

光学式文字認識のための手書き文字 C 6254-1979

(片 仮 名)

Handprinted Katakana Characters for Optical Character Recognition

1. 適用範囲 この規格は、光学式文字認識（以下、OCR という。）に使用する手書き片仮名文字の字形及びこれに関連する仕様を規定する。

2. 用語の意味 この規格で用いる主な用語の意味は、次のとおりとする。

- (1) 片 仮 名 アイウエオ カキクケコ サシスセソ タチツテト ナニヌネノ ハヒフヘホ マミムメモ ヤユヨ ラリルレロ ワヲン
・（濁点）、・（半濁点）、ー（長音記号）の49文字
- (2) 字 形 個々の片仮名の形状。
- (3) 字 体 一群の片仮名に対し規範として定められた字形の一組。
- (4) 線 素 個々の字形を構成する幅を持った線。
- (5) 線 幅 字形を構成する線素の幅。
- (6) 空 け き（隙） 字形で故意に保った線素間のすき間。
- (7) 突 出 し 長い線素にほぼ直交して伸びている短い線素。これは、ア、ウ、オ、カ、キ、ケ、サ、セ、タ、ナ、ス、ヒ、ホ、マ、ム、メ、ヤの字に見られる（4.の各字形を参照）。
- (8) か ぎ（鉤） 長い線素の始め又は終わりに、その線素とほぼ直交するように付ける短い線素。これは、ウ、カ、シ、セ、ソ、ツ、ヤ、ワ、ンの字に見られる（4.の各字形を参照）。
- (9) 短 線 素 複数の線素で構成されている字形において、比較的短い線素。これは、シ、ソ、ツ、ネ、ホ、ン、の字形に見られる（4.の各字形を参照）。
- (10) おう（凹）部の深さ かぎ（鉤）のある線素などで保つべきおう（凹）部の水平又は垂直に測った深さ（付図1参照）。
- (11) 帳票基準辺 機械読取りの際に基準となる辺（付図2参照）。
- (12) 記入わく 帳票基準辺に平行及び垂直な辺を有する、一つの片仮名を書く場所を示す長方形（内側）（付図2参照）。
- (13) 記入わくの幅 記入わくを構成する垂直線間の水平距離（付図2参照）。
- (14) 記入わくの高さ 記入わくを構成する水平線間の垂直距離（付図2参照）。
- (15) 記入帯 帳票基準辺に平行及び垂直な辺を有する、1行中の一組の連続した記入わくをすべて含む最小の長方形領域（付図3参照）。
- (16) 記入わく間隔 同一記入帯内の隣接する二つの記入わくの幅の垂直二等分線間の水平距離（付図2参照）。
- (17) 行間隔 列方向に隣接する二つの記入わくの高さの垂直二等分線間の垂直距離（付図2参照）。
- (18) 記入わく間余白 同一記入帯内で相互に隣接する二つの記入わくの水平距離（付図2参照）。
- (19) 文字境界 帳票基準辺に平行及び垂直な辺を有する、一つの字形のすべての線素をその内部に含む最小の長方形（付図2参照）。
- (20) 線素境界 ある文字を構成する一つの線素を字形とみなしたときにできる文字境界。
- (21) 垂直線素の傾き 字形を構成する垂直線素の線素境界における水平の辺と垂直の辺との比（付図4参照）。

- (22) 水平線素の傾き 字形を構成する水平線素の線素境界における垂直の辺と水平の辺との比(付図4参照)。
- (23) クリアエリア 帳票基準辺に平行又は垂直な辺を有する記入帯及びその回りの余白部を含む長方形領域(付図3参照)。
- (24) ドロップアウトカラー 人間の眼には周囲の余白部とはっきり区別できるが、当該OCRにはほとんど感知し得ない色相と濃さを持った色。

3. 字体の種類と呼び 字体の種類は1種類とし、その呼びはOCR-HKとする。

4. 字形 字形は、次のとおりとする。

アイウエオカキクケコサシスセソタチ
ツテトナニヌネノハヒフヘホマミムメ
モヤユヨラリルレロワヲン"。ー

5. 記入要項

- 5.1 一般的注意 一つの片仮名は、一つの記入わく内に4.に示された字形で書かなければならない。
- 5.2 空けき(隙) 4.の字形に示されている空けき(隙)は、1.0mm以上でなければならない。
また、4.の字形に示されている空けき(隙)以外の箇所に空けき(隙)ができるように書いてはならない。
- 5.3 突出し、かぎ(鉤)、輪、飾り 4.に示されている箇所以外に余分な突出し、かぎ(鉤)、輪、飾りなどを付加して書いてはならない。
- 5.4 字形に関する注釈 字形相互間の差異を明確にするために、特に注意して書くべき文字について、その変動許容範囲を附属書に示す。
- 5.5 字形の大きさ 文字境界の幅と高さとの許容値は、表1のとおりとする。

表 1 字形の大きさの許容値

字形の種類	エコニヘユロン	その他
幅	$\frac{2}{3} W \sim W'$	
高さ	$\frac{1}{2} H \sim H$	$\frac{2}{3} H \sim H$

- 備考 1. H は記入わくの高さ、 W はその幅である。
2. リ、トについては表1の幅の許容値は適用しない。
3. ーについては表1の高さの許容値は適用しない。
4. “。”、“”、“。”については表1の許容値は適用しない(附属書参照)。
5. 幅の最小値は2.5mm、高さの最小値は3mmとする。

5.6 垂直、水平線素の傾き 垂直線素の傾きは $\frac{1}{5}$ まで、水平線素の傾きは $\frac{1}{8}$ までとする。

5.7 線幅 線幅の許容範囲は、0.25mmから0.65mmまでとする。

備考 記入わくとその行配置の種類が1号の場合 線幅の最大値は、0.55mmとする。

6. 記入わくとその行配置

6.1 種類 記入わくとその行配置の種類は、表2の3種類に分類し、記入わくの幅、記入わく間隔の値は、表2のとおりとするのが望ましい。

表 2 記入わくとその行配置

単位 mm

種 類	I 号	II 号	III 号
記入わくの幅 (W)	$4 > W \geq 3.4$	$5 > W \geq 4.0$	$W \geq 5.0$
記入わく間隔 (P)	$P \geq 3.9$	$P \geq 4.5$	$P \geq 5.5$

- 備 考 1. 記入わくの幅の最小値は、3.4mm とする。
 2. 記入わく間余白の最小値は、0.5mm とする。
 3. 記入わくの高さ(H)は、幅(W)の1.1～1.4倍とする。

6.2 分光特性 記入わくは、ドロップアウトカラーで印刷するものとする。

7. 記入帯 記入帯は、記入わくを一まとめにして1行ごとに定めなければならない。ただし、1行を幾つかの認識対象片仮名の組に分割して、それぞれ別の記入帯としてもよい。

また、記入帯の行間隔は、表3のとおりとする。

表 3 記入帯の行間隔

単位 mm

種 類	I 号	II 号	III 号
行 間 隔 (LS)	$LS \geq 8.5$		$LS \geq 12.7$

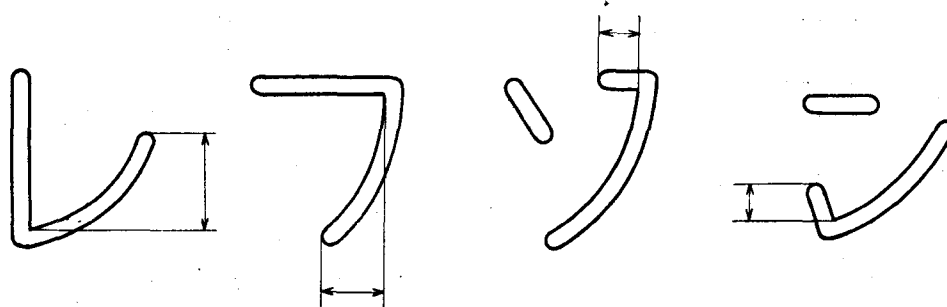
備 考 クリアエリアが他の行の記入帯のわく内に入らないことを条件とする。

8. クリアエリア クリアエリアの上下の境界と記入帯の上下の境界との垂直距離は4mm以上とする。クリアエリアの左右の境界と記入帯の左右の境界との水平距離は、6mm以上とする。

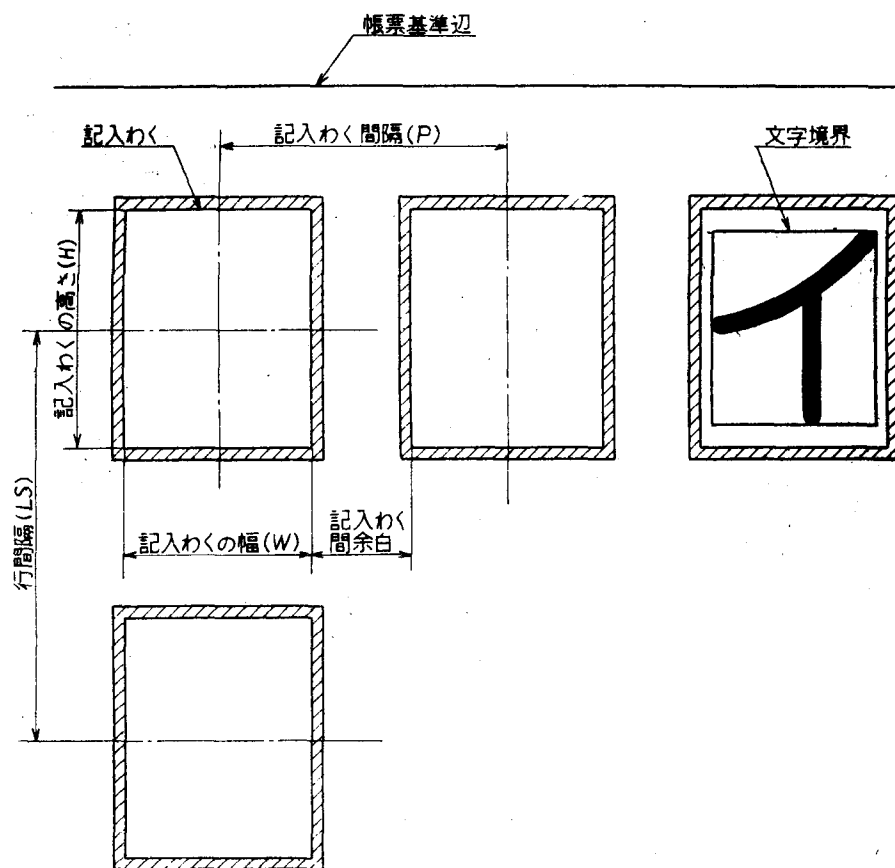
クリアエリア内には記入わく及び認識対象片仮名しか存在してはならない。クリアエリアは隣接する記入帯のクリアエリアと重なってもよいが、他の記入帯と重なってはならない(付図2参照)。

備 考 ドロップアウトカラーで印刷された文字は、クリアエリア内にあっても良い。

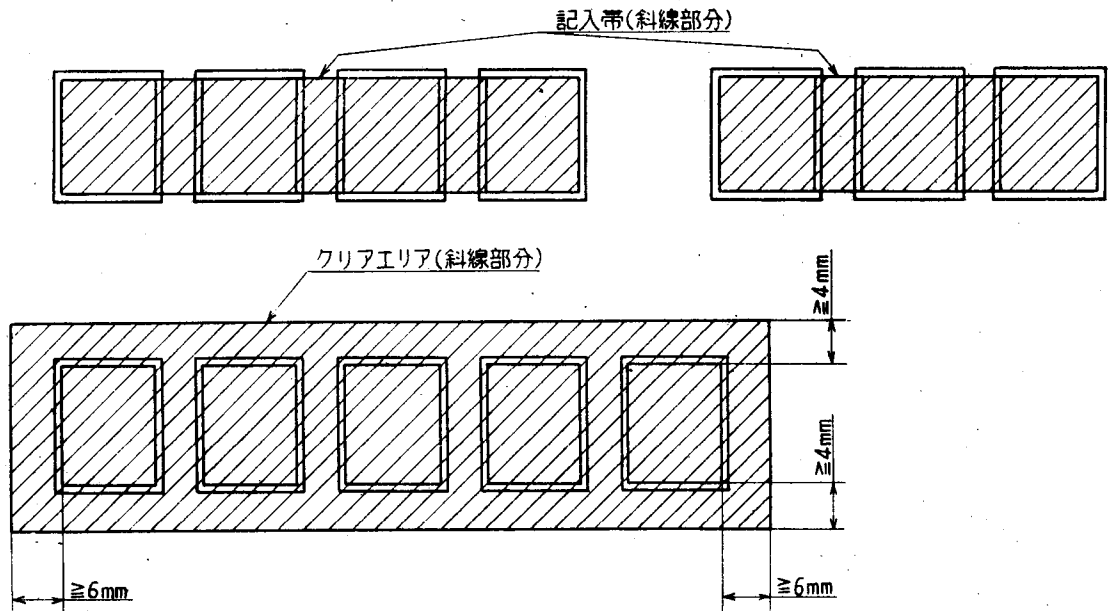
付 図 1 おう (凹) 部の深さの測り方



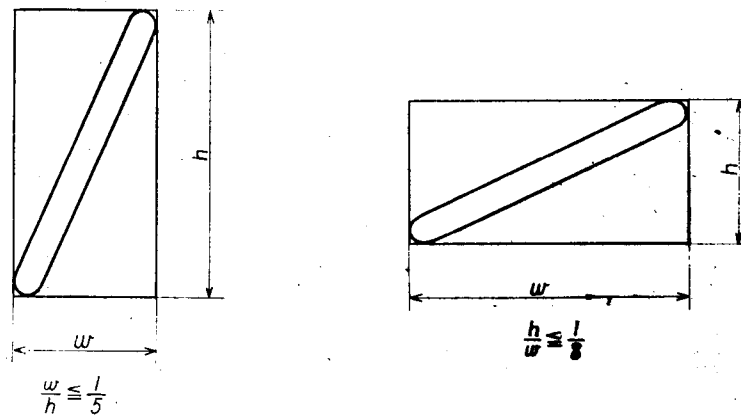
付 図 2 帳票基準辺、記入わく、文字境界



付 図 3 記入帯, クリアエリア



付 図 4 垂直水平線素の傾き

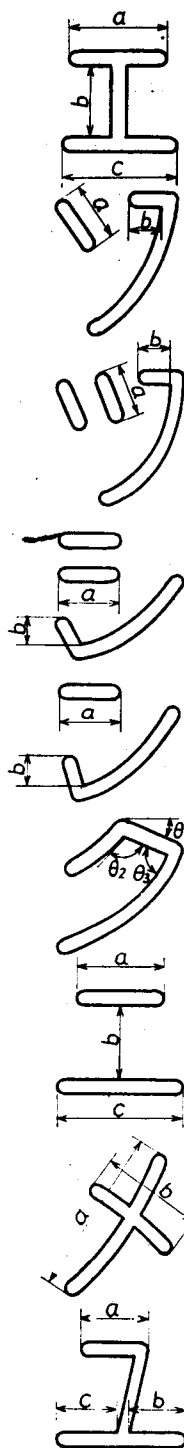


附 属 書 字形に関する注釈

字 形

条 件

備 考



$$a < c$$

$$\frac{a}{b} > 1$$

$$\frac{b}{a} \cong \frac{1}{2}$$

$$\theta_1 = 20^\circ \sim 30^\circ$$

$$\theta_2 \cong \theta_3 \cong 90^\circ$$

$$\frac{a}{b} \cong 1$$

$$\frac{a}{c} \cong \frac{2}{3}$$

$$\frac{1}{3} < \frac{b}{a} \leq \frac{1}{2}$$

$$a \cong b \cong c$$

光学式文字認識のための手書き文字 C 6255-1979

(数 字)

Handprinted Numerals for Optical Character Recognition

1. 適用範囲 この規格は、光学式文字認識（以下、OCR という。）に使用する手書き数字の字形及びこれに関連する仕様を規定する。
2. 用語の意味 この規格で用いる主な用語の意味は、次のとおりとする。
 - (1) 数 字 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 のアラビア数字。
 - (2) 字 形 個々の数字の形状。
 - (3) 字 体 一群の数字に対し、規範として定められた字形の一群。
 - (4) 線 素 個々の字形を構成する幅を持った線。
 - (5) 線 幅 字形を形成する線素の幅。
 - (6) 突 出 し 長い線素にはほぼ直交して伸びている短い線素。
 - (7) か ぎ (鉤) 長い線素の始め又は終わりに、その線素とほぼ直交するように付ける短い線素。
 - (8) 帳票基準辺 機械読取りの際に基準となる辺 (付図 1 参照)。
 - (9) 記入わく 帳票基準辺に平行及び垂直な辺を有する、一つの数字を書く場所を示す長方形 (内側) (付図 1 参照)。
 - (10) 記入わくの幅 記入わくを構成する垂直線間の水平距離 (付図 1 参照)。
 - (11) 記入わくの高さ 記入わくを構成する水平線間の垂直距離 (付図 1 参照)。
 - (12) 記 入 帯 帳票基準辺に平行及び垂直な辺を有する、1 行中の一組の記入わくをすべて含む最小の長方形領域 (付図 2 参照)。
 - (13) 記入わく間隔 同一記入帯内の隣接する二つの記入わくの幅の垂直二等分線間の水平距離 (付図 1 参照)。
 - (14) 行 間 隔 列方向に隣接する二つの記入わくの高さの垂直二等分線間の垂直距離 (付図 1 参照)。
 - (15) 記入わく間余白 同一記入帯内で相互に隣接する二つの記入わくの水平距離 (付図 1 参照)。
 - (16) 文字境界 帳票基準辺に平行及び垂直な辺を有する、一つの字形のすべての線素をその内部に含む最小の長方形 (付図 1 参照)。
 - (17) 線素境界 ある文字を構成する一つの線素を字形とみなしたときにできる文字境界。
 - (18) 垂直線素の傾き 字形を構成する垂直線素の線素境界における水平の辺と垂直の辺との比 (付図 3 参照)。
 - (19) 水平線素の傾き 字形を構成する水平線素の線素境界における垂直の辺と水平の辺との比 (付図 3 参照)。
 - (20) クリアエリア 帳票基準辺に平行及び垂直な辺を有する、記入帯及びその回りの余白部を含む長方形領域 (付図 2 参照)。
 - (21) ドロップアウトカラー 人間の眼には周囲の余白部とはっきり区別できるが、当該 OCR にはほとんど感知し得ない色相と濃さを持った色。
3. 字体の種類と呼び 字体の種類は、1 種類とし、その呼びは OCR-HN とする。

4. 字 形 字形は、次のとおりとする。

0123456789

5. 記入要項

5.1 一般的注意 一つの数字は、一つの記入わく内に 4. に示された字形で書かなければならない。

5.2 突出し、かぎ(鉤)、輪、飾り 4. に示されている箇所以外に、余分な突出し、かぎ(鉤)、輪、飾りなどを付加して書いてはならない。

5.3 字形の大きさ 文字境界の幅と高さとの許容値は、表 1 のとおりとする。

表 1 字形の大きさの許容値

幅	$\frac{1}{2}W \sim W$
高	$\frac{1}{2}H \sim H$

備考 1. H は記入わくの高さ、 W はその幅である。

2. 数字“1”については、記入わくの中央に記入することとし、表 1 の幅の許容値は適用しない。

3. 幅の最小値は 2.5 mm、高さの最小値は 3 mm とする。

5.4 垂直水平線素の傾き 垂直線素の傾きは $\frac{1}{5}$ まで、水平線素の傾きは $\frac{1}{8}$ までとする。

5.5 線 幅 線幅の許容範囲は 0.25 mm から 0.75 mm までとする。

備考 記入わくとその行配置の種類が I 号の場合 線幅の最大値は 0.55 mm とする。

6. 記入わくとその行配置

6.1 種 類 記入わくとその行配置の種類は、表 2 の 3 種類に分類し、記入わくの幅、記入わく間隔の値は、表 2 のとおりとするのが望ましい。

表 2 記入わくとその行配置

単位 mm			
種 類	I 号	II 号	III 号
記入わくの幅 (W)	$4 > W \geq 3.4$	$5 > W \geq 4.0$	$W \geq 5.0$
記入わく間隔 (P)	$P \geq 3.9$	$P \geq 4.5$	$P \geq 5.5$

備考 1. 記入わくの幅の最小値は 3.4 mm とする。

2. 記入わく間余白の最小値は 0.5 mm とする。

3. 記入わくの高さ(H)は幅(W)の 1.2 ~ 1.4 倍とする。

6.2 分光特性 記入わくはドロップアウトカラーで印刷するものとする。

7. 記 入 帯 記入帯は記入わくを一まとめにして 1 行ごとに定めなければならない。ただし、1 行を幾つかの認識対象数字の組に分割して、それぞれ別の記入帯とすることができる。

また、記入帯の行間隔は、表 3 のとおりとする。

表 3 記入帯の行間隔

種 類	単位 mm		
	I 号	II 号	III 号
行 間 隔 (LS)	$LS \geq 8.5$		$LS \geq 12.7$

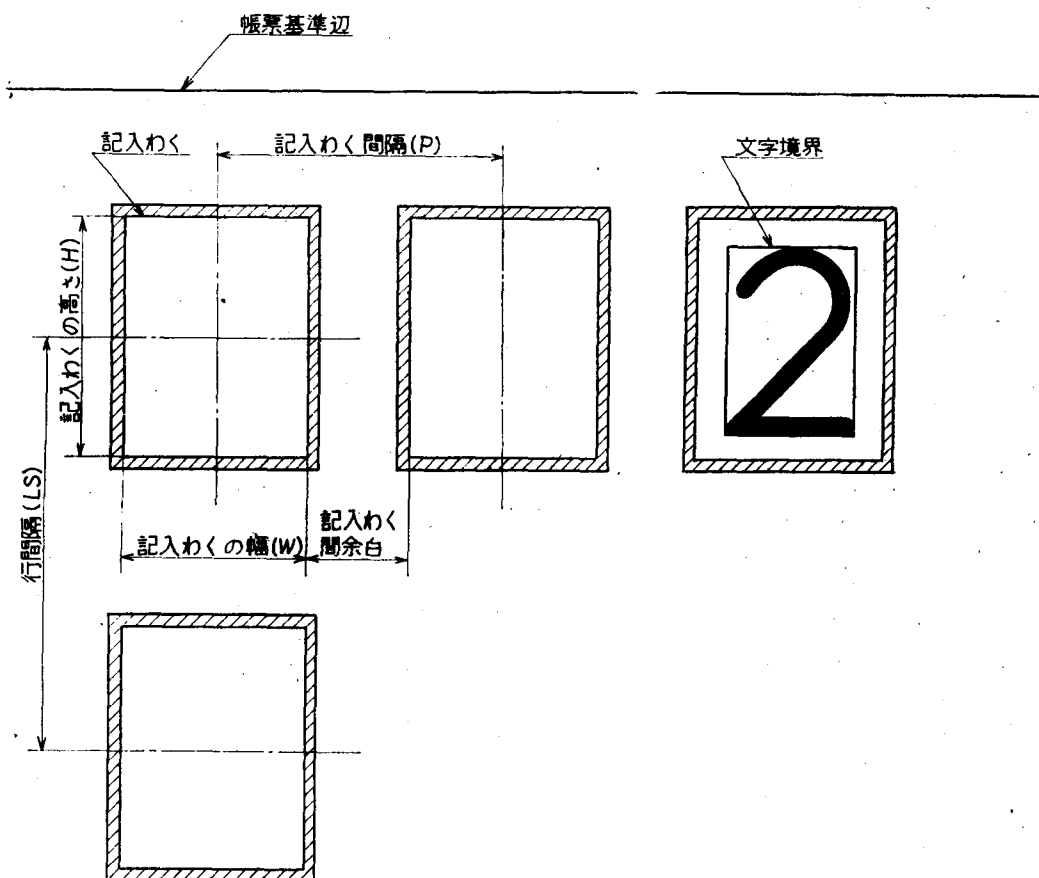
備 考 クリアエリアが他の行の記入帯のわく内に入らないことを条件とする。

8. クリアエリア クリアエリアの上下の境界と記入帯の上下の境界との垂直距離は4mm以上とする。クリアエリアの左右の境界と記入帯の左右の境界との水平距離は、6mm以上とする。

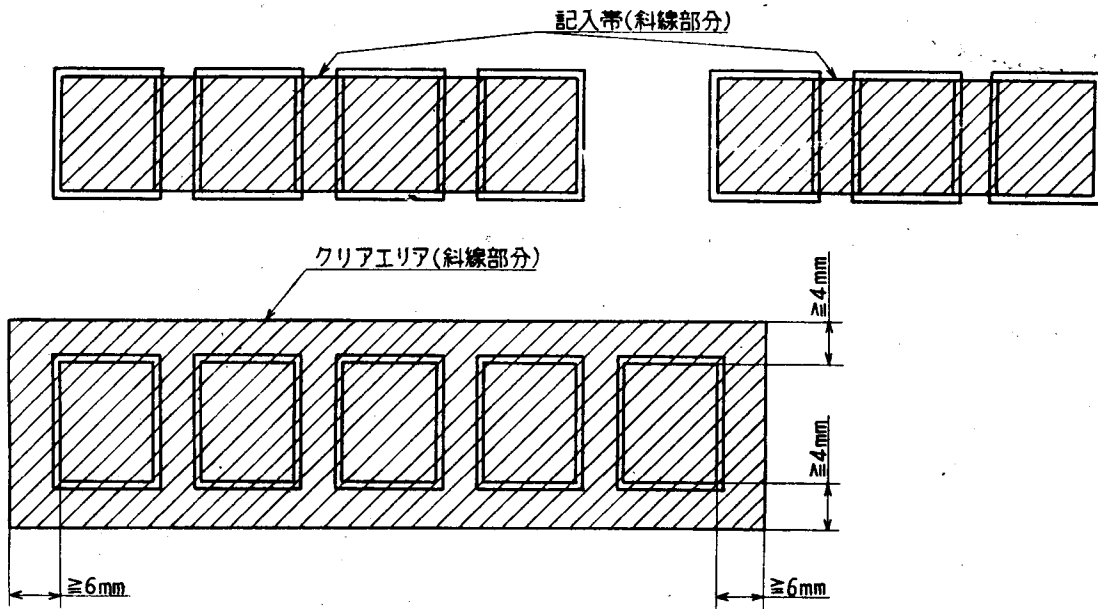
クリアエリア内には、記入わく及び認識対象数字しか存在してはならない。クリアエリアは隣接する記入帯のクリアエリアと重なってもよいが、他の記入帯と重なってはならない(付図2参照)。

備 考 ドロップアウトカラーで印刷された文字は、クリアエリア内にあっても良い。

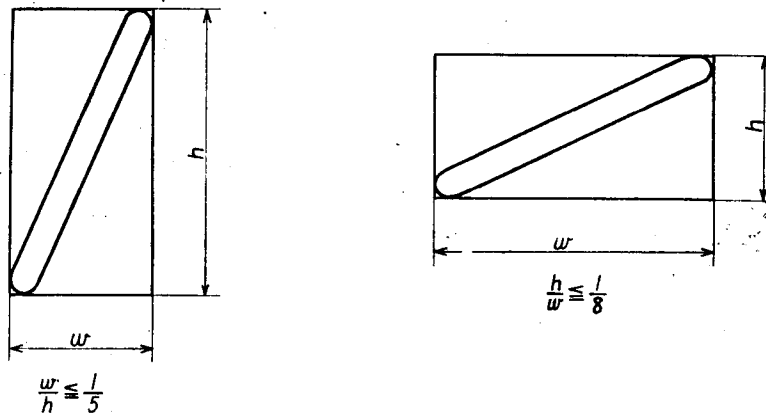
付 図 1 帳票基準辺、記入わく、文字境界



付 図 2 記入帯, クリアエリア



付 図 3 垂直水平線素の傾き



光学式文字認識のための手書き文字(数字) 解説

制定の主旨 文字は人間相互の通信手段として、長い歴史を持つものであるが、それが機械への命令、データとしても使用されるようになってきた。そこで機械にも分かるような明確な文字の形が必要不可欠となった。それに文字の集合、例えば、アラビア数字、英字、片仮名などは相互に無関係に発明、使用されてきたものであり、お互いの文字形で非常に似たものがある。典型的な例として、数字の零“0”と英字のオー“O”は完全に同一形である。このような場合には、人間同志の通信にも混乱を引き起こすので、実際プログラマーはそれぞれ種々の紛らわしい文字に対し、形を考えてそれらが慣用されているのが現状である。

特に日本は、和漢洋にわたり、上に述べたように、三つの文字の集合を日常使用している。人間の場合には機械にはまだできない状況判断から、これらが混同して使用されている場合でも、紛らわしい文字がほとんど分かるのであるが、現在の機械にはまだまだその能力はない。ここは正確に情報を伝達させる意味からも、やはり若干人間側が譲歩して機械に分かるよう、明確な文字形を書く必要があるわけである。逆をいえば、この JIS を満たすようにきちんと文字を書いていただければ、現在の OCR 技術は、ほとんど 100% 正しく文字を読み取ることができるということである。

しかし、文字は人間相互の約束事であるという側面を持つから、原理的には、同じ文字概念に対し、色々な形が考えられ、それを機械に学習させて、使用され得るわけであるが、これでは機械を含めた近代の文明社会が混乱を来すおそれがあり、ここに OCR 用手書き文字形の規格を制定することになった。この作業は、平松啓二教授を委員長としては5年間にわたり、関連機関の支援、協力を得、中央機関の学識経験者、ユーザ、メーカーの OCR のエキスパート、デザイナーによる活発な調査、実験、討議の下に制定されたものである。

1. 適用範囲 この規格の適用範囲はいわゆるアラビア数字であるが、しかし、単にそれだけが単独に使用される場合ではなくて、他に片仮名文字、英字記号とともに混用される場合を想定している。その意味で、この規格は若干厳しいものとなっている。

4. 字 形 上に述べたように特に“7”の右辺の線素のそりは、片仮名のクとワと区別するためのものであり、“5”の上部の突出しは英字の S と区別するためのものである。しかし、この場合は、実際にこのように突き出して書く人のほうが3分の2以上であるという統計データにも基づいている。一方“1”の単純な形は非常に重要である。それは、“1”と“1”のように書かれると、それらはそれぞれ“7”と“2”と非常に紛らわしい。この場合は、数字の集合の中で混乱が起きる。

5. 記入要項

5.2 突出し、かぎ(鉤)、輪、飾り これらの項目中、特に輪について説明する。ここで輪というのは通常ループといわれるもので、具体的には数字“2”で“ふ”のように書いた時に現れるループを輪と称し、このような輪を余分に書く事を禁じている。

5.3 字形の大きさ 表1で“1”は当然例外となる。それが備考2.に記されている。ここで記入わくの中央に記入するというのは、上又は下に偏して書くとそれぞれクォーテーションマークとコンマに間違えおそれがあるからである。

5.4 垂直水平線素の傾き これは確かに数字のみを読取り対象とする場合には厳しい規定であるが、1.適用範囲で述べたように混用を考慮してのことである。例えば“1”を傾けると片仮名の“ノ”と非常に紛らわしい。

6. 記入わくとその行配置

6.1 種 類 表2で種類としてⅠ号、Ⅱ号、Ⅲ号とあるが、これは6文字/インチ、5文字/インチ、4文字/インチにそれぞれ相当する。この種類に分けたのは便宜のためである。ただ、それぞれの種類を使う場合には、表2の規定に従うのが良いということをいっている。

C 6255-1979 解説

なお、記入わくの幅 (W)、記入わくの間隔 (P)、記入わく間余白 (S) の間には $P=W+S$ の関係がある。したがって S の最小値を 0.5mm としているので、例えば、I 号の場合 W の最小値が 3.4mm なので、 P の最小値は上式より 3.9mm となる。

参 考

規格原案の作成 この規格の原案作成は、社団法人 日本電子工業振興協会に委託され、同協会の OCR 手書文字専門委員会が行った。以下に原案作成委員会の構成表を示す。

	氏 名	所 属 (順不同)
委 員 長	平 松 啓 二	東京電機大学
幹 事	市 野 学	東京電機大学
	森 俊 二	工業技術院電子技術総合研究所
委 員	山 崎 一 生	工業技術院電子技術総合研究所 (現・新潟大学)
	木 下 昌 浩	郵政省郵政局
	瀬 川 彰	労働省労働市場センター
	山 村 修 蔵	工業技術院標準部
	上 田 陸 奥夫	社会保険大学校
	山 田 庸 宏	日本電信電話公社横須賀電気通信研究所
	河 西 芳 彦	日本国有鉄道
	石 黒 栄 一	株式会社三越
	大 塚 富士雄	日本電気株式会社
	星 野 幸 夫	日本電気株式会社
	坂 井 邦 夫	東京芝浦電気株式会社
	藤 本 好 司	株式会社日立製作所
	中 西 通 明	富士通株式会社
	南 部 元	三菱電機株式会社
	羽 下 雄之輔	沖電気工業株式会社
	石 橋 昇	株式会社東京計器
	安 田 道 夫	日立青梅電子株式会社
	石 井 信 行	日本アイ・ビー・エム株式会社
	石 橋 秀 雄	日本エヌ・シー・アール株式会社
	野 寄 雅 人	パターン情報処理システム技術研究組合
	楡 木 武 久	社団法人 日本電子工業振興協会