連続する誤認識時に音声入力からキー入力に切り換える最適な時点の決定法

中谷吉久† 守屋慎次††

音声認識システムが利用可能な対話型システムにおいて，音声入力時に誤認識が連続する場合，①音声入力からキー入力に切り換える最適な（すなわち，入力に要する時間を最短にする）時点が存在するか否か，②もしも最適な時点が存在するならば，それは何回目の発声か，を実験を通じて調べた．さらに，③その最適な時点は前もって予測が可能か否か，を検討した．結果として，①音声入力からキー入力に切り換える最適な時点は存在し，②それは利用者により異なることを明らかにした．また，③最適な切換え時点を予測する方法を提案し，その予測法によると，ある程度予測が可能であることを示した（被験者8人のそれぞれについて所要時間を予測し，その“誤差率”を求めたところ，2回目の発声の時点以降に切り換える場合の誤差率の平均値は10%以下だった）．この予測法は，以下のことに利用できる．利用者の“積算認識率”，音声入力時間，キー入力時間，および切換えのための移行時間を知ることができるならば，①利用者に対しては，その利用者に固有の音声入力からキー入力に切り換える最適な時点を予測し，教えることができ，②ユーザインタフェース設計者に対しては，設計時に，設計しようとする操作に音声を用いるべきか否かを決定する指針を与えることができる．

The Most Suitable Timing for Switching from Voice Input to Key Input during Consecutive Misrecognition

Yoshihisa Nakaya† and Shinji Moriya††

In this paper, we show through our experiments that there exists the most suitable timing for switching from voice input to key input during consecutive misrecognition and that the timing depends on users. Besides, we describes a method to estimate the timing. Through the method, a user can obtain his/her most suitable timing which minimize the time required to complete an input-task and a designer can obtain guidelines to take advantage of a voice recognition system.

1. はじめに

対話型システムにおいて，音声認識システムを有効に利用するためには，

（1） 音声認識システムにおける認識の速度および認識率をいかに上げ，それをどう評価するか，

（2） 音声認識システムをどこにどのように応用するか，

が重要な課題である．本論文は，上記(2)に焦点をあてる．

キーまたはマウスを使うより音声を使った方が操作に要する時間が短い場合がある，という実験結果が報告されている．たとえば，以下に示す場合である．

（1） コマンドをショートカットする[1]．

（2） 隠れたウィンドウを最前面に出す[2]．

（3） 単語単位や句単位で文章を入力する[3]．

しかし音声を使う際に，誤認識が連続して起こる場合には，キーやマウス等の他の入力機器を使わざるを得ない．

本論文では，音声使用時に誤認識が連続する場合，

① 音声からキーに切り換える最適な（すなわち，入力に要する時間を最短にする）時点が存在するか否か，

② もしも最適な時点が存在するならば，それは何回目の発声か，

を実験を通じて調べる．さらに，

③ その最適な時点は前もって予測が可能か否か，

を検討する．

† 神奈川県工業試験所技術管理部電子計算科
Computer Section, Technological Administration Division, Industrial Research Institute of Kanagawa Prefecture
†† 東京電機大学工学部情報通信工学科
Department of Information and Communication Engineering, Faculty of Engineering, Tokyo Denki University

文献1)では，キーのみを使う場合とキーと音声を併用する場合とを比較し，コマンドのショートカットを音声で実行することが有効であることを報告している．しかしそこでは，キーと音声を併用する場合，各被験者間で条件を統一するためにつぎのような制約がつけられていた．

(a) コマンドの入力には必ず音声を使う．

(b) 誤認識が3回連続したらキーを使う．

ここで，上記(b)の制約をはずしたら，すなわちコマンドの入力に際して誤認識が連続した場合に，キーに切り換える時点を4回目にするのではなく，他の時点（たとえば，3回目や5回目等）にしたら，音声の有効性の度合いがさらに高かったかもしれない．

文献2)においてSchmandtらは，被験者に2か月間にわたってウィンドウシステムを使用させ，どのような状況で被験者は音声を用い，どこで問題が生じ，ウィンドウ操作にどのような影響をもたらしたかを調べた．結果として，一部分もしくは全体が隠れているウィンドウに対しては，音声の方が速く操作できたこと，すなわち所要時間が短かったことを報告している．しかし，誤認識が連続した場合に何回目に音声からキーに切り換えたか，およびそれが所要時間にどう影響したかについては述べていない．

文献3)においてSchurickは，文章，数値，コマンドを文字単位，単語単位，句単位で入力する際のキー入力と音声入力を，所要時間と誤りの数で比較した．そして音声入力は，単語単位や句単位による文章の入力には有利であるが，他の場合には音声入力の利点は見られなかったと報告している．しかし，ここでもやはり誤認識が連続した場合については述べていない．

文献4)においてRudnickyらは，誤認識の結果を訂正する方法が作業の効率にどう影響するかを示すために，被験者に数字列（長さが1, 3, 5, 7, 9, 11の6種類）を入力させる実験を行った．その実験では，つぎの3つの入力方法が採用された．

① 音声のみで入力する．認識結果が正しくないときは，正しい認識結果が得られるまで，その入力を繰り返す．

② 音声で入力するが，認識結果が正しくないときはキーで訂正する．

③ キーのみで入力する．誤ってキー入力したときもキーで訂正する．

これら3つの入力方法を，入力動作，訂正動作，および確認動作のそれぞれに要した時間で分析した．結果として，入力動作に要する時間は，音声の方がキーより比較的短かった．しかし訂正動作に要する時間については，キーの方が比較的短かった．確認動作においては，キーの方がかなり速かった．全体の（すなわち，正しい数字列を入力し終わるまでの）時間は，数字列の長さが1, 3, 5の場合は，上記③の方法が最も所要時間が短く，数字列の長さが7, 9, 11の場合は，上記②の方法が最も所要時間が短かった．

上記②の入力方法は，音声による入力が誤認識だったときに，2回目の入力の時点でキーに切り換える場合である．ここでもやはり，音声からキーに切り換える時点を2回目にするのではなく，他の時点にしたら，所要時間がさらに短かったかもしれない．

本論文では，つぎのことを明らかにした．

① 音声からキーに切り換える最適な時点は存在する．

② 最適な時点は利用者により異なる．

③ 最適な時点の予測はある程度可能である（被験者8人のそれぞれについて所要時間を予測し，その誤差率（4.2節）を求めたところ，2回目以降に切り換える場合，誤差率の平均値は10%以下だった）．

音声認識システムを使用する際に誤認識が連続する場合，他の入力機器へ切り換える最適な時点が存在することを明らかにし，またその最適な切換え時点を求める方法および前もって予測する方法を提示したのは本論文が最初である．

本研究で得られた成果は，以下のことに利用できる．

(1) 利用者に対しては，その利用者に固有の最適な（すなわち，入力に要する時間を最短にする）切換え時点を教えることができる．

2) 音声を用いるユーザインタフェース設計者に対しては，どのような操作を音声で入力したらよいかを決定する指針を与えることができる．

2章では本論文で使用した音声認識システムの概要を述べ，3章では実験とその結果を述べる．4章では予測の考え方とその検証の結果を述べる．5章で本論文のまとめを述べる．

2. 使用した音声認識システムの概要

本論文で使用した音声認識システムは，パーソナルコンピュータ（NEC製PC 9801）用に開発された市

販の音声認識ボード（(株)リコー製)[5]とその制御用ソフトウェア，およびマイクロフォンから成る．この音声認識システムが認識できるのは，あらかじめ登録されている特定話者の発声語彙（発声に用いる語の集まり）で，登録可能な語数は最大 255 語である．1語は2秒以内に発声し終える長さでなければならない．なお，この制御用ソフトウェアは，このパーソナルコンピュータ上のワードプロセッサ「一太郎 Ver. 4[6]」または他のアプリケーションが稼動中に，そのキーストローク列をエミュレートするように設計されている．

この音声認識システムの使用法を以下に一太郎を例にして述べる．

- 発声語彙の登録法

① この制御用ソフトウェアを始動する．

② 発声語（発声に用いる語）を3回発声して登録する．その発声語に対応して一太郎が実行すべき動作を，一太郎をキー操作する際のキーストローク列によって定義する．

- 一太郎の実行中における発声語の使用法

① 一太郎を始動する．

② HOME キーを押す（これにより一太郎がキー入力モードから音声入力モードに切り替わり，この制御用ソフトウェアが始動する）．

③ 発声語を発声する（これによりその発声語が認識ボードによって認識され，認識結果の第1位の候補が制御用ソフトウェアに渡される．制御用ソフトウェアはその第1位の候補に割り当てられたキーストローク列を走査して一太郎を駆動する．駆動し終わると制御が一太郎に戻り，音声入力モードからキー入力モードに切り替わる．すなわち，ある発声語を発声すると，音声認識結果が正しい場合には，その発声語に割り当てられたキーストローク列をキーボードから打鍵した場合と同じ結果を得ることができる）．

3. 音声からキーに切り換える最適な時点を求める実験

本章では，(1)音声からキーに切り換える最適な時点が存在するか否か，(2)もしも最適な時点が存在するならば，それは何回目の発声か，を実験を通じて調べる．しかし，これらのことを一般的に示すこと，すなわち，これらのことがあらゆるシステムのあらゆる操作にわたって成立することを示すのは困難であると考える．したがってここでは，第一段階として，広く使われている1システム（一太郎 Ver. 4）の1操作を例にとって調べる．

3.1 実験で取り上げるコマンド

本実験では，文章の章や節の番号（たとえば，1. や 1.1 等．以下，これらを章節番号という）へジャンプするコマンド（一太郎では「検索」コマンドを用いる）を取り上げる．

一太郎において，検索コマンドをキーで入力する手順は以下のとおりである（ただし，検索はつねに文書の先頭から行うようにパラメータが設定されている）．

① ESC キーを押す（コマンドメニューが表示される）．

② Sキーを押すか，またはメニュー上のカーソルを「検索」の上へ移動し，改行キーを押す（「検索」コマンドが選択される）．

③ Sキーを押すか，またはメニュー上のカーソルを「文字検索」の上へ移動し，改行キーを押す（文字列を検索することが選択される）．

④ 検索文字列（たとえば，1.1 等）を入力し，改行キーを押す（検索したい文字列を決定する）．

⑤ 改行キーを押す（検索が開始される）．

⑥ ESC キーを押す（「次を検索」することをとりやめる）．

上記①〜③の手順をファンクションキーに登録し（すなわち，マクロコマンドとして登録し），そのファンクションキーを押すことにより，上記①〜③の手順をショートカットする方法もある．しかし一般に，ショートカット方式と音声とを比較した場合でも音声の方が優位な場合があることが文献1)で報告されている．したがって，上記①〜⑥に示す入力方法と上記①〜③の手順をショートカットする入力方法とを音声と比較しても音声の方が優位であることが予想できるので，本実験ではより一般的な入力方法である上記①〜⑥の入力方法と音声とを比較することに焦点をあてる．

実験で章節番号へジャンプするコマンドを取り上げる理由は，以下のとおりである．

(1) 音声入力は，利用者から直接見えない“もの”へのアクセスに向いている[2]．そこで，画面に表示されていない，すなわち見えない章や節（の番号）へジャンプするコマンドを音声で入力するのは価値があると考える．

(2) しかし，章節番号を音声で入力する場合には発声語に類似した部分が多く（たとえば，

1.	2.	3.	4.	5.	6.
1.1	2.1	3.1	4.1	5.1	6.1
1.2	2.2	3.2	4.2	5.2	6.2
1.3	2.3	3.3	4.3	5.3	6.3
1.4	2.4	3.4	4.4	5.4	6.4
1.5	2.5	3.5	4.5	5.5	6.5

図 1 実験で使用する章節番号 (36 種類)
Fig. 1 Chapter and section numbers used as vocal words in the experiments.

1.1 と 1.2 とでは 1. の部分が類似している), 誤認識しやすいと予想できる.

したがって, 誤認識が少なければ音声の方が有利であるが, 誤認識が連続する場合にはキーに切り換えることが必要になる, と予想する.

実験で用いる章節番号は, 論文や報告書等で見られる範囲の 36 種類 (1 章から 6 章までの各章が 1 節から 5 節まで) である (**図 1**). この 36 種類の章節番号を対象にすれば, かなりの数の論文や報告書等で本実験の結果が活用できると予想する. したがって, 図 1 に示す 36 種類の章節番号に限定した実験でも十分, 意義があると考える.

3.2 予備実験: 最適な切換え時点を調べる際の最大範囲の調査

音声からキーへ切り換える最適な時点は何回目かを調べる際の, その最大範囲 (すなわち, それ以降には最適な切換え時点は存在しないといえる範囲) を予備実験を通じて調べる.

被験者は 8 人で, 20 歳代前半から 30 歳代前半の全員が男性である. 「一太郎 Ver. 4」の使用経験は, 全くない者から約 4 年の者までさまざまである. 実験は, 2 人 (実験者 1 人と被験者 1 人, 役割は後述) で, 比較的静かな個室で行った.

パーソナルコンピュータは 2 台並べて使用した (以後, これらをパソコン 1 およびパソコン 2 と呼ぶ). どちらも NEC 製 PC 9801 RA で, クロック周波数は 20 MHz とした.

パソコン 1 では, そのキーボード上のスペースキーが 1 回押されるごとに, ランダムな順に並べられた 36 種類の章節番号がその順に 1 つずつ 1 回だけ表示される. パソコン 2 では, 一太郎と音声認識システム, および打鍵計測プログラム (打鍵されたキーとそのキーが打鍵された時刻を記録する) が同時に起動される.

実験をつぎのように行った. パソコン 1 の前に実験者が, パソコン 2 の前に被験者がそれぞれすわる. つぎに示す手順①, ②を 36 回 (すなわちランダムな順に並べられた 36 種類の章節番号のそれぞれを 1 回ずつ) 繰り返す.

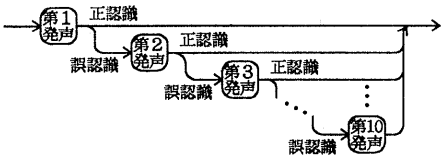

図 2 予備実験における発声のしかた
Fig. 2 The procedure which the subject utters a vocal word in the preliminary experiment.

① 実験者がパソコン 1 のキーボード上のスペースキーを 1 回押す (すると, パソコン 1 の画面上に章節番号が 1 つ表示される).

② 被験者は, パソコン 1 の画面上に表示された章節番号を見て, その章節番号を発声する (第 1 発声). **図 2** に示すように, それが正認識だったら, その章節番号の発声は終了する. 誤認識だったら, 再度, 同一の章節番号を発声する (第 2 発声). それが正認識だったら, その章節番号の発声は終了する. 誤認識だったら, さらに同一の章節番号を発声する (第 3 発声). 以下, 同様に行い, 誤認識が連続する場合には第 10 発声まで試みる. 第 10 発声のときは, それが正認識であっても誤認識であっても, その章節番号の発声は終了する.

実験を, 実験者と被験者の 2 人で上記のように行う理由はつぎのとおりである.

(1) 被験者には, 測定の対象とならない操作 (たとえば, パソコン 1 のキーボード上のキーを押して, パソコン 1 の画面上に章節番号を表示させる等) は行わせない.

(2) 章節番号が提示されてから, 被験者が音声入力またはキー入力 (予備実験では行わないが, 3.4 節で述べる主実験で行う) に移る際の手の動きにおいて, 一方が不利になることのないようにする.

測定するのは, 第 1 発声から第 10 発声までの各発声における積算認識率 $R_1, R_2, \cdots, R_{10}$ である. これらをつぎのように定義する.

$$R_1=\frac{v_1}{v_t}, R_2=\frac{v_1+v_2}{v_t}, \cdots, R_{10}=\frac{v_1+v_2+\cdots+v_{10}}{v_t} \quad (1)$$

ここで, $v_1, v_2, \cdots, v_{10}$ はそれぞれ第 1, 第 2, …, 第 10 発声において正認識した語数である. v_t は総発声語数 (ここでは, $v_t=36$) である.

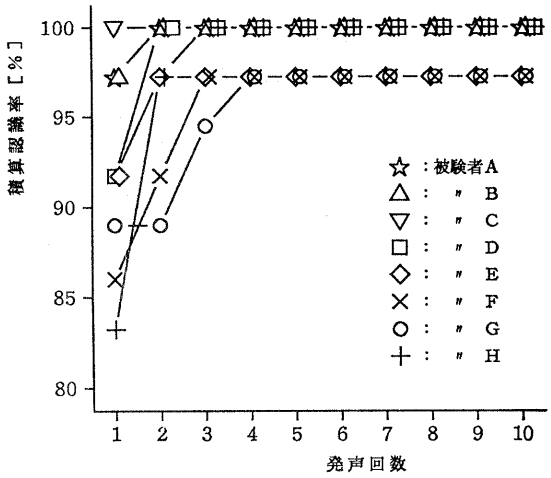

図 3 積算認識率
Fig. 3 Experimental results of the accumulated recognition rates of eight subjects.

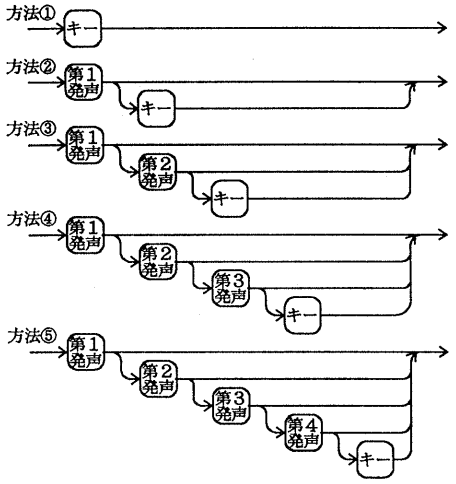

図 4 1コマンドの入力方法（5通り）
Fig. 4 Five ways ① to ⑤ for inputting a command which locates a cursor to a chapter or a section number in a document.

3.3 予備実験の結果

被験者8人（A～H）の積算認識率を図3に示す．横軸は第1発声から第10発声までの各発声であり，縦軸は積算認識率［%］である．

図3はつぎのことを示している．被験者Cは第1発声において，積算認識率が100%になった．被験者A，B，Dは第2発声において，被験者Hは第3発声において，それぞれ積算認識率が100%になった．ところが，被験者E，F，Gはそれぞれ，第2，第3，第4発声において積算認識率が100%未満で飽和している．このことは，被験者E，F，Gはそれぞれ，第3，第4，第5発声以降は発声しても正認識する見込みがないこと，すなわち発声することが無意味であることを示している．したがって，5回発声し，6回目にキーに切り換えること，またはそれ以降に切り換えることはどの被験者にたいしても無意味であることがわかる．

以上の事実より，音声からキーへ切り換える最適な時点は，5回目までを調べればよいと結論できる．

3.4 主実験：最適な切換え時点の調査

ここでは，音声からキーに切り換える最適な（すなわち，入力に要する時間を最短にする）切換え時点は何回目かを実験を通じて調べる．

実験に用いた装置，その設定，および被験者は予備実験と同じである．

音声からキーへ切り換える最適な時点は，5回目までを調べればよい（3.3節）．したがって，各被験者は，つぎの5通りのそれぞれの方法で章節番号へジャンプするコマンドを入力する（**図4**）．

方法① キーのみで入力する．

方法② 1回目は発声し，それが誤認識だったならば2回目はキーで入力する．

方法③ 1回目，2回目は発声し，それらが誤認識だったならば3回目はキーで入力する．

方法④ 1回目，2回目，3回目は発声し，それらが誤認識だったならば4回目はキーで入力する．

方法⑤ 1回目，2回目，3回目，4回目は発声し，それらが誤認識だったならば5回目はキーで入力する．

実験の手順は，予備実験の手順①②（3.2節）と同様である．ただし，手順②において被験者が発声する回数は10回ではなく，方法①～⑤のそれぞれに応じて異なる．また，最後はキーで入力する．

各被験者は，つぎの2通りのセッションを実施する．

セッションⅠ 上記の方法①②③④⑤をこの順に連続して行う．

セッションⅡ 上記の方法①②③④⑤をこの逆順に連続して行う．

セッションⅠおよびⅡの双方を行う理由は，方法①～⑤を実施していく上での慣れの影響を少なくするためである．

実験を行った期間は2日間である．第1日目に，被験者全員が発声語彙（**図5**の左側．右側は左側の発声

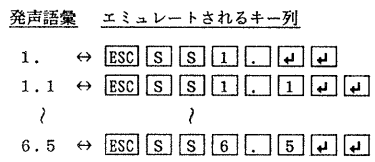

図 5 章節番号へジャンプするコマンドを音声で実行する場合の発声語彙とそれらが発声されたときに音声認識システムによってエミュレートされるキー列

Fig. 5 The left-hand side of each line shows a vocal word which the subject utters. The corresponding right-hand side shows a key string which is emulated when the subject utters the vocal word. Each vocal word is used to execute a command which locates a cursor to a chapter or a section number in a document.

語彙が発声されたときに音声認識システムによってエミュレートされるキー列）を登録し，3.2 節で述べた予備実験を行った．つぎに，被験者A，C，E，Gはセッション I，セッション II の順で実験を行った．被験者 B，D，F，H はセッション II，セッション I の順で実験を行った．第2日目（1週間後）は，発声語彙の登録と予備実験は行わずに，被験者A，C，E，Gはセッション I，セッション II の順で，被験者B，D，F，Hはセッション II，セッション I の順で，第1日目と同様に実験を行った．したがって，各被験者は2日間でセッション I を2回，セッション II を2回，すなわち方法①～⑤のそれぞれを4回ずつ行った．

測定する時間を，方法②を例にとり説明する（**図6**）．1種類の章節番号へジャンプするコマンドを入力する際には，つぎの2通りの場合が考えられる．

（1） 第1発声で正認識した場合．このとき，1コマンドの入力に要する時間（以下，これを1コマンドの入力時間と呼ぶ）は，1回の音声入力時間（HOME キーが押されてから検索コマンドを終了する ESC キーが押されるまでの時間）に等しい．

（2） 第1発声で誤認識したためにつぎにキー入力をする場合．このときの1コマンドの入力時間は，1回の音声入力時間と1回の移行時間（検索コマンドを終了する ESC キーが押されてから，つぎにメニューを表示する ESC キーが押されるまでの時間），および1回のキー入力時間（メニューを表示する ESC キーが押されてから検索コマンドを終了する ESC キーが押されるまでの時間，ただし誤ったキーを入力した場合はその入力に要した時間，およびその回復に要した時間も積算する）の和である．

実際に要する時間は，上記(1)，(2)のいずれかである．

方法②は4回行うので，したがって 36 種類の章節番号へジャンプするコマンドをそれぞれ4回ずつ，合計 144 コマンドを入力することになる．その 144 コマンドの入力に要する時間をすべて合計したものが，方法②について求める時間（以下，これを全コマンドの入力時間と呼ぶ）である．他の方法についても同様である．

3.5 主実験の結果

被験者8人の，方法①～⑤のそれぞれにおける全コマンドの入力時間を図7に示す．横軸はキーへの切換え時点であり，方法①～⑤にそれぞれ対応している．縦軸は全コマンドの入力時間［秒］である．

図7からつぎのことがいえる．

（1） 全コマンドの入力時間が最も短くなるのは，被験者A，B，E，F，G，Hについては

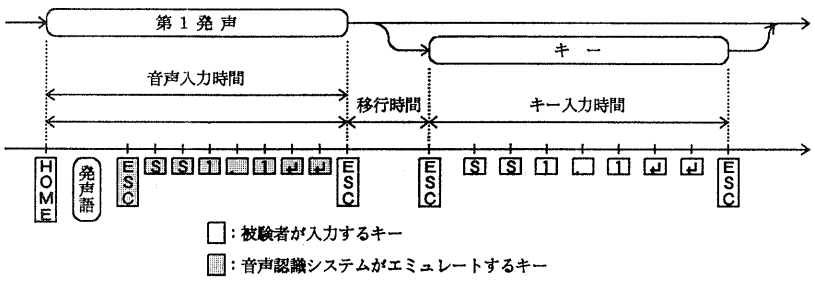

図 6 測定する時間（2回目に音声からキーに切り換える場合）

Fig. 6 This figure shows the elements of the needed time when inputting a command. This is the case when at first the subject takes a voice-input which results in a misrecognition and then he/she switches to a key-input.

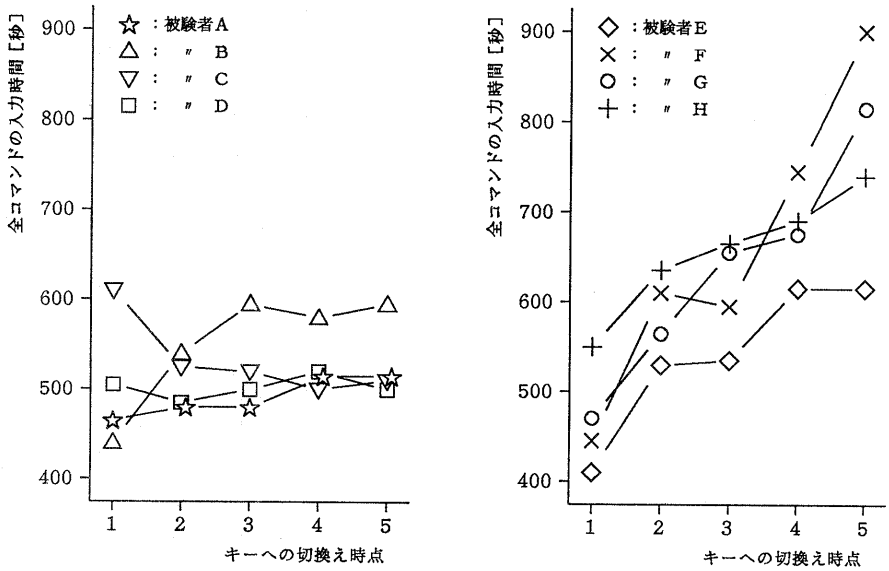

図 7 1～5回目にキーへ切り換える場合のそれぞれにおける全コマンドの入力時間

Fig. 7 Experimental results of the times which each of the eight subjects took in inputting four times the all of the vocal words shown in Fig. 1.
The horizontal axis represents when each of the eight subjects switched from a voice-input to a key-input. Each of the numbers 1 to 5 on the horizontal axis corresponds to each of the five ways ① to ⑤ in Fig. 4.

1回目に音声からキーに切り換える場合である．言い換えると，これらの被験者は音声を使用するべきではなく，キーのみで入力する方法が最適であるといえる．被験者Cは4回目に，被験者Dは2回目に音声からキーに切り換える場合が，全コマンドの入力時間が最も短い．したがって，被験者CおよびDは逆に音声を使用した方が効率がよく，誤認識が連続する場合の最適な切換え時点は，それぞれ4回目および2回目であるといえる．

上記(1)の結果は，被験者のそれぞれの音声入力時間，キー入力時間，移行時間および積算認識率に依存すると考える．そこで，方法⑤における被験者のそれぞれの音声入力時間とキー入力時間の度数分布を**図8**に，またそれらの平均値と標準偏差を**図9**に示す（4章で述べる予測と検証にこの平均値が用いられる）．

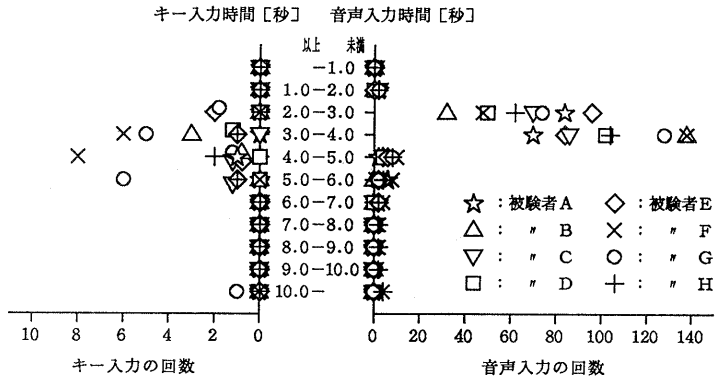

図 8 方法⑤における音声入力時間とキー入力時間のそれぞれの度数分布

Fig. 8 Histograms of the times which each of the eight subjects took to complete a single voice-input or key-input by the way ⑤ in Fig. 4.

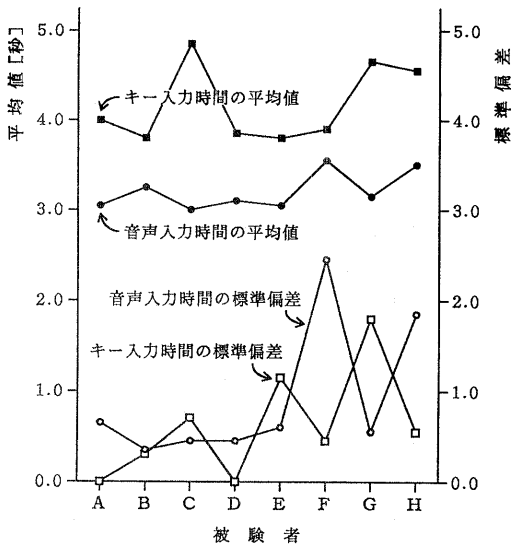

図 9 方法⑤における音声入力時間とキー入力時間のそれぞれの平均値と標準偏差

Fig. 9 The averages and the standard deviations of the times which each of the eight subjects took to complete a single voice-input or key-input by the way ⑤ in Fig. 4.

移行時間は，音声入力時間とキー入力時間の両者よりかなり小さいのでここでは省略する．積算認識率については，ほとんどの被験者が 3.3 節の予備実験の結果よりやや低くなる傾向がみられたが，ほぼ同様なのでやはりここでは省略する．

図 7，8，9 からつぎのことが読み取れる．

（2） キー入力時間と音声入力時間のそれぞれの平均値の差が，被験者 C と G は比較的大きく，他の被験者は小さい．差が大きい被験者 C は，図 7 を見ると右下がりになっており，音声の使用が有効であることがわかる．ところが，被験者 G は前述の差が大きいにもかかわらず，積算認識率が低いために（図 3），図 7 では右上がりとなり，音声の使用は有効ではないという結果になっている．つぎに前述の差が小さい他の被験者の場合は，被験者 F のように積算認識率が低いと図 7 は傾きが急な右上がりとなり，被験者 A のように積算認識率が高いと，傾きのゆるやかな右上がりとなるか，または被験者 D のようにほぼ横ばいになることがわかる．このように図 7 は，キー入力時間と音声入力時間のそれぞれの平均値の差が大きい場合と小さい場合，および積算認識率が高い場合と低い場合のおのおのが掛け合わさった場合の様相をよく表している．

（3） キー入力時間は 2〜6 秒の間に分布しており，ばらつきがみられるが，音声入力時間はすべての被験者が 2〜4 秒の間に集中している．これは，キー入力時間は被験者の熟練度に依存するが，音声入力時間はそれにあまり依存しないことが原因であると考える．

（4） 被験者 F および H の音声入力時間の標準偏差が大きい．これは，被験者が発声しても認識システムが反応せず，発声を繰り返したために時間を要したことがあり，それが影響したと考える．被験者 F のキー入力時間の標準偏差が大きい．これは，キーの入力ミスおよびその回復に時間を要し，それが影響したと考える．

上記（1）〜（4）の結果は，すべてつぎに依存すると考える．

① 使用するシステム（対話型システムと音声認識システム），および対象とするコマンド．

② 被験者のシステムおよびキー入力に対する熟練度（被験者の構成は，A，B，E，F，G，H が学部 4 年生，大学院生または大学院修了生で，C，D は学部 3 年生である）．

したがって，本実験の結果は 1 システムの 1 操作に対して，各被験者が固有に持つ最適な切換え時点を示したにすぎない．しかし，上記（1）〜（4）より，

① 音声からキーに切り換える最適な時点は存在し，それは本実験から求めることができること，

② 最適な切換え時点は利用者により異なるため，利用者ごとに調べなければならないこと，

が明らかになったと考える．

4. 予測と検証

本章では，音声からキーに切り換える最適な（すなわち，入力に要する時間を最短にする）切換え時点を，実際の作業に先だって，または実際の作業中に，予測する方法を述べ，つぎに 3 章の実験結果を用いてそれを検証する．

4.1 予測の考え方

予測の考え方を，例を用いて説明する．ここで用いる例は，3 回目にキーに切り換える場合のデータ（すなわち，第 1，第 2 発声のそれぞれの積算認識率，音声入力時間，キー入力時間，および移行時間）のみが

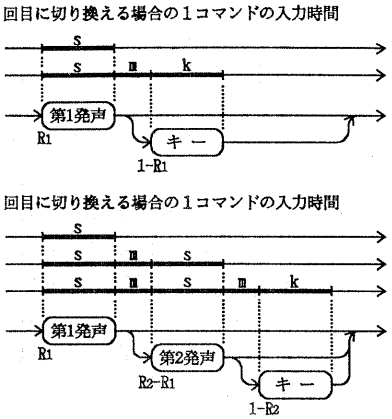

図 10 2回目または3回目に切り換える場合の1コマンドの入力時間の内訳
R_1, R_2 は第1，第2発声のそれぞれの積算認識率．s, m, k はそれぞれ音声入力時間，移行時間，キー入力時間

Fig. 10 Upper and lower figures show the elements of the needed time when inputting a command. Upper one shows the case when at first the subject takes a voice-input which results in a misrecognition and then he/she switches to a key-input. Lower one shows the case when at first the subject takes a voice-input which results in a misrecognition, when at second he/she takes a voice-input again which also results in a misrecognition, and when finally he/she switches to a key-input. The symbols R_1 and R_2 are the accumulated recognition rates of the first and second utterances respectively. The symbols s, m and k are the times which take in a voice-input, a switching from voice-input to voice-input (or key-input), and a key-input respectively.

既知であるときに，3回目に切り換える場合と2回目に切り換える場合とで，どちらが，全コマンドの入力時間がより短くなるかを予測する，という例である．

最初に，2回目に切り換える場合の全コマンドの入力時間の内訳を考える（**図 10**）．

（1） 第1発声で正認識した場合，そのコマンドの入力時間は1回の音声入力時間（s）だけである．第1発声の積算認識率を R_1 とすれば，全コマンドのうち，第1発声で正認識したコマンドの入力時間の合計は R_1sc（c はコマンド数，ここでは $c=144$）と表せる．

（2） 第1発声で誤認識したためにつぎにキー入力をした場合，そのコマンドの入力時間は1回の音声入力時間（s）と1回の移行時間（m），および1回のキー入力時間（k）の和である．キー入力をすることになる割合は $1-R_1$ であるから，したがって全コマンドのうち，第1発声で誤認識したためにキー入力をしたコマンドの入力時間の合計は $(1-R_1)(s+m+k)c$ と表せる．

2回目に切り換える場合の全コマンドの入力時間を T_2 とすると，これは上記（1），（2）の和であるから，つぎのように表せる．

$$T_2=\{R_1s+(1-R_1)(s+m+k)\}c \qquad (2)$$

同様にして，3回目に切り換える場合の全コマンドの入力時間 T_3 は，つぎのように表せる．

$$T_3=\{R_1s+(R_2-R_1)(2s+m)+(1-R_2)(2s+2m+k)\}c \qquad (3)$$

3回目に切り換える場合を実測することにより，式（3）における T_3, R_1, R_2, s, m, k の実測値を得ることができる．ここで得られた R_1, s, m, k を用いて式（2）の右辺を計算することにより，2回目に切り換える場合の全コマンドの入力時間の予測値 T_2 を求めることができる．

2回目に切り換える場合の全コマンドの入力時間の予測値 T_2 と3回目に切り換える場合のそれの実測値 T_3 とを比較したときに，（1）$T_2<T_3$ ならば2回目に，（2）$T_2>T_3$ ならば3回目に，それぞれ切り換えた方がよいといえる．

以上は，例を用いた予測の考え方の説明であるが，これを一般的に述べると，つぎのようになる．音声入力時間，キー入力時間，移行時間，および第 1〜n 発声のそれぞれの積算認識率が既知ならば，1〜n+1 回目の切換え時点の中で，どの切換え時点が，全コマンドの入力時間を最短にするか，を予測することができる．

上記の予測法の意義はつぎにある．実際の入力作業に先だって，または実際の入力作業中に，n+1 回目にキーに切り換える場合のみを測定することによって，1〜n+1 回目の切換え時点の中のどの切換え時点が最適かを予測することができる．

4.2 検　　証

4.1 節で述べた予測法を検証する．検証の方法はつぎのとおりである．

（1） 5回目に切り換える場合の実測値（3.5 節）を用いて，1〜4 回目に切り換える場合の予測値を求める．

（2） 上記（1）で得られた 1〜4 回目に切り換える場合の予測値と実験（3.5 節）で得られたそ

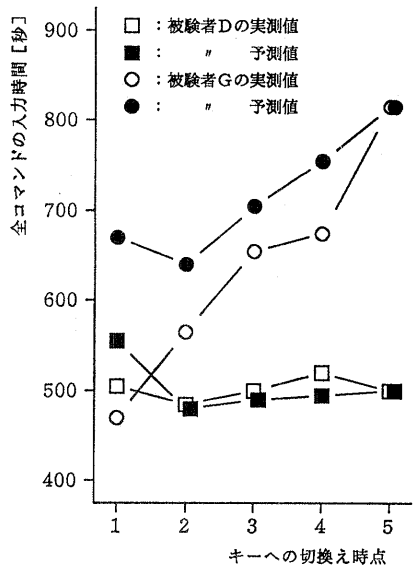

図 11 1～5回目にキーへ切り換える場合のそれぞれにおける全コマンドの入力時間の実測値と予測値（被験者DとGの場合）

Fig. 11 This figure plots the actually required times in the experiment by the two subjects D and G and the estimated ones by means of the equations derived in this paper. The horizontal axis represents when each of the eight subjects switched from a voce-input to a key-input. Each of the numbers 1 to 5 on the horizontal axis corresponds to each of the five ways ① to ⑤ in Fig. 4.

れらの実測値とを比較する．

3.5 節の実験結果において，被験者A，B，E，F，G，HとC，Dとでは対照的な結果が得られた．そこで，前者と後者のそれぞれの中で代表的な被験者GとDとについて，1～5 回目に切り換える場合のそれぞれの予測値と実測値を**図 11** に示す．ただし，5回目に切り換える場合の予測値は実測値をそのまま用いる．横軸はキーへの切換え時点である．縦軸は全コマンドの入力時間［秒］である．

図 11 からつぎのことが読みとれる．

（1） 被験者Dについて，全コマンドの入力時間が最小となる切換え時点は，実測値と予測値の両者とも2回目で一致した．

（2） 被験者Gのそれは，実測値が1回目，予測値は2回目で一致しなかった（その理由は下記(4)による）．

（3） 被験者D，Gとも，キーへの切換え時点が2～4回目については，予測値は実測値に近い値が得られた．

（4） 被験者D，Gとも，キーへの切換え時点が1回目のときの誤差（予測値と実測値の差）は，他の場合よりかなり大きい．

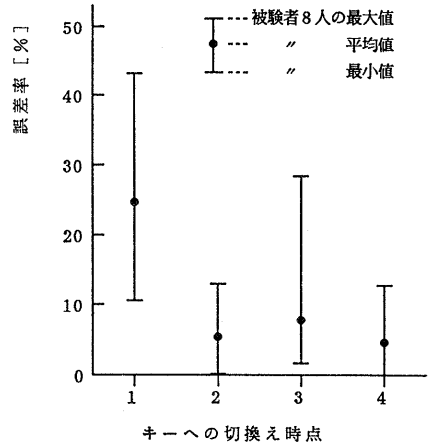

図 12 予測値の誤差率．被験者8人の最大値，平均値，および最小値

Fig. 12 Maximum, average and minimum values of the error rates of the estimated times for the eight subjects. The horizontal axis represents when each of the eight subjects switched from a voce-input to a key-input. Each of the numbers 1 to 4 on the horizontal axis corresponds to each of the four ways ① to ④ in Fig. 4.

ここで，被験者8人について，誤差率をつぎのように計算した．

$$\text{誤差率}=\frac{|\text{実測値}-\text{予測値}|}{\text{実測値}} \qquad (4)$$

被験者8人の誤差率の最大値，平均値，および最小値を**図 12** に示す．横軸はキーへの切換え時点であり，縦軸は誤差率［%］である．

図 12 はつぎのことを示している．

（1） キーへの切換え時点が 2～4 回目の場合，被験者8人の誤差率の平均値はそれぞれ 10% 以下で，比較的小さい．

（2） キーへの切換え時点が1回目の場合，被験者8人の誤差率の平均値は 24.8% で，かなり大きい．

上記(1)，(2)の結果は，つぎの理由によるものと考える．

（a） 切換え時点が 2～5 回目の場合のそれぞれのキー入力時間の実測値はほぼ等しい．しか

し，切換え時点が1回目の場合のそれは，2〜5 回目の場合より小さくなる傾向がある．

（b） 切換え時点が 2〜5 回目の場合，それぞれの音声入力時間の実測値はほぼ等しい．移行時間の実測値についても同様である．

これら2つの事実が，5回目に切り換える場合の実測値から，2〜4 回目に切り換える場合を予測すると誤差は小さいが，1回目に切り換える場合を予測すると誤差は大きい，ことの原因であると考える．

5. おわりに

本論文ではつぎのことを述べた．

（1） 対話型システムとの対話において音声を使用する際に誤認識が連続するとき，他の入力機器（たとえばキー）に切り換える最適な時点が存在することを初めて主張した．

（2） 上記（1）の音声から他の入力機器に切り換える最適な時点を求める方法を初めて提案し，実験を通じてその最適な時点を調べた（3.5節）．

（3） 最適な切換え時点を予測する方法を述べた（4.1 節）．そして，ある程度予測が可能であることを示した（4.2 節）．この予測法は，以下のことに利用できる．利用者の積算認識率，音声入力時間，キー入力時間，および移行時間を知ることができるならば，①利用者に対しては，その利用者に固有の音声からキーに切り換える最適な切換え時点を予測し，教えることができ，②ユーザインタフェース設計者に対しては，設計時に，設計しようとする操作に音声を用いるべきか否かを決定する指針を与えることができる．

上記（2），（3）は1システムの1操作を例にとって調べたにすぎない．これらを普遍性のある結論とするためには，さらに他の操作，他のシステムについても調べることが必要であると考えるが，本論文はその第一歩である．本論文のように具体的な操作を対象にして，1つずつ結論を積み重ねる以外に普遍性ある結論を導く方法はないと考える．

今後の課題はつぎのとおりである．

（1） 他の操作，他のシステム，他の被験者についても上記（2），（3）を調べる．

（2） 本論文で述べた予測法を実システムに組み込み，活用する方法を検討する．

参考文献

1) 中谷吉久，守屋慎次：文書編集における音声制御の一方式，情報処理学会論文誌，Vol. 33, No. 2, pp. 195-203 (1992).
2) Schmandt, C., Ackerman, M. S. and Hindus, D.: Augmenting a Window System with Speech Input, *Computer*, Vol. 23, No. 8, pp. 50-56 (Aug. 1990).
3) Schurick, J. M.: Efficiency of Limited Vocabulary Speech Recognition for Data Entry Tasks, *Proceedings of the Human Factors Society —30th Annual Meeting—*, Vol. 30, No. 2, pp. 931-935 (1986).
4) Rudnicky, A. I. and Hauptmann, A. G.: Multimodal Interaction in Speech Systems, In Blattner, M. M. and Dannenberg, R. B. (Editors), *Multimedia Interface Design*, pp. 147-171, ACM Press, Addison-Wesley (1992).
5) （株）リコー：NEC PC 9801 用音声認識ボード取り扱い説明書（MDB-SWR-01）〈Ver 1.1〉(1990).
6) （株）ジャストシステム：一太郎 Ver. 4 NEC PC-9800 シリーズ解説編 (1989).

（平成5年 1 月 6 日受付）
（平成5年11月11日採録）

中谷 吉久（正会員）

昭和57年東京電機大学工学部電気通信工学科卒業．昭和63年同大学院博士課程満期退学．現在，神奈川県工業試験所技術管理部電子計算科勤務．対話型システムにおけるユーザインタフェースの分析・検査・評価法および音声認識装置を用いたユーザインタフェースの研究に従事．電子情報通信学会会員．

守屋 慎次（正会員）

昭和48年東京電機大学大学院博士課程修了．工学博士．現在，東京電機大学工学部情報通信工学科および同大学院博士課程教授．昭和56年ニューヨーク州立大学，昭和57年イリノイ大学の各計算機工学科客員準教授．Interacting with Computers 誌の Special Editorial Board などを歴任．各種のペン入力システム（テレライティング，携帯入力，インフォーマルミーティング，講義・採点・自習支援），ペン操作技法，ペン入力データベース，ペン入力パタンの解析，音声入力の応用，インタラクションのモデル化と標準化の研究．計測自動制御学会ヒューマン・インタフェース部会のなかのペン入力研究会委員長．