

豊富な文字種を持つプリンタの実現性について

笠原 龍夫

(株) リコー

1. はじめに

近年、ディスプレイやプリンタに対する日本語処理機能の充実が著しいものがあり、解像度を向上させたり、キー入力方法に工夫をするなど、マンマシンインタフェースをより良くする手段が付加され単なるロギング端末からペンやノートの代りとなるワークステーションへとユーザの要求がより高度なものになってきた。

従って カナ漢字変換機能や文字認識、音声認識合成による入出力機能が盛んに研究開発されている。

しかし、人間が他に意志を伝えるときに重要な文字についての研究はあまりされていない。

特に和文の場合、欧米のようにタイプライタなどにより個人の文章を機械で作成する習慣が少ないので、機械により作成された文章への関心が薄いことが原因している。

そこで、将来、日本語処理機能の充実を図る上で、文字に係る問題点として、下の3点を仮定した。

- ・ 解像度の向上に伴うキャラクタジェネレータのメモリ容量の増加。
- ・ 人手によるフォント作成が必要に対応できなくなる。
- ・ キャラクタジェネレータにバリエーションのある文字を持たせることの可能性。

解像度の向上はレーザープリンタや高解像度ディスプレイの出現により始まったが、一般に普及させるには、ポータブルにする必要性があり、文字を和文対応にした場合、文字パターンメモリは莫大なものになってしまい、何らかの工夫が必要

と思われる。

次にフォント作成であるが、手書きの文字から筆者が判断できるように、例えば、明朝体でも使っている会社によって 個性がある。また、出力装置におけるドットの出力方法によっても、アウトプットされる文字パターンは少しずつ異なってくるので キャラクタジェネレータに入れる文字パターンを出力装置によって変える必要がある。従って手作業で文字の作成修理を行ってはいは需要をまかないきれなくなる可能性がある。線幅、ストロークのバランスを変化させ、要求に見合ったフォントを作成する機能が必要になってくると思われる。

次にレーザープリンタなどのノンインパクトプリンタでは自由な文字サイズ、文字種でアウトプットができるのでマルチフォント化は可能だが、1フォントずつ文字パターンをもつ方式ではメモリ容量が膨大になるので文字パターンを圧縮してメモリに貯め、アウトプット時に展開して使用するなどの機能が必要と思われる。

そこで、対応策として、

1. 合成文字による文字パターンメモリ容量の軽減 と 文字バリエーションの増加
2. 電子計算機導入による人の負担の軽減

を考えてみた。

文字合成¹⁾は、図-1. 1に示すように、文字の心線から基本ストロークを抽出し、骨格情報を得て、骨格辞書とし、これにストロークに関する太さの情報、ウロコ、ハネなどの飾り情報を付加することによりバリエーションに富んだ文字の作成が可能である。

図-1. 2はストロークに関する太さ情報と飾りの情報を変えることにより同一の骨格辞書から7種類のフォントを作成した例である。また、ストロークのサ

イズを変えることにより、変倍文字も、原理的には作成可能である。この方式では、文字パターンを特徴量で記憶されるのでメモリ容量も軽減できるし、ソフトウェア処理なので、ミニコンレベルでも、マイコンレベルでも応用が可能である。ただしキャラクタージェネレータとして応用する場合は、合成速度とメモリ容量がネックとなり、フォント作成機として応用する場合は、特徴抽出機能と、合成できるフォント種類の豊富さがネックになる。

ベアボン書体のストローク分解と文字エレメントの例

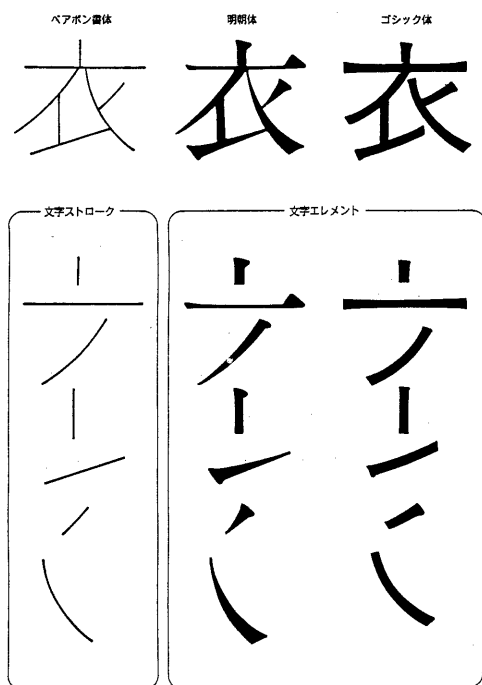
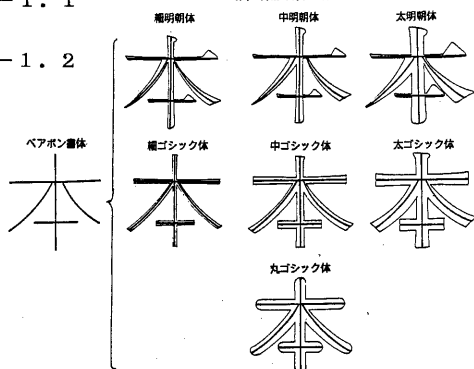


図-1.1

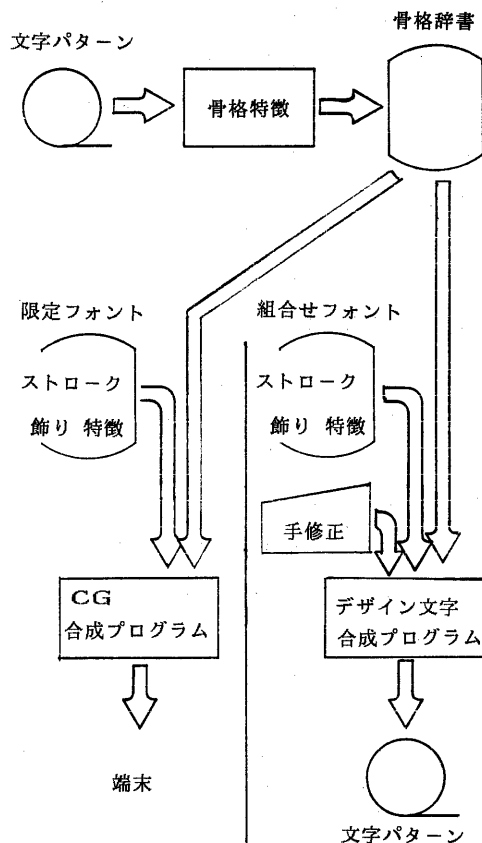
(同一画格を有する各種書体)

図-1.2



つまり、図-1.3に示すように、合成プログラムの生成できる文字に対する融通性により、適用範囲を決めることができるのでシミュレーションレベルでは同様に扱うことができる。

図-1.3



従って本研究では、合成文字による文字生成に主眼を置き、対照となる装置をプリンタ、文字をリコー明朝体として、豊富な文字種を持つプリンタ実現性について以下のような目標値を達成できるか否かシミュレーションしたので報告する。

- ・合成速度 0.1秒/文字
- ・メモリ容量 1/4以下
(ドットパターンと比較)
- ・特徴抽出作業 30文字/日

2. 合成文字作業方針

合成文字を作成する上で重要な文字構成要素の分類抽出、太線化手順などの処理の方式を決定した。

2.1 辞書構造

文字の構成要素を抽出する場合、一般的に部首、ストローク、飾りに分類して抽出するが、合成する際に少ない特徴量で高品質な文字が出力出来るように、構成要素ごとに調べてみた。

まず、部首であるが、メモリ容量の軽減には、有効な手段であるが、1つの部首が、すべての同じ部首に同様に使用できるかを”糸ヘン”、”草カンムリ”で調べてみた。

図-2.1.1

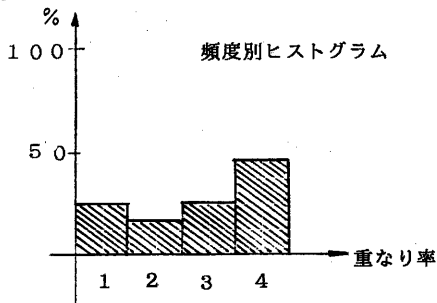
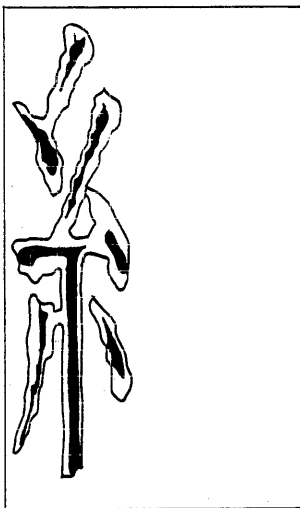


図-2.1.1が”糸ヘン”をもつ漢字8文字から”糸ヘン”を切りだし、同じサイズに変倍し、重ね合わせた結果である。黒い部分がすべての”糸ヘン”が重なった領域である。

次に”草カンムリ”の横棒と縦棒の交叉比率を調べた。図-2.1.2のようにAとB、CとDとEのように領域を決め、漢字8文字で頻度分布を取った。特に縦ストロークの分割比に大きなバラツキがあることがわかる。

このように”糸ヘン”、”草カンムリ”ともストロークのバランスが文字ごとに異なっていることがわかった。従い、部首を利用した場合文字全体のバランスが乱れてしまう危険があるので、部首の利用は取りやめた。

そこで辞書構造は下記のような3層構造にした。

- ・漢字辞書—文字の骨格情報を格納。

- 主にストロークの始点とサイズ

- ・ストローク辞書—

- ストロークの形、太線化情報を格納。

- ・飾り辞書—飾りの形を格納。

次にストローク、飾りの分類であるが、デザインの立場から見るとストロークは”はねあげ””点””はらい”...など42本に分類され、飾りは64x64ドットの明朝体の場合、48種類に分類されるので、これを利用した。分類は表-2.1.3, 2.1.4に記した。

また、ストロークを辞書に格納する場合、圧縮率と扱い易さからチェーンコードにして格納した。

飾りについてはドットで格納した。

図-2.1.2

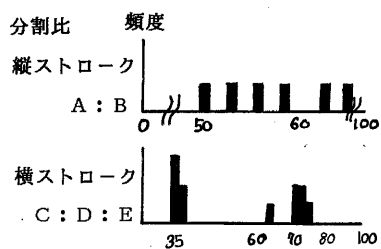
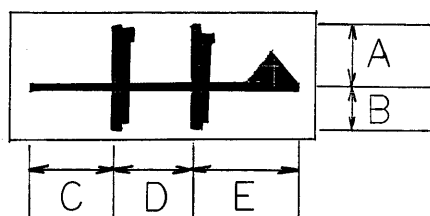


表-2.1.3

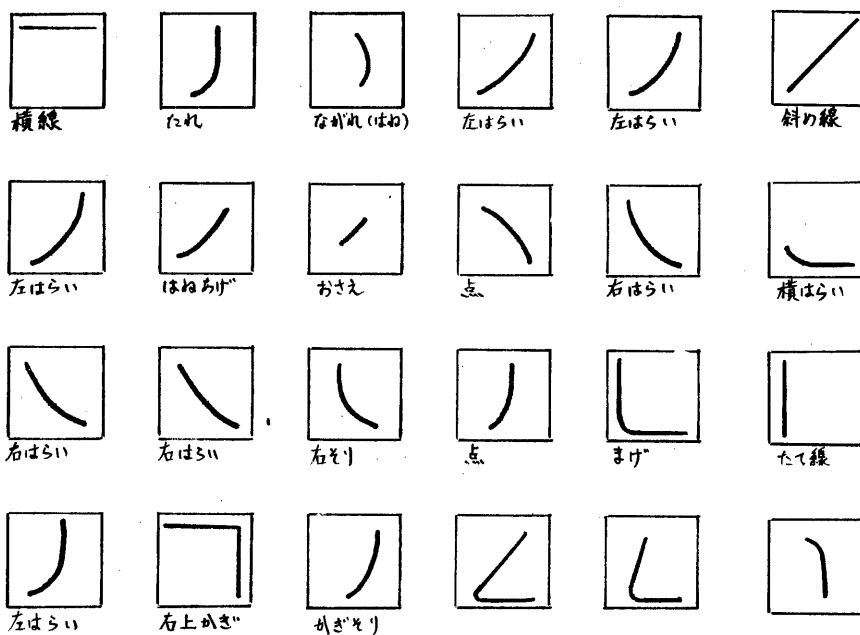


表-2.1.4

—	うろこ		┐			ㄣ	右はらい		┐	たてせん はねあげ	
			┐								
			┐								
┐	かとうろこ		┐	ほうろこ うろこ		／	はねあげ		┐	おさへり はねあげ	
										なみれ はね	
┐	あたま					／	おはらい				
┐	しめ					／	おさえ		┐	まげはね	

2.2 ストロークのバランス

文字は変倍するとストローク間のバランスが微妙に変わっていることに注目して、単純に変倍して作成した文字とデザイナーが作成した文字との誤差の算出を試みた。

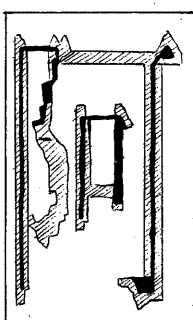
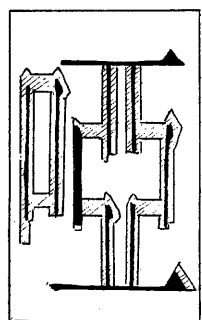
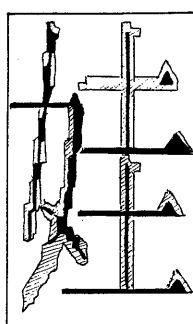
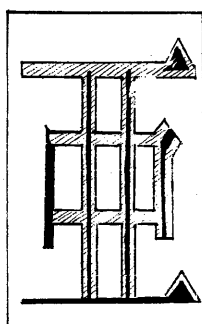
使用した文字は 64 X 64 ドットの明朝体から単純に変倍して作成した 40 X 40 ドットの明朝体とデザインした明朝体である。

2つの文字を重ね合せると 図-2.2.1 のように不一致画素が見られる。

図-2.2.1

黒領域：一致画素

斜線領域：不一致画素



この変倍処理における誤差の分布を 207 文字の漢字で調べてみると、表-2.2.2 のようになった。

(0, 0) を中心にストロークが一様に分布しているので相関関数の決定はむずかしい。

従って、変倍率ごとに変倍誤差をもつことは元文字パターンに忠実に再現する上で必要であるが、合成速度と辞書容量の点から採用は取りやめた。

表-2.2.2 デザイン文字と縮小パターンのストロークの変倍誤差と分布

X \ Y	-2	-1	0	1	2
-2	.	1	2	1	.
-1	.	6	15	8	.
0	1	12	26	7	2
1	1	5	8	1	1
2	.	.	1	1	.

. は 0.5% 以下

2.4 太線化方式

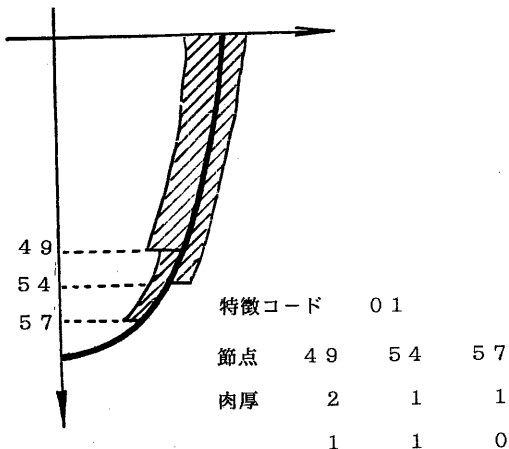
ストロークに対する肉付けは実際の文字の寸法に変倍したストロークの上を”平ペン”で書く方式を取った。ペンの回転は行っていない。

この方式は合成する書体が明朝体であることとストロークの分類をデザイナーの経験的法則から行っているために不安があり修正が簡単に行えることから採用した。

まず、”平ペン”の動かし方を決める為にストロークを、その方向、傾き、線幅、曲率によって5種類に分類し、分類に従う肉付け規則により、太線化処理を施した。

次に、太さが変化するストロークの場合、節点を設けた。太線化の方法は 図-2.4.1 に記した。

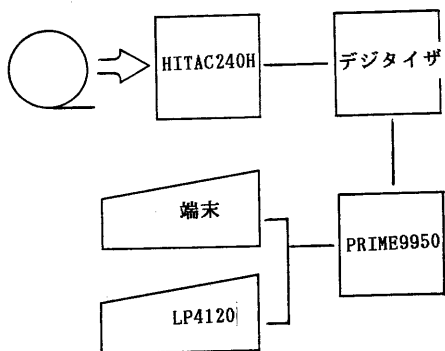
図-2.4.1



3. システム概要

主計算機はPRIME9950を用い、パターン修正はデジタイザ、辞書の修正入力は通常端末、アウトプットはレーザープリンタ (LP4120) を使用した。

またプログラムとして骨格抽出プログラム、合成プログラム、辞書修正プログラムがある。ステップ数は合成プログラムが 1000ステップ (フォートラン) である。

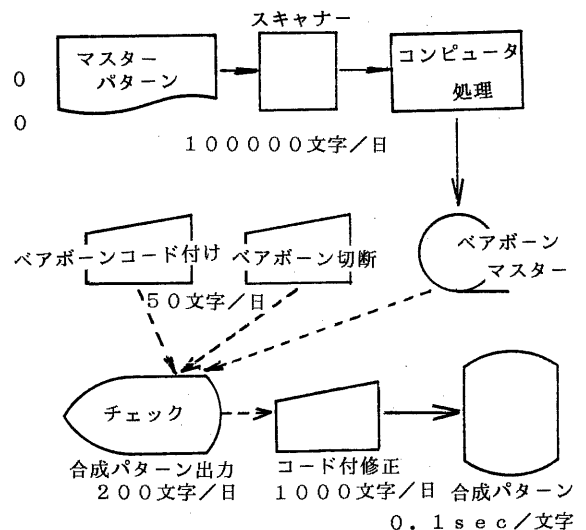


4. システムの運用

4.1 特徴抽出

処理速度は、辞書の作成が 平均40文字/日であるが、骨格情報は他のサイズの文字を変倍して作成する場合も利用できるのもので、この値は、総合してもっと大きな値になると思われる。

次に圧縮率であるが64x64明朝体 (500文字) でビットマップのものと比較した。



ビットマップ:

$$500 * 64 * 64 / 8 = 256 \text{ KB}$$

合成文字辞書:

漢字辞書:

$$\begin{aligned} & (\text{文字数}) * (\text{平均ストローク数} + 1) \\ & \quad * (\text{エレメント容量}) \\ & = 500 * (12 + 1) * 8 \\ & = 50.78 \text{ KB} \end{aligned}$$

ストローク辞書:

$$\begin{aligned} & (\text{ストローク数}) * (\text{エレメント容量}) \\ & = 42 * 124 \\ & = 5.78 \text{ KB} \end{aligned}$$

飾り辞書:

$$\begin{aligned} & (\text{飾り数}) * (\text{エレメント容量}) \\ & = 41 * 32 \\ & = 1.28 \text{ KB} \end{aligned}$$

$$\text{合計 } 57.14 \text{ KB}$$

したがって 圧縮率は 22%である。

4.2 文字作成

合成速度は文字を構成するストローク数に比例する。

従って合成速度は

$$(\text{平均ストローク数}) * (\text{ストロークの合成速度})$$

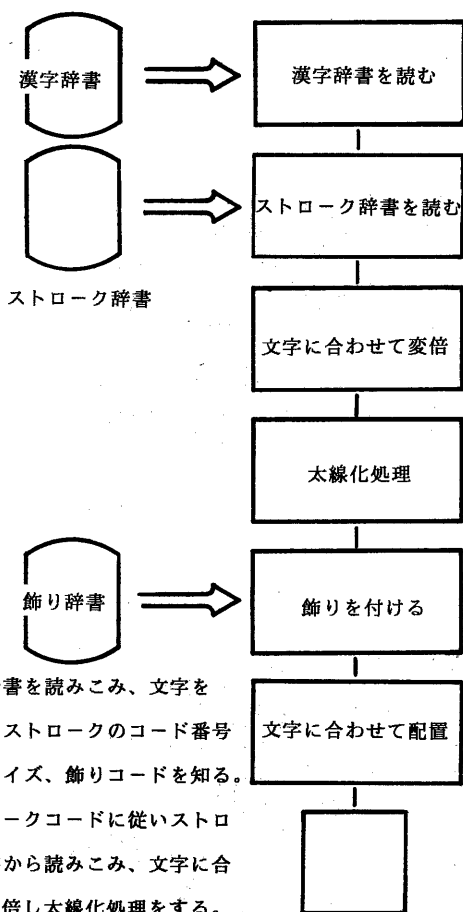
で算出した。

合成に使用した計算機はPRIME9950で

測定した結果 0.1秒/文字であった。

なお、500文字の平均ストローク数は

約13本であった。



1. 漢字辞書を読みこみ、文字を構成するストロークのコード番号、始点、サイズ、飾りコードを知る。

2. ストロークコードに従いストローク辞書から読みこみ、文字に合わせて変倍し太線化処理をする。

また、飾りが付いている場合は飾り辞書から読む。

3. 文字の対応する位置にストロークを配置する。

5. 考察

今回のシミュレーションで一番時間を必要としたのは辞書の評価、修正であった。

人間が評価するとプログラムでは、とうてい書けないようなことが起こり、辞書を修正することがたびたびあった。

また、変倍するとストローク間のバランスが微妙に変わる為、変倍誤差を測定したが、規則的な法則は発見出来なかった。

合成アルゴリズム、辞書構造は まだ、荒削りな部分が多いので、今後の開発ステップでもっと良い評価値が得られると思われる。

高解像度ディスプレイやレーザープリンタは高品質な画像をユーザに与えてくれる。しかし、搭載されているキャラクタジェネレータが高品質な文字をもっていないとは意味が無い。

文字作成の研究が進み、高品質な文字が簡単に作れるようになるよう精進してゆきたい。

6. 参考文献

- (1) The Concept Of
A Metafont
Donald E. Knuth
in Visible Language
vol. 16
- (2) マルチフォント合成器
阪本 久雄
National Technical
Report vol. 28 no. 05
Oct 1982
- (3) デジタル タイポグラフィ
C. ビゲロウ/D. デイ
サイエンス
- (4) 漢字 上下
佐藤 敬之助
丸善

印字サンプル

64 * 64 リコー明朝体

亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭葦芦梓庠
 幹扱宛姐虻飴絢綾鮎或粟裕安庵按暗案闇鞍杏
 以伊位依偉圉夷委威惟意慰易椅為畏異移維緯
 胃萎衣謂違遺医井亥域育郁一壹溢逸稻茨芋允
 印咽員因姻引飲淫胤蔭院陰隱韻吋右宇烏羽迂
 卯鵜窺丑碓臼渦噓唄蔚姥既浦瓜閨嚙云運雲荏
 餌營嬰影映曳榮永泳洩瑛盈穎穎英衛詠銳液疫
 益駢悅謁越閱榎厭円園堰奄宴延怨掩援沿演炎
 焰煙燕猿緣艷苑菌鉛鴛塩於汚甥凹央奧往応押
 旺横欧殴王翁襖黄岡冲荻億屋億臆臆乙俺卸恩

64 * 64 合成明朝体

亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭葦芦梓庠
 幹扱宛姐虻飴絢綾鮎或粟裕安庵按暗案闇鞍杏
 以伊位依偉圉夷委威尉惟意慰易椅為畏異移維
 緯胃萎衣謂違遺医井亥域育郁磯一壹溢逸稻茨
 芋允印咽員因姻引飲淫胤蔭院陰隱韻吋右宇烏
 羽迂雨卯鵜窺丑碓臼渦噓唄蔚姥既浦瓜閨嚙云
 運雲荏餌釵營嬰影映曳榮永泳洩瑛盈穎穎英衛
 詠銳液疫益駢悅謁越閱榎厭円園堰奄宴延怨掩
 援沿演炎焰煙燕猿緣艷苑菌遠鉛鴛塩於汚甥凹
 央奧往応押旺横欧殴王翁襖黄岡冲荻億屋億臆

64 * 64 合成ゴシック体

亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭葦芦梓庠
 幹扱宛姐虻飴絢綾鮎或粟裕安庵按暗案闇鞍杏
 以伊位依偉圉夷委威尉惟意慰易椅為畏異移維
 緯胃萎衣謂違遺医井亥域育郁磯一壹溢逸稻茨
 芋允印咽員因姻引飲淫胤蔭院陰隱韻吋右宇烏
 羽迂雨卯鵜窺丑碓臼渦噓唄蔚姥既浦瓜閨嚙云
 運雲荏餌釵營嬰影映曳榮永泳洩瑛盈穎穎英衛
 詠銳液疫益駢悅謁越閱榎厭円園堰奄宴延怨掩
 援沿演炎焰煙燕猿緣艷苑菌遠鉛鴛塩於汚甥凹
 央奧往応押旺横欧殴王翁襖黄岡冲荻億屋億臆