

# 堆積型文字データを用いた手書き 速記文字識別システム

1H-4

鎌倉 誠 河崎 徹 野中秀俊 伊達 惇

北海道大学工学部

## 1. はじめに

筆者らはこれまでに、石村式<sup>[1]</sup>の手書き速記文字を対象とした文字識別に関する、いくつかの研究・開発を行ってきた。幾何学的特徴抽出による手法<sup>[2]</sup>との性能比較のために、テンプレートマッチングを基調とした識別法についても研究を行ってきたが<sup>[3]</sup>、本報告ではその発展として、マッチングデータが持つ情報を加え合わせて利用する、堆積型文字データによる手書き速記文字識別法を提案する。

堆積型文字データと識別対象データの速記者が同じ場合と異なる場合の識別率の違いや、堆積させるデータセット数の増加に対する識別率の挙動から、このシステムの性能評価を行う。

## 2. 堆積型文字データ

テンプレートマッチングによる手書き文字の識別の場合、識別対象のデータには多少の揺らぎがあるので、いくつかの正解パターンを用意することが多い。正解パターンのセット数を増やせば、識別率は上がるが、その場合識別時間が線形的に増大する等の難点がある。

今回提案する堆積型文字データを用いた識別システムは、各文字について全ての正解パターンセットを堆積して1個のデータとする。よって正解パターンのセット数の増加により、識別時間を一定にしたままで識別率を向上させることが可能となる。

A System for Classifying Japanese Shorthand  
Characters by Accumulation Method

Makoto Kamakura, Tohru Kawasaki,  
Hiidetoshi Nonaka and Tsutomu Da-te

Faculty of Engineering, Hokkaido University

## 3. 実験

ここでは、実験により堆積型データと識別対象の被験者の違いによる識別率の変化、堆積データセット数と識別時間の関係、第三者のデータを混合堆積した場合の識別率の変化などを評価する。

異なる2名の速記修得者AおよびBの書いた文字をそれぞれ10セット用意した。そのうち8セットを堆積型文字データの作成用として用い、残りの2セットを識別することとした。これらのデータの正規化は、一律の大きさの正方形の中に相対的な大きさを考慮して書くことを要求すること、および文字の太さのに関する細線化を行うことによって実現した。

Figs. 1, 2に示す実験では、いずれか一方の修得者の堆積データのみを用い、Figs. 3~5に示す実験では、両方の修得者のデータの堆積比率を変えながら組み合わせた。これらの条件のもとで、それぞれ修得者A、Bの書いた文字を識別した。縦軸は識別率、横軸は堆積型データの内訳を示す。例えば、「A6B2」は、修得者Aのデータを6つ、Bのデータを2つ堆積したデータで識別したことを意味する。また、グラフはいずれも実線で示されたものが修得者A、点線で示されたものが修得者Bの結果である。平均識別時間は、修得者Aのデータの識別で $74 \pm 2$  ms、Bのデータの識別で $80 \pm 2$  msであった。

Figs. 1, 2を見ると、第三者の堆積型データを用いた場合、同一人の堆積型データで識別するときよりも識別率が低いという傾向が見られる。これは個人差による影響であるといえる。ただし、これらの図が示すように、堆積データセット数が比較的小さい場合には、第三者のデータでも堆積

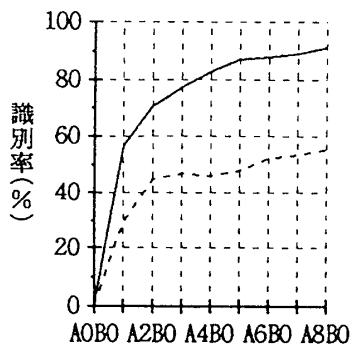


Fig. 1 Aの堆積データのみ

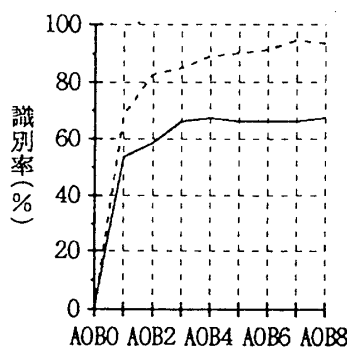


Fig. 2 Bの堆積データのみ

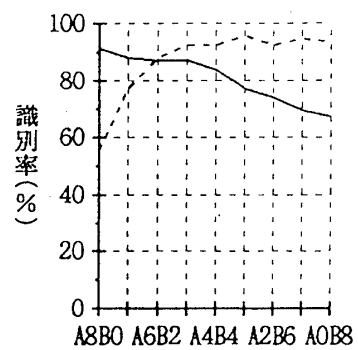


Fig. 3 堆積データ数一定

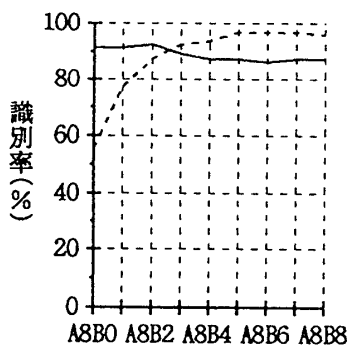


Fig. 4 Aの堆積にBの堆積を加えていく場合

[凡例]  
 実線—: 修得者A  
 点線…: 修得者B

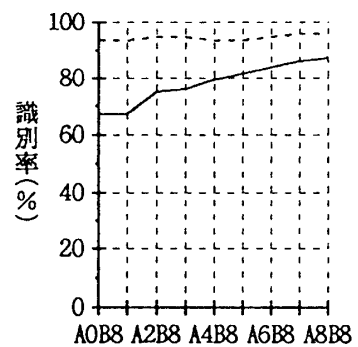


Fig. 5 Bの堆積にAの堆積を加えていく場合

数を増やすことによって識別率が向上している。一方 Fig. 3より、堆積するデータセット数が一定の場合、同一人の堆積データが多い方が良い結果が得られている。また、Figs. 4、5は、堆積数が十分に大きい場合、第三者のデータが混在することで識別率が下がる例と上がる例の両方を示している。

以上の結果から、堆積数が比較的小さく、識別率があまり高くない場合には、第三者のデータであっても堆積したほうが識別率が向上するといえる。

評価実験全体を通して、堆積数が大きくなって1文字あたりの識別処理時間は増加しないという性質が見られた。

#### 4. おわりに

今回は、2名の速記修得者のデータを用いて、堆積型文字データによる識別を行い、個人の「くせ」による識別率の差異、堆積データセット数・

比率を変えた場合の識別率の変化について報告した。

実験に使用したデータセットが少ないため、大規模な堆積を行った場合の挙動までは予測できなかったが、データ数をより多く集めてその結果について検証していきたい。

#### 参考文献

- [1] 石村 善左「石村式 速記講座基礎編 [改訂新版]」, 評論社, (1985).
- [2] 河崎 徹 伊達 惇「速記文字の個別識別システムの開発および特性評価」, 情報処理北海道シンポジウム'92, (1992).
- [3] 花田 将弘 河崎 徹 野中 秀俊 伊達 惇「テンプレートによる速記文字認識システムの開発と評価」, 情報処理北海道シンポジウム'93, (1993).