

文字盤構成の最適化と文字盤JIS案への適用について

渡辺 定久 (電子技術総合研究所)
 丸岸 洋 (松下電器産業株式会社)

1 はじめに

ペンタッチ型や外段シフト型の日本
 文入力装置あるは和文タイプライタ
 などに使用されている文字盤には

- ① 漢字を使用頻度によって分類し
 高使用頻度漢字(1級漢字)を
 文字盤の中央付近に、低使用頻
 度漢字(2級漢字)をその周辺
 に配置する階層構成型^{*}で、
- ② すべての漢字を一定の原則(代
 表音訓による50音順など)に従
 って配置する一様配列型

の2種類がある。

階層構成型文字盤の特長は、採字作
 業の大部分が非漢字領域と1級漢字領
 域だけで行なえ腕の移動がすくなく
 すむため、疲労がすくなく採字速度の
 向上が期待できる点にある。このため
 階層構成型文字盤の採用例は、作業能
 率を重視する職業的オペレータ向けの
 装置に多く、たとえば和文タイプライ
 タ用の文字盤において、つぎのよう
 な階層化が行なわれている。

1級文字	292字	} 文字盤内	
2級文字	896字		
3級文字	931字		
英教文字	102字	} 文字盤外	
予備文字	858字		
牌蔵文字	2520字		
		合計	2201字
		合計	3378字

これに対し一様配列型の文字盤は、
 職業的オペレータの存在が予想し難
 い分野で使用されることを前提とした
 装置での使用例が多く、ペンタイラ
 ーの商品名で登場して和文タイライ
 タのユーザ層を一部の専門家から一般

*文字盤によつては、1~3級、1~
 4級などに細分化する例もある。

へ拡大し、ワードプロセッサの普及に
 先駆的役割りを果たしたと考えられる
 型和文タイプライタや、文字盤型のワ
 ードプロセッサは、現在のところ、す
 べてこの形式の文字盤によって占めら
 れている。

職業的なオペレータにとつて都合が
 よいと考えられている階層型文字盤が
 ワードプロセッサに採用されている例
 のすくなくはないのは、階層型が初心
 のオペレータにとまどいを与えず採字
 能率の向上に障害とならぬと考
 へられているからに外ならない。実
 際、漢字の配置に関する記憶が充
 分でない初心者にとつて、目的
 文字が見えしを恐れて何回か
 くりめまして探索した1級漢字領
 域内ではなくてく黄に并するユウ、
 オウ、省に并するセイ、ショウな
 どまざらわしい代表音訓の存在は
 この傾向を助長する可能性がある
 がある)、2級漢字領域内にあつた
 ことを発見したりすることが全
 くなくはない切羽である。この
 ような経験をしたオペレータが
 将来にわたってこのいざつづけ
 るであろう不安感や拒否反応を
 考へると、ワードプロセッサに
 おける階層型文字盤の採用が、
 販買政策上の要請から、困難であ
 ることも無理のないことであら
 う。けれども、筆者らの経験から
 判断すると、初心者のオペレー
 タが使用する機を、熟練者向け
 の装置と区別して考へなければ
 ならない必然性は見出せない。
 この経験には、初心者がいつ
 までも初心者ではない、という自
 明なことのほかに、階層型文字
 盤による採字速度の向上は初心
 者の場合より顕著に認められる
 という点も含まれている。現在、
 文字盤の標準案作成作業が、

通産省工業技術院電気規格課からの委託により、(社)日本電子工業振興協会に設置された「日本語情報処理標準化委員会(委員長・元岡達東大教授-筆者のひとりも座談に参加)」による進められている。委員会活動は現在も継続中であり、どのような文字盤が標準案としてまとまるか未確定ではあるものの、階層型文字盤についての筆者らの見解を尊重していただいていること事情はある。従って、ワードプロセッサが主な適用範囲となるであろう文字盤標準案が規定する文字盤が現在のワードプロセッサのそれとは異なる階層型のそれとなる可能性はきわめて強い。

以上のような事情を考慮して、本稿では初心者による文字盤型入力装置の操作性に関する筆者らの経験を紹介し文字盤構成の最適化についての議論を進めることにする。それは同時に、近い将来制定されるであろう文字盤に関する工業規格の技術的背景のうちの説明にもなるはずである。

2 漢字の使用頻度について

階層型文字盤の良さは、使用頻度の高い漢字が文字盤の中央付近に集中して配置されていることによる。従って、使用頻度の高い文字セットが、入力装置などの利用分野に無関係に存在するのでなければ、階層型文字盤の汎用性は失われる。階層型文字盤に関する議論を進めるに当たっては、漢字の使用頻度がどのようになっているかを調べなくてはならない。

漢字の使用頻度調査については、国立国語研究所が新聞と雑誌について行った調査が良く知られている。新聞と雑誌における使用頻度分布として、表. 1が与えられている。

	新聞	雑誌
上位の 10字	10.6%	8.8%
50字	27.7	25.5
100字	40.2	37.1
200字	56.1	52.0
500字	79.4	74.5
1000字	93.9	90.0
1500字	98.4	96.0
2000字	99.6	98.6
2500字	99.9	99.5
3000字	99.9	99.9

表. 1 新聞と雑誌の使用度数分布

この表によると、新聞、雑誌でも頻度順位の上位約1000字によって全体の90%の漢字がカバーされている。上位集の中の高さが決まっているものの、同一の文字セットによるカバー率については明確でない。

同一文字セットによるカバー率を知るため、わが国では上の二つの資料のほか、(株)学習研究社がJIS C6226(情報交換用漢字符号系)の制定時に検討・照合の資料とするために作成した資料(以下「学研」、なお本資料の利用にあたっては(財)日本情報処理開発協会の協力を得た)、および「天声人語」「続天声人語」(源代博郎著・朝日新聞社、以下「天声」)、「情報処理 Vol. 17, No. 1~12(以下「情報」)における漢字の使用頻度順位表を準備した。

表. 2は上記5種類の資料における漢字頻度順位の相互比較を示したものである。各種資料において頻度順位が10位以上である漢字のうち何字が「新聞」における10位以上の文字セットにふくまれているかを示したものである。ただし、「新聞」と「雑誌」については、漢字が地名の表記に用いられている場合を除いて順位づけを行

なっている。

順位	雑読	天声	学研	情報
上位 500	427字	402字	371字	344字
1000	877	843	819	769
1500	1358	1354	1284	1234
全体	1642	1739	1792	1438

表. 2 各種資料における使用状況

「情報処理」は情報処理に関連した論文、論説など、まわめて限られた分野についてこの頻度調査があるため、「新聞」の重なりは他の資料の場合よりやや小さい（このことには、「情報処理」には個有名詞表記用の漢字が比較的多く使用されている、という点にも多少関係がある）。表. 2によると、そのような「情報処理」であったとしても、「新聞」の上位1000位までの漢字によって「情報」の上位1000位までの漢字の80%弱がカバーされること示されている。文字盤の用途を、通常の文章表記用と考える限り、各方面で共通して外用される1級漢字を約束することに無理はないといえる。

3 文字セットの大きさと採字速度

漢字の配置に関する記憶が充分でない初心のオペレータによる採字作業の大部分は、目的文字と文字盤中の文字との逐字的な比較を、目的文字を発見するまで続けることによって占められていると考えられ、もしそうだとすると、目的文字の大部分が、狭い1級漢字領域内で発見できる階層型文字盤による採字速度が、文字の探索を広い範囲について行なわなければならない一様配列型文字盤による採字速度を上まわるのは、当然であるといえる。

この節では、以上のことをは、きりさせるために行なった実験とその結果につ

いて説明する。

実験に使用した供試装置は3台の12段シフト型入力装置であり、文字盤の大きさが表. 3のようになっているほかは、同一のハードウェア構成による。

	漢字	非漢字	合計
10型	1008字	319字	1327字
15型	1392字	319字	1711字
22型	2016字	319字	2335字

表. 3 供試装置の文字盤

被験者は全員（WM, MR, SK, MH, UKの5名）学生である。経験者はいない。実験の結果が被験者の資質のちがいにによって不明確なものにならないようにするため、5名の被験者を二つのグループ（WM, MRのグループと、SK, MH, UKのグループ）にわけ、各グループとも3台の供試装置を逐次使用するようにした。入力用テキストは日刊工業新聞の社説であり、1個の実験では社説1編（約1050字）を入力するようにした。採字時間の測定は文字ごとに行なう文字コードと共に磁気ディスク装置に書きこむようになっていた。

文字盤に収容されている漢字についてこの採字速度の測定結果は、図. 1 図. 2 に示す通りであり、同一の供試装置によって行なった実験では、各グループの被験者の間の採字速度に大きな差が見られず、同時に、供試装置をかえて行なった2個目以降の実験では、例外なく小型の文字盤による方がすぐれた成績をあげていることがわかる。従って、小型文字盤による採字速度の上昇分が、文字盤が小型である程大量に必要となる2級漢字の採字による速度の低下

を補なって余りある程度であれば、10型あるいは15型文字盤の漢字部分を1級漢字とする階層型文字盤の方が、一様配列型文字盤である22型によるよりも、すじめた採字速度を示すことになる。

この実験では2級漢字を考える代りに、2級漢字に相当する漢字を16進4桁の数字コードによって入力するようにした。結果は図. 3, 図. 4に示す通りである。特に実験の初期の段階では、4桁コードによる外字入力という面倒な手間をかけたとしても、小型文字盤による方が大型の文字盤によるよりもすじめた採字速度が得られることが示されている。

被験者が最初に遭遇した外字の入力に要する時間は、短い被験者で平均20秒、長い被験者では35秒を越えている。それにも関わらず10型や15型の小型文字盤と外字入力併用方式の方が、外字入力をほめて必要としない22型文字盤を上まわす採字速度を、特に経験の浅い期間に示していることは、初心者による逐字的な採字作業に与える大型文字盤の影響の大きさを示しているものと考えられる。

4 階層型文字盤の場合

前節で取り上げた実験は、階層型文字盤の2級漢字の代わりに外字入力を考えたものであった。従って、この実験の結果が、階層型文字盤に対して初心者がいざうかもしぬ心理的な不安感による作業能率の低下や、目的文字の探索を実際に2級漢字領域から始めるかもしぬオペレータによって不当に顕在化される階層型文字盤の欠点を反映していない可能性はある。この疑問に答えるには、同じメカニズムと同じ大きさの文字セットを持つ階層型の文字盤と一様配列型の文字盤によ

る実験を、前節で述べたのと同じようにして行なえばよい。

表. 4はこのような実験に使用した3台のペンタック型の供試装置の概要であり、図. 5は2名の被験者OH, OHによる入力速度の測定結果である。

	1級漢字	2級漢字	非漢字
1S型	1008字	1934字	319字
2S型	1392	1550	319
3S型	2942	0	319

表4 供試装置の概要

図. 5は、被験者OH, OHが同じ供試装置1S型を使用した最初の実験では、OHより良い成績をあげたOHが、一様配列型の3S型を使用した2回目の実験では階層型の2S型文字盤を使用したOHより劣った成績しか残せなかったことを示しており、階層型による入力速度が一様配列型によるそれを上まわっていることを示していると考えられる。OHが一様配列型を6回が階層型を使用した3回の実験ではOHの成績が上まわっていることはもうまでもない。なお、この実験はOH, OHのほか3名の被験者によって行なわれた。その結果のなかには最終回に使用した3S型による成績が前2回の成績にくらべて極端に悪くなっている例もあるが(最終回の成績が最も良いのが普通である)、詳細は省略する。

5 階層構成の最適化について

この節では初心者による採字速度を最大にする階層化の方法を、文字盤サイズ一定の条件で考える。

初心のオペレータによる採字作業は次のようであるとする。

- ① 次に採字すべき文字を想起してその文字が漢字であれば代表音

訓などグループ分けをいこりるグループを1級漢字領域内から見つけ出す。

- ① 発見されたグループについて目的文字の探索を文字盤内文字と逐字的に比較しながら行なう。目的文字を発見したらその文字に対して入力操作を加える。
- ② 目的文字が1級漢字領域内で発見できなかったら、2級漢字領域で、①、②をくりかえす。
- ③ 文字盤内で目的文字の探索に失敗したら、外字入力操作を行なう。

階層型文字盤によつて2級漢字を入力する際には、漢字グループを発見する手順を2回くりかえす必要があるが、われわれの実験の結果は、2回に要する時間が無視できる程度に小さいものであることを示している。また、目的文字まで到達するのに要する時間は、逐字的な比較の対象となった文字盤中の文字に比例するとする。すなわち、1級漢字に対する索字時間を t_1 、2級漢字に対する索字時間を t_2 、1級漢字領域に収容されている漢字の数を N_1 、2級漢字領域に収容されている漢字の数を N_2 とすれば、

$$t_1 = K N_1 \quad \text{--- ①}$$

$$t_2 = K (N_1 + N_2) \quad \text{--- ②}$$

$$N_0 = N_1 + N_2 \quad \text{--- ③}$$

入力すべきテキストが N_1 字の1級漢字と N_2 字の2級漢字をふくんでいるものとすると、テキスト中のすべての漢字に対する索字時間 t_0 は

$$t_0 = K N_1 N_1 + K N_2 N_2$$

$$= K (N_1 + N_2) N_0$$

$$= K N_1 N_2 \quad \text{--- ④}$$

④式右辺の1項は、文字数 N_0 の1

様配列型文字盤による索字時間に外ならない。従つて④式は、階層型文字盤による索字時間が、1様配列型文字盤の場合よりも、右辺の2項だけ短くなることを示したものである。このことは、文字盤サイズ一定の条件の元での階層型文字盤の最適化が、 N_1, N_2 を最大にするように N_2 を決めることで実現できることを意味している。
 N_1 に対する N_1 、および N_2 に対する N_2 は、いわゆる漢字の使用頻度順位で、累積使用度数に対応する。従つて、入力すべきテキストの漢字に関する累積度数分布が与えられると、そのテキストを入力するのに最適な階層化を④式によつて行なうことができる。

結果の詳細を示すことは省略するけれども、入力すべきテキストにふくまれる漢字が「新聞」や「雑誌」におけるのと同じような累積度数分布を示すものとすれば、表.1を用いて、たとえば $N_0 = 1800$ 字のときの N_2 の最適値が、500字と800字の間にあることは簡単に求められるはずである。

6 文字盤の標準案について

現在作成作業が進行中の文字盤標準案は次のようになる可能性が強い。

非漢字	300字	} 合計2160字 (36行×60列)
1級漢字	780字	
2級漢字	1080字	

標準案における階層化が本稿での考へ方の叶によつて行なわれたいわけではないけれども(たゞ、よこの区切りの良さも重要な要素になる)、結果的には本稿での結論とほぼ一致した。

このことから考へて標準案は熟練したオペレータにはもちろん、初心者にも歓迎される文字盤になることが期待できよう。

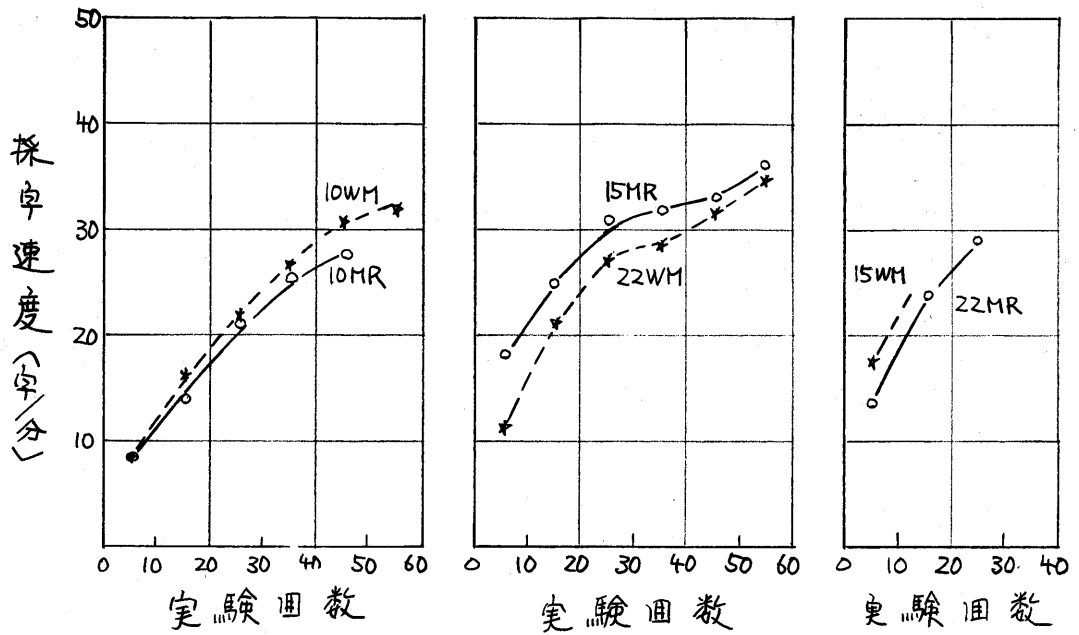


図. 1 15型文字盤と22型文字盤による盤内漢字の採字速度

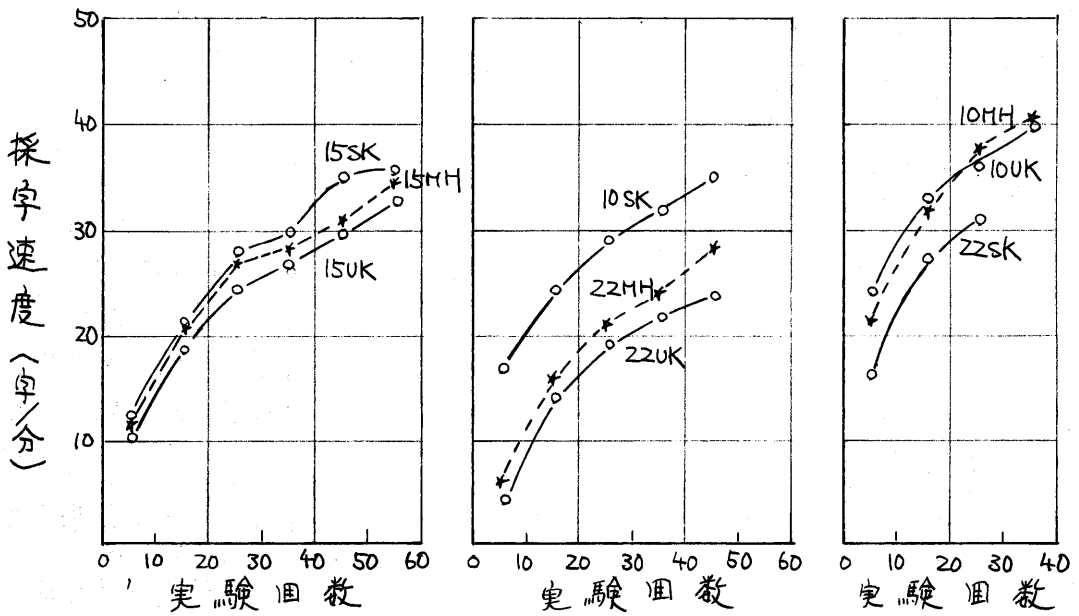


図. 2 10型文字盤と22型文字盤による盤内漢字の採字速度

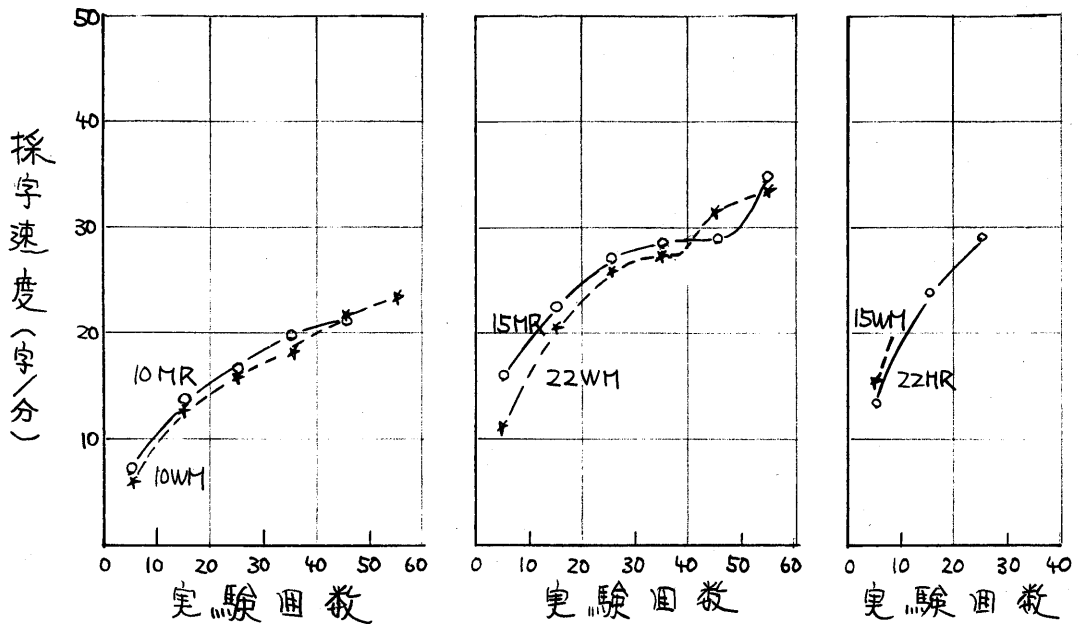


図. 3 15型文字盤と22型文字盤による漢字の採字速度

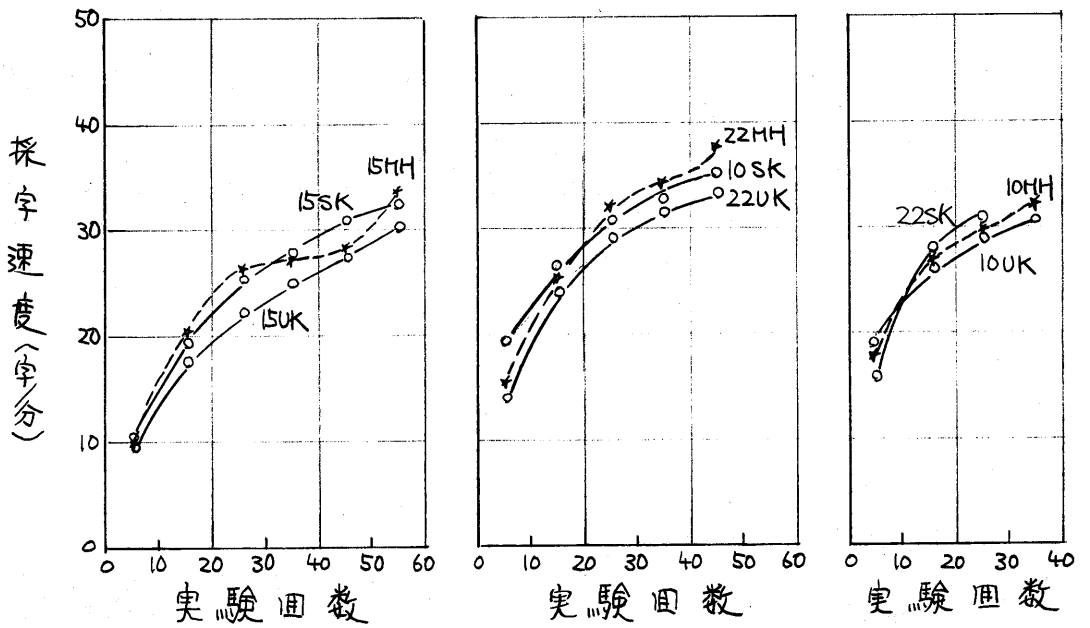


図. 4 10型文字盤と22型文字盤による漢字の採字速度

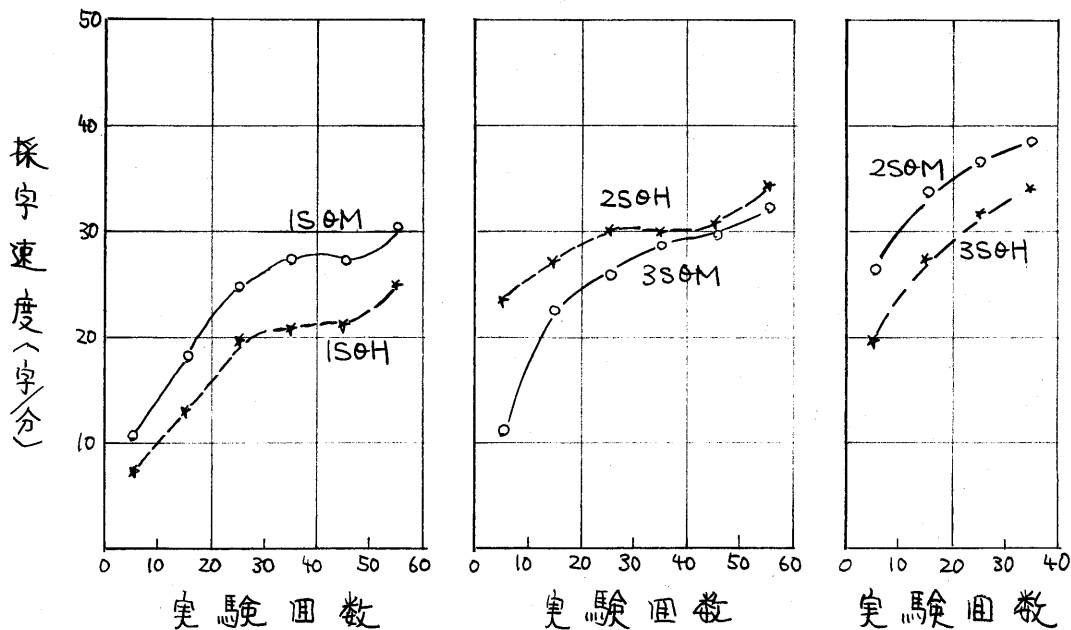


図.5 2S型文字盤と3S型文字盤による漢字採字速度

7 おわりに

本稿で使用した実験データは、電子技術総合研究所・人間機械システム研究室の設備によって収集したものであるけれども、その動機はほとんどは本文中でも触れた「日本語情報処理の標準化に関する調査研究」に負っている。このプロジェクトを起案された工業技術電気規格課の担当官、委員会を組織された理官にあたられている元岡達東大教授、事務局を分担されている(社)日本電子工業振興協会の関係者、属議に参加されている委員各位に感謝しとまわびを。

参考文献

- (1) 国立国語研究所：現代雑語九十種の用語用字(2)
- (2) 国立国語研究所：現代新聞の漢字
- (3) 日本語情報処理の標準化基礎調査，(社)日本電子工業振興協会(SIS)
- (4) 日本語情報処理の標準化に関する調査研究，(社)日本電子工業振興協会(昭和56年)
- (5) 渡辺定久：多段ニット形入力方法，「情報処理」Vol.23, No.6 昭和57年6月