

知的マルチモーダルユーザインタフェース を実現するための基本ソフトウェアの構成

難波 康晴[†] 青島 弘和[†] 堀 浩一^{††} 辻 洋[†] 絹川 博之[†]

[†] (株)日立製作所 システム開発研究所
^{††} 東京大学 先端科学技術研究センター

入出力手段の柔軟性やデータの意味の取り扱いを狙った知的なマルチモーダルなユーザインタフェース(UI)処理においては、実用化に向けて次の課題(1)~(3)がある。すなわち、(1)処理を分散・統合させる有効な仕組み、(2)UI処理の即応性を高めること、(3)動的に状況判断可能な対話的なシステムであること。この課題を解決すべく、マルチモーダルデータ処理のための分散アーキテクチャ、多重反射型インタフェース処理、および、両方向推論による状況依存型意味解析方法を提案し、実装した。個人情報環境上で実験したところ、同時入力によるマルチモーダルデータを補正したり、融合の意味を解析し知的な応答ができることを確認した。

Basic software for an intelligent multimodal user interface

Y. Namba[†] H. Aoshima[†] K. Hori^{††} H. Tsuji[†] H. Kinukawa[†]

[†] Systems Development Laboratory, Hitachi, Ltd.
^{††} Research Center for Advanced Science and Technology, The University of Tokyo

In order to embed an intelligent multimodal user interface(IMMoUI) to real systems, the following three problems should be overcome: an efficient architecture of distribution and synthesis for multiple multimodal data processing, high response in total system on run-time, and an interactive system which is able to understand the situation of the systems' status and users' status. We proposed the following three resolvents respectively: an agent-based architecture using a drip-drop model, a multiple reflex response mechanism, and a situation dependent semantic analyzer with bidirectional reasoning. This paper describes the problems, the resolvents, a configuration of the IMMoUI and several experiences.

1. はじめに

現実世界における人間同士のコミュニケーションでは、五感、言葉、そしてジェスチャなどのコミュニケーション手段を半ば無意識的に用いつつ、意味や意図を相手と授受している。計算機が関与する世界でも、究極のコミュニケーションの姿はこのような情報の授受形式を備えることで、計算機の狭い入出力バンド幅を意識することなく、また、計算機操作自体の難しさを廃した「人間中心型のユーザインタフェース(UI)」となるであろう。我々の研究の究極の目標はこのような人間中心型UIの実現ではあるものの、その実現は容易ではないと考えている。少なくとも、現在普及している極めて計算機中心型UIには、以下の2つの問題点がある。すなわち、

(A) 計算機との入出力手段の柔軟性の欠如

ユーザはプログラムの機能や仕様に従った入出力手段を押し付けられている。そのため、プログラム設計者の意図した利用方法や操作手順が、ユーザの直感、既得知識、使用している言葉の語順に食い違いがあると、馴染みにくく、覚えにくく、忘れやすくなり、使い勝手が悪くなる。

(B) データの持つ意味の取り扱い

人がコミュニケーションに使うデータには、通常「意味」があるのが普通であるにも関わらず、これまでのマルチメディア技術はスピード/ハイカラー/高圧縮率などを目指しており、扱われているデータ自体が持つユーザにとっての価値や役割の取り扱いは軽視されてきた。そのため、ユーザの命令や意図の計算機への伝達はあまり効率的ではなかった。

問題点(A)を改善する方法の一つとして、人が通常用いる音声、身振り、手書き、視線などのコミュニケーション手段を計算機インタフェースにも用いるマルチモーダルインタフェース(MMI)がある[1]。一方、問題点(B)を解決するためには、古くから自然言語処理の分野において研究されてきた意味や意図の把握に加え、機械学習やエージェント技術と組み合わせたUIの研究も始まり、これらは知的UIという潮流を形成しつつある[2]。

我々は、人間中心型UIの実現へ第一歩として、MMIと知的UIの長所を組み合わせた知的マルチモーダルユーザインタフェース(IMMoUI)を開発している。IMMoUIは、ユーザの都合によって実時間的には非同期的に入力される複数のモダリティのデータの相互関連性を解析し、タスクの進捗状況に応じて

意味や意図を把握し、出力メディアに適切な時間配分を行いつつ応答する計算機インタフェースである。このようなコンセプトは、これまでも研究されてきた[3][4][5][6][7]。ただ、依然として実用的なIMMoUIを実現する上で解決すべき次の(1)～(3)の課題が残されている。すなわち、

(1) 処理の分散と統合の仕組み

モダリティに特有なデータの性質や粒度に応じて処理可能な単位に分担しつつ、それらの処理結果をより高度な抽象レベルで複合できるアーキテクチャであること。

(2) UI処理としての即応性

処理の重い認識・合成処理を行いつつも、UIの処理として許される程度に素早く応答すること。

(3) 動的な状況判断

状況に依存する知識、アプリケーションに依存する知識、対話の流れに依存する知識などを用いて、高度に連携した知的な対話システムを可能とすること。

以下、本論文では、以上の問題点を解決するIMMoUIの設計方針(第2章)、その設計方針に基づく基本ソフトウェアの構成(第3章)、その実装を適用した個人情報環境での動作実験例と考察(第4章)について述べる。

2. 知的MMUIの設計方針

第1章で述べた解決課題に対し、我々は、それぞれ以下の方針で解決する。

(1) 処理の分散と統合の仕組みに対する設計方針

非同期到着データの統合を扱うのに好適なドリップドロップモデル[10]をベースに、解析時には各モダリティに依存する処理から依存しない処理へ(生成時は、その逆方向)への役割や機能を段階的に分担させつつ、各サブ処理が並行動作的／相互作用的に処理する分散アーキテクチャを提案する。(第2.1節参照) このようなアプローチに類するものとしてマルチモーダル処理をエージェントベースアーキテクチャで実現するOpen Agent Architecture (OAA)[8]が提案されているが、我々の提案する方式は、各構成要素の役割分担に関する指針を明確にすることで、黑板サーバなしでも、制御や優先権のランク付けや、知識処理の実行のスケジューリングを行うことが可能な仕組みである点異なる。

(2) UI処理としての即応性に対する設計方針

認識処理や合成処理の中には、UI処理として十分な即応性を満たないものがあり、事実上、(1)で考慮する分散処理だけでは対処しきれない。そこで、さらなる即応性を実現するために、継続作動処理、および、多重反射型インタフェース処理を導入する。継続作動処理とは、次々と非同期に到着するデータの受信割り込みを認め、それまでの解析処理の経過や結果と適切に混ぜて、あらたな結果を再構成する処

理方式である。多重反射型インタフェース処理とは、人間の反射(脊髄反射、大脳反射、…)をモデルにしたもので、軽い処理だけの結合で素早く応答する系と、重い処理を含めてもしっかり計算する応答系をハイブリッド的に結合することで即応性を高めている。(第2.2節参照)

(3) 動的な状況判断に対する設計方針

システムおよび対話の状況は実時間的に常に変化しており、前向き推論だけによる意味解析処理は爆発的な計算量を要する。そこで、まず、(A)状況依存型意味解析を行い[9]、(B)前向き推論で発生した未解決な十分条件のみ後向き探索を行い、停止条件としてシステム状態依存知識を用いる意味解析処理とする。(第2.3節参照)

2.1 マルチモーダル処理のための分散アーキテクチャ

非同期到着のデータの統合を扱う方法は、ドリップドロップモデル(DDM)をベースとする。DDMとは、マルチモーダルなデータ(以下、「MMデータ」と略す)を「雨粒」または「雨垂れ」で、それを抽象レベルに応じて扱う仕組みを「網」で喻えたモデルである[10]。そこでは、非同期的に到着するMMデータ「雨粒」は、その抽象レベル(雨粒の粒度)に応じた網に捕えられ、同一レベルの網の上で結合または分裂して新たな抽象レベルのMMデータ「雨垂れ」となって次の網に向けて落ちていく。抽象レベルに相当する処理単位が分散配置し、抽象レベル間処理順序を設定する。予めデータ処理の順序がある程度判明している処理系に対しては、抽象化階層構造を持つ黒板モデルよりも強く処理順序付けされている分、大域的データが減り通信効率や不必要な処理の起動の点で有利である。そのため、我々は、専用に通信制御を行うデータ通信支援プログラムを開発した。このプログラムの特徴は、

(1) マルチプラットフォーム対応非同期通信

TCP/IPのtcpプロトコルを用いて通信路を確保する通信デーモンをマルチプラットフォーム(Windows95およびUnix(SVR4))に実装し、ウィンドウイベントまたはシグナルによってMMデータの非同期到着が知らされ、通信デーモンへポーリングをかけることによって、非同期な通信を実現している。

(2) MMデータのパッケージ化

統一したデータ通信処理を行うためには、多様なモダリティに非依存なフォーマットで、MMデータを記述する必要がある。そのため、マルチモーダル特有属性[10]をパラメータ化し、抽象レベル内部で固有に発生する処理をパッケージ化する。「パッケージ化」を実現する手法は、DDMにおけるドロップ時にMMデータを出力する構成要素(=抽象レイヤ)が、次のレイヤに対するパッケージング(MMデータ自体をコンテンツとして包み込むようなMMデータを作成)

を行い、逆に、ドリップ時にMMデータを受け取る構成要素が外側のパッケージを1つ外し、内側のコンテナを本来のMMデータとして利用するのである。

2.2 多重反射型インタフェース処理

一般に高機能的な処理を行おうとすればするほど処理時間がかかる。例えば、インクモードで単純に応答表示を行うことはすばやいが、認識や解析処理を通すと時間がかかる。一方、実用的なインタフェース処理として十分なレスポンス性能を確保する必要がある。これらはトレード・オフである。そこで、「即答可能な応答経路で取り換えず応答し、遅い処理結果がより適切であると判断する場合は、先行する応答結果をキャンセルし、更めて応答することにする。」という多重反射応答処理モデルを提案する。すなわち、(1)まず、インタフェース処理を、データ入力ステージ、データ認識ステージ、意味解析ステージ、目的推定ステージ、戦略決定ステージ、データ合成ステージという7ステージ構成とする(図1)。これはNormanの人間の認知メカニズムは7ステージ・モデル[11]に似ているが、我々ののは、計算機が(計算機にとっては外界となる)世界と干渉するためのものであるところが異なる。

(2)次に、ユーザとシステムがデータをやり取りするループの中で、システム側が応答のためのデータパスを4つ用意する。すなわち、入力された素データを直接的に出力系に反映させる α パス、個別の認識系から直接的に合成系に行く β パス、認識系から意味解析系を経由して合成系に行くという γ パス、認識系・意味解析系・目的推定系・戦略決定系・合成系とすべてのパスを通過する δ パスである。応答処理時間の点では、 $\alpha > \beta > \gamma > \delta$ の順に優れ、適切な応答内容の点では、 $\delta > \gamma > \beta > \alpha$ の順に優れる。

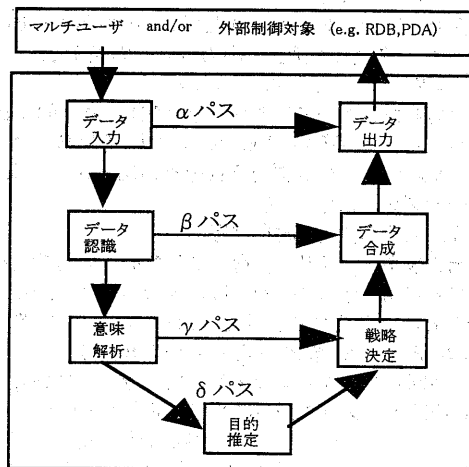


図1 多重反射経路 (αパス～δパス)

2.3 両方向推論を備えた状況依存型意味解析 意味解析の推論方式について説明する。

(1) イベント駆動とゴール駆動の併用

本実装方式による推論の駆動は、(a)ユーザからの入力によるイベント駆動と、(b)それまでの目的解析処理からの得られている目的に関するノードの発火によるゴール駆動である。

(2) 前向き推論と後向き推論の併用

(a)のイベント駆動に対して前向推論(Forward Reasoning)を行い、(b)のゴール駆動に対して後向推論(Backward Reasoning)を行うことで、両者の合流経路見つけるといった両方向推論を行う。ただし、ユーザからの入力イベントは(本実装におけるゴールの記述よりも)具象性に高いデータでありルールが早期に絞まれ易いこと、ある状況(仮想世界)を実現するまでの方法(つまり、ルール)の組み合わせは非常に多く後向きでは計算量が爆発してしまうこと、および、目的解析処理によって得られたゴールは必ずしも正しいものである訳ではないことといった理由のために、前向推論を主導とした推論戦略を行うことにした。すなわち、イベント駆動による前向推論を行い、ルールの前件の一部が満たされていない(これは、意味ネットワークの一種である複合機能連鎖構造[9]を用いて推論する場合、必須ノードが未発火であることに対応)場合に、その必須ノードを説明(証明)すべく後向推論を発火させる。システムの状態の変化に応じて知識ベース内で更新されるシステム状態依存知識(例えば、「文書アイコンが表示されているか否か」といった表示状態に関する知識など)は、事実(fact)として扱い、後向推論の停止条件、つまり、後向推論を始めたノードの説明(証明)として利用する。

(3) 仮想アクションという事実

ルールの前件に、システムが完了するべきタスクではなく、(推論時ではなく)実行時にユーザが完了させるべき条件内容(例えば、「宛先の記入」など)があることがある。このような条件内容は、推論時には、もちろん、その時点でのシステム状態依存知識で証明できないし、場合によっては実行時にも(ユーザのせいで)完了させられないかもしれない。このような条件内容に対して、後向き推論の最中にはこれ以上の後向きの探索はここで停止させるために、「仮想アクション」という種類の事実(fact)を用意する。仮想アクションは、(a)後向き推論時には停止条件として機能し、(b)制限付き前向き推論(後述)の時にはあたかもその時点では事実として実現しているかのように振る舞い、(c)実行時には、この仮想アクションが実際にユーザによって実行されるまで実行制御が停止するように処理することとする。機能連鎖構造を用いた知識表現上は、仮想アクションを表現するノードに対して、解析時に条件仮定として捉えるよう推論制御し、実行時にその条件を検証確認する

よう実行制御を行うという付帯条件を記述しておくことで実現する。

(4) 制限付き前向き推論

さて、後向き推論によって必須ノードの証明が済むことにより、その必須ノードを前件としていたルールの前向き推論を再開してもよいことになる。意味ネットワークの波及探索を利用して推論する場合に問題となるのは、ルールの後件のノードが必須ノードを持つ前件がすべて満たされたかどうかを知ることである。単純には、後向き推論を開始したノード毎にスタックを作りここに未証明の事実を載せ、すべての未証明事実がなくなった時点で、前向き推論を再開するという実装方法が考えられるが、この方法は、このノードを利用する前向き推論のたびに証明する必要がある、重複計算コストがかかる。そこで、本実装方式では、停止条件に用いた事実(仮想アクションも含む)をイベントとして見做し、この事実のイベント駆動による前向き推論を行うことで、推論の効率化を図っている。これによって、例えば、他の並列的に作動している後向き推論の証明事実として途中から利用できるようになり、重複計算がなくなる。なお、前向き推論が元の証明元に辿り着くまでの最中に再び未証明の必須ノードに遭遇するような合流経路へは探索を進むことがないように探索を制限しているの、この推論方式を、特に、制限付き前向き推論(Restricted forward Reasoning)と呼ぶことにする。

3. 知的MMUIの基本ソフトウェア構成

第2章の基本設計に基づくIMMoUIの基本ソフトウェアは、大別すると、マルチモーダル認識プログラム、マルチモーダル対話計画プログラム、マルチモーダル応答プログラム、マルチモーダル連携制御プログラム、の4つのプログラムからなる。

3.1 マルチモーダル認識プログラム

マルチモーダル認識プログラムは、テキスト、音声、印刷文字および手書き文字といったマルチモーダルデータの入力を認識し、その認識結果を相互的に補正する。

(1) 入力メディア制御プログラム

入力装置(イメージ入力装置、ストローク入力装置)をそれぞれ独立に制御する。

(2) 個別モダリティ認識プログラム

スキャナによる印刷文字、ペンによる手書き文字、マイクによる音声を認識する。具体的には、イメージデータやストロークデータから最小単位を切り出し、この最小単位、および、時間的／位置的に近接する最小単位の組み合わせ単位を認識する。なお、個々のモダリティ毎に好適な認識アルゴリズムや認識単位が異なるため、プログラムとしてはそれぞれ別の実装形態となる。

(3) モダリティ間認識調整プログラム

それぞれの認識結果を各モダリティに非依存なデータ構造である認識ラティスへ変換し、辞書(=モダリティごとの構成要素集)や文法(=構成要素の時間的／位置的な並び規則集)をもとに認識結果の評価値を調整する。そして、複数の認識ラティス間の認識結果から適切な組み合わせ候補を決定する。

3.2 マルチモーダル対話計画プログラム

マルチモーダル認識プログラムの認識結果および対象世界の知識に基づいて、ユーザからのマルチモーダル入力の意味を解析し、目的を推定し、ユーザへの応答すべき情報や、アプリケーションを動作させる機能的なシーケンスを導出する[9][10][12]。なお、必要に応じて、自然言語処理や推論処理を援用する。

(1) 意味解析プログラム

マルチモーダル認識プログラムの認識結果を操作指示内容として意味解析する。

(2) 目的推定プログラム

意味解析の結果、および、対象世界の状態などから、全体のタスクの流れの中におけるカレントの対話内容の位置や、最終ゴールを推定する。

(3) 戦略決定プログラム

応答戦略を決定し、ユーザへの応答すべき情報や、機能的なシーケンスを導出する。("What-to-Express"の決定)

3.3 マルチモーダル応答プログラム

導出された応答すべき情報を、どのモダリティ／メディアで、どの時間順序で出力するかを計画立案("How-to-Express"の決定)し、アプリケーションの実行に必要な制御シーケンスを生成し、その実行を管理する[13][14]。

(1) モダリティ間応答調整プログラム

複数の出力モダリティにまたがった出力順序や出力タイミング関係を調整し出力計画を構成する。

(2) 個別モダリティ生成プログラム

出力計画に沿って応答すべき情報を、最終的な制御シーケンスへ変換(生成)する。また、制御シーケンスを適切なタイミングでそれぞれの出力メディア制御プログラムに出力する。

(3) 出力メディア制御プログラム

出力装置(音声出力装置、文字表示出力装置、グラフィック表示出力装置)、および、アプリケーションをそれぞれ独立に制御する。なお、入出力の装置単体、あるいは、装置の組み合わせ単位ごとに、これを識別するためのニックネーム、配置位置情報、取り扱うデータのタイムスタンプ、利用ユーザ名といったマルチモーダルデータ特有属性[10]を取り扱う機能を与える。このように機能強化された装置の単位を、我々は、ユビキタス・コンピューティング・ユニット

(UCU)と呼ぶ。例えば、上記機能を搭載し、ペン入力と液晶表示出力とを備えたペンコンピュータや、タッチパネルによる画面上の位置入力と擬人化エージェントによる表情出力を備えたインテリジェントエージェントシステムなど。

3.4 マルチモーダル連携制御プログラム

IMMoUI全般に亘る管理機能や中継機能を司る。および、実装形態依存のドライバ機能である。

(1) 対象世界動的管理プログラム

仮想的な情報世界の知識、現実の物理世界の知識、および、両世界の対応関係に関する知識を司る。すなわち、それぞれの世界の操作対象、データ表現、表現態様、実装形態、その状態、それらの相互関係情報などを動的に管理する。

(2) UCUドライバプログラム

UCUごとに用意されたドライバで、マルチモーダルデータ(MMD)が構造的・意味的に組み合わせられた複合体データを、物理デバイスあるいは仮想メタファを用いて表示したり、操作可能とする。

(3) データ通信支援プログラム

UNIXおよびWindows95といった異なるプラットフォーム上で動作するプログラム間のデータの通信をサポートする。特に、マルチモーダルデータの通信に不可欠な非同期通信や、連続ストリーム型通信を支援する。また、LAN内で分散的に実行している各プログラムに対するネーミング・サービス(同一のサーバ・プログラムが複数起動している時に、クライアントへ適切なサーバを紹介するサービス)を提供する。

3.5 基本ソフトウェアの実装

以上の構成を持つIMMoUIの基本ソフトウェアを100キロステップ(但し、認識エンジンや形態素解析処理などを除く)を超えるプログラムとして実装してみた。この基本ソフトウェアは1台のワークステーション上でも稼働するが、複数台のワークステーションやパソコンを用いて、負荷分散させた方がリアルタイム性能が向上する。実際、通信負荷を軽減させるためにある程度ローカル処理が可能な単位であるUCUごとに1台の計算機を割り当てると良いようである。

4. 動作例と考察

4.1 個人情報環境への適用

第3章の基本ソフトウェアを、個人情報環境に適用した(図2)。この環境では、ユーザが複数の入出力メディアを同時的に使用することができる(具体例は次節以降参照)。また、従来のMMIは1つのモダリティ毎に1つの入出力デバイス(ペンパッド、マイク、ディスプレイなど)を使用していたけれども、この環境では、1つのモダリティに対して複数のUCUを割り当てることが可能である。このUCUが管理するマルチモ

ーダルデータ特有属性の情報には発生/出力場所情報があるので、複数のMMDの意味的な関係を解析し[10]、Location-awarenessのコンセプトを実現させることができる[15]。同様に、発生時刻によるMMDの融合を取り扱ったり、ユーザ(または、そのユーザグループ)に応じて用語の使い方などデータの解釈の違いを取り扱うこともできる。

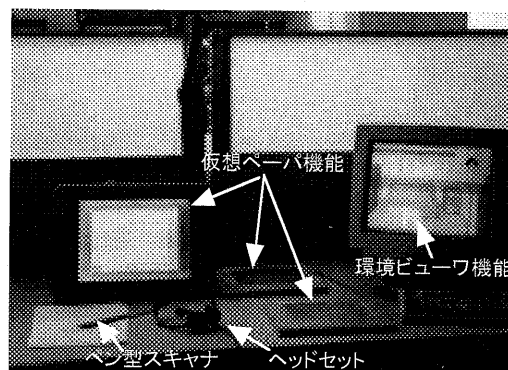


図2 個人情報環境 RVI-DESK

4.2 データ補正実験

イメージと音声と同時に入力し、あるモダリティのデータを別のモダリティのデータが補正する実験を行った。具体的には、アドレス管理ソフトなどへのデータ入力を行っている状況で、名刺上の電話番号(例えば、「3256」)のあたりにペン型スキャナを当てつつ、同時に「内線番号」と音声で補助しつつ、電話番号の入力を行う実験を行った。イメージからの認識結果の第1候補を暫定的に表示出力(この例では「た3ハ6」(図3))した後、モダリティ間認識調整プログラムは次の処理を行った[16]。

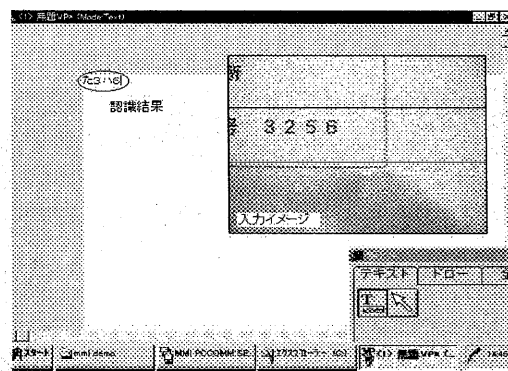


図3 印刷文字認識結果

- (1) 知識を用いた不的確な情報の削除
「内線番号は数字である」という知識から、数字の

並びを探し、数字の並びとして解釈できない部分を無視した。これは、名刺から読み取った電話番号前後の余計なイメージ部分をカットしたことに相当する。(この例では「た」を削除)

(2) 認識ラティス上の別候補

(1)と同じ知識を用い、第2候補以下の認識結果から数字として認識できるものに置換した。(この例では「ハ」を「25」に置換)(図4)

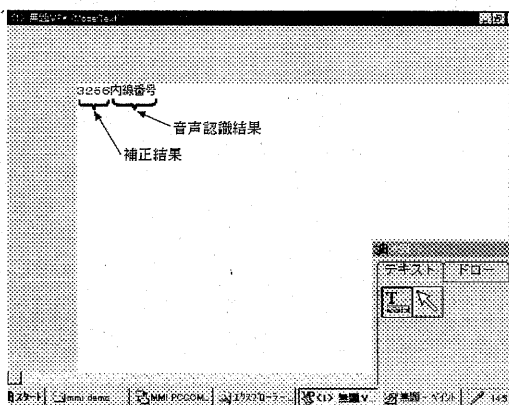


図4 データ補正結果

4.3 意味融合実験

ストロークと音声と同時に入力し、2つのモダリティが意味のレベルで融合し、1つの意味単位を構成する実験を行った。具体的には、電子メールソフトを操作している状態で、「返送」という文字をペンで指定しつつ、同時に音声で「この方法は？」とマルチモーダルな入力を行う(図5)。仮想的な紙を実現したUCU「仮想ペーパー」上には、インク形式の手書き文字と、コード化された文字が混在し、1つの意味のあるドキュメントを構成している。本実験では、このドキュメント内の情報と連携しつつ、アプリケーション操作のヘルプを応答してくる。

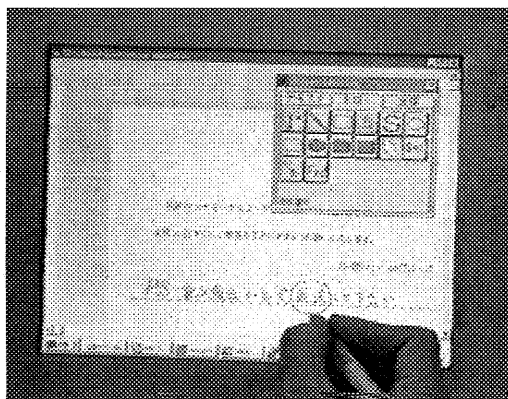


図5 仮想ペーパー(UCU-Paper)からの入力

それぞれの入力データをマルチモーダル認識処理し(この詳細は第4.2節とほぼ同様なので省略する)、意味解析プログラムは次の処理を行った[10][15].

(1) マルチモーダル属性の解析

2つのデータの入力時刻が近いのでデータ内容の意味的関連性は強いとして評価する。

(2) 自然言語やタスクの流れを利用した解析

自然言語処理により「この」は「返送」を指しうるので、ユーザからの要求は「返送の方法に関する何か」であると解釈する。

(3) 対象世界知識に基づく合理性の検証

このシステムが「返送」を実施可能(となりうる)かどうかを、対象世界知識および推論を用いて判定する。(この実験例では、推論の結果、「Yes」が得られる)

(4) 応答方法の可能性を検証

(3)同様にして、「返送する方法を説明する方法」を知っているかどうかを判定。(この実験例では、結局、「Yes」が得られる)

戦略決定プログラムは、(1)~(4)の結果に基づき、「返送の方法を説明する」ことを応答戦略として決定した。

さらに、この応答戦略に基づいて、アイコンの移動命令、音声合成装置への制御命令、電子メールソフトへの命令などへ具体化し、それらの実行をスケジュールリングし、実際に実時間において各出力メディアを制御して、ヘルプ応答という目的を果たした(図6)。仮想的なメタファをコントロールすることで現実世界のデスクトップとリンクする統合的なコントロール環境「環境ビュー」上で、合成音声、文書アイコン、矢印アイコンを同期させて動かし、「返送」の方法を説明している。アイコンの配置座標やソフトウェアの起動状態などは、操作状況に合わせて実時間的に変わるシステム状態依存知識(2.3節参照)である。これらを随時、知識ベースへフィードバックさせることで、命令を適切に具体化できている。

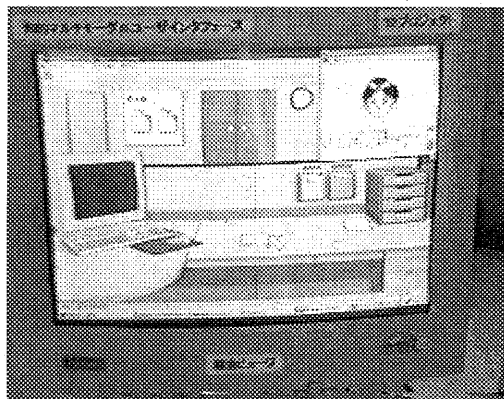


図6 環境ビューワ(EV)への出力

4.4 マルチモーダルデータ複合体の利用実験

マルチモーダルデータ複合体(MMDC)は、複数のモダリティによる複合関連的なデータの統一表現である。現時点の実装では、各処理要素が、通信処理系を介して授受されるMMデータフォーマット(2.1節(3)参照)によって記述されたマルチモーダルデータの集合としている。このデータの集合は、それぞれのモダリティに強く依存するデータだけでなく、各処理要素間の制御命令や、多重反射処理上発生している途中のデータも含んでいる。(2.2節参照)

このMMDCを介して、ユーザとシステムの状況に合わせたマルチモーダルデータのフレキシブルなハンドリング(すなわち、同時的入出力、意味的なモダリティ変換など)を可能とする実験を行った。具体的には、モバイル環境において入力されるマルチモーダルなデータをMMDCとしてシステム内部に存在せしめ、利用形態に応じて、文字列だけに変換させたり、マルチメディアによるリッチな情報表現に変換する。

(1) モバイル環境

図7は、ストローク、音声で操作できる地図システムである。横浜駅周辺の地図上でユーザがペンで移動経路を書き込む。ペンは地下階段の位置に動かしつつある時に、同時に「地下へ」と音声指示すれば地下の地図へ切り替えることもできる。このようにしてマルチモーダルデータの入力を続け、最終目的地までユーザがデータを入力すると、一連のMMDCがシステム内部に格納される。

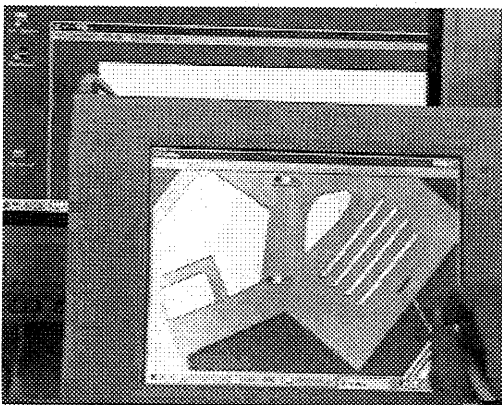


図7 地図システムへのマルチモーダル入力

(2) ナロー回線環境

(1)で入力したデータは、ある目的地への道筋を説明するためのものであるが、これをmpeg形式などの動画ファイルにすると、数メガバイトの大きさになる。このため、携帯端末やPHSやPDAなど移動体にデータを送ろうとすると、通話時間(したがって、通話料)が膨大になる。ところが、入力データのうち人

間に必要な内容は、実は簡易な文字列に変換でき得るものである。図8は、(1)での地図操作によって直接的に得られるストロークの座標列と地図データから最寄りのランドマークを抽出し、それを矢印で結んだ文字列と、「地下へ進む」などのキーワードをそのまま連結して文字列を作成した。これなら、携帯電話のショートメッセージサービスで遠隔地に送るのは簡単である。このように、意味内容を変化させずにデータの表現モダリティを変えてデータ量を削減することができる。

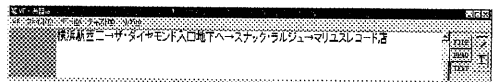


図8 文字列表現による意味的圧縮

(3) リッチなデスクトップ環境

(1)でユーザが入力しようとしている目的地は最後にデータを入力するまでわからないけれども、一旦入力されてしまえば、そのMMDCを解析することにより、出力時に効果的な演出ができる。強調表現やアクセント表現を駆使してリッチな情報表現ができる。入力は2次元地図の方が便利だが、出力は3次元のVR地図で表現する方が効果的な場合もある。

(4) 断片的で逐次的な情報授受環境

MMデータのすべてが一度に必要という訳ではなく、コンテキストに依存しつつ、断片的な情報を必要に応じて手に入れたい環境がある。例えば、日本語の対話による提案や誘導をおこなう携帯マン・ナビでは、ユーザの立っている場所に応じて、MMDCから「30メートル先の階段を降りてください。」などといった臨機応変なデータへ変換することもできる。

以上、利用形態(1)～(4)で紹介したように、マルチモーダルなデータを必要に応じて意味的な情報を解析し、意味内容を変化させずにモダリティの組み合わせやタイミングなどを工夫することで、通信帯域幅や出力デバイスの特性に応じたデータ表現へ変換することができた。

4.5 考察

第2章および第3章で提案したIMMoUIは、マルチモーダルデータ複合体、認識ラティス、マルチモーダル特有属性などのデータ構造/データ内容を導入し、各処理モジュール間の役割を明確に取り決めることで、各モダリティに依存するモジュール、モダリティに非依存的に処理できるモジュール、連携を支援するモジュールなどに分割している。このIMMoUIを個人情報環境に適用の結果、システム全体として統一的にマルチモーダルデータが扱えることが確認できた。

第4章での実験結果から、複数MMDの同時的な入力を統合して認識し、その入力の意味を解析できることを確認し、状況に依存する知識を用いた知的な対話システムが実現する見通しを得た。

5. まとめ

知的で、マルチモーダルなUIの実用化のために解決すべき課題、すなわち、(1)処理を分散させ、また統合する有効な仕組み、(2)ユーザインタフェース処理の即応性が高くなること、(3)動的に状況判断可能な対話的なシステムを実現できること、について議論した。本論文では、この課題を解決すべく、マルチモーダルデータ処理を可能とする分散アーキテクチャ、多重反射型インタフェース処理、および、両方向推論による状況依存型意味解析方法を提案した。これらのアイデアを実装し、知的マルチモーダルユーザインタフェースの基本ソフトウェアを開発した。これは、キーボード、ペン、マイク等の各モダリティによる入力を、単独あるいは統合して認識し、その入力の意味／意図を解析し、解析結果に基づき他アプリケーションを制御、あるいは、音、アニメーション、動画などのマルチメディアを用いて応答するための統一的なアーキテクチャを提供するソフトウェア群である。

このソフトウェア群を用いて、実際に個人情報環境へ適用実験した結果、ストロークと音声の同時入力によるデータを補正できること、イメージと音声の同時入力から意味や意図を解析し知的な応答ができること、そして、複数のモダリティによる複合関連的なデータの統一表現であるマルチモーダルデータ複合体によって、ユーザとシステムの状況に合わせたデータ処理ができることを確認した。以上より、入出力手段の柔軟性とデータの意味の取り扱いを兼ね備えた知的マルチモーダルユーザインタフェースが実現できる見通しを得た。

謝辞 終わりに、本研究のアイデアに関し、ご討論をいただいた電気通信大学大学院情報システム学研究科田野俊一助教授、本研究の機会を与えていただいた(株)日立製作所システム開発研究所片岡雅憲所長、実装や実験にも協力していただいた著者らの属する研究室の坂尾秀樹氏、富田民則氏に深く感謝致します。

参考文献

- [1] Bolt, R.A.: Put-that-there: Voice and Gesture at the Graphics Interface, *Computer Graphics*, Vol.14, No.3, pp.262-270 (1980).
- [2] Maes, P.: Agents that Reduce Work and Information Overload, *Communication of ACM*, Vol.37, No.7, pp.31-ff. (1994).

- [3] Neal, J.G. and Shapiro, S.C.: Intelligent Multimedia Interface Technology, *Intelligent User Interfaces*, Sullivan, J.W. and Tyler, S.W.(Eds), pp.11-43, ACM Press, Addison-Wesley, New York (1991).
- [4] Flanagan, J.L., Technologies for Multimedia Information Systems, 電子情報通信学会論文誌 D-II, Vol.J75-D-II, No.2, pp.164-178 (1992).
- [5] Koons, D.B., Sparrell, C.J. and Thorisson, K.R.: Integrating Simultaneous Input from Speech, Gaze and Hand Gestures, *Interligent Multimedia Interfaaces*, Maybury, M.T. (Ed), pp.257-276, AAAI Press, MIT Press, Cambridge, MA (1993).
- [6] Nigay, L. and Coutaz, J.: A Generic Platform for Addressing the Multimodal Challenge, *Proc. CHI'95*, pp.98-105 (1995).
- [7] 河野ほか: 仮説推論に基づくマルチモーダル入力統合方式, 情報処理学会ヒューマンインタフェース研究会 インタラクシオン'97, pp.33-40 (1997).
- [8] Cohen, P. R. Cheyer, A., Wang, M., Baeg, S.C.: An Open Agent Architecture, *Proc. of AAAI spring Symposium on Integrated Intelligent Architectures*, pp.1-8, March 1994.
- [9] Namba et al.: Complex Chained Function Structure for Human-Computer Interface, *Proc. HCI Intn'l 95 Poster Sess.*, p.32-32 (1995).
- [10] 難波ほか: マルチモーダルデータ特有属性の融合性を利用した意味解析, 情報処理学会論文誌, Vol.38, No.7, pp.1441-1453 (1997).
- [11] Norman, D.A.: Cognitive Engineering, User Centered System Design, Norman, D.A. and Draper, S.W.(Eds), pp.31-65, Lawrence Erlbaum, New Jersey, (1986).
- [12] 難波ほか: NLIにおける命令表現と実コマンドとのギャップ, 計測自動制御学会ヒューマンインタフェース部会 第10回ヒューマンインタフェースシンポジウム, pp.323-330 (1994).
- [13] 青島ほか: 機能連鎖構造に基づくヘルプ応答生成, 人工知能学会全国大会, pp.513-516 (1996).
- [14] 青島ほか: マルチモーダル応答生成における出力タイミングの決定, 情報処理学会全国大会, Vol.4, pp.15-16 (1997).
- [15] 田野ほか: 実世界・仮想世界を融合した知的ユーザインタフェースコンセプトの提案とその適用例, 計測自動制御学会ヒューマンインタフェース部会 第13回ヒューマンインタフェースシンポジウム, pp.109-116 (1997).
- [16] 富田ほか: ペン型スキャナを用いたマルチモーダル統合認識, 情報処理学会全国大会, Vol.4, pp.113-114 (1997).