

2J-4

ペン入力インタフェースとしての 漢字混じり仮名漢字変換評価

福永 幸弘* 井上 健** 鈴木 謙二**

*:東芝AVE(株) **: (株)東芝 情報処理・機器技術研究所

1. はじめに

ペン入力機器で日本語文書を作成するための入力形態としては、(1)平仮名で筆記し仮名漢字変換を行う、(2)漢字混じりで筆記し漢字混じり仮名漢字変換[1]~[5]を行う、(3)全ての文字を原文通り筆記の3種類が考えられる。このうち(3)の方法は、正しい漢字を書くことができない、複雑な漢字を書くのは面倒である等の理由から現実的ではない。

我々は、漢字文字認識と漢字混じり仮名漢字変換を有するペンワープロを用い、上記(1)と(2)の方式によって文字入力・変換候補選択を含む文書作成実験を行い、作成に要した時間、変換率等の評価を行なったので報告する。更に、被験者が仮名表記した文字や漢字表記した文字の調査結果を基に漢字混じり仮名漢字変換の変換率向上への指針を示す。

2. 評価システム及び方法

本評価実験で使用したシステムは、感圧型タブレット上に専用ペンを用いて予め設定される文字入力枠に筆記し、文字認識後に変換操作を行うものである。入力枠は画面上に6文字分が表示されているが、スクロールすることで最大20文字までを一度に入力することができる。次に評価システムの仕様を示す。

文字認識

- ・画数遵守(一部崩し字、略字は可)、筆順制限無し
- ・認識文字種→JIS第1水準(記号の一部を除く)

漢字混じり仮名漢字変換

- ・連文節変換
- ・基本単語数→85,800語
- ・入力表記数→296,000(「今日」など漢字単位に読みが付加できない語を除き制限無し)

評価方法は、前述のシステムに対し仮名表記で720文字程度、変換後の漢字で600文字程度となる原稿を平仮名筆記・仮名漢字変換、および漢字混じり筆記・漢字混じり仮名漢字変換でそれぞれ行い文書作成終了までの時間を測定することとした。また、被験者の使用方法による変動を抑えるために次のルールを設けた。

- ・カタカナ・英数・記号は常に直接筆記。
- ・変換指示位置は任意、無変換キーの使用も可。
- ・単漢字変換、辞書学習は使用しない。
- ・認識文字種は全文字種固定。

3. 評価結果

3-1. 文書作成に要した時間

被験者10人に同一文書を作成してもらった場合(A)と、数人に一人あたり10文書を作成してもらった場合(B)の時間測定の結果を表1に示す。

表1 文書作成時間測定結果

	作成時間(sec)		速度向上率 (%)
	平仮名	漢字混じり	
(A) 10人平均	2025.7	1796.5	+11.3
(B) 10文平均	2153.2	1901.5	+11.7

速度向上率を個別に調べると、(A)では被験者により約+2~+20%、同様に(B)では文書により約+1~+23%の範囲で、いずれも漢字混じり仮名漢字変換を行う場合の方が文書作成に要する時間が短かった。

このように文書作成時間が短縮された理由を調べるために、上記(A)の実験で得られたデータのうち平均的な例を選び、その詳細を表2に示す。

表2 文書作成時間測定結果(詳細)

筆記方法	入力文字数 括弧内:漢字	所用時間(sec)		
		入力	変換	総合
平仮名	748 (0)	1184	761	1945
漢字混じり	693 (70)	1174	499	1673

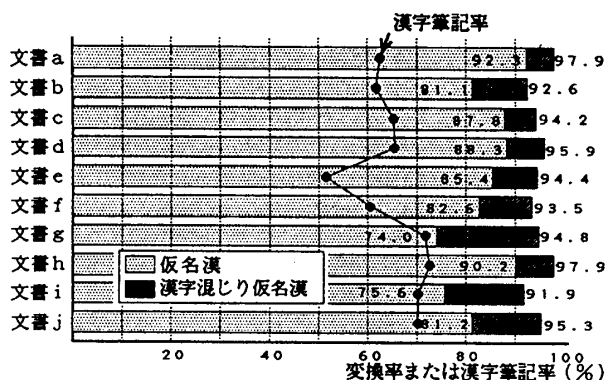
表2から、平仮名入力と漢字混じり入力では入力時間(筆記+認識)の差はほとんどなく、候補選択を含む変換時間が時間短縮に大きく寄与していることがわかる。

3-2. 変換率測定

表2で得られた漢字混じり仮名漢字変換が文書作成

An Evaluation of Handwriting Input
using Kanji-mixed Kana-to-Kanji Conversion.
Yukihiro FUKUNAGA, Takeshi INOUE, Kenji SUZUKI.
* Toshiba AVE Co., Ltd.
** Toshiba Corp.
Information Systems & Engineering Lab.

時間の短縮に寄与したことを検証するために、変換率を測定した。測定は表1(B)の文書作成実験でどの漢字を仮名で筆記したかを記録しておき、そのデータに基づいて作成した漢字混じりの入力文をあらためて変換システムに流し、第1候補と原文とを比較することにより行った。測定に用いた文書は何れも約150文節程度である。変換率および変換率に影響を与えると予測できる漢字筆記率(原文中の漢字が漢字で筆記された割合)をグラフ1に示す。



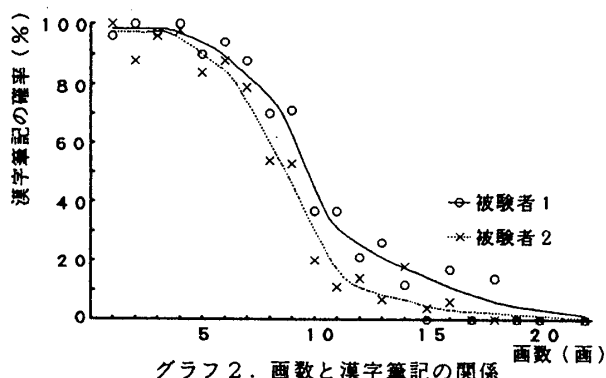
グラフ1. 変換率と漢字筆記率

例えば文書aでは仮名入力で92.3%であった変換率が漢字混じり入力を行うことにより97.9%まで向上したことが解る。10文書全てで変換率の向上が見られ、平均では11%の向上であった。変換率の向上分すなわち図中斜線分の候補選択処理に要した時間が不要となったことになる。また、文書により変換率の向上で差が生じるのは漢字で筆記された割合によるところが大きい、必ずしも一致しないのは漢字表記でも吸収しきれなかった未知語や辞書登録される単語頻度の付け方にも影響されているからである。

同様に表1(A)に関しても変換率の測定を行ったところ、被験者により+5~+13%の範囲で、平均8.6%の変換率の向上を確認することができた。同一の文書であっても、被験者個人の漢字習熟度や評価システムの漢字文字認識との相性等から、漢字筆記率は33~91%とかなりの幅があった。

4. 変換率向上への指針

今回行った漢字混じり入力の実験において、被験者が漢字表記で入力した文字および仮名表記で入力し漢字に変換した文字を調査した。その中で顕著に特徴が得られる例として入力者を限定した場合の漢字の画数と漢字筆記される確率との関係をグラフ2に示す。



グラフ2. 画数と漢字筆記の関係

被験者により若干の差はあるものの9画程度より低画数は漢字表記で、それ以上は仮名表記が急激に多くなるといった似通った線を描く。この閾値となる画数は、入力者の漢字習熟度、ペン入力に対する慣れ、文字認識性能等から決定されるものと考えられる。

更に、文字別に漢字・仮名のどちらで何回筆記されたかを調査すると、約9割の文字が漢字または仮名の何れか決まった筆記方法で入力されており、漢字または仮名の両方で筆記される文字は1割程度であった。

以上示したような入力者による筆記特徴を例えば変換候補の出力順位付けに利用することで変換率の向上が期待できる。同様に漢字文字認識の認識率向上にも有効であると予想できる。

5. まとめ

漢字文字認識、漢字混じり仮名漢字変換を持つペンワープロ上で仮名入力、漢字混じり入力の2方式による文書作成を行い、その両者の作成時間を比較することで、漢字混じり入力の方がより効率のよい文書作成が可能であることがわかった。また、変換率向上に伴う操作時間の短縮の影響が大であることもわかった。

更に、被験者の仮名書き、漢字書きのある種の特徴を得ることができた。次段階として、この筆記特徴を活用したシステムを構築し変換率の向上を目指したい。

尚、本評価実験は当社のXTEND™搭載のワープロ機能を用いて行った。

参考文献

[1]塩見 他, 情処論文誌, vol133, No7, pp. 920-927, (1992)
 [2]建石 他, 情処第45回大会, 4C-1, (1992)
 [3]金子 他, 情処第45回大会, 4C-4, (1992)
 [4]山田, 情処第46回大会, 5L-3, (1993)
 [5]建石 他, 情処第46回大会, 5L-9, (1993)