

コモンHIサービス環境の開発

2W-1

杉山博史 半田豊 長谷川保 住田一男 河野恭之 新居孝章 池田朋男 金澤博史 竹林洋一
(株)東芝 研究開発センター

1. はじめに

本稿では、種々のアプリケーションからパターン認識や自然言語処理等のメディア変換機能を有機的に利用し、ヒューマンインタフェース（HI）を高度化するための「コモンHIサービス環境」（HIウェア）について述べる。

インターネットやパーソナルコンピュータの普及により、様々な形態の大量のメディア情報にアクセスしたり編集したりすることが可能となったが、現状のシステムでは複数のデータを単に入出力しているに過ぎず、ユーザへの負担が増大している。HIの高度化を考えるとビットレベルの処理に加えて、各メディアの内容を理解する音声認識や文字認識等のメディア変換処理技術が必須となる。しかしながら、これまでのパターン認識や自然言語処理のシステムは他の応用システムへの考慮が十分にされていなかったため、オープン性に欠け用途は限定されていた。

マルチタスク環境下における音声認識機能をサーバ化した使い勝手の良いマルチモーダルインタフェース^[1]やネットワーク環境下における音声認識サーバ^[2]や機械翻訳サーバ^[3]、オンライン文字認識サーバについての報告がなされているが、種々のメディア変換処理機能を包括的に扱っている例はない。

2. コモンHIサービス環境の設計方針

実際のHIウェアの応用場面を考えると、応用システムが稼動するコンピュータの構成は様々なレベルが想定される。例えば、HIウェアを高性能なサーバ機で動かして複数のクライアント機で共有する場合や（図1）、単体の計算機に全ての機能を搭載して利用する場合もある（図2）。これらのシステム構成から要求されるHIウェアの仕様を以下に示す。

- ・ネットワーク対応
- ・マルチプラットフォーム対応
- ・種々のメディアに対応
- ・複数の機能を組み合わせて利用可能
- ・既存システムやアプリケーションからも利用可能

これらの課題に対し、様々なシステム構成に柔軟に対応できるクライアント・サーバモデルによるシステム構成を採用した。

クライアント・サーバモデルによるHIウェアのシステムモデルを図1に基づき説明する。各プラットフォーム上の応用システムは、音声認識/音声合成^[4]、文字認識^[5]などのサーバ化されたHI機能と各種入出力デバイスを利用してメディア変換処理を行う^[6]。各機能は統一的なAPI(Application Programming Interface)でアクセスを可能とし、複数の機能の組み合わせ利用を可能にするように設計した。

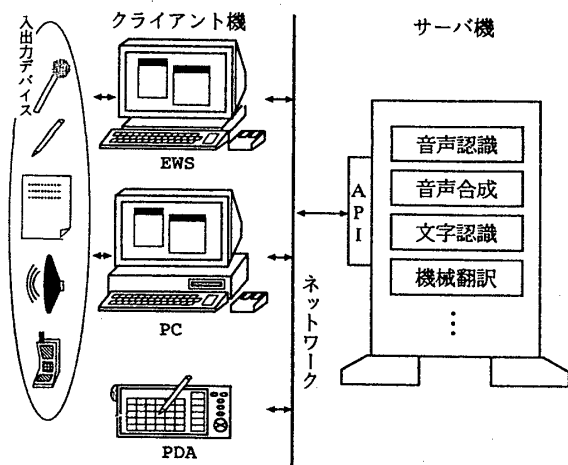


図1 ネットワーク環境下での構成

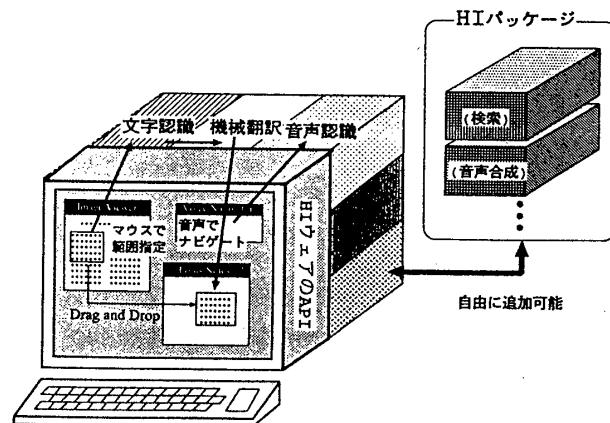


図2 マルチタスク環境下での構成

“Common Human Interface Services Environment”,
H.Sugiyama, Y.Handa, T.Hasegawa, K.Sumita, Y.Kohno, T.Nii,
T.Ikeda, H.Kanazawa, Y.Takebayashi : R&D Center Toshiba
Corp.

3. システムモデルの具体化

3.1 インタフェースの階層

個々のパターン認識や自然言語処理は扱うメディアの種類（音声、テキスト、画像等）と時間的連続性／非連続性、処理の同期／非同期等が異なるため、個別のAPIを備えるのが一般的である。アプリケーションの観点から考えると、機能の数だけAPIの種類があると複数の機能を組み合わせて利用するのは容易ではない。また、文字認識に限ってみても多くの認識アルゴリズムが提案されており、特定の認識方法に対応したAPIを直接利用するのでは認識アルゴリズムの変更や改良の度にアプリケーションの側も変更しなければならなくなる。

これらの問題を解決するため、HIウェアのインタフェースを3階層のAPIから構成した。上位層から順に

- (1) 複数の機能を組合わせて利用できるAPI
- (2) 個別機能のAPIを一般化したAPI
- (3) 個別機能毎のAPI

のようになる。(3)はエンジンのAPIであり、個々の処理やアルゴリズムに依存する。(2)は個別のAPIを一般化するために関数名や呼び出し手順を統一したAPIである。(1)は複数の機能を連携する際の同期処理などを隠蔽したAPIである。下位層になるほど細かい制御が可能であり、利用する機能レベルに応じてAPIを使い分ける。

以下に(2)の例として同期型APIを一部示す。

- ・HIウェアとの接続／切断 open/close
- ・HIウェアとの接続の確認 isopen
- ・パラメータの値の設定 set_param
- ・パラメータの値の取得 get_param
- ・メモリの解放 free_param
- ・処理要求 execute
- ・処理結果の取得 get_result

3.2 マルチプラットフォーム対応

様々なプラットフォームでの利用が可能となるように、各パターン認識と自然言語処理のエンジンは移植性の高いC言語で記述され、DLL(Dynamic Link Libraries)として実装されている。図3に示すように、ネットワーク上に分散して存在するリモートのエンジンはRPC(Remote Procedure Call)でアクセスする。

3.3 既存ソフトへの対応

応用ソフトとエンジンの間の接続は直接APIを利用する接続とOLEを介した接続の2種類がある。既

存ソフトからはAPIを直接呼べないので、Windowsの標準のアプリケーション間連携であるOLE,DDEインタフェースを備えたサーバを用意する(図3)。これにより、市販ソフトのマクロなどを利用して、HIウェアを利用するカスタムアプリケーションを簡単に構築することができる。

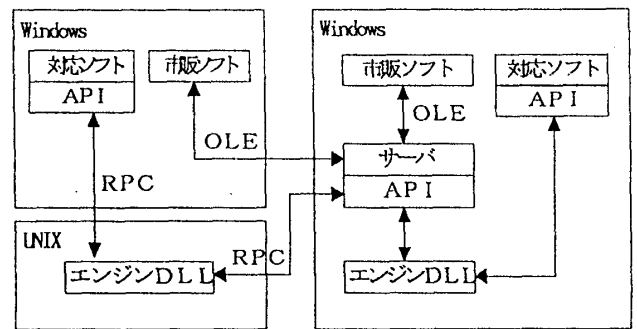


図3 システム構成

4. おわりに

ユーザが日常使用しているアプリケーションソフトから様々なHI機能を容易に組み合わせて利用可能なコモンHIサービス環境を実現した。今後、HI機能の品揃え拡充とメディア変換処理技術の高度化を図るとともに、新しい応用システムの開発に利用していく予定である。

謝辞

コモンHIサービス環境開発に際して御支援いただいた関係各位に感謝いたします。

参考文献

- [1]山田他, "実時間動作を考慮した音声認識サーバ", 音響学会講演論文集, 2-8-2, 1994
- [2]Rudnicky A. et al., "Spoken language recognition in an office management domain", IEEE, CH2977-7, 1991
- [3]伊藤, 野村, 武田, "クライアント/サーバ型機械翻訳システムの学習方式", 情処第50回全国, 5R-10, 1995
- [4]橋本他, "ワークステーションにおける音声認識インタフェースの検討", 情処学会HI研資, HC46-3, 1993
- [5]Tsujiimoto S., "Major Components of a Complete Text Reading System", IEEE, Vol.80, No.7, 1992
- [6]長谷川他, "コモンHIサービス環境の応用", 情処第53回全国, 2W-2, 1996