

マルチモーダル秘書エージェントシステムの開発

○金沢 博史 知野 哲朗 河野 恭之 屋野 武秀 池田 朋男
鈴木 薫 中山 康子* 福井 美佳* 真鍋 俊彦* 竹林 洋一*
(株) 東芝 関西研究所 *(株) 東芝 研究開発センター

〒 658 神戸市東灘区本山南町 8-6-26

* 〒 210 川崎市幸区小向東芝町 1

知識やノウハウを持っている本人に代わって問い合わせに答える、マルチモーダル秘書エージェントシステムを開発した。ユーザが音声やジェスチャ、自然言語で質問をすると、秘書エージェントはその意味内容を理解して、知識情報やノウハウを蓄積したデータベースから適切な回答を検索して提示する。本システムは知識ベースとノウハウベースの連携による知的検索技術、知識情報を検索・活用しやすくするための知識情報の構造化技術、従来の GUI に代わる音声やジェスチャを用いたマルチモーダル対話技術を統合して実現した。

Development of Multimodal Interface Agent System

HIROSHI KANAZAWA TETSURO CHINO YASUYUKI KONO
TAKEHIDE YANO TOMOO IKEDA KAORU SUZUKI
*YASUKO NAKAYAMA *MIKA FUKUI *TOSHIHIKO MANABE
AND *YOICHI TAKEBAYASHI

TOSHIBA KANSAI RESEARCH LABORATORIES.

* TOSHIBA R&D CENTER.

We have developed a multimodal interface agent system that supports information and knowledge sharing and inter-group communication. When a user enters a query using speech, deictic gesture and natural language, the agent system understands user's intention, retrieves the relevant information from its knowledge databases and presents the appropriate answer. The system is provided with intelligent information retrieval using knowledge and knowhow databases, knowledge structuring techniques and multimodal user interface.

1 はじめに

インターネットをはじめとするネットワークインフラの高度化と計算機の処理能力の増大は、個人や組織の情報収集・蓄積・発信の能力を向上させる一方で、情報洪水を招き、情報収集やコミュニケーションにおける人間の負担を拡大している。このような問題を解決するために、ユーザの要求をより自然に計算機に入力するためのインタフェースの高度化や、欲しい情報をできるだけ正確に提供する情報検索の高度化、広範囲にまたがる大量の情報の分類・組織化が求められ、多くの研究がなされている [1] [2] [3] [4]。

一方、仕事を進めるうえで必要な知識や情報は膨大であり、特に非定型業務は個人のノウハウに依存している部分が多い。しかし、個人が所有しているノウハウは暗黙知のままとどまり、貴重な知識が十分に利用されていないという問題がある。そこで、仕事の効率化を図るためには、組織や個人の暗黙知を形式知化し、それらの知識を共有化することが重要になって来る。

以上のような問題意識に基づき、個人の知識やノウハウを容易に獲得して、知識データベースに蓄積し、言語・音声・ジェスチャによる自然な対話でユーザの要求を理解し、適切な検索を行えるマルチモーダル秘書エージェントシステムを開発した。ここでは、本システム構築のアプローチとシステムおよび各処理の概要について述べ、最後に評価の結果明らかになった課題について述べる。

2 アプローチ

2.1 マルチモーダル入力統合

人間の認知や情動の性質に適合した計算機との対話の実現には、音声やジェスチャなどの複数のメディアを利用して、ユーザの意図を理解し、適切な応答を提示することが必要である。一般にマルチモーダルインタフェース (MMIF) は、音声やポインティングデバイスによる直接指示といった個々のモダリティの処理モジュールから非同期に与えられる認識結果を受け入れて互いに関連づけ、その統合・解釈を行う。即ち、ユーザが様々なモダリティを通じて入力した個々の断片的な情報の集合から、全体としてユーザが伝達を意図し

た内容を推定する。このマルチモーダル入力統合のためには、(1) 遅着データの統合、(2) 認識結果の曖昧性の解消、(3) 過去の問題解決データの参照といった課題の解決が必要である。これらの課題を効率的に扱える汎用の枠組を構築するために、筆者らは一貫性管理機構 ATMS (Assumption-based Truth Maintenance System)[5] を MM 入力統合に適用し、その制御メカニズムを定式化した。これにより、自然言語、音声、ジェスチャによるマルチモーダル入力を実現し、ユーザにとっては負担の少ない自然な対話が行えるようにした。

2.2 知識ベースとノウハウベースの連携

知識情報共有のキーとなる知識データベースのコンテンツの充実のためには、組織や個人の知識の形式知化が必須である。従来のエキスパートシステムでは、ドメインやアプリケーションを限定して、問題解決に必要な知識を分析して形式化した。が、オフィスの知識を分析してこのような知識ベースを構築するには膨大な手間がかかる。そこで、形式知化を知識とノウハウの連携で行うことにした。ここで、知識とは、意味のある抽出された情報同士を相互に関連づけたもので、用語の意味や物事の機能・構造・関係等を表すまとまりをもった概念構造であり、コンピュータが問題解決に利用できる形式で記述されたものである。一方、ノウハウは、人間が理解できる形式で記述された形式知と定義する。知識を体系的に記述した知識ベースと、体系化されていないが人間が理解可能なノウハウベースを構築して連携させることにより、非均質で記述形式が多様でも有益な知識を入力しやすくし、活用できるようにする。更に、蓄積されたノウハウから形式化できる知識を抽出して知識ベース化し、アドバイスの知的支援に発展させていく。

3 システム構成

システムは、図 1 に示すように、マルチモーダル入出力部、知識データベース部、検索部、知識情報構造化部から構成されている。マルチモーダル入力部では、音声、ジェスチャ、キーボードによる入力を受け付け、各メディアの認識結果を MM 入力統合解析部で解釈し、検索部へ送る。検索部では、知識ベースとノウハウベ

スを相互にアクセスして、所望のデータの抽出を行う。そして、検索結果をマルチモーダル出力部に送り、音声合成、テキスト、画像を用いた内容の提示と、擬人化エージェントによる身振り、顔表情といった非言語メッセージを提示する [6]。知識データベースは、各種問題解決、入力統合処理での曖昧性解消など幅広い意味レベルでの処理の基盤となるものである。また、知識情報の構造化部では、ユーザが登録したノウハウや報告文書等のマルチメディア情報に対してキーワード抽出、タグ付けを行い、自動的に構造化を行う [7]。

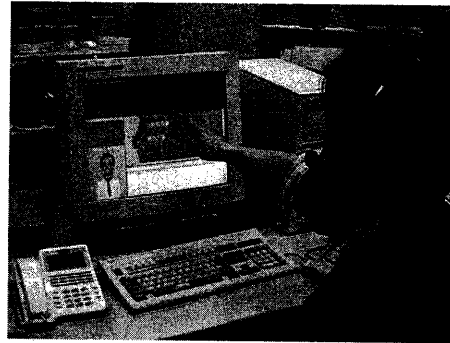


図 2: システムの全景

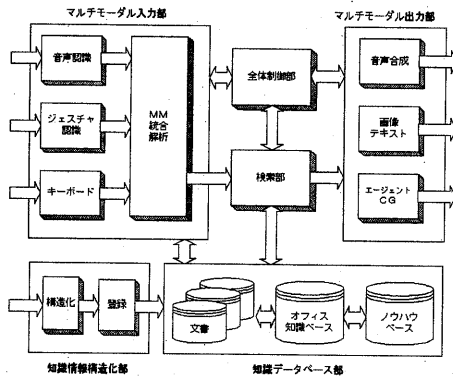


図 1: マルチモーダル秘書エージェントの構成

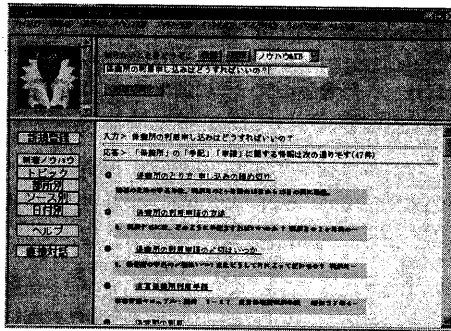


図 3: システムの動作画面例

図 2 に本システムの全景を、図 3 に動作画面表示例を示す。現在、本システムには事務手続き、研究関連、計算機の使い方などに関するノウハウが約 10,000 件登録されている。また、ノウハウの登録、参照したノウハウへコメントを付与する機能をもつ。

4 マルチモーダル入力統合

4.1 マルチモーダル照応解決

ポインティングジェスチャ入力と音声入力を利用するマルチモーダルインタフェースでは、それぞれのメディアで入力指示された対象を同定する照応解決が問題となる。図 4 は、本システムが行なうマルチモーダル照応解決処理の概要を示している。

ここでは、利用者が、画面に表示されたある人物の連絡先の提示を要求する場面での処理例が示されている。この音声とジェスチャを併用したマルチモーダル

入力に対して、システムはそれぞれ音声認識処理とジェスチャ認識処理を実施する。しかしこういった認識処理では一般に解釈の曖昧性が発生し、一意の認識結果ではなく、スコア付けされた複数の認識候補が、それぞれの入力モードについて得られる。そして、音声とジェスチャという異なるメディアからの入力を統合するために、同一の表現レベルへと変換を行ない、知識との照らし合わせによる曖昧解消統合が行なわれる。この知識を用いた統合解釈によって、複数のモードからの入力の統合が実現するだけでなく、各モードの解釈結果を相互に補完し合うことによって、各モードを単独に利用した場合には解決不可能であった認識誤りの発生が抑制される。

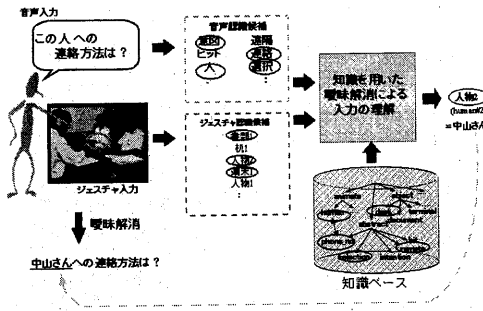


図 4: マルチモーダル照応解決処理

4.2 仮説推論に基づくマルチモーダル入力統合解釈

照応解決などのマルチモーダル入力統合解釈において、(1) 遅着データの統合、(2) 認識結果の曖昧性、(3) 時間の取り扱い、が大きな問題となる。その解決のためには、「遅着したデータに関係のない解析の途中データを再利用する」、「他の候補の解析結果のうち、再利用できる部分は再計算しない」といった制御が必要となる。そこで、筆者らは MM 入力統合解釈処理の制御メカニズムの基盤として、ATMS を選択した。これは、多重世界問題を取扱う問題解決器に対する知的キャッシュであり、複数の世界観で情報を最大限に伝達して効率良く問題を解決する上で有用な機能を提供する。この ATMS ベースの問題解決プロセスを定式化することで、上記の課題の前 2 点、すなわち遅着データの統合と各モダリティの認識結果の曖昧性の課題に対処し、効率的に適切な解釈結果を導く制御メカニズムを構築した [8]。また、時間経過の取扱いについても、入力時刻情報を付した ATMS の仮定を生成することで対処可能とした。この枠組みにより、(1) 問題解決の途中過程を可能な限り生かしながら、適切な解を導ける候補の集合を探索することができ、(2) 遅着データを取り込んで MM 解析を再開することができるようになった。

5 知識ベースとノウハウベースの連携

知識情報源は、オフィス知識ベース、ノウハウを蓄積するノウハウベース、既存のマルチデータベースで、これらのデータはオフィス知識ベースにより相互に関

結びけられて利用される。

5.1 オフィス知識ベース

汎用の知識ベースとしては、CYC、WordNet、EDR 辞書等が知られているが [9]、筆者らは知識情報共有に焦点を絞った知識ベースを構築中である。オフィス知識ベースは図 5 に示すように、概念体系や推論の基本処理を持つクラス階層と、オフィスの実体や相互関連を記述するインスタンス階層を OODB(Object-Oriented DataBase) 上に実装した。クラス階層は、人・組織、文書、業務のオフィスの事物の相互関連や機能、業務の手順、概念・語彙体系を記述したものである。基本オブジェクトは人、組織、業務、文書で、下位オブジェクトが階層関係で記述されている。また、オブジェクト同士は各種関係のリンクを持つ。

オブジェクトは、基本オブジェクト Task、Document、Group、Person の下位オブジェクトとして定義する。property は属性、関係等を定義し、type で属性値の型を定義する。オブジェクト自体や property は同義語を持つことができ、syns で定義する。たとえば文書の人への関係 document'of を作成者、執筆者、記入者等と呼ぶ。

5.2 ノウハウベース

ノウハウベースには、記述方式や粒度を限定せずに、組織内で共有すると有益であると思われるものを個人が提供して蓄積した。ノウハウデータは、自然言語対

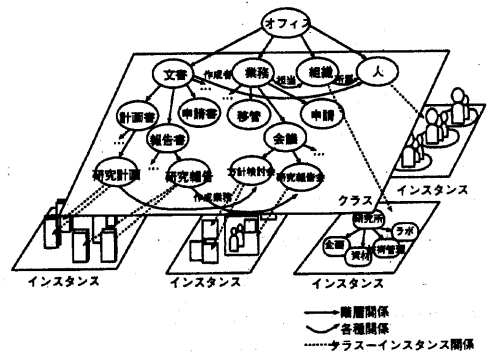


図 5: オフィス知識ベースの構成

話で検索できるように、対象、行為、形態、状況等の13種類の属性や、ノウハウ同士の関連を記述する浅い構造を持つ。これらの構造は、ノウハウデータ登録に際してユーザが記述したタイトルと本文から抽出する。ユーザからの問合せの解析結果と、この構造情報を参照して、マッチングを行なう。また、ノウハウ同士の関連（詳細、理由、事例等）のリンクづけがされており、たとえば「もっと詳しく」「なぜ」等の問合せに回答できるようになっている。ノウハウベースはORDB (Object-Relational DataBase) 上に実装し、ノウハウの内容は、事務手続きのやり方、書類の書き方・事例、研究関連のノウハウ、計算機や周辺機器の使い方、福利厚生や日常生活に関わるノウハウ、役に立つページへのリンク、用語集等で、現在約12,000件登録されており、継続的に増加している。

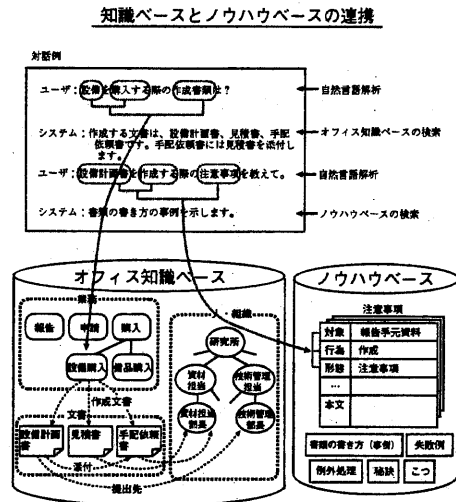


図6: 知識データベースの検索対話例

5.3 知識データベースの検索

システムはマルチモーダル入力による質問を解釈し、オフィス知識ベースまたはノウハウベースから適切な回答を検索して提示する。人・組織の役割や関連文書など、体系的に整理された知識を用いて回答できる質問に対しては、知識ベースを参照して回答し、こつや秘訣、注意事項、事例、例外処理、失敗談など、体系的に記述されていないものはノウハウベースを検索して回答する。図6の例では、「設備を購入する際の作成書類は?」という問い合わせに対して、自然言語解析の結果、知識ベースに記述されている「設備購入」オブジェクトと関係属性「作成書類」のペアがマッチし、作成する書類として「設備計画書」、「見積書」、「手配依頼書」を回答する。一方、「設備計画書を作成する際の注意事項は?」という問い合わせに対しては、知識ベースでは「設備計画書」オブジェクトはマッチするが、「注意事項」に該当する属性や関係が存在しないのでノウハウベースの構造情報とマッチングを行い、対象が「設備計画書」、行為が「作成」、形態が「注意事項」であるノウハウデータを検索して提示する。

5.4 知識情報の構造化

ユーザが登録したノウハウや報告文書等の、テキスト、イメージ、映像情報等の多種多様のマルチメディア

情報は、検索・活用しやすいように構造化して蓄積する。入力されたマルチメディア情報は、HIウェア[10]の機能であるドキュメントリーダ、音声認識、英日翻訳等により日本語テキストに変換し、キーワード抽出、タグ付けを行うことで自動的に構造化する。キーワード抽出は、オフィス用語辞書を参照し、文書の種類毎に抽出キーワードの優先順位を決めるルールを用いて行う。優先順位は出現頻度に語の重要性の重みづけをして計算する。オフィス用語は重要語として重みを高くし、一般名詞は重みを低くし、さらに一般的な名詞でキーワードとしてふさわしくない語を不要語として登録しておき、重みを0として計算する。計算の結果、評価値の高いものからキーワードとして抽出する。ノウハウや文書の登録に際しては、自動的に抽出されたキーワードをユーザに提示するようにし、未知語は新たにキーワードとしてユーザに指定してもらうようにした。これにより、用語辞書のキーワードを順次拡張している。テキスト以外の情報については、キーワードが抽出できないので、タイトルと補足説明文を入力するようにし、これによりキーワードを抽出するようにした。

6 評価

現在、社内の1,000人規模のオフィスで本システムを日常業務に使いながら、実証評価を行っている。同時に、コンテンツを限定して音声による検索要求やジェスチャによる指示等のマルチモーダル入力を適用し、ポストGUIに向けたマルチモーダルインタフェースの評価を行っている。評価実験の結果、以下の課題が明確になった。

(1) ユーザの意図に合わせた情報の提示

ユーザのレベルや利用時の目的により、知識情報の提示の仕方を変えるべきである。特に、基本的なことは分かっているが、さらなるノウハウを知りたいというアドバイス型では対応する情報だけでなく、関連情報などのより広い範囲の情報を提示する等の戦略が必要であることがわかった。

(2) システムのロバスト化

各メディア処理の高精度化と、対話履歴やコンテキストを利用した知的な対話処理による頑健なインタフェースの構築が重要である。また、どのような入力ができるのかや、検索失敗の理由をユーザに通知するなどして、システムの能力と限界をユーザに明示的に提示することと、擬人化エージェントを利用して顔の表情や合成音の声質などの非言語メディアで情報をユーザに提示することもシステム全体のパフォーマンス向上につながる。

(3) 知識情報の提供の動機づけ

組織における知識情報共有を促進するには、個人が積極的にノウハウを提供する動機づけが必要である。提供したノウハウが他人の問題解決等に役に立ったことがわかるフィードバック機能が必要である。

7 まとめ

マルチモーダル秘書エージェントシステムの概要およびその構成要素について述べた。本システムの特徴は、深い構造を持つ知識ベースと浅い構造を持つノウハウベースの連携による検索と、時間遅れや認識の曖昧性解消のために導入した仮説推論によるマルチモーダル入力統合解釈にある。今後は、知識ベースの大規模化と検索の高精度化、さらにマルチモーダルインタ

フェースのロバスト化を行い、アドバイスやヘルプなどの適切な応答を行える知的支援機能を構築する。

参考文献

- [1] K.J.Hammond et al., FAQ Finder: A Case-Based Approach to Knowledge Navigation, *Proc. IJCAI-95*, pp.2071-2072, 1995.
- [2] G.A.Miller, WordNet: A Lexical Database for English, *Commun. ACM*, **38-11**, pp.39-41, 1995.
- [3] A.Kobsa, Combining Deictic Gestures and Natural Language for Referent Identification, *Proc. COLING86*, pp.356-361, 1986.
- [4] D.B.Koons, C.J.Spaarrell, & K.R.Thorisson, Integrating simultaneous input from speech, gaze, and hand gestures, in M.T.Maybury(ed), *Intelligent Multimedia Interfaces*, pp.267-276., 1993.
- [5] J.de Kleer, An assumption-based TMS, *Artificial Intelligence*, **28**, pp.127-162, 1986.
- [6] 知野 他, 音声入出力, タッチジェスチャ入力, およびエージェントCG出力を持つマルチモーダル対話試作システム, *SLP17-20*, 情報処理学会, pp.115-120, 1997.
- [7] 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 - オフィス知識ベースとノウハウベースの構築 -, *インタラクション97*, 情報処理学会, pp.103-110, 1997.
- [8] 河野, 屋野, 池田, 知野, 仮説推論に基づくマルチモーダル入力統合方式, *インタラクション97*, 情報処理学会, pp.33-40, 1997.
- [9] D.Lenat et al., CYC, WordNet, and EDR: Critiques and Response, *Commun. ACM*, **38-11**, pp.33-48, 1995.
- [10] 杉山 他, コモンHIサービス環境の開発, 情報処理学会, 全国大会論文集 (6), pp.193-194. 1996.