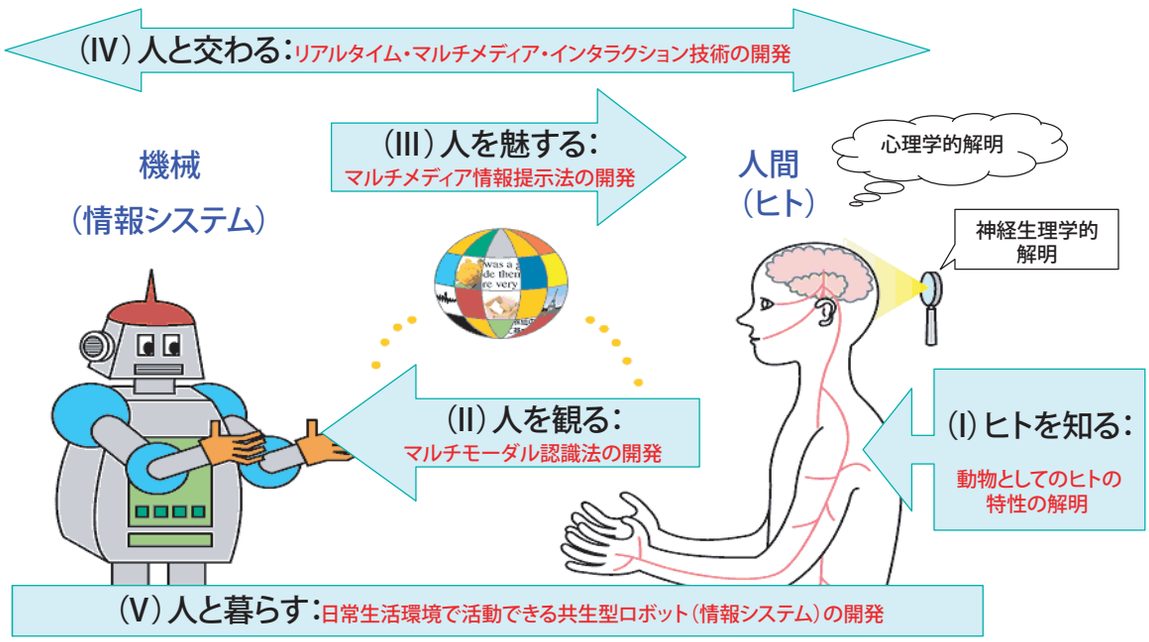

図 -2　人間と共生する情報システム実現のための研究課題

現在ではマウスやマルチウィンドウ・ディスプレイを使った GUI (Graphical User Interface) が主流になっている．また，これからのインタフェースとして，力学メディアなど視覚以外のモダリティも含めた MMUI (Multi-Modal User Interface) や，MMUI に人間のジェスチャや音声を認識する機能を付加した PUI (Perceptual User Interface) が提唱されている．図 -1 中央下端に書いたように，PUI では「さようなら，キーボード，マウス」がスローガンになっている．

しかし，これらのインタフェース研究は，人間と機械とは主従の関係にあるという立場，すなわち，人間からの命令とそれに対する機械の反応・応答というコンセプトに立脚している点ではすべて同一の「命令―応答インタラクション・モデル」であり，その意味では，産業革命以後のヒューマン・マシン・インタフェースの延長線上に位置することには変わりない．

一方今後の社会が情報社会として発展するためには，ネットワーク結合されたコンピュータ群と人間とが一体となってさまざまな活動を行うことが必要であり，そのためには従来の命令―応答モデルを超えた新たなインタラクション・モデルの導入が必要となる．

そうした次世代のインタラクション・モデルとして我々は，「人間と共生する情報システム」というコンセプトを提案してきた．すなわち，従来のインタフェース研究と異なり，日常生活環境において「人間と共生する情報システム」を実現するには，システム自身が主体性，自律性を有し，パートナーとして人間との間で双方向の動的インタラクションを行うというモデルが必要となると考えられる．すなわち，図 -1 右端に示したように，人間と共生する情報システムは，

- 人間からの命令を受けて動くだけでなく，
- システム自身が自律的，能動的に人間の行動や意図を理解し，
- それに基づいて人間への情報提供を proactive に行うといった双方向の(主従関係ではない)動的インタラクションを実時間で行うことが必要である．つまり，人間と共生する情報システムでは，「コマンドよ，さようなら」がスローガンとなる．

また，「人間と共生する情報システム」を実現するには，人間がシステムを見てパートナーとして認知することが重要となり，システムの持つ身体性（物理的な身体形状とその動作機能）が情報処理機能とどのような関係を持てばよいのかを解明することも重要な研究課題となる．

以上述べた問題意識に基づき我々は，日常生活環境において人間と共生する情報システムを実現するための基礎研究として，以下の研究課題に焦点を当てて研究を推進してきた(**図 -2**)．

(I)「ヒトを知る」：ヒトの特性や行動パターンを解明し，得られた知見に基づいて共生のために必要な機能を求める．

(II)「人を観る」：システムがリアルタイムかつ能動的に人間の状態，動作を観測し，その意味・意図を理解

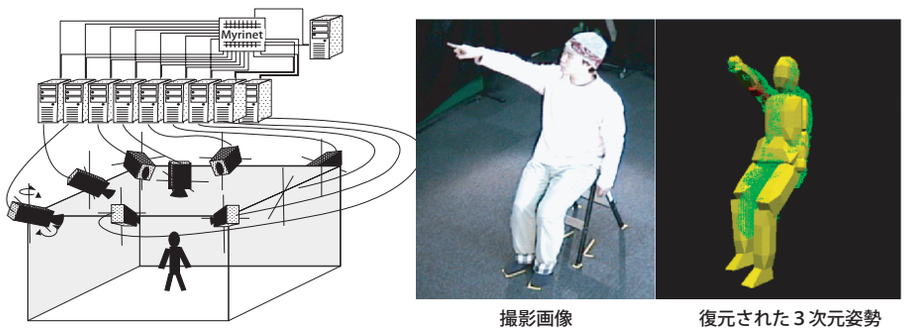

図 -3　環境埋め込み型カメラネットワークと指差し動作の 3 次元解析

する.

(III)「人を魅する」：システムが proactive に人間に働き掛けを行うとともに，人を魅了するような情報提示を行う.

(IV)「人と交わる」：「息の合った」（意識的，無意識的）コミュニケーションを実現する.

(V)「人と暮らす」：人と共に暮らすには，システムが日常生活環境にうまく馴染まなければならない.

以上の考え方に基づいて我々の研究室で行った研究の概要を以下に紹介する．まずはじめに，「ヒトを知る」，「人を観る」の実現を目指して開発した，(1) 環境埋め込み型カメラネットワーク（図 -3）および (2) 装着型能動視覚センサ（図 -5）を用いた研究について述べ，これに続く章では「人と交わる」の実現を目指した (3) 自律ダイナミクスに基づく動的イベントのモデル化（図 -8）について述べる.

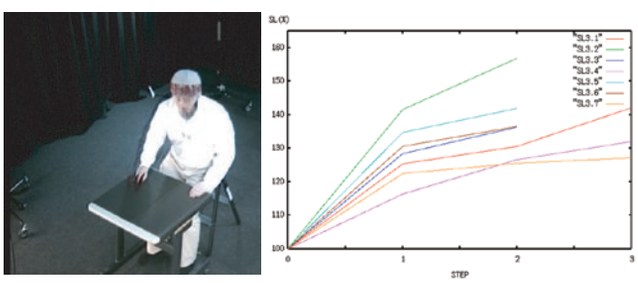

図 -4　作業への集中度の増加と観測される自己負荷量の増分との関係

環境埋め込み型カメラネットワークによる人間の姿勢・動作の 3 次元理解

人間の姿勢や動作からその心的状態を推定し，適切な情報を提示することは，人間と共生する情報システムの 1 つの重要な機能となる．日常生活における自然な（マーカなどをつけることなく）人間の行動を 3 次元的に計測することを目指して開発した環境埋め込み型カメラネットワークを**図 -3**（左）に示す．このシステムは，ネットワーク結合された多数のカメラによって撮影した多視点映像から，人間の詳細な（8mm 程度の空間解像度の）3 次元動作をビデオレートで計測することができ，ゴルフスウィングやダンス・体操の分析，ジェスチャの定量的解析などに利用できる.

研究当初は，人間の指差し動作の特性を 3 次元的に分析し（**図 -3**（右）），右手を使っている場合は，「右目と右肩の中点」と「人差し指の指先」を結ぶ 3 次元直線上に指示対象があることが分かり [2]，その特性を利用して「指コン・ロボット」を開発した.

最近では，人間の姿勢や顔の向きの動的変化パターンから，作業に対する集中度や対象に対する興味の度合いを推定する手法について研究を行っており，その一例を以下で紹介する [3].

まず，人が完全にリラックスした状態を基準に，体の形状や筋肉のモデルを参考にして計算される位置・運動・姿勢保持エネルギーの総和を，その時点において投入されている単位時間あたりのエネルギー（自己負荷量）として定義した.

机に座りながらコンピュータで作業する人の自己負荷量の時間変化を計測した結果，要求された仕事に求められる集中度と計測された自己負荷量の増分との間には強い正の相関が認められ（**図 -4**），提案した自己負荷量が人の集中度の変化を表していることが確認された．なお，自己負荷量の内訳を分析すると，主要な成分は上体の傾きとなっており，人が集中すると作業対象（コンピュータ画面）のほうへ上体が強く引き寄せられることが分かった.

図 -5　装着型能動視覚センサ

図 -7　視体積交差法による手持ち物体の 3 次元再構成

図 -6　装着型能動視覚センサによる注視対象の映像獲得．注視点（赤い点）に応じてパン・チルト・ズームが制御される

装着型能動視覚センサによる状況理解の共有

情報システムが人間と共生するには，人間が見ている状況を同じ視点で捉え，人間と状況理解を共有することが望ましい．

人と同じ視点から対象の映像を記録する手段として装着型カメラがある．我々は，人の視線を計測し，それに応じてカメラを能動制御することによって，解像度と視野のトレードオフを解決しつつ，人が興味を持った対象に関する 3 次元的な情報を得ることを目指して，視線測定装置と 2 台の首振りカメラで構成される（ヘルメット型）装着型能動視覚センサ（**図 -5**）を開発し，以下の機能を実現した．

注視点の 3 次元位置計測

奥行きが場所ごとに大きく変わる日常生活環境でも，装着者の注視している 3 次元位置を正確に検出する手法を考案した．また，計測された装着者の視線情報から，その人が注視しているかどうかを判断し，注視している場合には注視点の奥行き情報をもとに注視対象の詳細な映像をズームアップ撮影することによって，人間の意図が込められた Life Log 映像撮影ができることを示した（**図 -6**）．

手持ち物体の 3 次元ディジタル化

物体を手で持ちながらいろいろな方向から見回している様子を装着型能動視覚センサによって撮影した映像から，物体の 3 次元形状・表面模様を計測する方法として，視体積交差法とステレオ解析を組み合わせた方法を提案した（**図 -7**）[4]．このシステムを使えば，陶器やバッグなどの商品を手で持って眺めるだけで，それらの 3 次元像が自動的に記録され，日常生活の支援に活用できる．

自律ダイナミクスに基づく動的イベントのモデル化

情報システムが人間とスムーズにコミュニケーションするためには，言葉による意味的意思疎通に加え，間合いの取り方といったインタラクションのダイナミクスが重要となる．我々は，インタラクション・ダイナミクスのモデル化と間合いの取れたコミュニケーションの実現を目指して，「自律ダイナミクスを持つ情報システム」という考え方を提案し，以下のような研究を行った．

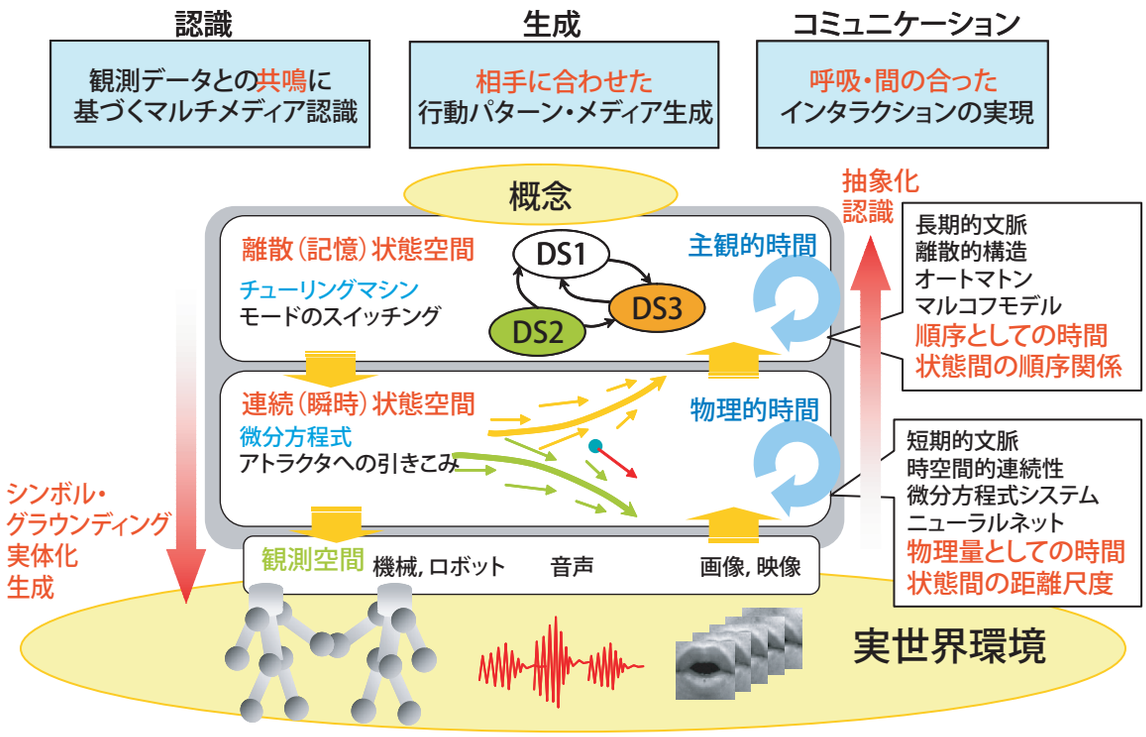

図 -8　自律ダイナミクスを持つ情報システムによる動的イベントの表現・学習・認識・生成

動的イベントの分節化・学習・認識のための Hybrid Dynamical System

人の動きは，身体各部位の運動リズムや，環境との相互作用のタイミングなど，時間軸上での複雑なタイミング構造を持つ．たとえば，手話などでは左右の手の形状と動作の時間的関係によって単語や文の意味が決まる．また，人の行動・動作においては，関節制御のタイミング構造が，あるタスクを達成するための本質的な構造となる．このような構造化された複雑な動的イベントと，それらの間の相互作用を記述・認識・生成するための新たなモデルとして，力学系モデル（距離空間における連続的な状態遷移を記述する微分方程式系）と情報系モデル（順序構造を持つ集合上での離散的状態遷移を記述する記憶書き換え系）を統合したハイブリッド・ダイナミカル・システムを提案した（**図 -8**）[5]．

実世界におけるエネルギーを伴った連続的な変化を表現するには力学系が適しているが，複雑な構造を扱うには適していない．一方で，「口を広げた後に発話する」のように，意味づけられた情報の間の関係（同期関係や順序関係等）を表現するには，情報系が適しているが，あらかじめ記述単位(記号)を定義する必要がある．我々が提案したハイブリッド・ダイナミカル・システムは，複数の線形システムと有限状態オートマトンからなり，オートマトンの１つ１つの状態が，それぞれ異なる線形システムに対応している．このシステムにより，入力された観測データが単一の線形システムで表現できる時区間に分節化され，どのような順序でこれら線形システムが活性化していくかをオートマトンによって表現することによって，複雑なイベントの記述・認識が可能となる．逆に，オートマトンで状態遷移が生じると，各状態に対応する線形システムが活性化され，時系列データが生成される．

こうしたモデルに基づき，まず我々は，ハイブリッド・ダイナミカル・システムの学習法として，大量の時系列データを与えるだけで，データの時間的分節化と線形システムの同定・クラスタリングを同時に行う手法を提案した．これによって，与えられた観測データの記述に適した線形システムおよびオートマトンを自動的に抽出・組織化することが可能となる．

タイミング構造に基づく表情の記述・生成・認識

人はコミュニケーションにおいて，生得的・発達的に獲得した表情の変化パターンを利用することで，自分の心理状態を表出するとともに，表情から相手の心理状態を読み取ることができる．このため，情報システムが，表情から人の心的状態を推定できるようになれば，proactive な情報提供を効果的に行うことができる．

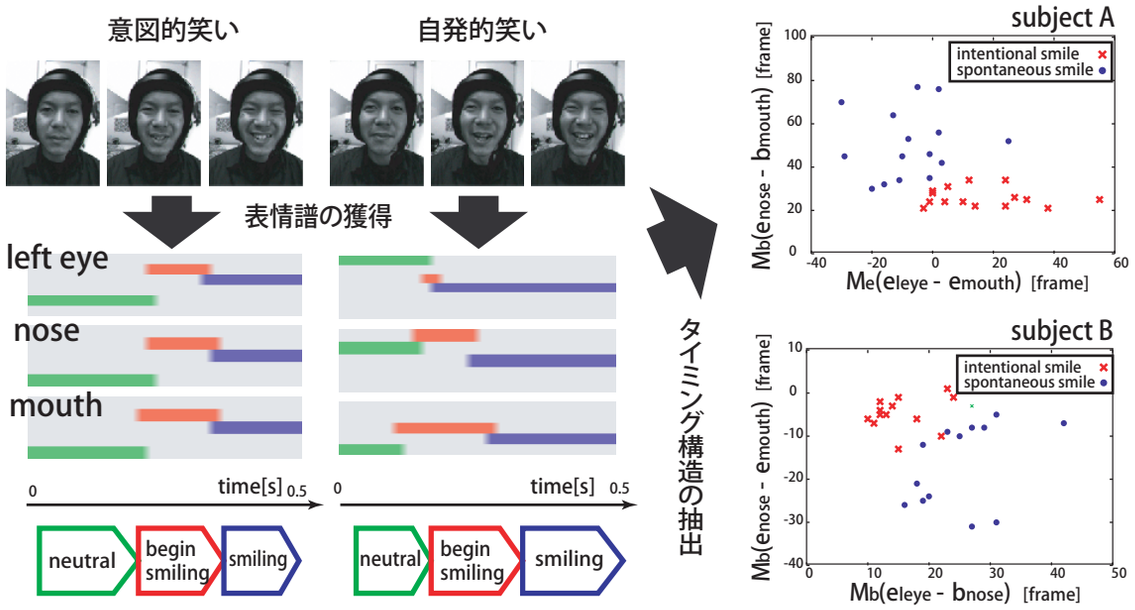

図 -9　タイミング構造の分析による表情の分類

これまでの表情認識・生成研究では，主として感情に基づく基本的カテゴリ（喜び・驚き・恐怖・怒り・嫌悪・悲しみ・軽蔑）に表情を分類することが検討されてきた．しかし，実際の表情は，意図的に制御されて生成されるものや，感情・気分などによって自発的に表出されるものがあり，これらが同時に混ざり合って表情が生まれることも多い．つまり，人は刻々と変化する相手の表情の微妙な動きを観察することで，基本的な感情カテゴリよりもずっと粒度の細かな分類を行い，相手の内部状態およびその変化を推定していると考えられる．

本研究では，表情は，顔の構成要素群（顔パーツ）の運動のタイミング構造によって生じるものと考え，「表情譜」という新たな表情の記述方法を提案した [6]．

先のハイブリッド・ダイナミカル・システムのモデルに基づき，ある線形システムの活性化によって表現される運動の時間範囲を時区間と呼ぶことにすれば，タイミング構造とは，複数の時区間がどのような時間関係で発生し終了するのかといった構造として定義することができる．そこで，各パーツの運動をそれぞれハイブリッド・ダイナミカル・システムによってモデル化することで得られる時区間の集合を，「表情譜」と呼ぶことにする．

表情譜によって表される，口・鼻・目といった顔パーツの運動開始・終了タイミングやその持続時間といったタイミング構造を解析することで，１つの「笑い」という表情であっても，「意図的な笑い」と「自発的な笑い」というより粒度の細かな２タイプに判別・分類できることを実証した（**図 -9**）．

本研究で提案した，タイミング構造に基づく動的イベントのモデル化は，人の動作や視線による意図の推定や，ロボットの行動制御への応用が考えられる．

漫才の動的構造の分析

人間同士の対話は，単に言語的な意味を伝えるだけでなく，意欲を高める，楽しい雰囲気を作るなどといった，より上位の目的・意図を持つことが多く，その伝達には非言語的な情報が大きな役割を担っていると考えられる．たとえば漫才では，冗談を言い出すタイミングにより，客に伝わる面白さがまったく変わってくる．本研究では，複数人の対話を自律ダイナミクスの考え方でモデル化することの利点について基礎的検討を行うために，漫才を例にとり，その動的な対話構造をさまざまな時間スケールで分析した．

分析結果の１つとして，一方の発話終了時刻に対する他方の発話開始時刻（発話移行区間）の傾向を，ボケ役，ツッコミ役ごとに調べたものを**図 -10** に示す．ボケ役の分布の中心がほぼ発話移行区間の長さ０にあるのに対し，ツッコミ役では負，すなわち相手の発話が終了する前に発話を開始するというオーバラップの傾向が強く見られた．ツッコミ役は，話題の切り替えや盛り上げといった対話の流れを制御する役割を持つため，自然な対話を実現するためには，負の発話移行区間が重要な役割を持つのではないかと考えられる．

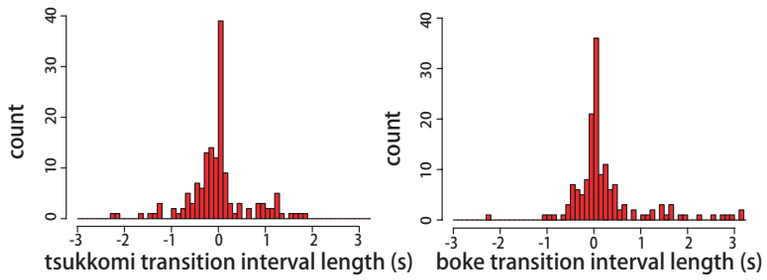

図 -10　役柄ごとの発話移行区間長分布．ツッコミ（左），ボケ（右）

おわりに

我々が提唱している「人間と共生する情報システム」という考え方は，ヒトに関する心理学，生理学的知見に対する情報学的解釈，活用法を開発することが，今後の情報システムでは必要不可欠であるという考え方に基づいており，そうした研究を基盤とする情報学と従来の（人間とは独立に成り立つ）計算機科学・工学との本質的違いを浮かび上がらせていると言える．また，人間と共生する情報システムは，「人にやさしい」，「バリアフリーな」情報社会の構築に大きく寄与するものと考えられ，IT 社会の深化が人間中心の社会の構築を目指したものであることを広く社会にアピールすることができ，社会的にも大きな意義があると考えられる．

「人間と共生する情報システム」という考え方は，「人とは何か」といった哲学的課題および，「IT 技術と人間の暮らし・活動」といった今日的，近未来的課題に対する取り組みとしての側面を持ち，今後も継続的かつ学際的な幅広い研究活動を展開することが必要である．幸い，本年度より 5 年間の計画で開始された「情報爆発時代に向けた新しい IT 基盤技術の研究」（領域代表：喜連川優　東京大学教授）における（新）A03 柱「情報爆発時代におけるヒューマンコミュニケーション基盤」では，コミュニケーションの観点から「人間と共生する情報システム」に関する研究がさらに深められることになっており，その成果が期待される．

謝辞　研究会などを通じて，日頃から議論していただいた A03 柱の班員および研究室のメンバに感謝する．本研究の一部は，科学研究費補助金 13224051 の補助を受けて行った．

参考文献

1）松山隆司，杉本晃宏，佐藤洋一，川嶋宏彰：人間と共生する情報システムの実現を目指して，人工知能学会誌，Vol.19, No.2, pp.257-266 (2004).
2）田中宏一，和田俊和，松山隆司：3 次元人体形状計測に基づく指差し動作の解析，情報処理学会研究報告，CVIM-137-017 (2002).
3）Sumi, K., Tanaka, K. and Matsuyama, T. : Measurement of Human Concentration with Multiple Cameras, Proc. 9th. Int. Conf. on Knowledge-Based Intelligent Information & Engineering Systems, pp.129-135 (2005).
4）Tsukizawa, S., Sumi, K. and Matsuyama, T. : 3-D Digitization of a Hand-held Object with a Active Wearable Vision Sensor, Proc. ECCV Workshop on Human Computer Interaction, pp.129-141 (2004).
5）Kawashima, H. and Matsuyama, T. : Multiphase Learning for an Interval-based Hybrid Dynamical System, IEICE Trans. Fundamentals, Vol.E88-A, No.11, pp.3022-3035 (2005).
6）川嶋宏彰，西山正紘，松山隆司：表情譜：タイミング構造に基づく表情の記述・生成・認識，FIT2005 情報科学技術レターズ，pp.153-156 (2005).

（平成 18 年 6 月 12 日受付）