

コンピュータは巨大なデジタル情報空間の入口的存在となり、世界中のマルチメディア情報を“いつでもどこでも”利用することが可能となってきた。ネットワークを行き来する音声、言語、映像などのマルチメディア情報が人間に届き、内容を理解したとき、はじめて価値が生まれる。現在の GUI (Graphical User Interface) の情報アクセス能力は貧弱なので、ユーザの意図や情報の内容を理解する“Ask and Tell”型のマルチモーダル インタフェースへの移行が必要である。人間は、言語理解、情報検索、問題解決などの知的活動を行っており、これらを人間中心の立場で支援するのがマルチモーダル インタフェースである。ネットワーク対応のヒューマン インタフェース (HI) の高度化により、マルチメディア機器や情報サービスなどを創出していきたい。

With computers serving as the means to access vast amounts of information, various forms of multimedia information, including speech and graphics, are increasingly available anywhere at any time. To enhance the usefulness of this multimedia information, it is necessary to introduce the “ask and tell” type multimodal interface. This type of interface grasps the user’s intention and the contents of the information, enabling it to assist human intellectual activities, such as language understanding, information retrieval and problem solving, from a human-centered point of view.

1 まえがき

オフィスや家庭で使用する情報機器のマルチメディア化とネットワーク化が進み、一般のユーザがインターネットやモバイル コミュニケーションを手軽に利用できる時代となった。コンピュータは仕事や遊びの道具ではなく、巨大なデジタル情報空間の入口的な存在となり、世界中のマルチメディア情報を“いつでもどこでも”利用したり、“だれとでも” コミュニケートできるデジタル情報環境が実現されようとしている。今後出版、放送、広告などの融合が加速され、情報サービス、教育、エンタテインメント関連の新しい産業や文化が生まれるであろう。“人間や社会のために何を支援するのか”という社会学的側面が技術開発に求められる時代になったと言えよう。人間 (ユーザ) 中心の視点で“使いやすさ、満足感、豊かさ”を追求する HI 技術への関心が世界的に高まっている理由はここにある。

HI は、人間と機器との“界面としてのインタフェース”のことでありと誤解されがちである。HI 技術は、人間と機器システムのインタラクション (対話/相互作用) を対象とするが、インタフェースを人間工学的に設計評価するだけではなく、マルチメディア機器やコミュニケーションシステムなどを創造する分野でもある。システム開発者とユーザの視点とを合わせもち、応用/ニーズ指向で機器システムやデバイスの開発を旨とする点が特徴である。このため、音声、画像、言語などのメディア情報処理、入出力デバイス、認知科学などの広範な技術との連携が必須(す)である。

ここでは、マルチモーダル インタフェースを中心に HI の技術動向について述べる。

2 HI の進化

コンピュータと HI の進歩に Alan Kay は大きな足跡を残した。1970 年代後半に Alan Kay は、コンピュータが“メタメディア (Personal Dynamic Media)”であるという点に着目し、ノートブックサイズのコンピュータ Dynabook (ダイナブック) のコンセプトを提唱した⁽¹⁾。“デジタルの世界”では、音声認識、翻訳、画像理解などのメディア理解や、音声、言語、画像などのメディアの編集や検索が可能であると考え、人間の創作活動を支援するためのメディア (技術) として Dynabook の構想を示した。また、GUI とビットマップディスプレイを装備した対話型コンピュータを開発した。さらに、コンピュータネットワーク環境を構築し、先端要素技術を統合して電子出版、教育、アニメーション関連の



図1. HI の進化 コマンド言語入力方式の第一世代から GUI 方式の第二世代、そしてマルチモーダル インタフェースの第三世代の HI へと進化していく。

Evolution of human interfaces

魅力的な応用ソフトも開発した。

HI は図 1 のように進化している。第一世代のインタフェースは、“Remember and Type” 型であり、キーボードから文字列で指示を与えるコマンド言語入力方式である。専門家がコンピュータを独占していた時代に作られた。複雑なコマンドを覚える必要があるのが初心者向きではないが、効率的なのでソフトウェア開発者を中心に現在も根強く使われている。

第二世代は、現在の“See and Point” 型の GUI であり、机上の作業環境を模擬したデスクトップメタファ（隠喩）に基づいている。マウスの直接操作でメニューやアイコンを選択するだけで、コマンド入力と同等のことが行えるという利点がある。しかし、GUI の直接操作では膨大なデジタル情報を迅速に検索できないことが明らかとなってきた。

Alan Kay は 70 年代後半に、すでに GUI の限界に気づき、大規模ネットワーク時代には、ユーザの意図や情報の内容を理解できる知的なエージェント（Agent）技術が必須となることを予見していた。

第三世代の HI は“Ask and Tell” 型のマルチモーダルインタフェースである⁽²⁾。マルチモーダルインタフェースは、視覚、聴覚、触覚、言語などの複数の感覚器（センサ）や認知処理様式（モダリティ）を考慮してメディア情報を処理し、コンピュータと人間の高度な対話（インタラクション）を目ざすものである。ネットワーク環境下でコンピュータと自然に対話できるポスト GUI として期待されている。エージェントを介したマルチモーダル HI は音声やジェスチャ理解、意図状況理解、知的対話処理が必要であるが、最近の音声入出力や三次元グラフィックスの性能向上は目覚しく、第三世代の HI の萌（ほう）芽期に入ったと考えられる。

3 マルチモーダルと HI

3.1 メディア変換とメディア理解

人間どうしのコミュニケーションでは、送信した音声や言語などの情報が、受け手の人間に届き、情報の内容を理解したり感動したとき、はじめて価値が生まれる。つまりメディアの価値は、単なる情報伝達ではなく理解が重要であり、自然で快適な HI を実現するにはメディア理解の高度化が必須である。

人間が言葉や文書を認識し理解するという事は、膨大なパターン情報を、情報量の少ない意味情報（表現）に変換することである。集約された意味表現により複雑高度な知的処理が可能となる。しかし、メディアの変換エラーやあいまい性は本質的な問題があり、快適なインタラクションの実現には、頑健な認識理解技術と認識エラーやあいまい性への対処が欠かせない。

各種メディア変換処理と認識理解の技術の関係を図 2 に

示す。横軸は、文字、音声、画像・映像、位置動作情報などのメディアの種類を表す。一方の縦軸は、信号レベル、記号レベル、意味・意図レベルの処理の深さを表す。複数のメディア変換処理技術を組み合わせて、高度なマルチモーダル対話が実現できるわけである。

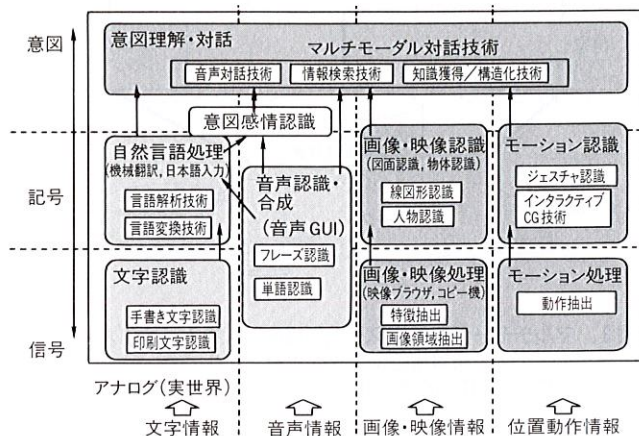


図 2. メディア変換処理技術と認識理解技術の関係 複数のメディア変換処理技術を組み合わせることによりマルチモーダル対話が実現できる。

Relationship between media conversion technologies and recognition/understanding technologies

3.2 マルチメディア情報の構造化とマルチモーダル対話

インターネット/イントラネットを介して送られてくるフロー情報と組織や個人が保有するストック情報が飛躍的に増大し、情報洪水の問題が深刻化してきた。また、音声や映像データもネットワーク上を行き来するようになり、マルチメディア情報が氾（はん）濫してきた。

図 3 は、マルチメディアネットワーク環境でのマルチモーダルインタフェースの位置づけを示す。ユーザとコンピュータとの検索対話インタフェースとマルチメディア/マルチモーダル情報の構造化（インデックス付け）インタフェースが両輪となっている。メディア理解やメディア変換処理は氾濫するマルチメディア情報を収集構造化して整理する際にも有用である。

多様なマルチメディア情報から役にたつものを収集し、内容を理解して構造化し、知識データベースに蓄積すると、意味内容レベルでの検索や再利用がしやすくなる。映像メディアの場合、シーンの変わり目や特定の人物が映っているシーンを検出して構造化すれば、映像データに付加価値が付く。

検索対象となるマルチメディア情報と対応させて、種々のセンサ情報をマルチモーダル情報として取り込み、構造化することで、“いつ”、“どこで”、“だれか”といった使用状況や動作環境に関する情報を抽出でき、自然なコンピュ

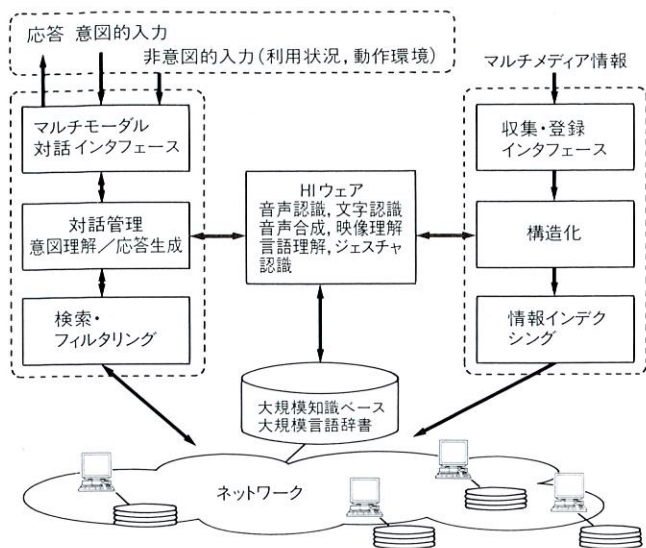


図3. マルチモーダルHIシステムの構成 検索対話インタフェースとマルチメディア情報構造化技術が両輪となり、マルチメディアネットワーク環境のHIを高度化する。

Multimodal HI system

ータとの対話に活用できる。例えば、“先月のはじめに電車の中で読んだヒューマンインタフェースに関する記事”といった、状況や内容レベルでのマルチメディア情報の検索が可能となるわけである。

4 HI技術の応用と実用化

当社は、HIコア技術開発に横ぐしを通し、製品への適用を促進するために、95年にHI技術センターを設立した。パーソナル用から、家庭、産業、公共用のさまざまな機器システムにユーザ指向/ニーズ指向で付加価値を提供するために研究開発に取り組んでいる。

図4に当社のHIコア技術と製品への適用を示す。HIコア技術として、使いやすさを追求する人間工学設計技術、大量のデータをデジタル化し構造化するための文書画像理解技術、マルチメディアの中心となる言語メディア処理技術を取り上げ、その高度化に注力している。

文書画像理解技術は、郵便番号の自動読取り機、ドキュメントリーダー、帳票読取りシステムに応用している。ネットワーク環境で流通する大規模マルチメディア情報の構造化の重要性は高い。

言語処理技術は、日本語ワープロのかな漢字変換、インターネット翻訳、文書検索、情報フィルタリング、知識情報共有システム、ヘルプデスク、音声合成などの高度化に必須である。

インタフェースの設計評価に関しては、携帯機器のGUI設計やATM(無人預出金機)のユーザインタフェース設計

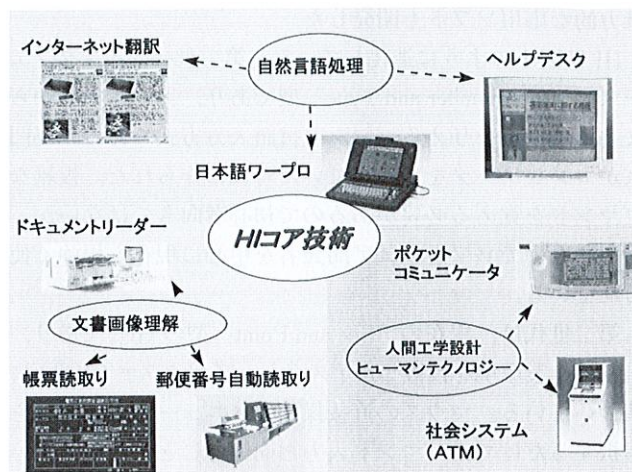


図4. HIコア技術の製品への適用 人間工学設計技術、文書画像理解技術、言語メディア処理技術が種々の製品に適用される。

Application of HI core technologies to Toshiba products

のガイドライン策定やノウハウの共有を図り、全社関連部門が連携してユーザに価値と満足感を提供できるように研究開発を進めている。

マルチモーダルインタフェース実現のため各種メディア変換処理を容易に利用できるHIウェア(コモンインタフェースサービス環境)の枠組を開発した⁽³⁾。メディア変換処理の高性能化と処理機能の拡充を進めている。

5 あとがき

ネットワーク上を行き来する映像、音声、言語などの情報は、人間に届き、情報の内容を理解したり感動したとき、はじめて価値が生まれる。ネットワーク環境下における知識情報の共有やコミュニケーションの円滑化など、マルチモーダルインタフェースの活躍の場は多い。ユーザ指向/ニーズ指向で種々の先端メディア変換処理技術やデバイス技術などを統合し、デジタル機器や情報サービスの研究開発を進めていきたい。

文献

- (1) A. Kay: IEEE COMPUTER, pp.31-41 (1977)
- (2) 竹林洋一: 電子情報通信学会論文誌, J77-D-II, 8, pp.1417-1428 (1994)
- (3) 辻本修一, 他: 情報処理学会HI研究会, 72-3, pp.13-18 (1997)



竹林 洋一 Yoichi Takebayashi, D.Eng.

研究開発センター 情報・通信システム研究所 HI技術センター長, 工博。
HI技術の研究に従事。情報処理学会会員。
Communication & Information Systems Research Labs.