

ヒューマンインタフェースの進化

竹林 洋一†

音声言語の観点からヒューマンインタフェースの GUI から PUI への進化について述べる。ユビキタス情報環境下ではコンピュータの小型化とネットワーク環境の整備により、広域に分散した多種多様な情報が氾濫することになる。このため、人間支援のための価値の創出に向けて、マルチモーダルとエージェント指向のインタフェース技術の高度化が必須となり、センシング技術とマルチモーダル知識コンテンツ技術の重要性が増す。また、音声言語がユビキタス情報環境下のマルチモーダル知識コンテンツの表現とヒューマンインタフェースの高度化において中心的役割を果たすことを示し、今後の音声言語インタフェースの発展について述べる。

Evolution of Human Interfaces

Yoichi TAKEBAYASHI†

This paper gives an account of the evolution of human interfaces from GUI to PUI, focusing on spoken language processing. Thus far, multimodal human interfaces and interface agents have been studied as a strong candidate for post GUI interfaces which can deal with the flood of information in ubiquitous computing environments. In that sense, sensing technologies and multimodal knowledge content technology, in particular, the technology for spoken language, are very important and valuable. This paper shows that spoken language plays an important role not only in the representation of multimodal knowledge contents but also in human-computer interaction and further points out future research problems concerning spoken language.

1. はじめに

コンピュータや携帯電話の小型化・高機能化と情報通信インフラの急速な整備により、ライフスタイルや仕事の仕組みが大きく変わってきた。デバイス技術の進歩と情報通信機器のプロードバンド化とワイヤレス化の進展は、人間、モノ、金、家庭、教育、病院、スポーツなどの情報化を進め、多様な技術や産業を発展させながら、人間社会を変革していくことは確実である。多数のコンピュータが日常の仕事や身の周りの生活環境に浸透するユビキタスコンピューティングの世界の萌芽期に至ったとも言えよう。

このようなユビキタス情報社会では、いつでも、どこでもネットワークにアクセスできるデジタル情報基盤が前提となる。「箱屋(機器)」や「土管屋(ネットワーク)」などの狭い専門という呪縛にとらわれていると、豊かな価値は創れない。「人間支援のためにコンピューティングパワーを活用する」という「ヒ

ューマンインタフェース研究」的発想、さらには社会インフラ整備、プライバシー確保、美しさの追求、ビジネス開発など、「土建屋」「法律家」「デザイナー」「経営者」的発想までもが必要とされる時代になったのである。

筆者は、音声認識、音声対話、ウェアラブルコンピュータ、ユビキタスデバイス、ナレッジマネジメント、マルチモーダル知識などの研究と実用化に取組んできた。その結果、デバイス、信号からパターン、ナレッジまでを視野に入れ、「人間支援のためのバウンダリーレスなヒューマンインタフェース(HI)研究」を経験することができた[1][2][3]。企業のトップや専門家(通信や音声など特定分野の研究者)に、未踏分野の「価値を創る」研究の重要性を説明する機会が何度もあったが、その本質を理解させることは難しかった。そんな折に、産総研の吉川氏の「第2種の基礎研究」のコンセプトに出会った[4]。筆者の研究姿勢に合致するものであり、心強く感じた次第である。第1種の基礎研究は「普遍的な知識(理論、原理)を発見・解明する従来型の基礎研究」とし、第2種の基礎研究は「応用研究」と

† 静岡大学 情報学部 Faculty of Information,
Shizuoka University

差異化し、「普遍的な知識の選択、融合、適用を繰り返す、試行錯誤しながら知見（価値）を得る基礎研究」と定義した。社会的ニーズと各種技術の進化を洞察して、研究テーマを将来シナリオの中で位置づけ、多様な研究分野や研究者とのインタラクションを繰り返すとのこと。企業の目的基礎研究の定義を明確化したとも言える。

ユビキタス情報環境下の音声やヒューマンインタフェース研究を第2種の基礎研究として推進するには、応用やデモシステム開発だけでは不十分である。経済・社会のニーズや第1種の基礎研究にも関心を持ち、各種知識（ナレッジ）や技術の選択・融合・適用を行い、それらの研究成果を蓄積、共有、発展させ、将来の複数の技術やビジネス開発に連続的に貢献していくことが必要である。

上述したように、ヒューマンインタフェースは人間や社会と複雑に関わっているため、技術の進歩を直線的に論じることはできない。このため本文では、音声言語情報処理の研究者に向けて、ヒューマンインタフェースの多面的かつ多層的な特徴を、筆者の研究事例を用いて検討する。特に、ユビキタス社会で成長が期待されるマルチモーダル知識コンテンツと産業の活性化について詳しく述べ、ヒューマンインタフェースの望ましい進化と音声言語情報処理研究の発展について述べる。

2. GUI から PUI へ

2.1 GUI：机上の仕事を模倣

コンピュータは、過去50年にわたり「紙の文化」を模倣し、計算や事務処理の効率化に使われてきた。コンピュータのヒューマンインタフェースは図1に示すように、計算機室で専門家が特殊なコマンドで操作するCUI(Character User Interface)から、卓上で個人が直感的に操作できるGUI(Graphical User Interface：グーイ)へ、そして、音声言語で気軽に対話できるPUI(Perceptual User Interface：プーイ)へと進化している[5][6][7]。

	Yesterday	Today	Tomorrow
Where	計算機室	デスクトップ	どこでも
Who	エキスパート	個人	グループ
How	Remember & Type	See & Point	Ask & Tell
	CUI	GUI	PUI

図1 ヒューマンインタフェースの進化
CUI,GUI,から PUI へ

現時点で主流のGUIは、初心者にも操作が容易だったので、パソコンや携帯電話の普及を促した。GUIの画面レイアウトと対話（インタラクション）は、人間の認知のメカニズムを考慮してデザインされている。机上の仕事を模倣し、コンピュータの内部状態をユーザに視覚的に提示することにより、直感的かつ直接的なコンピュータ操作を可能とした。

ところが、有線や無線のネットワークが生活の隅々まで浸透し、コンピュータや情報にアクセスする機会が増えてくると、人間（ユーザ）からコンピュータへの意図の伝達は“See and point”の操作では柔軟性と即時性が不十分であることが明らかとなってきた。また、マイク、スピーカ、カメラがコンピュータに装備され、音声認識や画像認識も利用できる環境が整ってきた。このため、コンピュータに処理を任せる比率を高めて、直接操作を極力減らすというエージェント指向のPUIの研究が活発化している。音声言語処理とは無縁であったヒューマンインタフェースの研究者にとっても、音声やエージェントは懐疑的ではあるものの身近な存在となり、音声とヒューマンインタフェースの研究交流が増えてきた。音声言語インタフェースの研究が実際の応用場面で利用できれば、飛躍的に発展すると思われる。

2.2 Usable と Useful

ヒューマンインタフェースの研究は、「使い易い（usable）」と「役に立つ（useful）」という二つの側面を追求している。前者は操作性に関わるものであり、既存の機器システムや製品のGUIなどの操作性を評価改良し、ベンチマークで製品の競争力を強化するという研究である。音声認識の認識性能を競うベンチマーク型の研究はusableの研究煮に該当する。「使い易さ」や「認識性能」の改善はニーズが顕在化しており、研究者にとっては目標が明確で研究コミュニティがでやすく、評価尺度があるので論文も書きやすい。ところが、マクロな視点で考えると、使い勝手や性能競争のベンチマーク型研究だけでは、勝ち組と負け組はできても、新しい産業や市場は創造できない。

これに対して後者の「役に立つ(useful)」という観点でのヒューマンインタフェースの研究は、人間社会の多様な個人の多様な要求に応える「価値」を創る研究である。「人間に役に立つなら何でもあり」という研究が可能であり、成長可能な研究分野である。ユビキタス情報環境下では、各種デバイスやソフトウェアを有機的に統合利用することにより、多種多

様な機器システム、応用、サービスを開発できる。しかし、研究者、研究分野、産業分野ごとに、異なる価値判断基準があるので、広い視野で研究の価値を判断することは難しい。

最近では、多様な入出力デバイス、音声、画像、3DCG、位置情報などが簡単に扱えるようになり、表層的なデモやコンセプト提案の研究発表が増えてきた。「論文のための研究」や「研究のための研究」も見受けられる。社会的ニーズと各種技術を十分理解して、1章で述べた第2種の基礎研究としての取り組みが必要である。各種知識や要素技術の選択、融合、適用を繰り返し、「人間支援のための価値を創るヒューマンインタフェース研究」を推進するような研究風土の実現を目指したい。

音声言語処理の研究者は、多様な研究者とのインタラクションが可能である。それを積極展開し、また、UsableとUsefulを志向することにより、人間支援と経済・社会の発展への貢献が可能である。

2.3 Alan Kay と Dynabook

GUIの生みの親であるAlan Kayは、「人間がメディアを形づくりメディアが人間を形づくる」「未来を予測するベストの方法は、それを発明することだ」という言葉を実践してきた[7]。1970年代後半に、McLuhanのメディア論を4ヶ月かけて読破し、半導体などの技術の進歩を徹底的に考え、人間の創造力を増幅するパーソナルコンピュータ“Dynabook”のコンセプトを示した。コンピュータのシミュレーション能力に着目し、個人の創造力向上のためのメタメディア(Personal Dynamic Media)として、

Dynabookを発案した。その途中段階として、ハイパーテキストなどの各種ソフトとビットマップディスプレイ、マルチウィンドウ、マウスを装備した、ワークステーション“Alto”を開発し、ソフトフォントを用いた電子出版などの画期的アプリケーションも開発した。Altoを開発時点で、GUIの限界を見抜き、ポストGUIの必要性を示唆するなどヒューマンインタフェースの進化に多大な影響を与えてきた。

筆者は1985年にMITメディアラボでAlan Kayのゼミに参加する機会を得て、ヒューマンインタフェース研究の方法論を学んだ。Alan Kayの「性能競争する狩猟民族から、価値を創造する農耕民族へ」という最近の言葉は、音声やヒューマンインタフェ

ース研究のあるべき姿を示している。

2.4 PUI：人間の行動を支援

人間の日常活動で用いる“Ask and tell”型の対話で巨大なデジタル情報空間を自由に行き来できるようなヒューマンインタフェースの実現が望まれている。図2はポストGUIとしてのPUIの概観を示している。ここで重要なのは、氾濫する情報の内容を理解したり、人間の言葉や行動を理解したり、人間の意図や状況を理解し、適切に応対し、人間への負荷を軽減する機能の実現である。様々なデジタルコンテンツの内容と、ユーザの意図や状況を理解するには、コンピュータの「耳」「目」「頭」を良くすることが必須であり、これまでヒューマンインタフェースの研究コミュニティと別の世界で個別メディアや専門領域の研究に専念してきた、音声、ビジョンなどの研究者とHI研究者の連携が必須である。

このような考えでTurkと筆者は1997年にPUIというワークショップを発案した[2]。PUIとは人間の五感を扱うマルチモーダルインタフェースの概念を包含し、人間中心の次世代ヒューマンインタフェースの実現を目指し、GUIに匹敵する研究分野への成長を期待し名付けた。PUIは、音声などのマルチモーダル情報を対象とし、パターン認識や知識処理などの要素技術や、社会心理および認知科学とも関わる発展途上の研究分野であると言える。

図3にPUIで利用する各種メディア変換処理(認識・理解)技術を示す。横軸は、音声、画像などのメディア(センサ情報)を示し、縦軸は、デバイス、信号、記号、意味の処理のレベルを示す。これらの各種メディア情報処理(変換処理)の研究は個別のコミュニティがある。音声や自然言語の研究も種々の専門的な研究分野があり活動は盛んである。

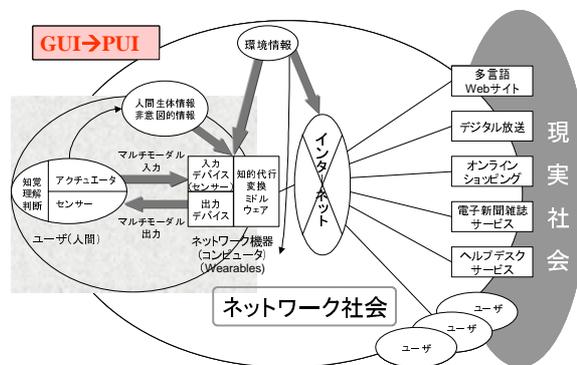


図2 ポストGUIとしてのPUIの概観

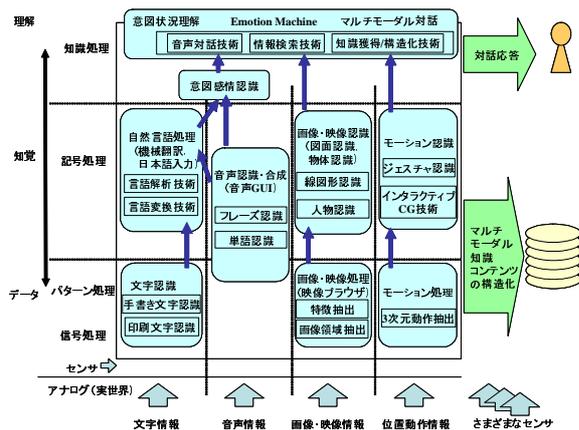


図3 PUIで利用する各種メディア変換処理

ところが、専門という壁に阻まれ、連携は少ない。ところが、ユビキタス情報環境下に様々なセンサのネットワークが構築されてくると、膨大なマルチモーダル情報がネットワークを行き来することになる。各種メディア変換処理により、これらのマルチモーダル情報を圧縮、構造化して、付加価値を付けて蓄積することが可能となる。また、マルチモーダル情報を統合することにより、ユーザの意図状況や感情の理解も可能となる。さらに、複数のセンサとメディア変換処理を組み合わせ、意味・意図のレベルで統合することによりマルチモーダル対話が可能である。

HI 研究者は、GUI や種々の周辺機器やデバイスを活用して、魅力的な研究や役に立つ応用システムを開発できるので、音声認識や AI の有効性に懐疑的である。音声言語に基づく PUI の高度化を加速するには、音声言語研究者のソフトウェアの公開や評価実験を共同で行い[9]、様々な知見やノウハウを蓄積し、頑健で実用性の高い技術を次々と開発できるような研究開発の仕組みを検討すべきである。

3. PUI の高度化に向けて

3.1 インタフェースエージェント

MIT メディアラボを創設した Negroponte は、理想的なヒューマンインタフェースとして、気の利いた英国人の執事 (Butler) をあげている。電話的に確に應對し、訪問者を見分け、主人のためなら嘘をつくことも厭わない。タイミングを心得ていて、個別の事情にも配慮する。そのようなエージェントは、生まれつきのセンスに加えて、膨大な訓練や知識・ノウハウが関係していることは明らかである。PUI や音声言語インタフェースの高度化に向けて、様々

な理論やアルゴリズムに加えて、本格的な常識や知識の研究が必須である。

PUI の深く関係するインタフェースエージェントは、1958 年に LISP を開発中の McCarthy が、知識を探索する "Advice Taker" というプログラムで提案したものである。常識を、「人間、あるいは、機械が、アドバイスの意味を解釈しながら受け取ることのできる知識」と定義した。当時から知識や常識を研究対象にしていたのである。当時の人工知能の研究とヒューマンインタフェース研究との垣根は低く、インタフェースエージェントに関する融合研究が行われていたと言える。

Minsky の「心の社会」では、心のはたらきをモデル化するための基本構成要素を、「エージェント」と定義した。しかし、インタフェースエージェントの定義と区別するため、執筆中の "The Emotion Machine" では、エージェントではなく、リソース (Resource) という用語を使っている。

「2001 年宇宙の旅」に出てくる HAL は、音声認識、音声対話や画像理解の能力があり、高度な知能を有していた。1980 年代末の Apple 社のコンセプトビデオの "Knowledge Navigator" は、擬人化エージェントを介して、音声でスケジュール調整、電子メール、講義の準備などの様々な仕事の代理をさせ、人間の知的能力を増幅することの重要性を示していた。しかし、これらはあくまでも将来ビジョンを示しただけである。エージェントの気の利いた処理や気配りを実現するためには、ユーザや様々な事物に関する広く深い知識や常識が必要であり、本格研究が必須である。

3.2 音声自由対話システム

音声を含むマルチモーダルインタフェースや擬人化エージェントの研究が盛んである。様々な音声対話システムが開発されているが、1992 年に筆者も参画して開発した TOSBURG II は、「不特定ユーザに何ら制約を設けずに動作する」音声自由対話システムである[3]。Knowledge Navigator のインタフェースエージェントを具現化したシステムであるとも言える。

当時の音声認識研究は HMM のモデルの詳細化と語彙数と認識性能のベンチマークに終始していた。これに対して、筆者らは実際の応用場面を想定し、ユーザに何ら制約を与えずに動作するロバストな音声対話システムの実現を目指すこととした。当初は音声関係者に賛同者が全くいなかったが、その後の「雑音」、「不用語」、「単語スポッティング」、「感情」、「対話」、

「マルチモーダル」、「擬人化」が音声やヒューマンインタフェースのキーワードとなり、現在の音声言語インタフェースの先駆けとなった研究だと考えている。文献[1]に開発の経緯を詳しく説明してあるので、ここではその要点を列挙する。

- 不特定ユーザが利用し評価できるようにファーストフードの注文をタスクにした。単語数が少なくても、音声言語の特徴を生かし、実用化の展望が見えるシステムとして選定した。
- ユーザに何も制約を与えずに動作する音声対話システムを開発することとした。実時間処理を実現するため DSP ボードを自前で開発し、発話のタイミングを決めるための感圧フロアマットを導入した。
- 雑音や言い淀み、言い直しなど不用意な発話に対応できるように、雑音免疫学習に基づくワードスポティング法とキーワードに基づく話し言葉の理解などの要素技術を開発し、それらを統合し発話理解を行うことにした
- リアルタイムのユーザ主導型の対話や自由発話を可能とするための、ファーストフード店に向くなどして、自由発話やタスクおよび対話制御を分析し、関連する知識と辞書を整備した
- 将来の本格的実用化を意識し、音声対話システムの構成が発話にどのように影響するかを調べるために、音声対話のデータ収集とコーパスの構築支援システムを開発した。

このようにして開発された TOSBURG II は、音声応答を遮って発話可能な実時間音声自由対話システムである。表情付のマルチモーダル応答、アクティブ音声応答キャンセルなどの要素技術も開発し、「話し言葉」の意味内容理解をベースにし、音声理解の曖昧性はマルチモーダル応答と対話処理で吸収し、実用的で円滑な実時間音声対話を実現した。また、音声対話研究を効率化するために、対話データを簡単に構造化できる GUI を利用した音声対話コーパスも試作した。長期間にわたって実験を行い、多数の人々に音声対話の可能性を示した。その後、不特定話者の発話でタスク達成率が 100%に近い音声対話システムの開発の報告がないのが残念である。

音声自由対話システムの開発経験から、インタフェースエージェントの高度化には、話し言葉や人間の振る舞いに関する知見や知識の活用が必須であることを肌で感じた。しかし、最近の研究でも、言い淀むと助けてくれるコンピュータ[10]のように、音声言語の

本質を捉え、役に立つような応用を考えている研究があるが、音声やエージェント関係の研究者は、このような地道な研究よりも、モデルやアルゴリズムの提案や、社会的側面や身体的な側面の研究を志向しているようである。コンセプト主導やデモ中心の研究に偏っており、実世界の応用を意識し、ロバストな音声理解や音声対話技術の開発、地道な言語辞書や知識ベース構築に関する基礎研究や実世界で役立つ応用開発が必須である。

3.3 ウェアラブルコンピュータ

コンピュータや入出力機器を装着するウェアラブルコンピュータの利点は、音声言語インタフェースを使うことにより、両手が自由となり、スポーツ中や歩行中にコンピュータを迅速に操作できる点にある。コンピュータの応用を拡大するので、産業的かつ社会的意義は大きい。音声言語インタフェースは、ウェアラブルのコア技術であり、音声言語研究者は技術とビジネス動向に関心を持っていただきたい。1998年に MIT と東芝他が共同で開催したファッションショウを通じて、筆者はウェアラブルやユビキタス技術のライフスタイルへの影響を真剣に考える機会を得た。時計、メガネ、オーディオ機器は、ウェアラブル型に進化するのは確実であり、その際の音声言語の役割についてのシナリオを考えるのも一案である。ファッションを経験して、ファッションやデザインは人間の気持ちを左右し、電子産業、情報産業、サービス産業の創出にも寄与できることを実感した。

ウェアラブルコンピュータでは装着型インタフェースが主役となる。コンピュータ本体に代わって、Peripheral (周辺機器) が Central (中心) となり、センサ、入出力デバイス、無線、音声認識、エージェントなどの要素技術開発が進み、医用機器からファッション、サービスまで、新規事業が生まれそうである。

WearableTokyo で試作した「ソムリエエージェント」は、買い物客は、小形パソコンとメガネ型ディスプレイとマイクを装着して、話し言葉でソムリエエージェントと対話し客の嗜好や料理に合ったワインを選択し購入できる。音声による対話は歩きながらも自然であり、必要な時に、必要な知識や情報を素早く取得でき、快適であった。実用化の際には、今回の例のように、両手が自由で音声入力が出来、インタフェースエージェントの知識やノウハウを活用した応用が有力であると思われる。

3.4 コピキタスヘッドセット

音声言語インタフェースの研究はハードウェアやネットワークのことを意識して欲しい。コピキタス情報環境下では、「コンピュータの周辺機器が中心となりつつある」と同時に、ネットワークも基幹ネットワークよりも、末端の毛細血管に相当する人間に近い無線ネットワークが重要となる。近距離無線技術 Bluetooth は、小型・低価格・低消費電力で人に優しいという特徴があり、「人とコンピュータ」、「人と環境」、「人と人」、「atom と bit」をつなぎ、マルチモーダルインタラクションを実現するコア技術として有望である。東芝が製品化したコピキタスヘッドセットは、ハンズフリーで音声認識によるコピキタス機器のコントロールや高音質なオーディオ入出力などを可能とする人間に最も身近なウェアラブル機器である[11]。音声言語を中心とするコピキタス情報環境を実現するためのウェアラブル機器である。パソコン、PDA、オーディオ機器などを音声で操作でき、記憶メディアへの記録や口述音声の文章への書き下ろしも可能である。さらに、博物館や美術館などのオンデマンド・ガイダンスシステムや、健康福祉や教育に関するヘルプデスクシステムへの応用が考えられる。

4. マルチモーダル知識が拓くコピキタスH Iの世界

ヒューマンインタフェースを第2種の基礎研究として進化させるには、社会や経済のニーズに対する洞察力が必要である。ここでは、本格研究の対象となるコピキタスヒューマンインタフェースとマルチモーダル知識コンテンツについて述べる。

4.1 マルチモーダルとナレッジの広がり

図4はマルチモーダル知識コンテンツの広がりを示す概念図である。現在の情報メディア産業が対象としているコンテンツには、音楽、映画、TV番組、新聞、ゲームなどがあり、ビジネスに直結する魅力的なコンテンツをめぐって凄まじい争奪戦を行って

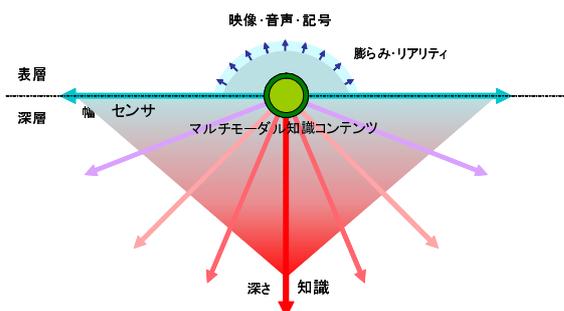


図4 マルチモーダル知識コンテンツの情報の広がり

いる。

映像や音声で構成されるマルチメディアコンテンツはリアリティに溢れ、視聴覚に訴える力が強いがある意味では表層的である。これに対してテキスト、自然言語、知識(ナレッジ)情報は奥が深く含蓄がある。これらの種々のメディアは、それぞれ異なる特徴がある。筆者等はこれらを統合して扱うマルチモーダル知識コンテンツを提案した。映像や音声情報に他のセンサ情報を加えて豊かさや広がりを増大させることができる。さらに、テキストや知識を付加することで、深みを増すわけである。すなわち、従来型の映像、音声、記号のコンテンツにセンサ情報と知識を付加することで、豊かで深みのあるコンテンツに統合させることが可能となる。

4.2 マルチモーダル知識配信システム(MKIDS)

ポスト GUI を担うマルチモーダルインタフェースとして、筆者らは、1991年に音声自由対話システム TOSBURG II[3]を開発し、1997年に知識情報共有を促進するマルチモーダル秘書エージェント KIDS (Knowledge and Information on Demand System) [14]を開発した。これらのマルチモーダル技術をコピキタス時代のビジネス伸張に向け発展させ、2001年に開発したのが MKIDS (Multimodal Knowledge and Information on Demand Service) [12]である。

映像・ナレッジ・Bluetooth 技術を融合し、マルチモーダル知識(MPEG4映像、音声、テキスト)を、オンデマンドでコピキタス機器に配信するシステムである[12]。音声対話でマルチモーダル知識コンテンツをサーバから Bluetooth アクセスポイント経由でコピキタス機器に配信できる初めてのコピキタスシステムである。マルチモーダルナレッジコンテンツは MPEG7 の枠組みを利用して構造化し、音声対話で膨大なナレッジデータベースから適切なマルチモーダル知識コンテンツにアクセスできる。映像/ナレッジを作成できるオーサリングツール(MKIDS オー

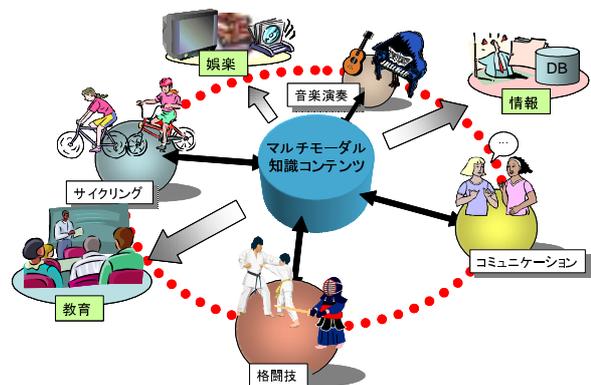


図5 マルチモーダル知識コンテンツを中心とするコピキタスヒューマンインタフェースの研究

サ)は、映像、音声の豊かな表現力のある情報メディアと奥の深い知識を表現できる知識メディアとを簡単に統合し、次世代コンテンツとして制作可能とした。

4.3 マルチモーダル知識コンテンツ

筆者らは「世界(人間・環境)はコンテンツの宝庫」という思想に基づき、マルチモーダル知識コンテンツの文工融合研究に取り組んでいる。個人が、いつでも、どこでも、大量の情報を利用(消費)できるようなユビキタス情報社会では、生活や仕事の様々な場面で大量・多様・複雑な情報やサービスの提供が可能となる。多様なセンサやデバイスが人間・社会・環境に埋め込まれてネットワーク化され、人間が装着するウェアラブル機器とともに、互いに様々な情報をセンシングし、理解し、発信し、コミュニケーションすることが可能となってくる。

ユビキタス情報社会の高度化に向けて、多様なセンサを開発し、センサと Bluetooth を統合した LSI を開発すれば、世界中の環境や人間の身体から様々なセンサ情報を簡単にワイヤレスで伝送できる、センサ情報の流通・編集・蓄積が容易になれば、新しいサービスが創出でき、新しい技術やビジネスモデルが生まれるであろう。人間のリアルな生活のサポートを指向したユビキタスシステムやマルチモーダル知識コンテンツを開発することにより、新産業・新技術の創出が可能である。

図5はマルチモーダル知識コンテンツを中心とする、種々のユビキタスヒューマンインタフェース技術の例を示している。筆者らは、ユビキタスコミュニケーション)、楽器演奏、格闘技、情報教育、英語学習、演劇、義太夫節などの様々な分野を対象に、マルチモーダル知識コンテンツを開発中である[13][15]。表現力の豊かな映像やセンサ情報と、奥の深い知識やテキストから構成される新しいタイプのマルチモーダル知識コンテンツの開発を加速し、応用場面で役立つレベルになるように開発を進めている。

図6は筆者らが検討中の空手の心・技・体の3つのモデルと遍在流空手学習環境を示す[15]。ユビキタス環境下のHI研究において、最近'consciousness'や'awareness','emotion'が注目されている。筆者らは、Minskyの"The Emotion Machine"[16]の感情モデルに基づき、恐怖心や不快感の生成過程についての6階層モデルを検討している。ウェアラブルセンサとユビキタスセンサ情報に

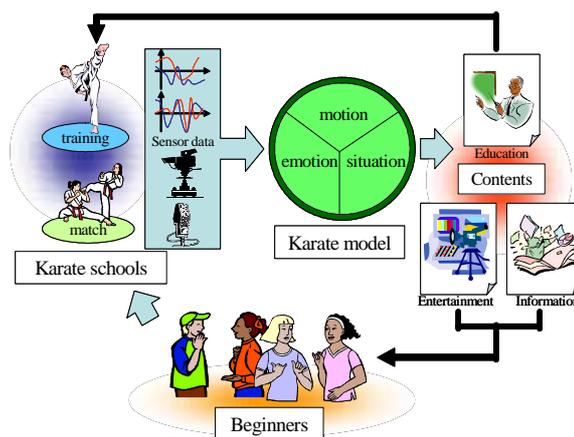


図6 Emotion マシンモデルに基づく心技体を総合した遍在流空手学習環境

対し、信号処理、パターン処理、知識処理を行い、「動作」、「位置」、「間合い」、「心理状態」を抽出する。空手自体が、Emotionの研究にとって格好の題材であり、空手を通じてMITのMinskyと空手におけるEmotionに関する共同研究を進めることになった。各種センシング技術を使った空手上達のためのマルチモーダル知識コンテンツと遍在流空手道場の研究は多様な可能性がある。

図7は、筆者らの研究チームが開発中の各種マルチモーダル知識コンテンツを互いに関連付けることにより、それまで無関係であった知識が図7のように結びつくようになり、マルチモーダルなセマンティックWebが構築できることを示す図である。セマンティックWebの研究が盛んになってきたが、設計や方法論だけでは価値は生まれない。自然言語処理の辞書と文法に相当する知識コンテンツの地道な構築に注力することが大切である。ユビキタス情報環境における豊かな価値の創出を目指してマルチモー

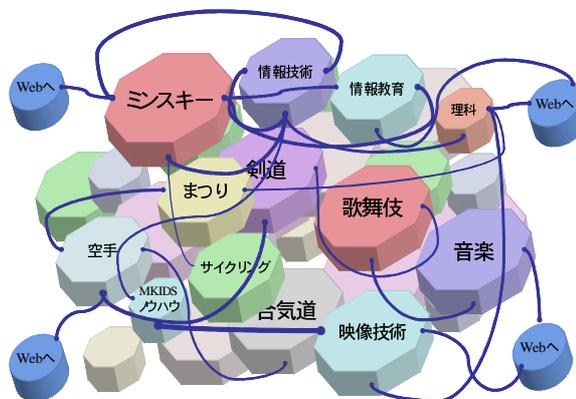


図7 マルチモーダル知識コンテンツのセマンティックWeb化

ダル知識コンテンツを中心とし、デバイス、音声、映像、言語、からナレッジ(知識)感情までを視野に入れて文工融合のPUIの研究に取り組んでいる。

5. むすび

本稿では音声言語の研究者の皆さんに向けて、ヒューマンインタフェース研究の特徴と進化の動向を論じてきた。音声言語の観点でヒューマンインタフェースの今後の研究課題をまとめる。

- コピキタス情報環境の下では、音声や言語の研究者は優位なポジションにいる。社会や産業への貢献を念頭に、「第2種の基礎研究」として音声言語とヒューマンインタフェースの基礎研究をバウンダリーレスに展開すべきである。
- 人間支援型の産業創出を念頭に音声やヒューマンインタフェースの研究を推進し。ビジネス貢献を継続できれば、産業への寄与と音声やHI技術の健全な発展につながる
- ウェアラブルやコピキタスの世界では、音声言語とPUIおよびエージェント技術が重要な役割を担うことになる。音声研究者が、他のセンサ信号処理、パターン処理、認識理解との融合技術を主導すべきである
- コピキタスコンピューティングのキー技術はセンシングと理解。世界はコンテンツの宝庫であり、センサ、信号処理、パターン処理とAI技術で新産業の創出が可能である。

これからコピキタス情報社会の実現は本格化する。音声と自然言語は、人間・社会・環境をコンテンツとして表現し、種々のインタラクションを高度化するためのコア技術である。音声言語処理の研究者が、産業や社会への貢献を視野に入れて、バウンダリーレスに研究を進め、多くの成果が生まれることを期待したい。

最後に、様々な場面でヒューマンインタフェース関連の研究と一緒に取り組んできた東芝と静岡大学の皆様に感謝いたします。また、本稿を執筆する機会を与えていただいた早稲田大学の小林哲則教授と、本稿の内容に関して貴重なコメントを多数頂戴した産総研の後藤真孝氏に感謝いたします。

参考文献

- [1] 竹林洋一: 人間支援のためのバウンダリーレスなAI研究に向けて, Vol.16, No.4, pp.550-559 (2001).
- [2] Yoichi Takebayashi: Integration of Understanding and Synthesis Functions for Multimedia Interfaces,

Multimedia Interface Design, pp.2233-2256, ACM Press Book (1992).

- [3] 竹林洋一: 音声事由対話システム TOSBURG -ユーザ中心のマルチモーダルインタフェースの実現に向けて-, 電子情報通信学会論文誌, Vol.J77-D-, No.8, pp.1417-1428 (1994).
- [4] 吉川他: 新しい産業技術研究方法論 -第2種に基礎研究と本格研究, 産総研技術情報部門 (2003).
- [5] M. Turk and Y. Takebayashi (eds.): Proc. of the Workshop on Perceptual User Interfaces, Banff, Canada (1997).
- [6] 竹林: ヒューマンインタフェースの観点から見た気の利いた情報システム, 信学誌, Vol.82, No.4, pp.310-318 (1999).
- [7] Alan Kay: Personal Dynamic Media, IEEE Computer, pp.31-41 (1977).
- [8] 竹林洋一, 金沢博史: ワードスポッティングによる音声認識における雑音免疫学習, 電子情報通信学会誌, Vol.74-D11, Vol.2, pp.121-129 (1991).
- [9] 連続音声認識コンソーシアムホームページ, <http://www.lang.astem.or.jp/CSRC/>
- [10] 後藤 真孝: 解説 "音声補完: 言い淀むと助けてくれる音声インタフェース", 情報処理(情報処理学会誌), Vol.43, No.11, pp.1210-1216, November (2002).
- [11] 金澤博史, 友田一郎, 高島由彰, 竹林洋一: コピキタス社会に向けた Bluetooth ヘッドセットの開発, 音響学会春期講演論文集, Vol.1, pp.219-220 (2002).
- [12] 竹林洋一, 鈴木優, 岐津俊樹, 浦田耕二, 網淳子, 宮澤隆幸, 金澤博史: コピキタス環境における音声対話システム MKIDS の開発, Vol.1, pp.99-100 (2002).
- [13] 山本剛, 坂根裕, 竹林洋一: コピキタスミーティングからのマルチモーダル知識獲得に関する研究, インタラクション 2003 (2003).
- [14] 中山, 真鍋, 竹林: 知識情報共有システム (Advice/Help on Demand)の開発と実践, 情報処理学会シンポジウム論文集インタラクション'97 (1997).
- [15] 坂根裕, 高島政実, 大谷尚史, 竹林洋一: コピキタスセンシングによる格闘技の身体および心的モデルの検討, 情報処理学会研究報告 2003-UBI-1, pp.65-72 (2003).
- [16] Minsky Home Page, <http://web.media.mit.edu/~minsky>