

ソフトウェア音声認識インタフェースの検討

1E-8

橋本秀樹[†] 永田仁史^{††} 竹林洋一^{††}

[†]東芝ソフトウェアエンジニアリング(株)

^{††}(株)東芝 研究開発センター 情報・通信システム研究所

1.はじめに

計算機に対する入力手段として、キーボードやマウス以外のペンや音声などの入力メディアが注目されてきている。その中でも音声メディアは人間にとって自然な入力手段であるため、様々な研究開発が行われてきたが、音声認識には多大の計算量が必要であり、また音声入力手段を既存のインタフェースに統合する枠組みが確立されていないなどの問題があるため、広範に普及していない。

筆者らは、既開発の高速アルゴリズムを応用した高精度なソフトウェア音声認識技術[1]に基づき音声認識機能をサーバ化することにより、音声入力メディアの統合化の問題の解決をはかった。本稿では、ソフトウェア音声認識サーバとその応用について述べる。

2.従来の音声認識インタフェース

音声入力をキーボードやマウスのように標準的な入力手段とするには、認識性能とともにインタフェースの実現方法が重要である。従来の実現方法としては、応用プログラムの中に認識機能を一体化して持つ方法、音声認識ドライバを介在させる方法、音声認識結果によってキー入力をエミュレートする方法などがあるが、いずれの方法にも音声認識機能を特定のプログラムが占有するという問題がある。マルチタスクを前提とした音声認識インタフェースに関しては、Schmandtらの Xspeak[2]や、Rudnickyらの CM-SLS[3]があるが、リアルタイム処理やパーソナルなワークステーション上での利用形態についての検討が十分ではない。

以上のことを考慮し、筆者らは高い認識性能を確保できる単語音声を認識対象とし、応用拡大が図りやすいソフトウェア音声認識技術を採用し、音声認識機能のサーバ化によって、マルチタスクへの音声入力の実現を図った。

3.音声認識インタフェースの実装

3.1.ソフトウェア音声認識

音声認識インタフェースの音声認識機能の実現には、既開発のソフトウェア音声認識[1]を応用している。ソフトウェア音声認識は、汎用ワークステーションに標準装備されているCODECを利用し、リアルタイム音声認識を、高速アルゴリズムに基づいてソフトウェアのみにより実現しているため、DSP等の音声認識専用のハードウェアが不要となる。また、複合類似度法をベースに、特定・不特定話者に対応した高精度の音声認識が可能である。ソフトウェア音声認識の応

用によって、ハードウェアに依存しない高性能で安価な音声認識機能が提供できる。音声認識インタフェースは、この技術に基づき、SPARC Station上に実装した。

3.2.音声認識サーバ

マルチタスクへの音声入力を可能にするために、クライアント・サーバモデルを応用し、音声認識機能をサーバ化した[4]。音声認識サーバには音声認識の処理機能を集中して持たせる。クライアントは、サーバと通信を行って音声認識のサービスを受ける。複数のクライアントがサーバの介在によって音声認識用の資源を共有し、音声認識機能を利用できる。

図1に音声認識インタフェースの構成を示す。複数のクライアント(AP#1-#3)が音声認識サーバに接続されている。クライアントは図2に示す手順でサーバと通信を行い、音声認識機能を利用する。図1の例では、サーバは、AP#2が依頼した認識対象語彙を利用して認識処理を行い、結果をAP#2に送信する。クライアントは認識結果をイベントとして受け、応用に従った任意の処理を行う。

認識結果の送信先は音声フォーカスによって指定する。音声フォーカスの役割は、認識対象語彙を特定クライアントに関するものに限定して認識誤りを少なくし、また認識結果の送信を特定クライアントに限定することである。図1において音声フォーカスの当たっているのはAP#2である。

応用プログラムの試作にあたり、クライアントと音声認識サーバ間の通信プロトコルを策定し、プロトコルに対応したC言語で呼びだし可能なクライアント・プログラミング・ライブラリを作成した。通信プロト

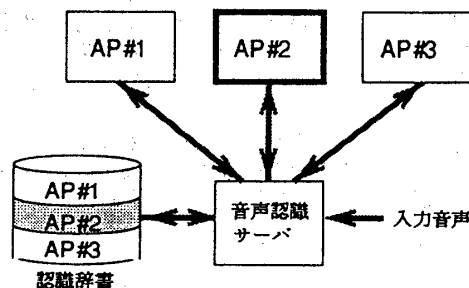


図1.音声認識インタフェースの構成

- (1)音声認識サーバとの通信路を確保する。
- (2)認識対象語彙をサーバに通知する。
- (3)認識結果の送信をサーバに依頼する。
- (4)音声認識サーバからのメッセージを待つ。
- (5)受信メッセージに従って処理を行う。
- (6)(4)以降の処理を繰り返す。
- (7)サーバとの通信路を閉じる。

図2.クライアントのサーバとの通信手順

Speech Recognition Interface for General Purpose Workstation.

Hideki Hashimoto[†], Yoshifumi Nagata^{††}, Yoichi Takebayashi^{††}

[†]Toshiba Software Engineering Co., Ltd.

^{††}Toshiba Corporation, Reserch and Development Center.

コルを標準化することにより、特定のハードウェアに依存しない音声の応用が可能である。

サーバ化に伴う音声認識機能の仮想化と、ソフトウェアのみによる音声認識機能の実現によって、ハードウェアに依存しない標準的な音声認識の利用が可能となった。

4.マルチメディア入力への対応

マルチウィンドウ環境における音声認識の自然な利用をはかるために、図3に示すように音声認識サーバとウィンドウシステムを組み合わせ実装した。

キーボードと音声という2種類の入力メディアを、1つのタスク(ウィンドウ)に結びつけるだけではなく、キーボード入力と音声入力を別個のタスクに結びつけ、複数の応用プログラムを別個の入力チャンネルを通じて同時に制御可能にした。分離した入力対象タスクをユーザが判別できるように、入力手段毎にactiveなタスクの表現を変える。図3では、キーボードフォーカスをウィンドウ枠を太くして表現し、音声フォーカスをウィンドウタイトルの色の変更で表現している。音声フォーカスの設定は、マウスおよび音声(ウィンドウ名)によって行う。

このように、音声入力対象を音声フォーカスによって管理しユーザに提供することによって、複数の応用プログラムの動作するマルチウィンドウ環境における音声入力の利用を可能とした。

5.電子メールシステムへの応用

音声認識インタフェースの評価を目的にいくつかのクライアントを試作した。そのうちここでは音声メールツールについて述べる。

音声メールツールは図4に示すようなXウィンドウベースのプログラムであり、音声操作により受信メールの内容を確認したり、返事を送信できる。

ツールの上部がリスト表示部、中央が受信メール表示部、下部が送信メール編集部である。リスト表示部において指定したメールを受信メール表示部に表示し、送信メール編集部でメールを作成し送信する。音声メールツールの認識対象語彙を図5に示す。「上司」や「緊急」は、音声マクロコマンドとして実装されており、メールのヘッダに対してボタン照合した結果を利用して受信メールのリストを限定・表示するものである。

受信したメールを検索し、返事を送るといった複雑な作業も、音声入力を使うことによって、キーボードから殆ど手を離すことなく簡単に行うことができる。

6.むすび

ワークステーションにおける標準的な音声認識機能の実装法として音声認識サーバを検討し、試作した。本稿で述べた音声認識機能のサーバ化によって、複数の応用プログラムの音声による操作が可能となった。また、実時間処理が可能な高精度のソフトウェア音声認識技術により、ハードウェアに依存しない広範な応用が可能である。今後、音声メディアを利用したより使い勝手のよいHIの確立を目指し、本システムの評価と改良を進めて行く予定である。

参考文献

- [1] 永田, 竹林: 「ワークステーションにおける音声認識機能の開発」, 信学技報, HC-9119 (1991)
- [2] C.Schmandt, M.Ackerman and D.Hindus: "Augmenting a Window System with Speech Input", IEEE COMPUTER, No.23, pp.50-58 (1990)
- [3] A.Rudnicky, J.Lunati and A.Franz: "Spoken language recognition in an office management domain", Proc. ICASSP'91, pp.829-832 (1991)
- [4] 橋本, 永田, 竹林: 「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」, 情報学会HI研究会, HI-46-3 (1993)

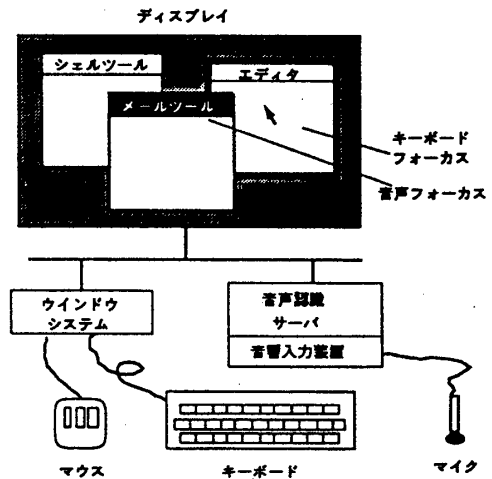


図3.マルチメディア対応の音声認識インタフェース

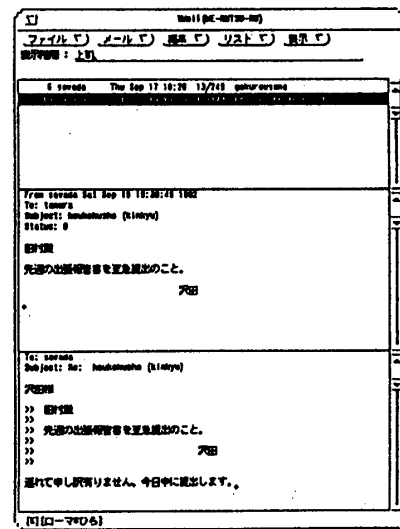


図4.音声メールツール

オープン	クローズ	セーブ	終了
送信	返事	アンドゥ	コピー
カット	ペースト	クリア	引用
サイン	前	次	先頭
最後	昨日	今日	全選択
アイコン化	同様	上司	連絡会
会議通知	緊急	本社	研究所

図5.音声メールツールの認識対象語彙例