

# 追加指示式カナ鍵盤漢字入力装置

KANJI INPUT DEVICE USING MACHINE DIRECTED "TOUCH METHOD"

吹抜 敬彦      福島 紘一      島崎 由美子

Takahiko FUKINUKI      Kouichi FUKUSHIMA      Yumiko SHIMAZAKI

( 日立製作所 中央研究所 )

( Central Research Lab., Hitachi, Ltd. )

## 1. 緒言

日本において、計算機が普及し日常業務に深く入るにつれて、漢字の問題は避けて通ることのできない問題になりつつある。すなわち、計算機が理工学の計算にのみ用いられた初期の時代を過ぎ、広く社会活動に活躍するようになり、登録業務に、ドクメンテーションに、人事管理に、日本語情報処理に、出版印刷に、銀行業務にと広く利用されなければならなくなつた結果、米国育ちの計算機も日本の風土に順応することが要求されるようになったのである。

漢字の問題は、入力、内部処理、出力の3つに分けられるが、内部処理に関しては、一旦、漢字がコードに変換されていれば、英数字の場合とそれほど本質的な差のあるわけではなく、問題は少ない。出力に関しては、漢字プリンタ、漢字ディスプレイなどにハードウェア的大問題はあるものの問題点は比較的明確になっているものと思われる。

これらに対して、入力は人間の操作性の問題など人間工学的要素が多く、問題は単純ではない。そして本文において詳述するように、単に計算機入力機器というだけの機器としての見方だけでは不十分で、そのバックにある社会習慣、日常業務形態、勤労労働状態などまで広く関係してくる。そして、これまでいくつかの方式が提案されているが、人間の訓練による上達度や、人間の記憶容量など、方式の優劣の判定の条件も複雑である。このため、各方式の評価も人によってまちまちのようである。

筆者らは先年、この漢字入力の方法として「あいまいさ種別のフィードバックを有するカナ鍵盤漢字入力装置」を発表した(別に、追加指示式とも云う)。そして計算機によるリアルタイムのシミュレーションを行なった。しかし、これだけでは使い勝手や装置の規模などを実感として把握できなが、たうらみもある。しかし、これらのマンマシン性の高い機器では重要なポイントである。

本報告はこのような観点から試作された装置の概要を述べるものである。

## 2. 漢字入力装置の一般概要

### 2.1 漢字入力装置の備えるべき条件

以下にこれら条件を列挙する。

- (1) 安価であること(使い易さに応じて多少の価格の階層性がある方が望ましい<sup>(1)</sup>)。
- (2) 特殊な練習が不要なこと(カナタイプ、英文タイプ程度)。OJTTが可能なこと。
- (3) 高速打鍵が可能なこと(このためにはタッチメソッドが可能なこと)。

- (4) 素人も専門家も同一インタフェイスで操作でき、専門家には熟練の程度に応じてそれなりの高速化の道が開けていること。<sup>5)</sup>
  - (5) 誤入力率が低いこと。
  - (6) 漢字のみでなく、カナの入力も可能なこと(文章中の60%はカナである)。
- 以上の主要な条件のほか、さらに考えられる条件をいくつかを列挙すると、以下のようになる。
- (7) 国際標準的な機器が流用できること。<sup>5)</sup>
  - (8) 社会的バックアップがあること(具体的に規定しがたいが)。
  - (9) その他機器として一般的に備えるべき条件(無騒音、故障が少ないこと、省電力、体積重量床面積などが小さいこと)。

## 2.2 漢字入力装置の分類<sup>3)</sup>

筆者らはこれを次のように分類している。

- (1) 非マンマシン対話式
  - (a) フルキー式(漢テレ、和文タイプ、これに準ずるもの)
  - (b) 記憶コード式(記憶コードまたは数字コード)
- (2) マンマシン対話式
  - (c) カナ入力表示選択式(1次入力として読みカナを入力し、同音異字がある場合は機械がそれらの漢字を表示し、操作者がその中から1つを選ぶ)
  - (d) カナ入力追加指示式(1次入力として読みカナを入力し、同音異字がある場合には機械がいかなる種別の追加情報が必要かを指示する)-----日立提案の新しい方式(第2.3節以下で詳述する)
- (3) 翻訳式
- (4) 文字認識式

図2.1に上記(c)および(d)のブロック構成を示す。

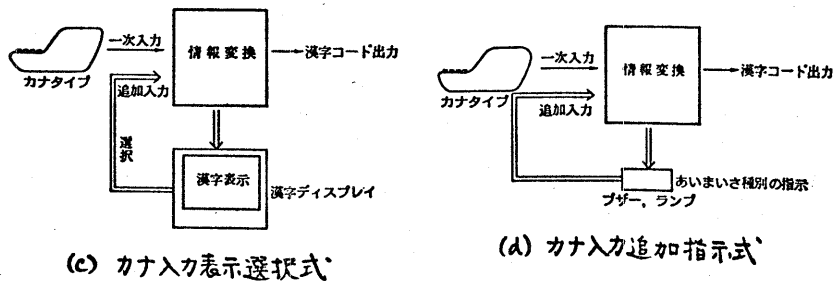


図2.1 対話式入力装置のブロック構成

## 3. 新方式の提案

以上述べてきた従来方式(a), (b), (c)の問題点をいろいろ検討すると、漢字入力装置として望ましい条件として下記のような項目が考えられる。<sup>1,2)</sup>

- (i) 特殊な訓練を要しないためには、人間と機械が長短相補な「マンマシン対話式」に向うべきである。
- (ii) 誤打鍵率を下げるため、ならびにタッチメソッドを可能にするため、腕の動きを排し、指の動きのみとし、ホームポジションのあるカナタイプ(英文タ

イブ)による入力とする。

(Ⅲ) 漢字ディスプレイは高価かつ二度手間の原因となり、かつタッチメソッドを不可能にするため使わない。

これにより、問題は結局「あいまいさ」(同音異字)があつたとき、このマンマシンシステムとして望ましいフィードバックは何かということに帰着する。これがポイントである。考えてみると我々は過去、知らず知らずのうちに「マンマシン対話」=「グラフィックディスプレイ」というように考えがちであつたと思ふ。前述の(c)カナ入力表示選択式はこの発想に基づくものであつた。

これに代る一つの解答が、音(音色の異なるブザーなど)、色ランプなどの利用である。しかし、この方法ではフィードバックの情報量や種類に限度があり、前述の(c)の方法のように該当する文字を呈示する方法などはとれない。

したが、て、この情報の利用方法として考えられるのは、追加すべき情報の種類のみをフィードバックすることである。このようにして新たに当社で開発された方式が前記の(d)の方式である。

次に具体例を示す。

かなはそのまゝ打つ。漢字を入力したいときには、まず漢字であることを表示するため機能キー(スペースキーで代用する)を打鍵しておく。そして、1次入力として、カナ2字をキー入力する。このカナ文字情報のみであまいさが残る場合には機械の側から2次入力として追加情報を入力するように指示する。

(例1) : 3字目情報の追加指示

オト(男, 弟, 音, 訪, 劣, 衰)に対して3字目情報(コ, ウ, ち, ズ, ル, ロ)の追加を指示する。

(例2) : 音(オン)情報の追加指示

ナミ(波, 浪, 淡, 並)に対しては音(オン)情報(ハ, ロ, ル, ハ)の追加を指示する(音の1字目)。

(例3) : 部首情報の追加指示

シン(臣, 信, 娠, 紳, 審, 診, 浸, 真, 芯)に対しては部首情報(ハ, ニ, オ, イ, ウ, ゴ, サ, ち, フ)の追加を指示する(部首の最初の読み)。

(例4) その他(つくり, 特殊, リジェクト)

なお、訓読みの場合には、同一訓に対して平均1.2字の同音異字があるのに対し、音(オン)の場合には平均8.7字という統計がある<sup>6)</sup>。これは中国の漢, 吳の発音から由来するもので、2字目には、ン, イ, ツ, ウ, … など約10通りの音韻しか存在しないためである。このことから、訓読みが容易な漢字については、訓入力のみ限定することにより、入力システムの単純化, 簡易化を図るようにした。

問題となる点としては、部首の定義がある。この原因の1つは康熙字典に由来するものと、最近の略字に由来するものがある。前者の例としては、「和」、「相」がそれぞれ「口」、「目」の部に入っており、後者の例としては、「声」、「帰」、「匠」が古い字体の關係からそれぞれ「耳」、「止」、「酉」の部に入っている。ここでは、これらの矛盾を避けるため、字体の左上部を含む部分を部首と定めた。ただ、部首のは、きりしない字については極力この情報を使うことを避けて、3字目や音の情報を使うようにすることが望ましい。

次に入力コードの統計を示す。

これまで速バタ方式に従って、リアルタイム計算機シミュレーションを行なった。そして、今回その見直しを行なった。そして、追加入力は2回までという条件で以下のような結果を得た(ただし、現在なお改良作業を続けているので、あくまで暫定値として見て頂きたい)。

(1) 入力コード形式の統計

読みカナ2字のオ1次入力と、追加入力種別がどのようになっているかを表3.1に示す。入力コードの決定にあたっては、使用している間に自然に覚えてしまえるようなものにするよう努めた。また「つくり」の情報(字体の右下部を含むもの)は比較的なじみが薄いため、極力これを避けるようにした。

(2) 入力回数

漢字1字を入力するのに、どの程度の追加入力をすればよいかを表3.2に示す。これは表3.1を単純に集計したものである。ただ、次項に示すように、1つの漢字について入力方法が複数個あるものについては、熟練度が上るにつれてだんだん入力回数の少ない方法を使用するようになること、入力コードの決定に際しては使用頻度の高い漢字はなるべく追加回数が少なくなるように配慮したので、これらを加重した平均値はもう少し少なくなると思われる。

(3) 1字あたりの入力方法

前述の如く、本入力コードシステムにおいては、訓読みが容易なものについては訓読みに限ることによってシステムの単純化を図った。しかし、訓読みについては個人差があることから音読みと訓読みの両方の入力を許容したものも多い。また、複数の訓読みを有する漢字もある。その結果、表3.3に示すように1字あたり複数個の入力方法を許すことにした。

表3.1 入力形式の統計

1次入力	オ1回追加	オ2回追加	漢字数	例
XX			425	兄(ア=), 絹(キヌ)
XXX			507	(牛潮後失)
XXXX			6	(試志快心)
XXX音			32	(思想)
XXX部			25	(阪坂)
(XXXつ)*			(4)	
XX音			585	(跡後痕)
XX音音			50	(集暑)
(XX音部)*			(28)	
(XX音つ)*			(3)	
XX部			902	(臣信媛神醫診没真心), (塩潮)
(XX部音)*			(37)	(送還)
XX部部			62	(簡環)(粗阻)
(XX部持)*			(36)	(裁裁)
(XX部つ)*			(123)	(望怨)

( )\*は避けたい組合せ

X: 読みかな, 音: 音(オン), 部: 部首, つ: つくり, 持: 持殊

表3.2 入力回数

追加なし	15%
1回追加	71%
2回追加	14%

表3.3 1字あたりの入力方法

1通り	1527字
2通り	481
3通り	85
4通り	13
5通り	3
6通り	1
11通り	1
計	2111字

## 4. 漢字入力装置 CTT の開発

### 4.1 目的

以上述べてきた方式を具体的にハードウェア化する。この試作の目的とするところは次の通りである。

- (i) マンマシン性の高い機器は実際に人間が使ってみないと本当のことがわからない。また入力装置としての規模を知ることも重要である。
- (ii) 漢字処理機器というと、大型コンピュータ(あるいは小型+大容量外部メモリ)という先入観が強い。これがわずかのハードウェアで実現できることを実証することは意義深い。

我々はこの機器に CTT (Chinese character Typewriter by Touch method), または Chi-Ton-Ton との愛称を付けている。

### 4.2 漢字入力装置 CTT

漢字入力装置 CTT は、その主要部の構成ならびに周辺機器によって、表 4.1 のような形態をとることを考えている。このうち、今回試作したものは CTT-PM (P-ROM を主要記憶装置とし、制御はマイクロコンピュータで行なう。周辺機器はプリンタをとりあえず接続できるようにする) である。

主要な用途別の形態は次のようになろう。

研究開発用	CTT-PH	(入力装置の開発)
訓練用	CTT-RM	
専業用	CTT-RT	(漢字入力請負業)
業務用	CTT-RM	(一般事務用)
盲人用	CTT-RT	(後天性失明者用として)
通信用	CTT-RC	(遠隔地からの日本文の伝送用)
印刷用	CTT-RM(RH)	(Cold Type System 用)
オンライン用	CTT-R	

Table 4.1 A Series of Kanji Input Devices

	Peripheral (s)					Memory (kByte)	Control	Editing*2
	*5	C*6	M*7	T*8	H*9			
CTT-P			*1			P-ROM(11)*4	μ-Com*3	yes(option)
CTT-R						ROM(11)*4	μ-Com*3	yes(option)
CTT-D						disk(~20)	wired-logic	no

\*1 under development

\*2 RAM(1-2kByte) added

\*3 e.g. INTEL-8008

\*4 including Program

CTT; Chinese character Typewriter by "Touch method"

\*5 Input to Computer

\*6 Communication Use

\*7 Printer (monitor)

\*8 Tape (magnetic/paper)

\*9 Tape and Printer

### 4.3 試作装置の概要 ---- マイクロコンの利用

当初、本方式の特徴の一つは、処理が簡単で計算機などが不要なことであった。しかしその後、マイクロコンピュータが普及し、wired-logic で構成すると同程度になってきた。また、これによりメモリの使用方法をさらにキメ細かくし、メモリ全体の使用量を減らせることから総合的には経済化できること、編集機能

などの付加が容易で発展性に富むことなどから、マイクロコンピュータを使用することに決めた。

使用した〔I-8008〕は、最大記憶容量(内部メモリ)が16KByteであり、先のシミュレーションによつて判明した必要ビット数12万ビット(データのみ)に近い。このため、必要バイト数の節約をさらに進めれば、余裕をもつて收容できると判断された。

また、〔I-8008〕は、命令のサイクル時間がかかり長い。しかし、この種のマンマシンシステムには充分過ぎるほどの速度であることがわかった。

以上の結果、〔I-8008〕によるデータの作成、ソフトウェアの試作を進め、種々の改良の結果、プログラム込みで11KByte(うちプログラム約270ステップ、450バイト)に収めることが可能になった。そして、今後予定している各種の機能(編集などのWord-processing)にも充分の余裕を残している。図4.1に基本構成を、図4.2にその外観を示す。

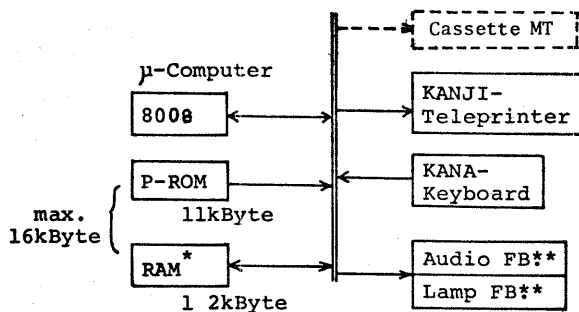


Fig. 4.1 Block Diagram

- \* for editing(option)
- \*\* feed-back to indicate the class of additional information

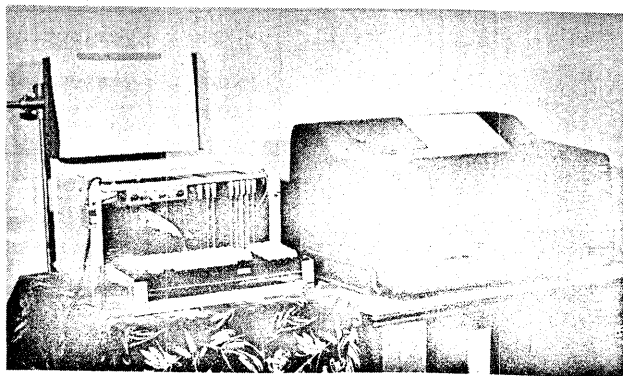


図4.2 漢字入力装置CTT外観

(マスクROMを使用するとき、電子回路部分は1枚のプリント板となり、キーボードの下に収まる。)

カナキーの配列はJIS規格によつた。これを図4.3に示す。ここでは、初心者で部首の定義のあいまいな人の便宜のために、部首を併せ刻印してある。ただ前述のように、たとえば「才」(テ崩)は「テ」のところにあるので慣れてしまえば不必要なものである。

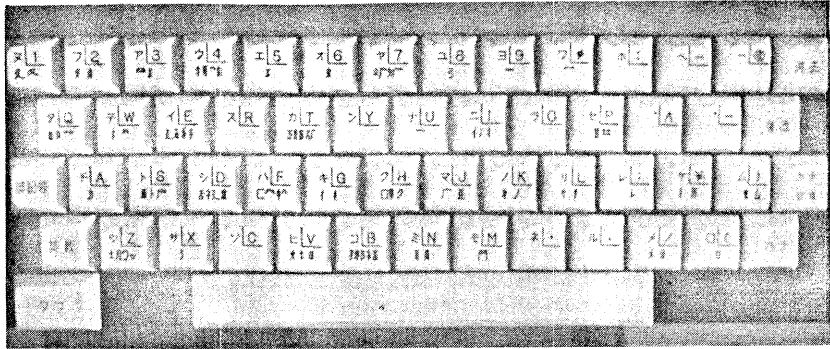


図4.3 キーボードの配列(初心者用)

次に、完成時の予想図を図4.4に示す。

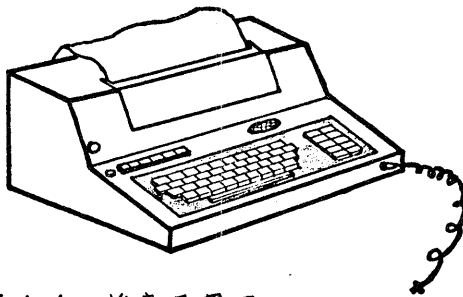


図4.4 将来予想図

#### 4.4 入力例

入力例を下に示す。

色は句へど散りぬるを、我世たれぞ常ならむ。  
 いろニ彩 シル {ワガ}ヨビセ ツネジ  
 {ワカ}

ういの奥山今日越えて、浅き夢見し酔ひもせず。  
 ウヰヤマヒノコ {コエ} アヒ ムシル ヨウ  
 {コエ}

- : 音尋末
  - △: 3字目末
  - x: 部首末
- } 機械からのあいまいさの指示(ワザ-おまじランブ)

#### 4.5 多種の漢字コードへの対処とハードウェアの開発

現在、たまたま手持ちのO社製他速漢テレプリンタ(約1.5字/秒)を利用し、ここで用いられている漢字コードを基準としてP-ROMに記憶させた。記憶装置としてP-ROMを使用している間は、その使用対象に合わせてP-ROMの内容を決定すればよいが、将来これの経済化、小型化のためマスクROMを使用するようになれば多種コードにどう対処するかは大きな問題になる。国家的見地から早急に決定されることが期待される。

なお、現在マスクROMでは、16Kbit(2KByte)のものが出ている。これを使用すれば記憶装置の大きさは、 $\frac{1}{8}$ となり、個数にして現在の40個以上のP-ROMがわずか6個程度で済んでしまう。そして、画期的な経済化が図られる。

この結果は、いわゆる入力装置というほどのものとはならず、カナキーボードの下側にプリント板を一枚とりつけた位の大きさになりうる。

#### 4.6 入力打鍵速度に関する検討

本方式における入力速度は、フルキー式、表示選択式に比し高速であることは本質的に間違いないと思われる。

本方式と速度の点で対抗しうるのは記憶コード式であろう。すべてのカナ、漢字にカナ(あるいはアルファベット)2字からなる記憶コードが与えられて、これを入力する方法と比較してみたい。このコードを人間が記憶できるかどうかはいろいろの意味で異論もあるが、現に訓練すれば可能であることが実証されている。

漢字を打鍵する際には、本方式ではスペースキーを、上記方式ではシフトが必要であるが、これを無視して、漢字率 $m$ の文章を入力する場合の必要な1字あたりの平均打鍵数を示すと、

$$\text{本方式} \quad 3m \text{ (漢字の場合)} + (1-m) \text{ (カナの場合)} = 2m + 1$$

$$\text{上記方式} \quad 2$$

となる。したがって漢字が50%を下まわる文章の場合には本方式の方が高速に打鍵できることがわかる。

#### 5. 漢字研究に関連して

今後、国内において漢字処理の研究が盛んになるものと思われるが、これまでの研究経過を顧りみてこの研究の特殊性、問題点を述べてみたい。

やってみて感じることは、予想以上に難しい問題であるということである。この意味は4つある。

- (1) 音声認識、文字認識、自動翻訳などの研究に共通して云えることであるが、才の赤児にも可能な音声認識が計算機には非常に不得手なように、現在の機械は人間臭いことはあまり向いていない。このあたりの限界をよく見極め、場合によってはかなり泥臭い方法を採用することが必要な場合もあろう。
- (2) 我々には、対象物が日本文という日頃見なれたものであることから来る錯覚がある。ある先進コンピュータメーカが和文組版という比較的機械的な処理に取り組んだが、当初の難易さについての予想が完全に間違っていて計画変更し苦しんだという<sup>(4)</sup>。
- (3) 漢字というものが他の対象と異なり数が特に多いため、安易に取組めない。



一旦取組んだら、一つ一つの漢字を検討してゆくという泥臭く、かつ骨の折れる作業を根気よく続けてゆかねばならない。

- (4) そして最後に感ずることは、この研究は単なるハードウェアの問題でなく、国民の習慣の問題、文化体系との関連など、単なる工学では解決しない問題を含んでいるということである。

## 6. 結言

従来の漢字入力装置の問題点を解決する新しい入力装置の開発を行なった。この入力方式は前述したように、人間と機械が長短相補う「マンマシン式」であり、昔々色のフィードバックという新しいマンマシンインタラクションによりタッチメモッドによる入力を可能にした。

そして、従来「漢字処理」というと大型計算機（あるいは大容量外部メモリ）という先入観を打破り、小型のマイクロコンピュータでこれを実現した。また、必要な記憶容量も、種々の工夫の結果、約11KByteに収まった。

今後、各種の方面で漢字の処理は重要になって来ると思うが、これまでこのための基本要素である入力装置には問題も多く、漢字処理全般、ひいては日本における広義の情報処理の発展を阻害して来た。この発展に本試作装置が役立つことを期待したい。

残された問題、今後の問題としては、入力コードの改良、ハードウェアの改良などがあり、さらに Word-Processing 機能の付加なども考えてみたい。

## 7. 参考文献

- 1) 吹抜, 福島, 島崎; あいまいさ種別のフィードバックを有するカナ鍵盤漢字入力装置, 信学会電子計算機研究会 EC 72-5 (昭和47年5月)
- 2) 吹抜; ユニークな機能をもつカナ鍵盤漢字入力装置, エレクトロニクス, 昭和48年1月号, pp 130-136
- 3) 日本情報処理開発センター; 日本語情報処理の技術動向調査報告書, (昭和48年3月)
- 4) 山本; 印刷とエレクトロニクスの歩み寄り——コンピュータ組版の概要——テレビジョン学会誌, Vol. 28, No. 6, pp 447-450 (昭和49年6月)
- 5) 喜安; 日本語の表現をどう考えてゆくか (昭和46年電気連大 No. 172)
- 6) ——; 漢字情報処理特集号, 情報処理学会誌, Vol. 10, No. 5 (1969. 9月)