音声言語情報処理 7−17
（1995. 7. 22）

マルチモーダルインタフェースを持つ住所入力システムの評価実験

荒井和博 吉岡 理 管村 昇 嵯峨山 茂樹

ＮＴＴヒューマンインタフェース研究所

〒238-03 神奈川県 横須賀市 武 1-2356

arai@nttspch.hil.ntt.jp

概 要 本稿では，マルチモーダルインタフェースを持つ住所入力システムの評価実験結果について報告する．実験では25名の被験者に単語音声認識機能を持つマルチモーダルインタフェースと，音声認識処理ができない以外は全く同一のユーザインタフェースとを利用させ，それぞれにおいて合計約1万1千回の住所入力を行なわせた．実験により 1) 音声認識を用いた場合，候補単語数の操作時間に対する影響は少なく，操作時間を約12%削減できる，2) 約12%の発声に言い淀み，未知語発声などが現れ，83.6%の認識率が得られた などの結果が得られた．

Evaluation of an Address Entry System Utilizing a Multimodal User Interface

Kazuhiro ARAI, Osamu YOSHIOKA, Noboru SUGAMURA, and Sigeki SAGAYAMA

NTT Human Interface Laboratories

1-2356 Take, Yokosuka-shi, Kanagawa 238-03, JAPAN

arai@nttspch.hil.ntt.jp

Abstract This paper describes a performance evaluation for an address entry system utilizing a multimodal user interface. Twenty-five subjects used two user interfaces, one utilizing and the other excluding isolated word speech recognition in order to evaluate the time necessary for address data entry in both interfaces. Approximately eleven thousand entries for each interface revealed the following results: 1) Speech recognition decreased the dependency of time on the vocabulary size and reduced by 12%, the overall time necessary to enter data. 2) A rate of 83.6% speech recognition has been achieved for speech data containing 12% of inappropriate utterances.

1 はじめに

音声認識技術の発達に伴い, 音声認識を計算機端末上で利用するための研究が進められている [1]-[6]. マルチモーダルインタフェースに関する検討は, 音声認識の計算機端末上での利用を実現する上での中心的課題といえる. マルチモーダルインタフェースの構築により, キーボードやマウスといった既存の入力手段と音声入力とが相補的に統合され, 効率的なユーザインタフェースが実現されると考えられる. 本稿では, 我々がこれまでに開発してきた住所入力のためのマルチモーダルインタフェース [7]-[11] の評価実験とその結果について報告する.

マルチモーダルインタフェースの評価実験は, これまでにもいくつかの機関で行なわれているが [2]-[5], 現在までのところ少数の被験者が行なった実験の報告にとどまっている. また, 計算機端末にマルチモーダルインタフェースを導入した場合の効果を測る観点としては, マルチモーダルインタフェース構築の目的により, 操作時間, 操作回数, タスク達成率, ユーザ選好度などさまざまな尺度が用いられている.

本研究では, 計算機端末にマルチモーダルインタフェースを導入した場合にデータ入力の効率化がどの程度図られるのかに注目して評価実験を行なった. 実験では, 25 名の被験者に単語音声認識機能を持つマルチモーダルインタフェースと, 音声認識処理ができない以外は全く同一のユーザインタフェースとを利用させ, それぞれにおいて合計約 1 万 1 千回の住所入力を行なわせた. また, 本実験では住所入力に対する効率を測る目的から, 評価の尺度として操作時間を用い, それぞれのユーザインタフェースを用いた場合の操作時間を比較した. 更に, 1) 音声入力操作誤り, 2) 発声誤り などについても分析を行なった.

以下, 2 章では住所入力システムの概要について述べる. 3 章では操作時間の比較を目的とした評価実験の条件について詳述する. 4 章では実験結果を示し, 操作時間の比較, 音声認識率について考察する. 更に, 4 章では, 音声入力操作誤り, 発声誤りなどに関する分析結果についても言及する.

2 住所入力システム

電話番号の市外・市内局番は, 住所に基づいて割り当てられている. 本システムは, 入力すべき住所に必ず電話番号が併記されていることを前提としており, 電話番号の市外・市内局番と住所の関係を記述したデータベースを利用している.

電話番号	03-3206-2314
住　　所	東京都 中央区 新富
じゅうしょとうきょうと ちゅうおうく しんとみ	

図 1: 被験者に与えた問題の例

住所入力の際には, まず住所に併記されている電話番号を入力し, 電話番号の市外・市内局番から住所の候補を検索する. 検索された住所候補は, 都道府県, 市区郡, 町村などの住所区分ごとに住所単語に分解され, それぞれの入力欄へ候補単語として格納される. 候補単語の選択は, 1) マウス操作によるメニュー選択, 2) キーボード操作による候補単語選択, 3) 単語音声認識による候補単語選択 の各手段によって行なわれる.

認識処理は, 音声認識サーバ [12][13] において行なわれる. 音声認識は, 認識対象語彙, 音声データの送信, 及び認識結果の受信により実行される. また本システムは, 候補単語間の階層関係を利用してユーザが確定した単語を基に他の入力欄の確定, 候補単語の絞り込みといった入力補完処理を行ない, ユーザ操作回数の低減を図っている.

3 評価実験

音声認識の導入により, 住所入力に必要な時間が短縮されるか否かを明らかにすることを目的に評価実験を行なった. 実験では音声入力の経験を持たず, 計算機へのデータ投入を業務とする 25 名の女性を被験者とした. 評価実験では各被験者に単語音声認識機能を持つマルチモーダルインタフェースと, 音声認識処理ができない以外は全く同一のユーザインタフェースとを利用させた. 各被験者には, 決められた時間ごとに 2 つのユーザインタフェースを交互に利用させ, 音声認識に関する経験が実験結果に及ぼす影響を相殺するよう配慮した.

被験者に与えた問題の例を図 1 に示す. 問題には入力すべき電話番号, 住所が表示してあり, 住所には読みが付与されている. 被験者は与えられた電話番号, 住所を入力するものとし, 音声認識機能を持つマルチモーダルインタフェースを利用する

表 1: 被験者ごとの住所入力回数

被験者番号	1人あたりの住所入力回数	人数	計
1, 2	225	2	450
3-6	465	4	1,860
7-25	480	19	9,120
合計	—	25	11,430

場合には，必ず少なくとも一度は音声入力を行なうよう指示した．また，メニューに表示される候補単語数は 10 とし，候補単語数が 10 を越える場合はマウス操作により 10 づつ表示するようにした．

実験は 1) 操作説明・練習，2) 本実験 の順に行なった． 操作説明・練習は被験者ごとに 1 時間とし，音声認識機能を持つマルチモーダルインタフェースを利用して 24 住所の入力を行なわせた．操作説明・練習及び本実験は，電話応対が頻繁に行なわれているオフィス環境下で行なわれ，いずれの場合においても，操作説明者が被験者からのシステムに関する質問に応じた．本実験において被験者が入力した住所数を表 1 に示す． 住所リストの作成に際しては，電話番号の市外・市内局番から検索される住所候補数の分布比率が，関東 1 都 6 県における分布比率と同じになるよう配慮した．本実験において被験者は，3 日間で各ユーザインタフェースごとに 11,430 住所を入力した．被験者への住所の提示順序は，ユーザインタフェース及び被験者間でランダムとし，住所の提示順序が実験結果に与える影響を無くすよう配慮した．

本実験における被験者の入力音声は音声データファイルに保存された．また被験者のマウス，キーボード操作及び音声認識結果は，各操作が行なわれた時刻と共に操作履歴ファイルに保存された．

4 実験結果

操作履歴ファイルに保存された操作の分析及び音声データの聴取により，操作時間及び音声認識率を求め，発声内容の分析を行なった． 以下ではこれらの結果について述べ，結果の考察を行なう．

4.1 操作時間の比較

本システムにおける住所入力操作は，1) 電話番号の入力，2) 各入力欄ごとの候補単語の選択 の順で行なわれる (図 2 参照)． 各入力欄においては，

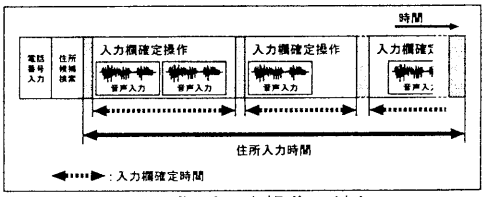

図 2: 住所入力操作の流れ

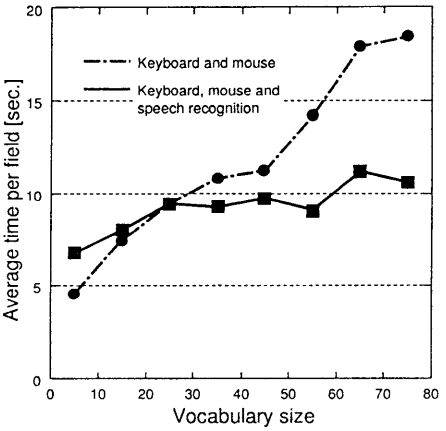

図 3: 各入力欄を確定するための時間

誤認識などにより音声認識が複数回利用される場合もある．本実験では 1) 各入力欄を確定するための時間 (以下，入力欄確定時間と呼ぶ)，2) 住所入力を完了するための時間 (以下，住所入力時間と呼ぶ) の 2 つの観点で操作時間を比較した．

各入力欄に割り当てられる候補単語数は，与えられた電話番号や他の入力欄が確定されているか否かによって変化する [7]． このため，候補単語数と操作時間の関係を得るには，入力欄確定時間を比較する必要がある．一方，住所入力時間は，住所の階層構造を利用した候補単語の絞り込みなど本システム固有の処理を反映しており，本タスクに対する音声認識の有効性を測る目的から比較した．図 3 に各候補単語数における入力欄確定時間を示す． また，電話番号から検索される住所候補数に対する住所入力時間を図 4 に示す．11,430 住所に対する住所入力時間の平均値は，音声認識を用いた場合 13.9 秒，また用いなかった場合 15.8 秒となり，関東 1 都 6 県を対象とした今回の実験では操作時間が約 12% 削減できた．

図 3 及び図 4 より，音声認識を用いなかった場合，候補数の増加に従っていずれの操作時間も

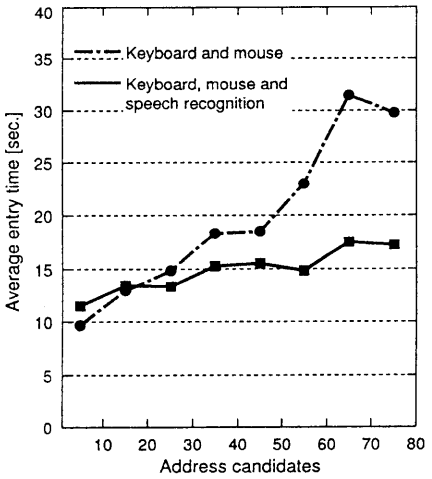

図 4: 住所入力を完了するための時間

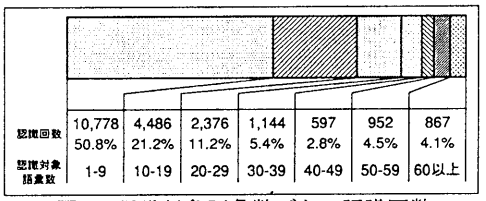

図 5: 認識対象語彙数ごとの認識回数

増加していることが分かる．この原因は以下のように考えられる．即ち，キーボード操作によって候補単語の中から目標の単語を選択する場合には，キーを押下するごとに候補単語を 1 つづつ入力欄に表示させた．また，メニュー操作によって選択する場合には，10 の候補単語を表示し，候補単語数が 10 を越える際にはマウス操作により 10 づつ表示するようにした．このような候補単語の表示・入力方法は広く一般に採られているが，目的とする単語を選択するために多くの操作及び確認が必要となり，候補数の増加に従って操作時間が増加する．

一方，音声認識を用いた場合，尤度順に並べられた認識結果を確認するので候補単語数が多い場合でも少数の候補単語を扱う場合と同程度の操作時間で入力欄を確定できる．即ち，音声認識の 1 位認識結果が正解であった場合には，入力欄に目的の単語が表示される．また，1 位認識結果が正解でなかった場合でも，尤度の順に従って認識結果が表示されるので，少数の上位認識結果を確認するだけで目的の単語をキーボード・マウス操作により選択・確定することが可能である．

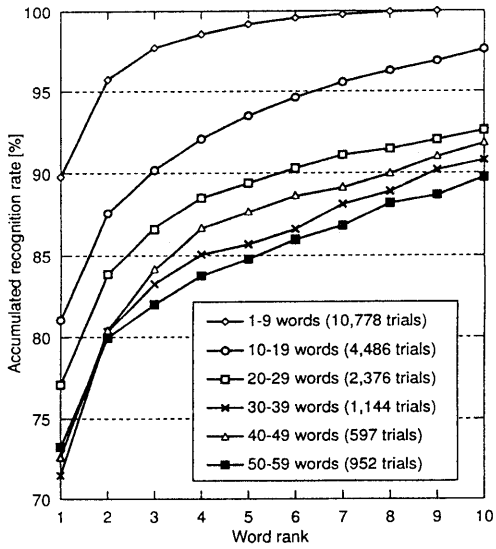

図 6: 認識対象語彙数ごとの認識率

4.2 音声認識率

音声入力においては，認識対象語彙外の発声，言い直し，言い淀みあるいは語頭，語尾の途切れなどさまざまなノイズが現れる．認識アルゴリズムの性能評価を行なう場合，認識対象語彙の 1 つが発声され，しかもノイズを含まない音声データファイルのみを対象として認識率を計算する方法が考えられる．しかし本実験では，音声認識のユーザインタフェースとしての性能を測る目的から，保存された音声データファイルのすべてを対象として認識率を計算した．

認識対象語彙数ごとの認識回数を図 5 に示す．図 5 では，認識対象語彙数が 1[1]から 9 まで,10 から 19 までといった区分で認識回数の比率を示している．図 5 で用いている認識対象語彙数の区分に沿って求めた認識対象語彙数ごとの認識率を図 6 に示す．各区分ごとの 1 位認識率と認識回数から求めた平均 1 位認識率は 83.6% であった．図 5 より，本実験では 50 単語未満の音声認識が認識処理の 90% 以上を占めている．50 単語

[1]候補単語数が 1 であっても，被験者が音声入力を行なえば認識処理が行なわれたものとみなした．

表 2: 音声データの分類区分

区　分	定　　義
順位内	認識対象語彙を発声し, 目的の単語が認識結果として得られた.
順位外	認識対象語彙を発声したが, 目的の単語が認識結果として得られなかった. (例) 目的の単語に対する尤度が低い
リジェクト	認識結果が全く返ってこなかった.
未知語	認識対象語彙外の単語発声. (例) 漢字の読み間違い, 濁音・半濁音と清音の取違い
発声誤り	誤った発声. (例) 語頭, 語尾の途切れ, 言い直し, 言い淀み

表 3: 音声データに含まれるノイズ等

語中ポーズ	リップノイズ
聞き取り難い発声	呼吸音
無音	

未満の音声認識の場合, 10 位以内の認識率はほぼ 90% であり, マウス, キーボード操作によって誤認識の訂正を行なえば, 少数の候補単語を扱う場合と同程度の操作で入力欄の確定ができたと考えられる.

4.3 発声内容の分析

収録された音声データの聴取により, 全音声データを表 2 に示す 5 つの区分に分類した. また, 表 3 に示すノイズ等が聴取された場合には, 対応するラベルを音声データに付与した. 音声データの聴取作業は, 1 名の作業者が行なった. 「未知語」, 「発声誤り」の区分, 及び各種ノイズに対するラベル付与は, 聴取による判断に基づいて作業者が行なった. また「順位内」, 「順位外」及び「リジェクト」への分類は, 操作履歴ファイルに記録されている認識結果と発声内容を照合することにより行なわれた. 発声がまったくおこなわれておらず背景雑音のみが収録されている音声データファイルは, 発声誤りに分類された. 表 2 の区分に従って分類した音声データを図 7 に示す. 図 7 より約 82% の音声入力が認識処理されており, また約 12% の音声入力は未知語や発声誤りといったユーザ側の誤りに起因するものであった.

表 3 の区分に従ってラベル付与された音声データ数を図 8 に示す. 図 8 において, 例えば「呼吸音」のグラフは, 「呼吸音」のみ, あるいは「呼吸音」と他のノイズを含む音声データファイルの数が 4,397 であったことを示す. 更にこれらノイズが認識率に及ぼす影響を測るため, 認識対象語彙を発声し認識結果の得られた発声 (「順位内」に分類) であって, 表 3 に示したノイズを含む音声データを対象として音声認識率を求めた. 各ノイズを含む音声データを対象とした 1 位認識率及び 10 位までの累計認識率を図 9 に示す. 図 8 と図 9 から以下の知見が得られる.

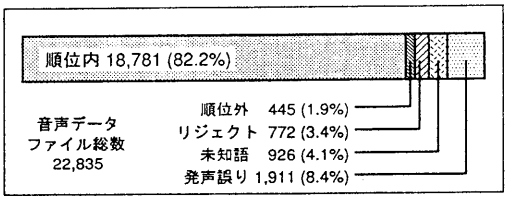

図 7: 音声データの分類

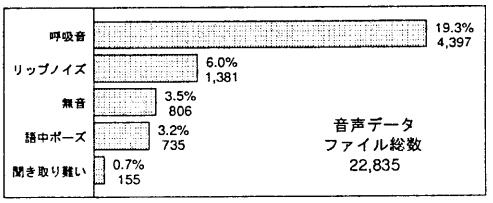

図 8: ノイズを含む音声データの比率

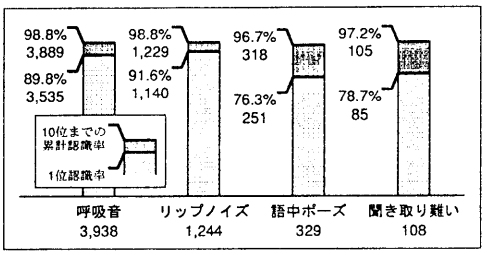

図 9: 「順位内」に分類され
ノイズを含む音声データの認識率

約 19% の音声データに「呼吸音」が現れているが, 「呼吸音」を含む音声データの 1 位認識率は 89.8% であり, 全音声データを対象に求めた平均 1 位認識率 83.6% (4.2節参照) より高い. また, 「リップノイズ」を含む音声データの認識率もほぼ同様の傾向を示している. これらの結果から, 本実験において現れた「呼吸音」及び「リップノイズ」は, 認識率に大きな影響を及ぼさないノイズと考えられる. 一方, 「語中ポーズ」や「聞き取り

難い」を含む音声データでは1位認識率が著しく低下しているが，これらのノイズが含まれた音声データが現れる比率は低い．

5 おわりに

本稿では，計算機端末にマルチモーダルインタフェースを導入した場合にデータ入力の効率化がどの程度図られるのかに注目して行なった評価実験について報告した．評価実験では住所入力に対する効率化を測る目的から，評価の尺度として操作時間を用いた．それぞれのユーザインタフェースを用いた場合の操作時間の差を比較対象とし，また発声内容の分析もおこなった．評価実験の結果，以下の点が明らかとなった．

1. 候補単語数が30以上の際に音声認識を用いた場合，操作時間が短縮できる．
2. 音声認識を用いた場合，住所候補数及び単語候補数の操作時間に対する影響は少ない．
3. 関東1都6県を対象とした住所入力の場合，全体で操作時間が約12%削減できた．
4. ノイズ，発声誤り等を含む全音声データを対象とした平均1位認識率は，83.6%であった．
5. 全音声データの約12%に言い淀み，未知語の発声などユーザの誤りが含まれていた．
6. 「呼吸音」，「リップノイズ」は発声中によく現れるが音声認識性能への影響は少ない．一方，「語中ポーズ」及び「不明確な発声」を含む音声データでは平均に比べ1位認識率が約5～7%低下する．

今後は，1) 候補単語の先頭カナ1文字を用いた候補の絞り込み，2) 連続単語音声認識による複数入力欄への同時入力 などの検討を行ない，他のタスクへの適用を検討する．

謝辞

本実験に御協力頂いたＮＴＴテレマーケティグ株式会社 有山裕孝氏，寺山貞光氏，小玉佳宏氏並びに被験者になって頂いた皆様に深謝致します．単漢字辞書構築に尽力頂いたＮＴＴアドバンステクノロジ株式会社 角田美幸氏並びに音声データファイルの聴取確認作業に尽力頂いた庄司貴代美氏に深謝致します．研究の機会を与えて頂いたヒューマンインタフェース研究所音声情報研究部北脇信彦部長と本システム構築に関して御討論頂いた音声情報研究部の諸氏に感謝致します．

参考文献

[1] Robert Strong, "CASPER: A Speech Interface for The Macintosh," Proceedings of *European Conference on Speech Communication and Technology, EuroSpeech'93, Vol.3, pp.2073-2076*, (1993.Sep).

[2] Lewis R. Karl, Michael Pettey, and Ben Shneiderman, "Speech versus mouse commands for word processing: an empirical evaluation," *International Journal of Man-Machine Studies*, **39**, *pp.667-687*, (1993).

[3] Alexander I. Rudnicky, "Factors Affecting Choice of Speech Over Keyboard and Mouse in A Simple Data-Retrieval Task," Proceedings of *European Conference on Speech Communication and Technology, EuroSpeech'93, Vol.3, pp.2161-2164*, (1993.Sep).

[4] 志田 修利, 西本 卓也, 小林 哲則, 白井 克彦, "音声・マウス・キーボードを併用した作図システムS-tgifとその評価", 信学技報, **SP94-29**, (1994.Jun).

[5] 野口 淳, 坂井 信輔, 畑崎 香一郎, 磯 健一, 渡辺 隆夫, "パソコン音声認識ソフトウエアを用いた音声ダイヤラの試作", 信学技報, **SP94-54**, (1994.Nov).

[6] 坂井 信輔, 畑崎 香一郎, 渡辺 隆夫, "音声入力を用いたパソコンネット旅客機空席案内システム", 日本音響学会 平成7年度 春季研究発表会 講演論文集, **3-P-25**, *pp.211-212*, (1995.Mar).

[7] 荒井 和博, 吉岡 理, 嵯峨山 茂樹, 山田 智一, 野田 喜昭, 井本 貴之, 管村 昇, "音声認識機能を持つ住所入力システム", 情報処理学会 音声言語情報処理研究会, **5-10**, (1995.Feb).

[8] 荒井 和博, 吉岡 理, 嵯峨山 茂樹, 山田 智一, 野田 喜昭, 井本 貴之, 管村 昇, "音声認識機能を持つ住所入力システム", 電子情報通信学会 総合大会講演論文集, **SD-9-7**, *pp.379-380*, (1995.Mar).

[9] 吉岡 理, 荒井 和博, 嵯峨山 茂樹, 山田 智一, 野田 喜昭, 井本 貴之, 管村 昇, "音声入力機能を持つ住所入力システム", 日本音響学会 平成7年度 春季研究発表会 講演論文集, **3-P-26**, *pp.213-214*, (1995.Mar).

[10] Kazuhiro Arai, Osamu Yoshioka, Shigeki Sagayama, and Noboru Sugamura, "A Prototype of an Address Input System with Speech Recognition," Proceeding of *ESCA Tutorial and Research Workshop on Spoken Dialogue Systems -Theories and Applications-, pp.213-216*, (1995.Jun).

[11] Kazuhiro Arai, Osamu Yoshioka, Sigeki Sagayama, and Noboru Sugamura, "An Operation Analysis of An Address Input System with Speech Recognition," Proceedings of *International Conference on Human-Computer Interaction, HCI'95*, Submitted for publication, (1995.Jul).

[12] 山田 智一, 野田 喜昭, 嵯峨山 茂樹, "実時間動作を考慮した音声認識サーバ", 日本音響学会 平成6年度 秋季研究発表会 講演論文集, **2-8-2**, (1994.Oct).

[13] 山田 智一, 野田 喜昭, 井本 貴之, 嵯峨山 茂樹, "クライアント・サーバ構成のHMM-LR連続音声認識システムとその応用", 情報処理学会 音声言語情報処理研究会, (1995.Feb).