音声言語情報処理 7－18
（1995．7．22）

音声とペンを入力手段とする マルチモーダルインタフェースの構築

菊池 英明　安藤 ハル　畑岡 信夫

(株)日立製作所 中央研究所
〒185 東京都国分寺市東恋が窪 1-280
E-mail:hkikuchi@crl.hitachi.co.jp

あらまし

本稿では，音声とペンを入力手段とするマルチモーダルインタフェースを適用した携帯情報端末のユーザインタフェースにおいて，操作性評価と，情報統合方式に関して述べる。まず，Wizard of Oz方式を用いてユーザインタフェースの操作性評価実験を行なった結果，音声とペンの複合入力形態が有効であることを確認した。さらに，実験の際に収録した被験者の操作例を分析した結果，明確になった音声入力とペン入力の非同期性の問題に対して，時間同期性を用いる情報統合方式を考案した。この情報統合方式を導入したプロトタイプシステムを構築し，評価実験を行なった結果，従来の順序情報を用いる情報統合方式と比較して性能の向上が認められた。

A Construction of Multimodal Interface using Speech and Pen as Input Modalities

Hideaki KIKUCHI, Haru ANDO, Nobuo HATAOKA

Central Research Laboratory, Hitachi Ltd.,
Kokubunji-shi, Tokyo, 185 Japan
E-mail:hkikuchi@crl.hitachi.co.jp

Abstract We have developed a multimodal window system using speech and pen as input modalities. In this paper, we describe an evaluation of usability and a method of integrating input information from speech and stylus pen. At first, we confirmed efficiency of the proposed multimodal interface by the Wizard of Oz experiment. In addition,we proposed a method for integrating input information to cope with asynchronous inputs and speech recognition errors.

1.はじめに

PCやWSにおけるユーザインタフェースとして，ウインドウやアイコンなどを用いたGUI環境を提供するウインドウシステムが広く利用されている。ウインドウシステムに対して，ユーザはマウスなどのポインティングデバイスを用いて，視覚化された文字情報を直接操作できる。しかし，キーボードを文字入力手段の主体とした従来のウインドウシステムでは，キーボードの特殊なキーとポインティングデバイスとを組み合わせた操作など，ユーザの熟練を要する操作が少なくない。また，携帯型機器にウインドウシステムを適用する場合には，キーボードの利用が困難である。そこで，操作の自然さおよび携帯性の観点から，ペンあるいは音声といった新たな入力手段が注目されている。

一方，マンマシンインタフェースの研究においては，人間が使い慣れた複数の入出力手段（モダリティ）を同時に用いることにより，より自然な入力を可能にするマルチモーダルインタフェースの研究が盛んである[1]。

本稿では音声とペンを入力手段とするマルチモーダルインタフェースを有するウインドウシステム（以下，マルチモーダルウインドウシステム）の構築と評価について報告する。なお，具体的なアプリケーションとしては，マルチメディア社会のプラットホームとして期待される携帯情報端末を選び，特に基本的なファイル管理機能をタスクとして設定した[2]。

まず，第2章ではマルチモーダルウインドウシステムの有効性の予測と操作方法分析を目的としたWizard of Oz実験について示す。次に，第3章では，実験により得た操作例を時間同期性の観点から分析し，時間同期性を用いる情報統合方式を提案する。第4章では，提案した情報統合方式を導入し構築したプロトタイプシステムの構成と評価結果について報告する。

2.マルチモーダルウインドウシステムの有効性評価

マルチモーダルウインドウシステムにおいて，発話内容，ペンのジェスチャ，および音声入力とペン入力の組合わせ方などのユーザインタフェース仕様は明らかでない。ここで，理想的なユーザインタフェースの仕様を明確にするためには，現状の音声認識の認識精度の低さによる入力効率の低下を回避し，さらにシステムの実時間応答とフレキシブルな操作を実現した擬似システムを用いて実験を行なう必要がある。従って，まず，Wizardと呼ばれる操作者がシステムの一部を代行するWizard of Oz方式[3]を用いた擬似実験にて，音声とペンの複合入力形態の有効性を評価すると同時に，ユーザ操作の様子を収録し，操作例の分析を行なった。本章では，まず，複合入力形態の有効性評価について述べる。

2.1 Wizard of Oz実験の概要

まず，Wizard of Oz実験の環境を図1に示す。図1において，被験者に操作者の存在を意識させないために，被験者と操作者は別室にする。また，操作者が被験者の意図を解釈して操作を行なうために，被験者の音声入力情報はマイクを通じて操作者のヘッドホンに出力され，被験者が電子ペンを用いて入力したジェスチャなどのペン入力情報は，被験者が用いるペンPCよりネットワークを通じて操作者側のワークステーションの画面に描画される。ワークステーションの画面には，ペンPCに表示されているウインドウシステムの画面と全く同じ状態を表示する。さらに被験者の行動はテレビカメラにより撮影され，操作者側のモニタに映される。その際，後に被験者の行動を分析するために，被験者の行動と音声をビデオテープに記録する。この様にして，操作者は，音声入力情報とペン入力情報およびモニタ情報から被験者の操作の意図を解釈し，マウスを用いてメニュー等を操作することにより，音声認識部を中心としたウインドウシステムの処理の一部を代替する。

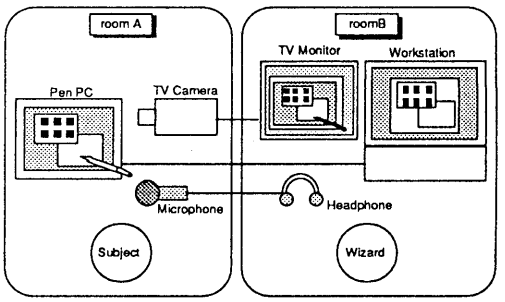


図1.Wizard of Oz実験の環境

表1.実験条件

実験方法	Wizard of Oz方式
被験者数	12名（男性6名，女性6名）
タスク	ファイルの移動/削除/複写
入力方法	音声/ペン/音声とペンの複合入力
評価項目	各入力方法の操作性　5段階
解析項目	音声とペンの複合入力形態

ここで，実験の条件を表1に示す。実験において，被験者は，ファイル管理の基本的な機能であるファイルの移動，削除，複写の3タスクを実行する。

実験の手順として，まず，できるだけ実環境での操作方法に近いデータを収集するために，実験を行う前にタスク毎の状況設定を被験者に与える。例えば，ファイルの複写をタスクとする実験では，ファイルのアイコンを表示する２枚のウインドウの他に，残りディスク容量を表示するウインドウを提示し，残りディスク容量に注意しながらファイルを一方のウインドウから他方のウインドウへ一つずつ複写させるという設定を与える。同様に，各タスク毎に状況設定を被験者に与える。

実験ではそれぞれのタスクを，ペンのみ，ペンと音声，音声のみという順に被験者に実行させる。操作者は別室で被験者の操作を観察し，被験者の意図に対応した処理をシステムに替わって行う。

また，被験者には，ペン入力用のペンを持たせ，ヘッドセットタイプのマイク（接話マイク）を付けさせる。単語発声／連続発声の発声形態や，ペンと音声の組み合わせ方など，操作方法は自由とする。

2.2 複合入力形態の有効性評価結果

各入力方法の５段階評価を元に求めた，ペン入力に対する複合入力の相対的な操作性評価を図2に示す。図2において，"使いやすさ"に関して，複合入力をペン入力より高く評価した被験者は12名中8名いた。また，各作業毎の操作性において，特にファイル指定作業とコマンド指定作業に関して，複合入力をペン入力より高く評価した被験者が過半数を占めた。これらの結果から，ペンと音声を入力手段とする携帯情報端末において，ペンと音声の複合入力形態の有効性を確認した。

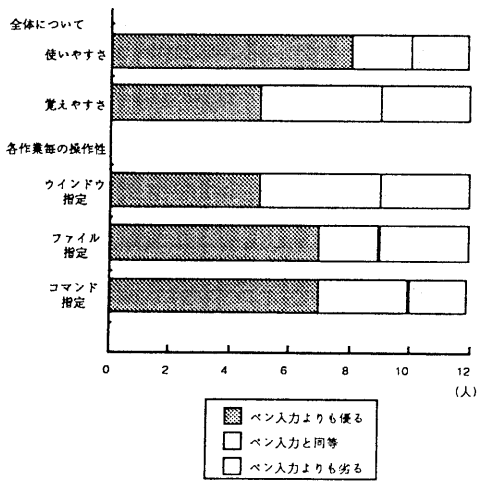


図2.ペン入力に対する複合入力の相対的操作性評価

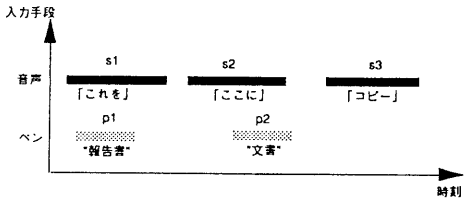
3.時間同期性を用いた情報統合方式の提案

3.1 操作例の分析結果

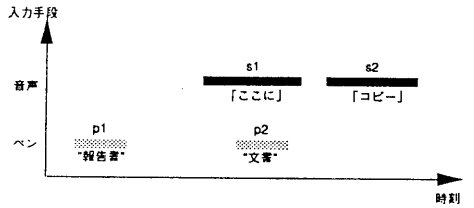
ビデオテープに記録した被験者の操作例を分析した結果，様々な操作方法が見られた。図3に操作例の一部を示す。図3(a)の例は，画面上に表示されたアイコン"報告書"をペンで指しながら，同期して「これを」と発声し，次に"文書"のアイコンをペンで指しながら「ここに」と発声し，最後に音声のみで「コピー」と発声する操作を示す。この例においては，音声の指示表現と，ペン入力が１対１に対応しているため，操作の意図を解釈する際に曖昧性は存在しない。

また，図3(b)の例は，画面上に表示されたアイコン"報告書"をペンで指し，次に"文書"をペンで指しながら同期して「ここに」と発声し，最後に「コピー」とコマンド名を発声する操作である。この場合，二つのペン入力に対して，音声の指示表現の数が一つであるため，操作の意図を解釈する際に，曖昧性が存在する。このように，一続きの操作において音声の指示表現とペン入力が１対１に対応しない例が，3タスクにおいて12人中それぞれ11，9，4例と高い頻度で存在した。

ここで，"Put that there"[4]を始めとする従来のマルチモーダルインタフェースでは，音声入力情報の指示表現とポ



(a)ペン入力が音声入力に同期しているケース



(b)ペン入力が非同期的に現われるケース

図3.Wizard of Oz実験により得た操作例

インティング入力情報の座標値を出現順序に応じて統合させる情報統合方式が用いられることが多かった。この様な手法は，２つの入力手段が常に相補的に組み合わせて用いられるという仮定に基づいている。しかし，図3(b)の例においては，音声入力が部分的に省略されているため，単純に出現順序に応じて音声入力情報とペン入力情報を統合させることができないという問題が明確になった。

3.2 時間同期性を用いた情報統合方式

3.1で述べた問題は，入力情報を出現時間情報と切り離し，かわりに近似的に出現順序を用いて解釈しようとするために生ずる。つまり，異なる入力手段による入力情報を統合して解釈するには，厳密には入力情報の出現時間情報が必要であると考えられる。従って，本研究においては，出現時間情報を用いる情報統合方式を用いる。

まず，時間的に同期して現われる音声入力情報とペン入力情報は同一対象を示すと仮定する。この仮定により，システムは入力情報の同期性を判定し，同期している音声入力情報とペン入力情報のみを統合すればよい。ここで，システムの情報統合部では時間情報を伴った入力情報に対し，図4に示すAllenの時間関係表現[5]に基づいた時間同期性の判定基準によって音声入力情報とペン入力情報の同期性を判定する。例えば，図3(b)の例の場合，時間的に同期しているs1とp2を統合し，p1とs2をそれぞれ単独に入力された情報と判断する。

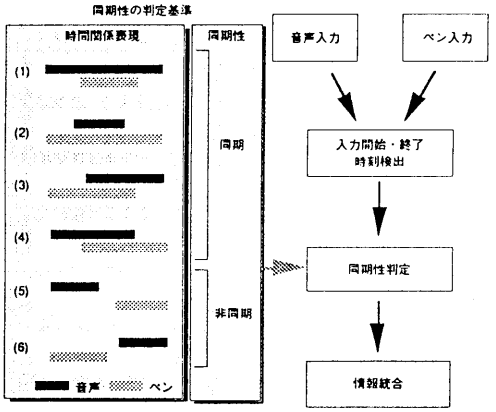


図4.同期性の判断基準と情報統合方式

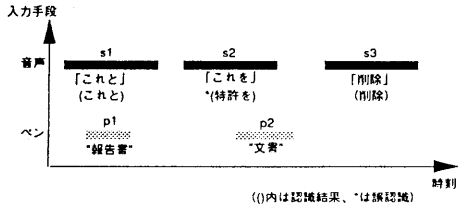


図5.音声誤認識が生じた場合の操作例

また，情報統合において，時間同期性を用いることにより，音声誤認識によって生じる曖昧性を解消することができる。例えば，図5に示す操作例において，「これとこれを削除」というユーザの発話に対し，システムが「これを」を「特許を」と音声誤認識した場合について説明する。図のs1とp1，s2とp2が時間的に同期しているため，それぞれの組み合わせについて統合を行なうが，s2の「特許」とp2の"文書"が異なる対象を指すため，前述の仮定に反する。従って，このように音声入力とペン入力が競合する場合には，モダリティの信頼性から，ペン入力の結果を用いることにする。

次章では，ここで提案した情報統合方式を導入したプロトタイプシステムの構成と，情報統合方式の評価について述べる。

なお，時間情報を用いた情報統合方式として，時間的距離を用いた方式[6]や，時間的交差をルールにより判定する方式[7]が提案されているが，いずれも評価結果を得るまでには至っていない。

4.マルチモーダルウインドウシステムのプロトタイプ構築と評価

4.1 システム構成

図6に示した構成のプロトタイプシステムを構築した。

本システムでは，ワークステーションを音声認識用に１台，アプリケーション駆動用に１台使用している。音声認識部(Speech Recognizer)として連続音声認識サーバ[8]を用い，認識可能語彙数は43単語，発声形態は離散発声とした。また，ペンによる入力情報を取り込むために，液晶タブレットと電子ペンを入力デバイスとするペンPCを用いる。

音声認識部が出力する認識結果は，各発話の入力開始時刻と終了時刻とともにネットワークを介して情報統合部(Information Integrator)に送られる。また，ペンの入力イベントは，ペンPCからネットワークを介してウインドウマネージャ，情報統合部に送られる。情報統合部は，音声入力情報とペン入力情報を統合して，ユーザの操作意図を解釈し，コマンドに変換した上でウインドウマネージャにコマンドを実行させる。

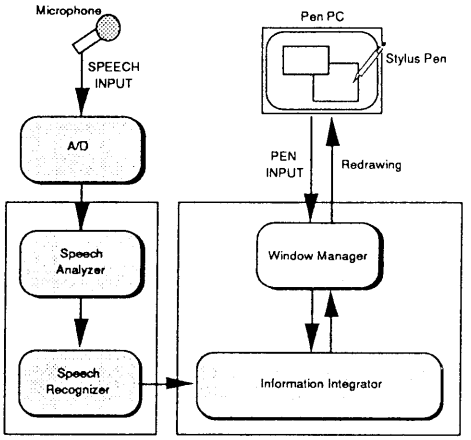


図6.システム構成図

4.2 時間同期性を用いた情報統合方式の評価

3.2において提案した情報統合方式の性能を評価するために，5人の被験者による評価実験を行なった。5人の被験者がファイルの移動，削除，複写の3タスクを行なった後，各タスクにおいて，記録した操作方法について情報統合成功／失敗の割合を解析した。解析した結果を従来の方式での結果とともに，図6に示す。

この図において，従来方式の性能評価として，2.2に述べた非同期的な入力を含む操作に対して，情報統合失敗とした。このようなケースは，のべ15操作において12回見られた。一方，提案した方式では，このうち11回において情報統合に成功しており，少ないサンプル数であるが，大幅に性能が向上したと言える。

なお，提案した方式は，非同期的な入力を含む操作，含まない操作のそれぞれにおいて1回ずつ情報統合に失敗しているが，この原因は，いずれも一つのペン入力情報が二つの音声入力情報と時間的に交差して現われ，時間

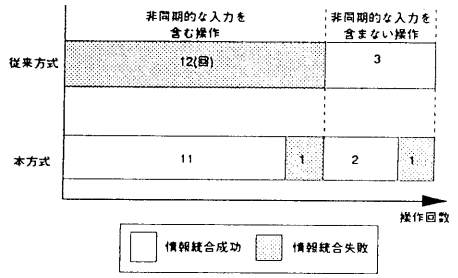


図7.情報統合方式の性能評価結果

同期性を用いて統合する際に曖昧性を解消できなかったためである。今後，高次の意味処理を導入する等の対策を検討すべきと考える。

5.まとめ

本研究では，音声とペンを入力手段とするマルチモーダルインタフェースのプロトタイプシステムの構築を行なった。音声とペンを組み合わせて用いる複合入力形態の有効性を確認し，さらに非同期的な入力に対処するために時間同期性を用いた情報統合方式を提案した。提案した方式を導入したプロトタイプを構築し，情報統合方式の性能に関して評価した結果，大幅な向上が見られた。今後は，複雑なタスクにおけるシステムの有効性評価，及びプロトタイプの拡張を行なう予定である。

参考文献

[1]例えば，安藤,北原,畑岡:"音声とポインティングジェスチャを入力手段としたインテリアプランニング支援システムの評価,"第9回HIシンポジウム予稿集，pp.37-42(1993)

[2]菊池,安藤,畑岡:"マルチモーダルウインドウシステムの構築,"第10回HIシンポジウム予稿集，pp.547-554(1994)

[3]J.M.Francony et.al.:Towards a methodology for Wizard of Oz experiments, Third Conference on Applied Natural Language Processing, Tront,Italy, 31 March-3 April, pp.277-296(1992)

[4]R.A.Bolt:"'Put-that-there,'Voice and Gesture at the Graphics Interface,"Computer Graphics, vol.14, no.3, pp.262-270(1980)

[5]J.F.Allen:"Maintaining Knowledge about Temporal Intervals",Comm.of the ACM,vol.26,No.11,pp.832-843(1983)

[6]Y.Bellik,D.Teil:"A Multimodal Dialogue Controller for Multimodal User Interface Management System Application:A Multimodal Window Manager," INTERCHI '93,Amsterdam,24-29 Apr 1993

[7]L.Nigay,J.Coutaz:"A Generic Platform for Addressing the Multimodal Challenge," CHI'95,pp.98-105(1995)

[8]天野他：音声対話システムの試作，日本音響学会，講演論文集，1-1-20(1992)