

遠隔端末からの実時間文字認識*

花木真一**・天満 勉**・木地和夫**

Abstract

Characters printed by hand on an economical input device are recognized immediately after their completion by a computer software. A total of 86 characters, viz. Japanese Katakana characters, 10 numerals, and some symbols are recognized by the scheme.

A character pattern is analyzed and approximated by pieces of lines called segment. The data on segments are stored in a data structure, a bi-directionally linked linear list. Features are extracted from the list and by the existence of specific features, the input pattern is identified by a decision tree.

A demonstration was carried out from a remote terminal (about 10 kilometers away from the computer) which was connected to the computer via commercial telephone lines. There was no serious problem of noises caused by a long distance transmission. This real time character recognition scheme was also applied to a man-machine communication system called CRIARS which recognizes hand-printed characters and responds in synthesized speech.

1. ま え が き

最近、計算機に文字列を入力する一つの方法として、実時間文字認識の手法が研究されてきている^{1)~4)}。従来、文字自動認識機械は、印刷にせよ手書きにせよ、主として形成されてしまった文字図形を認識対象としてきた。これに対し、実時間文字認識は、文字図形が描かれつつある時々刻々のペンの座標を計算機に入力し、描かれた文字を認識する点の特徴である。ここでは、人間が図形入力装置上に線図形として描く文字をオン・ラインで処理し、しかも、比較的短時間内に認識するような文字認識の手法を実時間文字認識と呼ぶことにする。もちろん、認識対象は、たとえば、フローチャートのシンボルのような一般図形であっても差支えないが、この論文では、それらの代表として対象を文字図形だけに限定した。

これまでの実時間文字認識の研究の多くは、認識の

* Realtime Character Recognition from a Remote Terminal, by Shin-ichi Hanaki, Tsutomu Temma and Kazuo Kiji (Central Research Laboratories, Nippon Electric Company, Ltd.)

** 日本電気株式会社・中央研究所

† どの程度までを短時間とするかは一概にはいえぬが、1文字あたり1秒以下程度で認識を終了することが望ましく、これに比し認識処理に極端に長い時間を必要とするものは、単にオン・ライン文字認識と呼ぶのが適当だろう。

アルゴリズムに重点を置いており、実際の応用のために解決されなければならない問題、たとえば、遠隔端末から文字を描いて実時間で認識させることや、簡便で経済的な図形入力装置を利用するといったことには、ほとんど注意が向けられなかった。しかし、将来遠隔地に設置されているセンターの大型計算機を、誰でも、いつでも、どこからでも利用できる状況を想定すると、経済性、信号伝送の容易さ、既成の通信回線の利用などの諸問題は、きわめて重要になってくる。

われわれは、簡便さ、経済性および信号伝送の容易さを考慮し、電話回線を利用して信号を伝送することができる既製の遠隔描画装置テレメール^{††}を図形入力装置として用いることによって、遠隔端末からの実時間文字認識を試みた。さらに、一つの応用として、実時間文字認識によって計算機に情報を入力し、合成音声によって計算機からのメッセージを出力する man-machine communication system "CRIARS^{†††}" を作り、実際に、日本電信電話公社の専用電話回線を利用して、計算機から直線距離約 10 km の地点に設置された遠隔端末から会話形式のやりとりを行なった。

†† 日本電気株式会社の遠隔描画装置の商品名。

††† Character Recognition Input & Audio Response System の頭文字の組み合わせ。

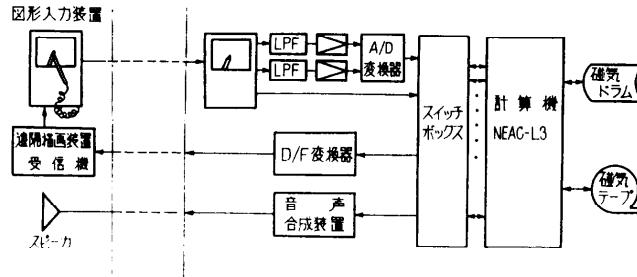


Fig. 1 The hardware system of CRIARS

2. 実時間文字認識の方法⁵⁾

2.1 ハードウェア

この実時間文字認識のハードウェア・システムが、CRIAR システムの一部として図1に示してある。システムの中となる計算機は NEAC-L3 である。デジタル計算機 NEAC-L3 は、内部記憶容量 8,192 語、1 語 36 ビットのワード・マシーンであり、サイクル・タイム 6.5 μ s、演算速度は加減算の平均約 20 μ s、乗除算平均約 40 μ s である。外部記憶装置として磁気テープのほか、記憶容量 65K 語の磁気ドラム装置、外部機器との入出力用にリアルタイム・チャンネル用スイッチボックス（以下単にスイッチボックスと呼ぶ）を備えている。スイッチボックスは、8 組の入出力チャンネルを持ち、ソフトウェアの指令によって、どれか 1 つのチャンネルを電子的に選択し、計算機と特定の外部機器との入出力の接続を行なう選択装置である。

図形入力装置として用いた遠隔描画装置⁶⁾は、本来は送信機と受信機とを対にして閉じた系で、図形による人と人との通信のために用いられるものである。すなわち、送信機・受信機とも、図2に示されるような

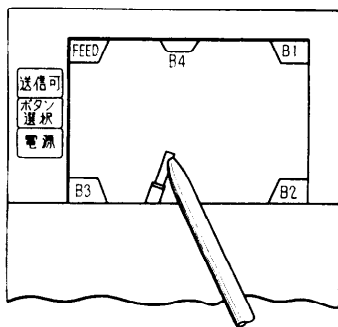


Fig. 2 The drawing area of the input device and control buttons.

約 9×13 cm の横に長い画面を持ち、送信機の画面内に通常のボールペンを組み込んだ送信ペンで図形を描くと、送信ペンの運動が受信機の画面上に再現され、受信ペンによって同一の図形が描かれる。送信ペンはその運動を伝達・分解するためのパントグラフと呼ばれる特殊なリンク機構の先端に結合されている。送信ペンを動かすと、パントグラフによって、ペン位置の座標が2つの回転角度に変換され、これらの角度が回転型の可変インダクタを用いた発振器により2つの周波数に置換される。これらの周波数は、画面の上下に対して1,310~1,490 Hz、左右に対して2,060~2,340 Hz の範囲で変化する。送信ペンには、筆圧によって開閉するマイクロ・スイッチが組み込まれており、ペンの上下信号を発生する。2種類の周波数成分を含む座標信号は、このペン上下信号と混合されて回線に送出される。受信機は送信機と同一寸法のパントグラフ機構および画面を持ち、サーボ・モータを用いて、受信ペンの位置が常に座標信号の2つの周波数成分に対応した位置にあるように追従動作をし、図形を再現する。

送信機から回線に送り出された信号には、送信ペンの座標、ペンの上下の情報などが含まれているが、計算機に入力するためには、これらの情報を直接ハードウェアで抽出するのはそれほど容易ではない。この実時間文字認識システムにおいては、図1に見られるように、この過程に遠隔描画装置の受信機を改造して利用した。すなわち、受信機中でペンを動かすサーボ・ループを切り離し、受信ペンの位置を固定したまま、サーボ・モータに加わる直流電圧を外部的に取り出すことによって、誤差電圧として送信ペンの座標の信号を得、さらに、受信ペンの上下動を制御しているリレーを利用してペンの上下の信号を得ている。座標信号は、低域ろ波器を通った後増幅され、アナログ・デジタル変換器によって、縦・横各 10 ビットのディジ

タル信号に変換されて、スイッチボックスから計算機に入力される。座標入力のサンプリング間隔は、主として、座標入力プログラムによって支配されており、必ずしも一定ではないが、約 10 ms である。

一方、計算機からの出力として、送信可ランプとボタン選択ランプの点滅の制御は、つぎのようにして行なわれる。すなわち、NEAC-L3 から加えられる制御用デジタル信号に応じて、デジタル・周波数変換器²⁾(D/F 変換器)を利用してランプ点滅の信号を送り、端末側の受信機内でペン上下制御および紙送り制御を行なうための2つのリレーをそれぞれ作動させ、それによって送信可ランプおよびボタン選択ランプを点滅させている。

2.2 オペレーション

文字は、図2に示されるような図形入力装置の画面内の紙の上に送信ペンで描かれる。画面の左上には、この遠隔描画装置に付属した FEED ボタンがあり、ペン先をこの FEED ボタンに近づけると、ハードウェアによって検出が行なわれ、画面の紙が送り出される。FEED ボタンの他に、画面の隅および縁の部分に、ソフトウェアで監視されている4個の制御用ボタンを設け B1~B4 の記号をつけた。これらの制御用ボタンは、ソフトウェアによって適宜その機能を変更して使用される。

実時間文字認識では、各文字の終わりおよびまとまったいくつかの文字の書き終わりをプログラムが検出する必要がある。このシステムでは、動作を確実にするために1文字書き終るごとにペン先を制御ボタン B1 に近づけて文字の区切りを合図し、この動作を繰り返して入力する文字全部を書き終わったら、ペン先を制御ボタン B2 に近づけて、プログラムに認識処理開始の合図をする。認識開始の合図を受けると、プログラムは直ちに文字ごとに認識を行ない、文字名、入力された文字図形の画面内での位置と大きさの情報とを出力バッファに出力する。

1度に認識できる文字数は、バッファ・メモリの大きさによって制限されるが、現在のところ最大9文字である。認識の対象となる文字は、片仮名(濁点、半濁点つきを含む)、アラビア数字、記号 +, -, ×, ÷, = の合計 86 文字で、文字の大きさは、約 8×8 mm 程度の小さな文字から、画面一杯の大きな文字まで自由な大きさで、しかも、画面内の任意な位置に書かれた文字を認識することができる。

画面の左側にある送信可ランプとボタン選択ランプ

とは、それぞれ、ソフトウェアの指令によって点滅し、文字を書くべきタイミングとボタン選択の必要性を知らせるためのものである。

2.3 ソフトウェア

(1) 前処理と折線図形の抽出

計算機に入力された座標は、使用した遠隔描画装置のリンク機構の特性と受信機の増幅器の非直線性などのために、通常の直交座標での座標とはかなり異なっており、このままでは利用できない。そこで、まず、あらかじめ蓄えられている約 250 点分の座標変換表を用い線形補間して、入力座標データを直交座標系の座標データに変換する。この結果得られた座標 (X, Y) (0 ≤ X ≤ 1300, 0 ≤ Y ≤ 900) の系列を、以下では入力データとして考えることにする。

(X, Y) の系列は、まず、Groner の方法²⁾によって平滑化される。点を、その X, Y 座標を成分とする2次元ベクトル P または S で表わす。図3に示されるように、ペンが画面について新しい座標データが入力されるたびに、この新しい入力点 P_i の座標データと、それまでのデータから得られている軌跡の平滑点 S_j との荷重平均

$$S_{j+1} = (m+n)^{-1} (m S_j + n P_i) \quad (1)$$

(i=0, 1, …, j=0, 1, …)

をとって軌跡の平滑点を伸ばして行く。ただし、ペンが初めて画面につく点 P₀ については、S₀ = P₀ とする。また、ペンが停止またはきわめてゆっくり移動する場合に、不要な荷重平均の計算を行なわないために、一定の正数 r を設け

$$|S_j - P_i| < r$$

となる場合には、P_i は無効データとして S_{j+1} を求める(1)式の計算は行なわない。たとえば、図3で平滑点 S₂ に対して、入力点 P₃ は S₂ を中心とする半径 r の円 C の内部にあるから、P₃ に対しては平滑点を求める計算はされず無効データとなる。実際のプログラムでは、m : n = 1 : 3, r = 5 に選んだ。

このようにして得られた平滑点の系列から、つぎの

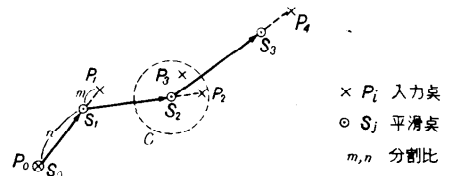


Fig. 3 Smoothing of the input points by weighted sum averaging technique

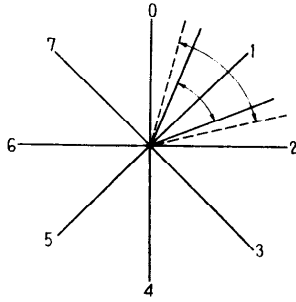


Fig. 4 Eight quantized directions and the ranges for the quantized direction "1"

ような Bernstein の方法¹⁾によって、逐次的にセグメントを抽出する。

- 1) ペンが画面についたらセグメントの始まりとし、入力点を最初の平滑点 $S_0(j=0)$ とする。
- 2) 平滑点を入力し S_{j+1} とする。
- 3) ペンが画面についていれば、手順4を実行する。そうでなければ、終了する。
- 4) 図4のように番号をつけた8量子化方向 (0, 1, ..., 7) の中心線 L_d に対し、互いに重ならない $L_d \pm 22.5^\circ$ の扇形領域 (例: 量子化方向1に対して実際の範囲) を考え、新たに入力された平滑点 S_{j+1} と1つ前の平滑点 S_j とから、これら2つの平滑点を結んだベクトル $S_{j+1}-S_j$ の方向が含まれる領域に対応する量子化方向 d を定め、 S_{j+1} に対応させる。
- 5) $k=j+1$ にする。
- 6) 平滑点を入力し S_{k+1} とする。もしペンが画面についていれば、手順7を実行する。ペンが画面についていなければ、入力済みの最後の平滑点までを1本のセグメントとして登録し、終了する。
- 7) S_{k+1} と1つ前の平滑点 S_k とを結ぶベクトル $S_{k+1}-S_k$ の方向が、 $L_d \pm 35^\circ$ の範囲 (例: 量子化方向1に対して点線の範囲) にはいっていれば、 S_{k+1} に量子化方向 d を対応させる。もし、この範囲にはいっていないければ手順9を実行する。
- 8) $k=k+1$ として手順6へもどる。
- 9) S_k までを1本のセグメントとして登録する。 $j=k$ とし、新たなセグメントの始まりとして手順4へもどる。

実際のプログラムにおいては、平滑化の処理と、セグメント抽出の処理とは組み合わされて逐次的に実行

されている。

このようにして、8量子化方向のいずれかをその方向として持つ抽出されたセグメントは、入力された文字図形を近似する折線図形の各線分を構成する。セグメントの情報は、1セグメント分が15語の1ブロックに蓄えられる。認識プログラムの中には90ブロック分のメモリ領域が設けてある。1つのセグメントが完結するごとに、空いているブロックにそのセグメントの始点、終点の座標、量子化方向、セグメントの長さの情報が書き込まれ、連続しているセグメントのブロックどうしが、両方向性のポインタによってリンクされリニア・リストになる。数字の「4」を書くときに生ずるような文字中の「とぎれ」に対しても独立したブロックが1個与えられる。このようなリニア・リストが入力された数個の文字に対してそれぞれ形成され、各リストの先頭のブロックの位置が認識プログラム内のレジスタに保持される。

図5に、実際の入力図形と、それから抽出された折線図形とをCRTディスプレイの上に表示したものを示す。さらに、そのうち数字「5」に対する実際のリニア・リストをラインプリンタで打ち出して表1に示す。この表では、横1行がセグメントまたはとぎれの1つのブロックを表わしている。ADRは各ブロックの先頭の語の番地、TYPEはブロックのタイプ、 X_0, Y_0, X_1, Y_1 はそれぞれ始点、終点の座標、 $\Delta X, \Delta Y$ は座標の増分、 l は長さ、 d は量子化方向、BPはバックワード・ポインタ、FPはフォワード・ポインタ、 L は相対長 (全長に対する百分率) である。

(2) 簡単なリスト処理と認識樹木

使用者が、制御用ボタンB2に触れることによって、入力したい数個の文字をすべて書き終わったことを合図すると、認識プログラムは各リストの処理を開始する。まず、1つの文字を構成している線の全長を計算し、各セグメントととぎれの長さを全長によって

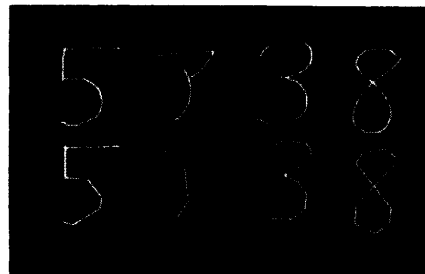


Fig. 5 Input patterns and their corresponding patterns (below)

Table 1 A linear list for the figure "5" displayed in Fig. 5

ADR	TYPE	X_0	Y_0	X_1	Y_1	ΔX	ΔY	l	d	BP	FP	L
4,500	1	187.0	767.0	187.3	702.7	0.3	-64.3	64.3	4	0	4,515	17
4,515	1	187.3	702.7	230.8	699.3	43.5	-3.4	43.7	2	4,500	4,545	11
4,545	1	230.8	699.3	259.2	664.1	28.4	-35.2	45.2	3	4,515	4,560	12
4,560	1	259.2	664.1	256.3	640.9	-2.9	-23.2	23.4	4	4,545	4,575	6
4,575	1	256.3	640.9	202.3	608.1	-54.0	-32.9	63.2	5	4,560	4,590	16
4,590	1	202.3	608.1	181.5	615.0	-20.8	6.9	22.0	6	4,575	4,620	5
4,620	2	181.5	615.0	181.0	764.0	-0.5	149.0	149.0	0	4,590	4,635	39
4,635	1	181.0	764.0	294.9	768.0	113.9	4.0	113.9	2	4,620	0	30

正規化し、相対長を該当するブロックに書き込む。つぎに、リストの前後で相対長が2%以下のものは、文字の形にほとんど影響しない微細な雑音としてリスト中から除く。さらに、それぞれのセグメントの量子化方向を、その両端点の座標から計算しなおし、その結果、同一量子化方向のセグメントが連続してしまう場合には、それらを1つのブロックにまとめる。これは平滑化された軌跡から量子化方向を出した際に、初めの2点によって決定された量子化方向が大きな影響を持ちすぎるのを補整する。このようにして修正されたリストに対して、連続したセグメントのブロックをまとめてストロークと呼び、1文字中のストロークの数を数える。1つのストロークは、ペンが画面についてから離れるまでの部分に対応し、たとえば、「2」は1画、「ア」は2画、「ホ」は4画である。

入力された図形が2画以上の場合には、それが濁点または半濁点付きの片仮名である可能性があるので、そのリストの最後の部分を調べて濁点または半濁点の有無を検査し、もし、存在する場合にはその種類を記録したのち、その部分をリストから切り離して、濁点、半濁点のつかない片仮名部分のリストのみを残す。

このようにして得られたリストから文字の特徴を抽出し、これらの特徴の有無によって、樹木状の認識論理を構成して認識を行なっている。特徴は、いくつかの基本的な性質を抽出するサブルーチン群を組み合わせで記述される。基本的なサブルーチンとしては、たとえば

- ・指定されたストローク内で、最初の、または最後の、または最大の折れ（連続した2セグメント間の量子化方向の差）を求める。
- ・指定されたストロークの最初と最後のセグメントの方向差を求める。
- ・指定された2つのセグメントの交点を求める。ただし、実際に交わらない場合には、それぞれのセグメントを延長して得られる仮想交点を求める。

- ・指定されたストローク中の最大長のセグメントの量子化方向を求める。
- ・指定されたセグメントに対し、別与えられた点と、このセグメントのどちら側にあるかを調べる。

などのサブルーチン群が用意された。これらのサブルーチン群を組み合わせ、入力図形を近似した折線図形データをもとに、人間が直観的な方法で各文字を区別する特徴の記述を樹木にプログラミングし、得られた樹木をさらに実際データによって修正するという手順を繰り返して認識樹木を作り上げた。

認識プログラムはアセンブリ言語を用いて書かれ、ワーキング・メモリを除いて約8K語の大きさになった。類似した文字、クと7、ノと1、ニと=、メと×、ソとリ、ナと+、などの書き方に注意を与え、10人の被験者に対しそれぞれ86種の文字を書かせたところ、約95%の文字が1回で正しく認識できた。認識処理に要する時間は、最大9文字に対して約1.3秒であった。

3. 遠隔端末からの使用

遠隔端末からの使用実験は、2章で述べた実時間文字認識方式を入力手段として応用したCRIARシステムで行なった。CRIARシステムは、手書きの片仮名と数字とを認識して文字列を入力し、入力された情報に基づいて計算機内で処理を行ない、さらに、適当なメッセージを合成音声で出力して人間との情報のやりとりをする man-machine communication system である。合成音声は、図1に示されるように、NEAC-L3 計算機によって制御されるターミナル・アナログ形音声合成装置^{8),9)}によって発生される。音声合成プログラム (NESS) に合成すべきセンテンスを音韻記号列として与えると、これに基づいて NESS プログラムが音声合成装置を制御する信号列を順次発生し、1回に約7秒以下のセンテンスを合成することができ

る。CRIARS の応用プログラムの1つは、計算機が合成音声でクイズの問題を出し、それに対し人間が文字で答えを書くと、計算機がその答えを処理して、いろいろなコメントを発して、つきつきとやりとりを続ける会話形のプログラムである。

遠隔端末からの実験のため、計算機の設置場所から直線距離で約 10 km 離れた地点に端末用の図形入力装置および音声周波増幅器とスピーカとを設置し、それぞれ計算機側の装置と日本電信電話公社の専用電話回線を介して接続した。伝送は回線損失約 20 dB 図形入力装置の送出レベル -5 dBm の状態で行なわれた。受信側に抵抗減衰器をそう入して調べた結果、10 dB のマージンがとれることが確認された。図形の伝送に関しては、遠隔描画装置の送信機と受信機とを対にして、文字、基本図形などを描いて調べたところ、忠実に伝送されることが確認された。しかしながら、遠隔描画装置の受信機のサーボ機構の特性のために、描かれた受信図形には現われないが、高い周波数成分を持ったパルス性の雑音が混入しているおそれがあった。そこで、さらに、伝送された図形信号を計算機に入力して実時間文字認識プログラム中の前処理ルーチンによって認識の基礎となる折線図形を抽出して調べたところ、問題となるようなパルス性の雑音はなく、専用電話回線を介さずローカルに図形入力装置から計算機に図形を入力した場合に比べてとくに差異は認められなかった。認識結果も、遠隔端末からの場合とローカルな場合との差異は認められず良好であった。この遠隔端末からの計算機の利用は、CRIARS の運用として、延べ 50 時間以上にわたって、全くの使用未経験者を含む 30 名以上の人によって行なわれたが、文字認識を主として動作は良好であり、伝送に伴う雑音の影響は問題なかった。

CRIAR システムのプログラムは、ワーキング・メモリを除いて約 17 K 語の大きさがあるため、プログラムを適当な単位ごとに分割して磁気ドラムに蓄え、必要なプログラムのみを順次コア・メモリにロードして使用した。このため、人間が1つの動作を行なってからそれに対する応答が得られるまでの待ち時間が、不都合なほど長くなる可能性があったが、最長の、文字認識と音声応答の組合せに対して約 3~4 秒程度であり、人間と計算機とのやりとりのテンポが幾分ゆっくりしはしたが、この待ち時間は不愉快なほど長くはなかった。

4. む す び

従来、計算機のそばに図形入力装置を置いてしか行なわれなかった実時間文字認識が、専用電話回線を利用して、計算機から直線距離約 10 km 離れた地点に設置された簡便な遠隔端末から行なわれた。さらに、この実時間文字認識方式は音声応答と組み合わせられ、man-machine communication system CRIARS が実現され、手書き文字入力-音声応答による計算機との会話に利用された。図形入力装置として、既製の遠隔描画装置が計算機に接続して利用されたが、十分に耐え、実用的であることが判明した。この遠隔描画装置は、図形入力装置として安価であって、経済的にもすぐれている。

しかしながら、文字認識能力をさらに向上させることや、文字と文字との区切りが、毎回合図することなく、プログラムによって自動的に行なわれるようにして、操作を簡単にする必要がある。また、ソフトウェアによって、すべての処理を行なっている現在の認識方式は、いわば、シミュレーションであり、実用化のためには、前処理の部分をハードウェア化し、多重化することによって、経済性を向上させる必要がある。これらは、今後に残された課題である。

終わりに、CRIAR システムの実現の機会を与えられ、終始ご指導いただいた当社中央研究所関口 茂 参与に感謝する。音声合成装置とその制御プログラムとは、加藤康雄研究マネージャーを中心とする音声研究グループによって開発されたもので、同装置の利用に対し、深く感謝する。専用電話回線の使用に対してお世話になった日本電信電話公社施設局専用設備課関係各位に深謝する次第である。

参 考 文 献

- 1) M. I. Bernstein: Computer Recognition of Online, Hand-written Characters. Memorandum RM-3753-ARPA (1964)
- 2) G. F. Groner: Real-time Recognition of Hand-printed Text. Proc. FJCC '66, pp. 591~601.
- 3) J. G. Simek & C. J. Tunis: Handprinting Input Device for Computer Systems. IEEE Spectrum, Vol. 4, No. 7 (1967), pp. 72~81.
- 4) G. M. Miller: On-line Recognition of Hand-generated Symbols. Proc. FJCC '69, pp. 399~412.
- 5) S. Hanaki, K. Asai, and K. Kiji: Real Time Recognition of Handprinted Characters—

- Japanese Character "KATAKANA" Recognition— NEC Research & Development, No. 14 (1969), pp. 80~86. この論文の summary が、同一標題で下記にある。
- Proc. of the Second Hawaii International Conference on System Sciences. (1969), pp. 345~348.
- 6) 奥村宣彦, 川瀬 進, 窪田潤二郎, 森 克彦: テレメール, NEC, No. 70 (1965), pp. 2~10.
 - 7) K. Asai, S. Hanaki, and K. Kiji: A New Remote Graphics Terminal. Proc. of the Third Hawaii International Conference on System Sciences. (1970), pp. 718~721.
 - 8) 加藤康雄, 落合和雄: ターミナルアナログ型音声合成装置の試作, 日本音響学会講演論文集, 1-2-14, (1966), pp. 61~62.
 - 9) Y. Kato, K. Ochiai, and S. Azami: Speech Synthesis by Rule Supplementarily Using Natural Speech. Reports of the 6th International Congress on Acoustics. (1968), pp. B-199~B-202.

(昭和 45 年 7 月 28 日受付)