(12) UK Patent Application (19) GB (11) 2 208 556 (13) A

(43) Date of A publication 05.04.1989

(21) Application No 8818639.0

(22) Date of filing 05.08.1988

(30) Priority data
(31) 8719115 (32) 12.08.1987 (33) GB

(71) Applicant
Linotype Limited

(Incorporated in the United Kingdom)

Chelham House, Bath Road, Cheltenham, Gloucestershire, GL53 7LR, United Kingdom

(72) Inventors
Timothy Holloway
Michael Stewart Fellows
Fiona Georgina Elisabeth Ross

(74) Agent and/or Address for Service
Wynne-Jones Laine & James
22 Rodney Road, Cheltenham, Gloucestershire, GL50 1JJ, United Kingdom

(51) INT CL⁴
G06K 15/02

(52) UK CL (Edition J)
G4H HTS H1A H13D

(56) Documents cited
GB 2184876 A **GB 2175723 A** **GB 2140946 A**
GB 2016187 A **GB 2007413 A** **GB 1461413 A**
EP 0120481 A2 **US 3938099 A**

(58) Field of search
UK CL (Edition J) **G4H HTS**
INT CL⁴ **B41J, G06K**

(54) Printing

(57) A computerised typesetting apparatus copes with alphabets such as Nastaliq (the written form of Urdu) in which character shape selection depends on position in the word and contiguity to other characters, a sloping calligraphic script has to be simulated, and symbols such as dots and diacriticals have to be added in appropriate relation to the characters. Digital coded information is stored on all the different character shapes and on all the symbols together with rules on their inter-relationships. A keyboard with basic character and symbol selection keys is operable in conventional fashion and when each word or character/symbol combination has been keyed in, the computer makes the appropriate selections from a font and governs their arrangement.

GB 2 208 556 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

	i	m	f	u
ALIf1 3011				
ALIFu1 4011				
ALIf2 3012				

FIG.1.

	i	m	f	u
BEi1 1021		BEm1 2021	BEf1 3021	BEu1 4021
BEi2 1022		BEm2 2022	BEf2 3022	
BEi3 1023		BEm3 2023	BEf3 3023	
BEi4 1024		BEm4 2024	BEf4 3024	
BEi5 1025		BEm5 2025		
BEi6 1026		BEm6 2026		
BEi7 1027		BEm7 2027		
BEi8 1028		BEm8 2028		
BEi9 1029				

FIG.2.

i		m		f		u	
JIMi1 1031	ـجـ	JIMm1 2031	ـجـ	JIMf1 3031	ـج	JIMu1 4031	ج
JIMi2 1032	ـجـ	JIMm2 2032	ـجـ				
JIMi3 1033	ـجـ	JIMm3 2033	ـجـ				
JIMi4 1034	ـجـ	JIMm4 2034	ـجـ				
JIMi5 1035	ـجـ	JIMm5 2035	ـجـ				
JIMi6 1036	ـجـ	JIMm6 2036	ـجـ				

FIG.3.

i	m	f		u	
		DALf1 3041	ـد	DALu1 4041	د

FIG.4.

i	m	f		u	
		REf1 3051	ـر	REu1 4051	ر
		REf2 3052	ـر		
		REf3 3053	ـر		
		REf4 3054	ـر		
		REf5 3055	ـر		

FIG.5.

3/18

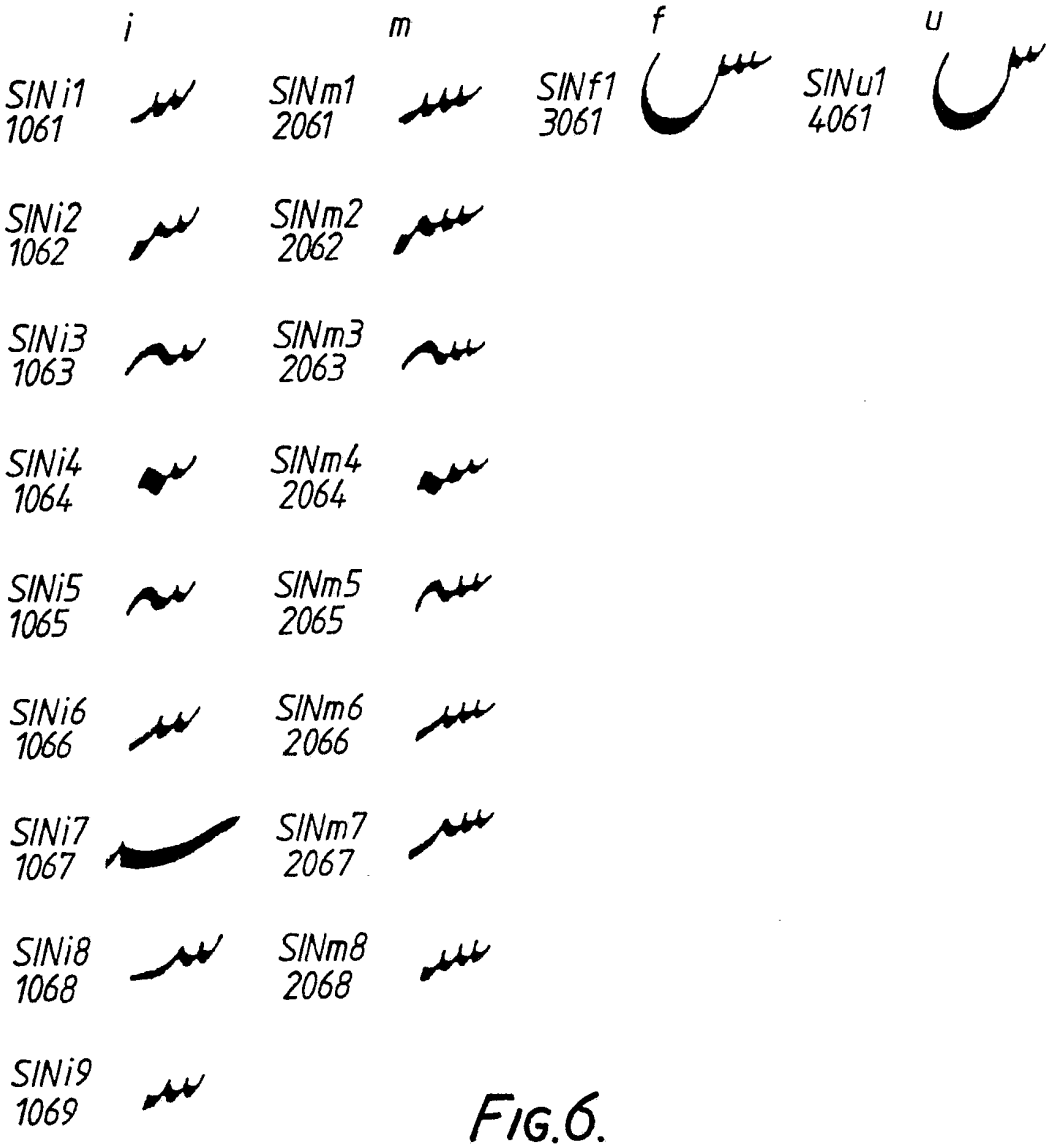

i
m
f
u
SINi1
1061
SINm1
2061
SINf1
3061
SINu1
4061
SINi2
1062
SINm2
2062
SINi3
1063
SINm3
2063
SINi4
1064
SINm4
2064
SINi5
1065
SINm5
2065
SINi6
1066
SINm6
2066
SINi7
1067
SINm7
2067
SINi8
1068
SINm8
2068
SINi9
1069

FIG.6.

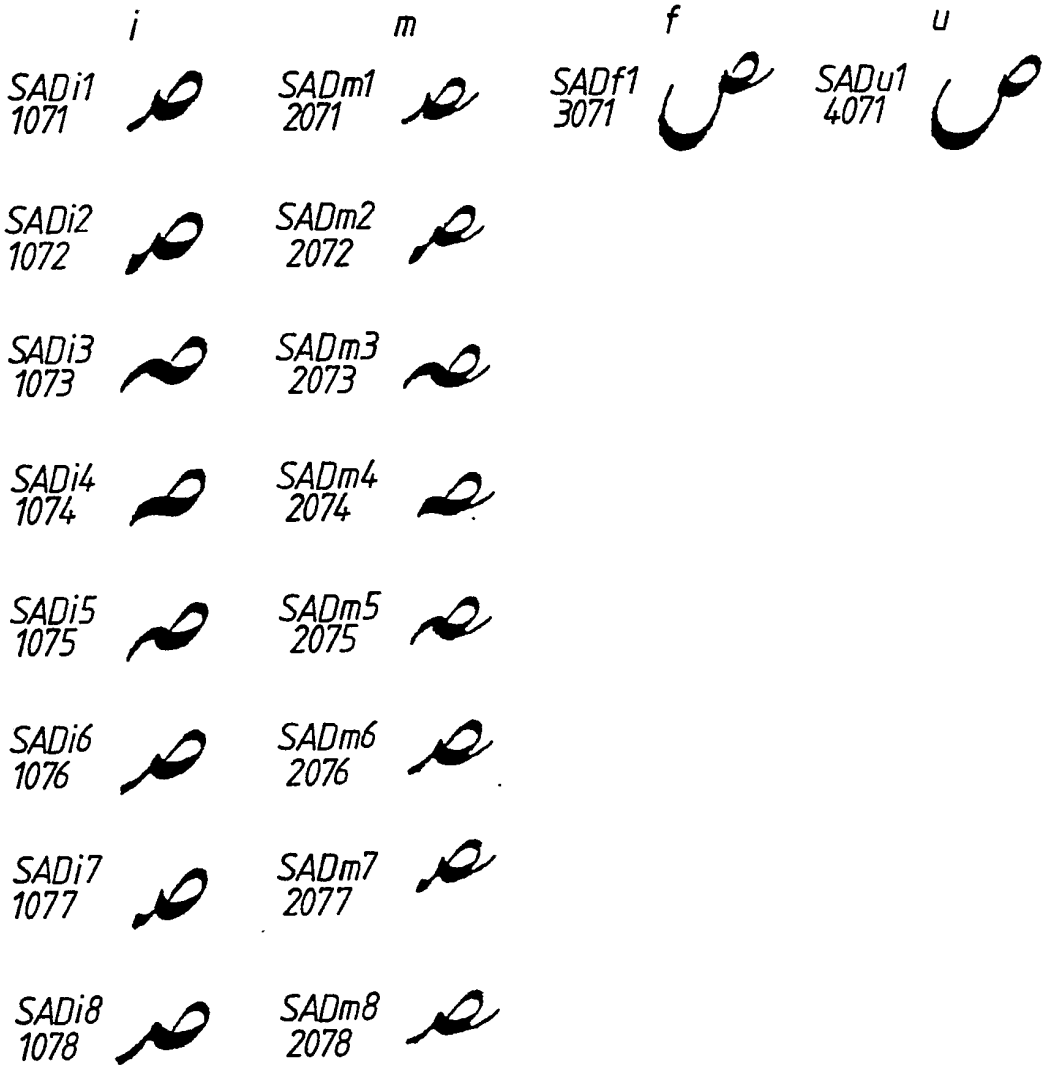

i
m
f
u
SADi1 1071
SADm1 2071
SADf1 3071
SADu1 4071
SADi2 1072
SADm2 2072
SADi3 1073
SADm3 2073
SADi4 1074
SADm4 2074
SADi5 1075
SADm5 2075
SADi6 1076
SADm6 2076
SADi7 1077
SADm7 2077
SADi8 1078
SADm8 2078

FIG. 7.

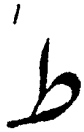
	i	m	f	u
TOE i1 1081		TOE m1 2081		TOE f1 3081		TOE u1 4081	
TOE i2 1082		TOE m2 2082					
TOE i3 1083		TOE m3 2083					
TOE i4 1084		TOE m4 2084					
TOE i5 1085		TOE m5 2085					
TOE i6 1086		TOE m6 2086					
TOE i7 1087		TOE m7 2087					

FIG.8.

	i	m	f	u
AIN i1 1091		AIN m1 2091		AIN f1 3091		AIN u1 4091	
AIN i2 1092		AIN m2 2092					
AIN i3 1093		AIN m3 2093					
AIN i4 1094		AIN m4 2094					
AIN i5 1095		AIN m5 2095					
AIN i6 1096							

FIG.9.

6/18

i		m		f		u	
FEi1 1101	ـو	FEm1 2101	ـهـ	FEf1 3101	ـفـ	FEu1 4101	ف
FEi2 1102	ـو	FEm2 2102	ـهـ				
FEi3 1103	ـو	FEm3 2103	ـهـ				
FEi4 1104	ـو	FEm4 2104	ـهـ				
FEi5 1105	ـو	FEm5 2105	ـهـ				

FIG.10.

i	m	f		u	
		QAFf1 3111	ـقـ	QAFu1 4111	ق
		QAFf2 3112	ـقـ		
		QAFf3 3113	ـقـ		

FIG.11.

	i	m	f	u
KAFi1 1121		KAFm1 2121		KAFf1 3121		KAFu1 4121	
KAFi2 1122		KAFm2 2122					
KAFi3 1123		KAFm3 2123					
KAFi4 1124		KAFm4 2124					
KAFi5 1125		KAFm5 2125					
KAFi6 1126		KAFm6 2126					
KAFi7 1127		KAFm7 2127					
KAFi8 1128		KAFm8 2128					

FIG.12.

8/18

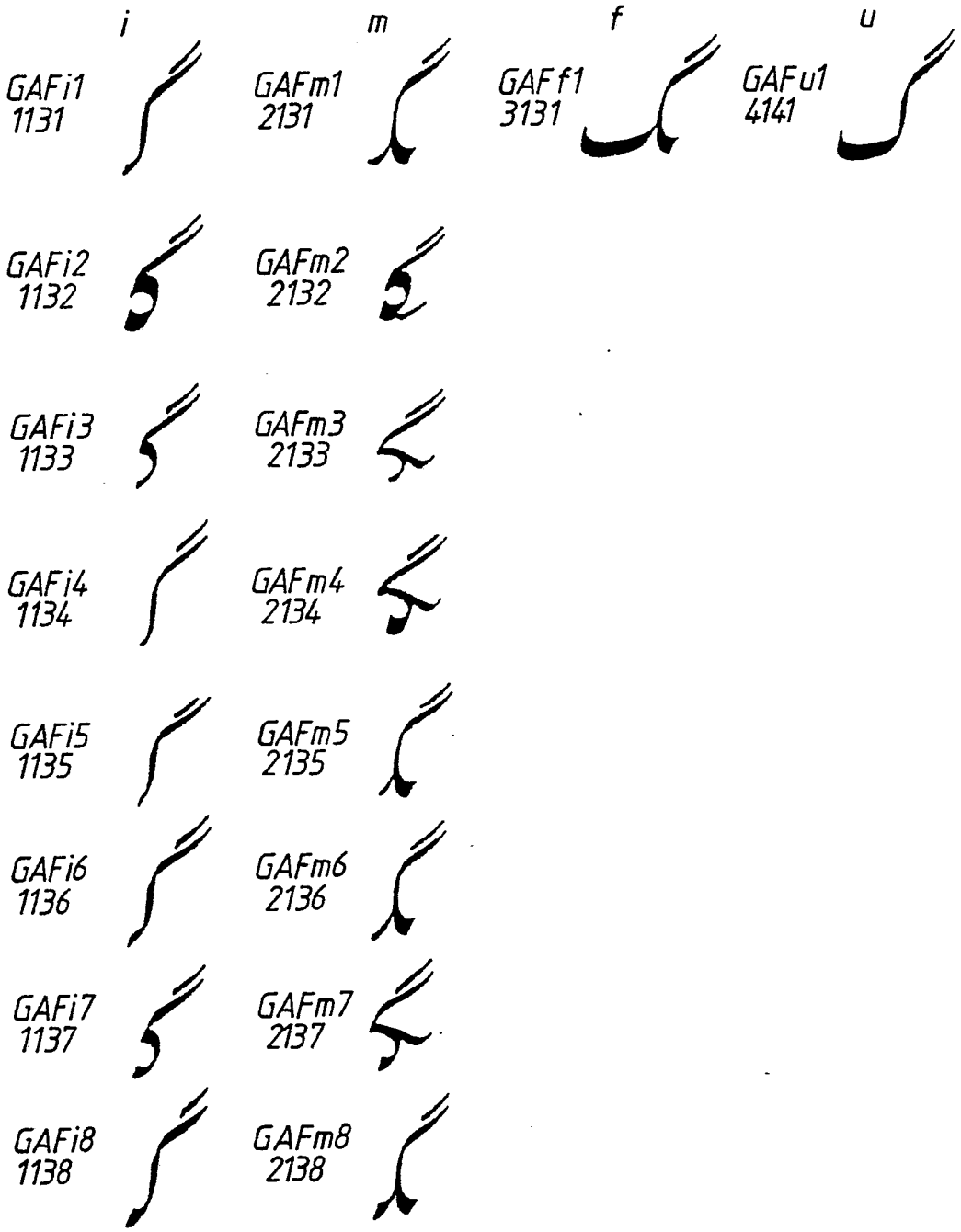

i
m
f
u
GAFi1 1131
GAFm1 2131
GAFf1 3131
GAFu1 4141
GAFi2 1132
GAFm2 2132
GAFi3 1133
GAFm3 2133
GAFi4 1134
GAFm4 2134
GAFi5 1135
GAFm5 2135
GAFi6 1136
GAFm6 2136
GAFi7 1137
GAFm7 2137
GAFi8 1138
GAFm8 2138

FIG.13.

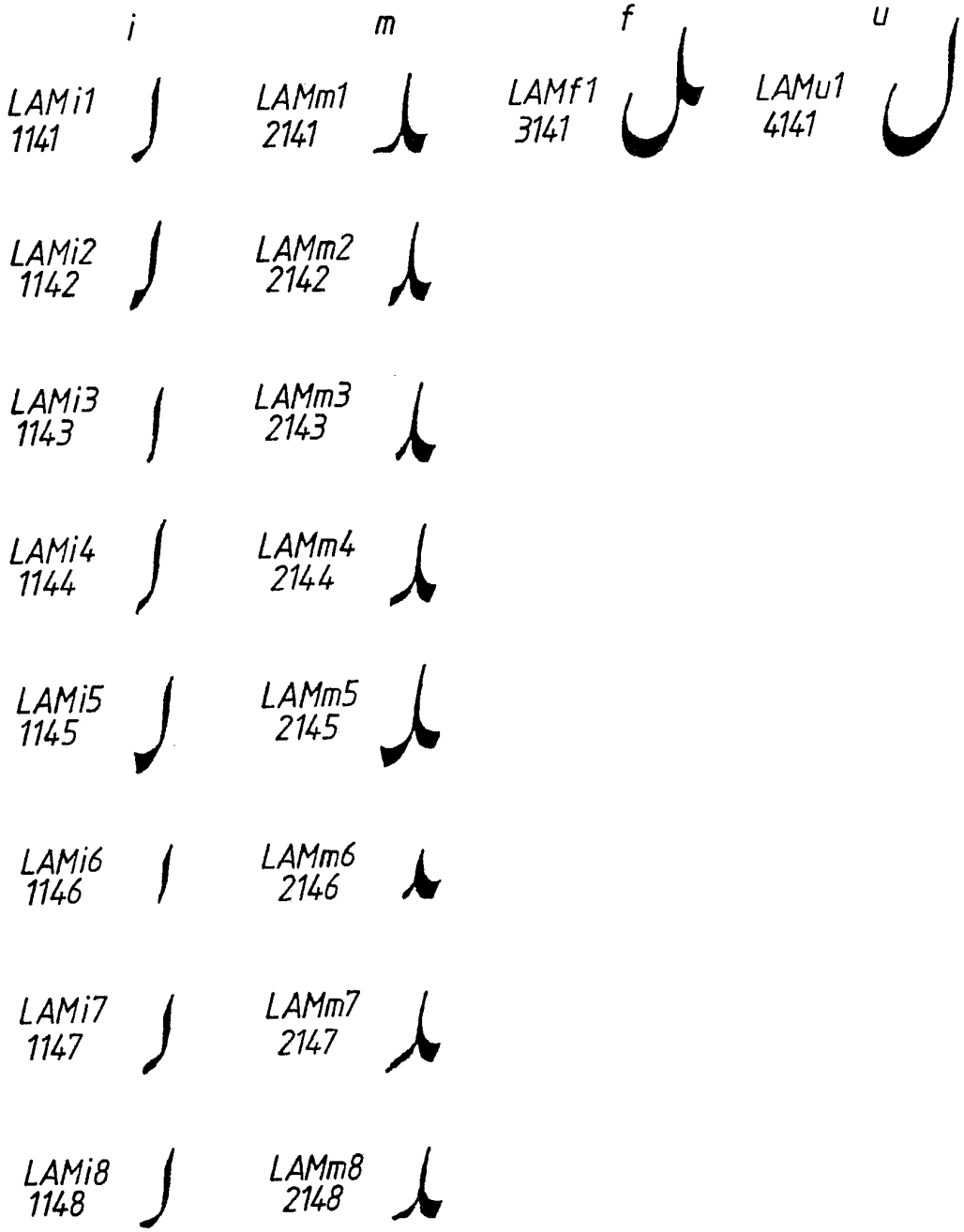

i
m
f
u
LAMi1
1141
LAMm1
2141
LAMf1
3141
LAMu1
4141
LAMi2
1142
LAMm2
2142
LAMi3
1143
LAMm3
2143
LAMi4
1144
LAMm4
2144
LAMi5
1145
LAMm5
2145
LAMi6
1146
LAMm6
2146
LAMi7
1147
LAMm7
2147
LAMi8
1148
LAMm8
2148

FIG.14.

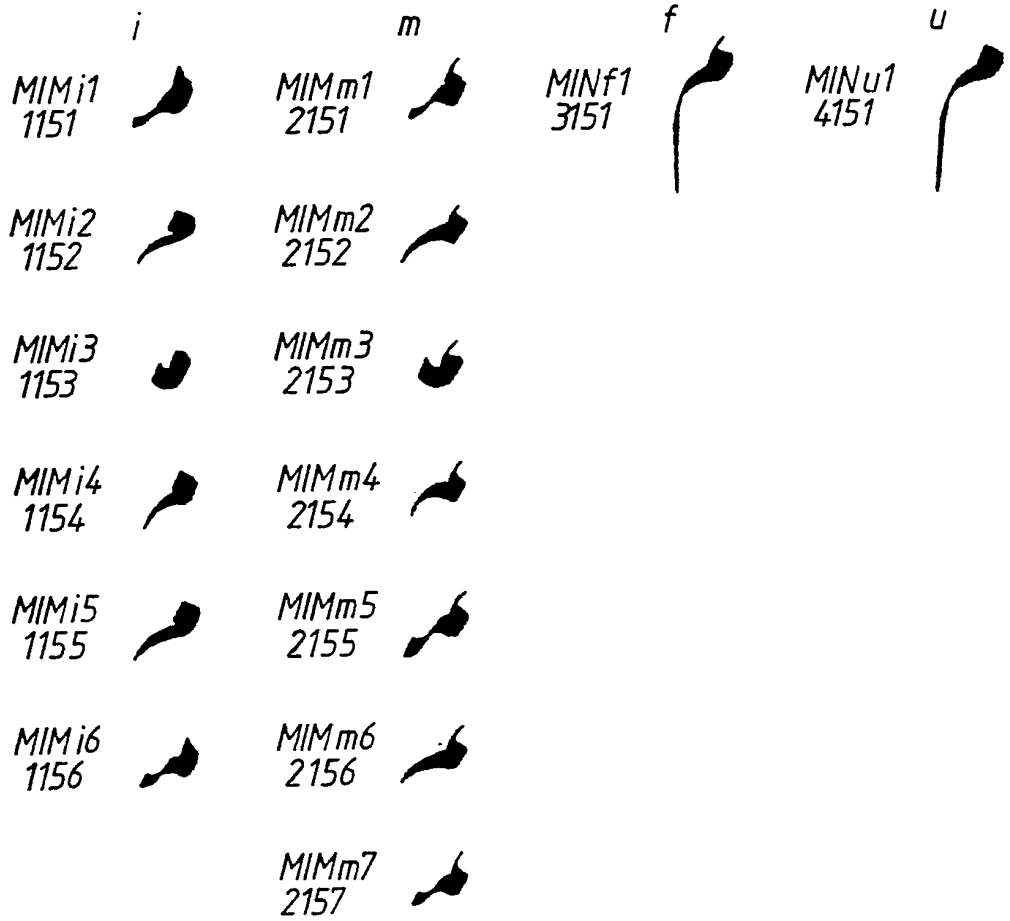

i
m
f
u
MIMi1
1151
MIMm1
2151
MINf1
3151
MINu1
4151
MIMi2
1152
MIMm2
2152
MIMi3
1153
MIMm3
2153
MIMi4
1154
MIMm4
2154
MIMi5
1155
MIMm5
2155
MIMi6
1156
MIMm6
2156
MIMm7
2157

FIG.15.

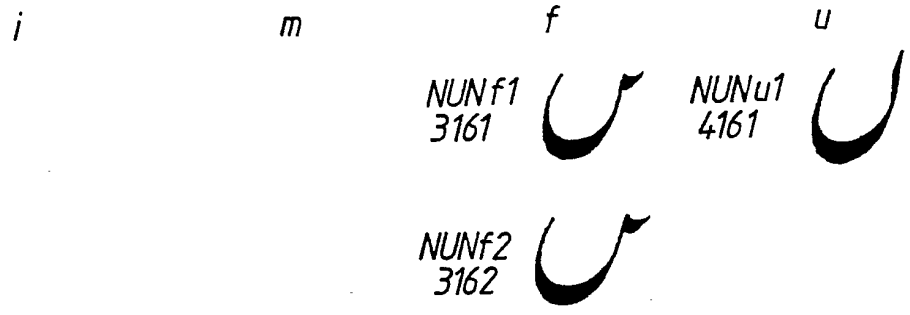

i
m
f
u
NUNf1
3161
NUNu1
4161
NUNf2
3162

FIG.16.

i	m	f	u
		VAOf1 3171 	VAOu1 4171 
		VAOf2 3172 	
		VAOf3 3173 	

FIG.17.

i	m	f	u
HAYCi1 1181 	HAYCm1 2181 	HAYCf1 3181 	HAYCu1 4181 
HAYCi2 1182 	HAYCm2 2182 	HAYCf2 3182 	
HAYCi3 1183 	HAYCm3 2183 	HAYCf3 3183 	
HAYCi4 1184 	HAYCm4 2184 	HAYCf4 3184 	
HAYCi5 1185 	HAYCm5 2185 		
HAYCi6 1186 	HAYCm6 2186 		
HAYCi7 1187 	HAYCm7 2187 		

FIG.18.

i	m	f	u
HAYAi1 1191	HAYAm1 2191	HAYAf1 3191	HAYAu1 4191
HAYAi2 1192	HAYAm2 2192		
HAYAi3 1193	HAYAm3 2193		
HAYAi4 1194	HAYAm4 2194		
HAYAi5 1195	HAYAm5 2195		

FIG.19.

i	m	f	u
			HAMZAu1 4201

FIG 20

i	m	f	u
		CHHOTI YEf1 3211	CHHOTI YEu1 4211
		CHHOTI YEf2 3212	

FIG.21.

i	m	f	u
		BARI YEf1 3221	BARI YEu1 4221

FIG.22.

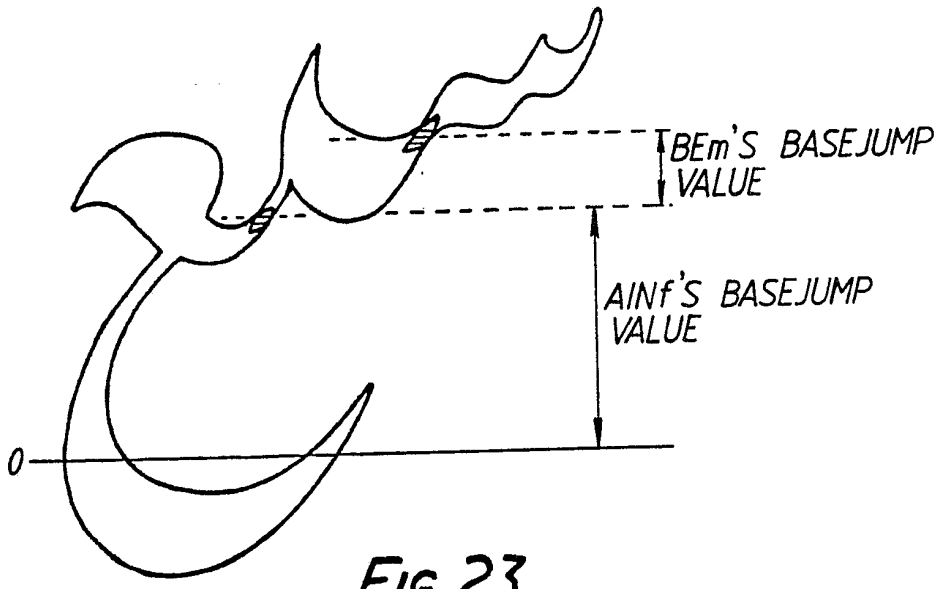


FIG.23.

b) DOTS

SUPERSCRIPTS: 

1 2 3 4 5

SUBSCRIPTS: 

FIG.24.

DIACRITICALS

SUPERSCRIPT:  etc

SUBSCRIPT:  etc---

FIG.25.

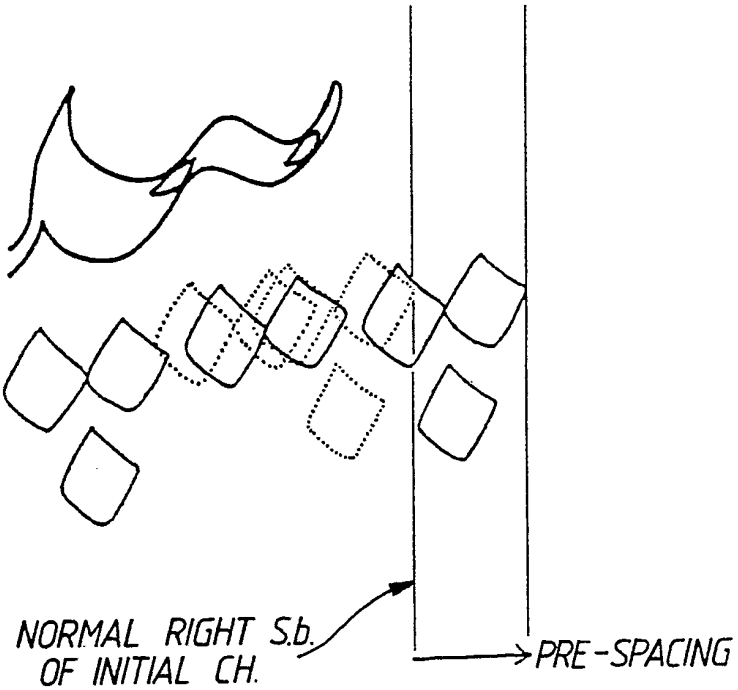


FIG. 26.

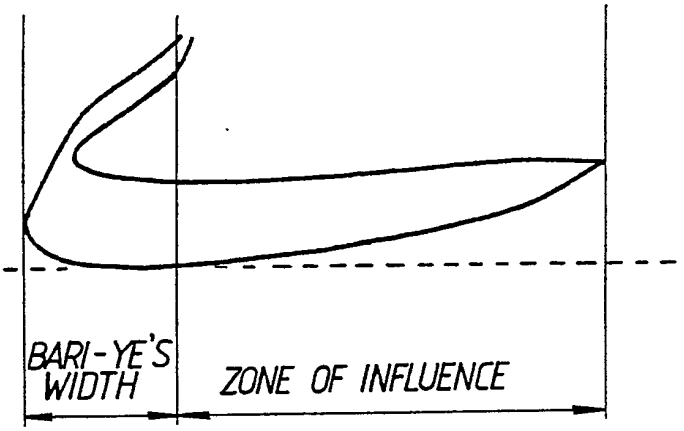


FIG. 27A.

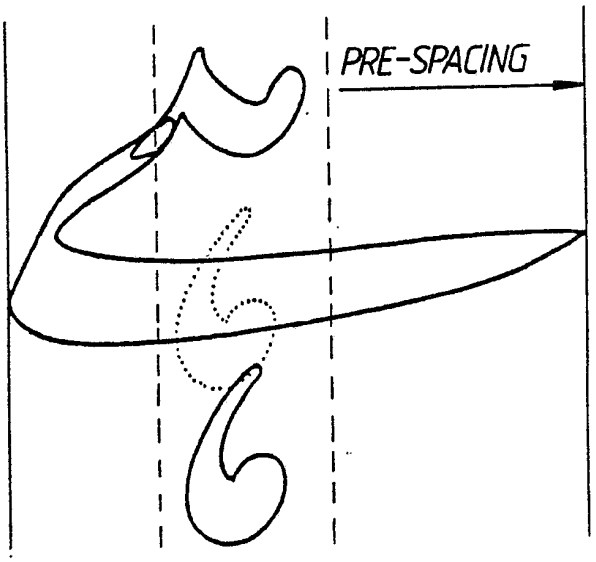


FIG.27B.

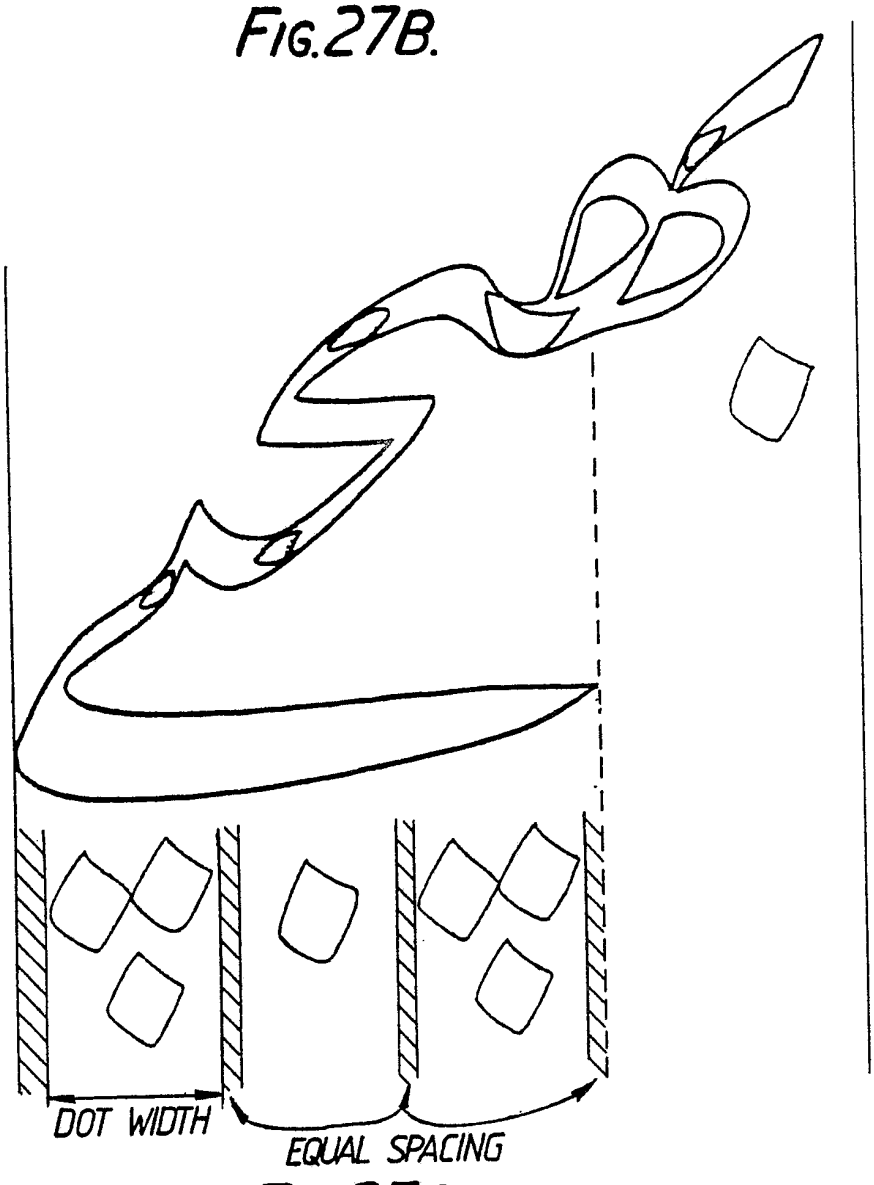


FIG.27C.

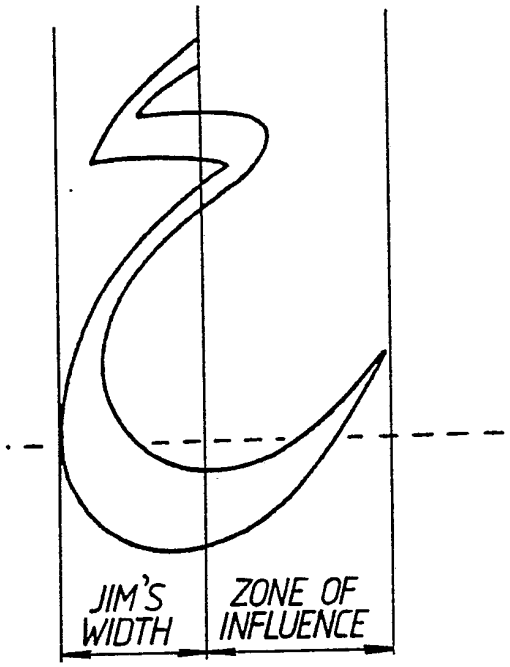


FIG. 28A.

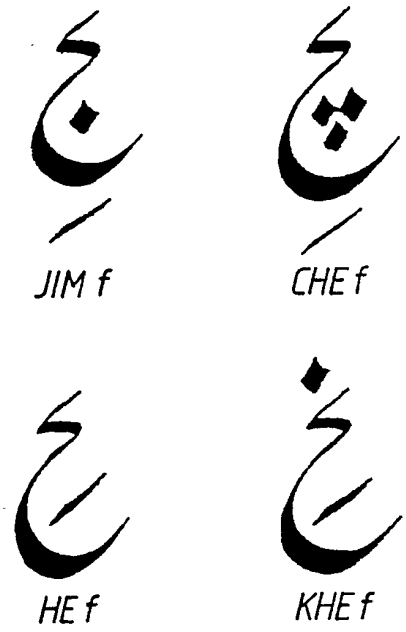


FIG. 28B.

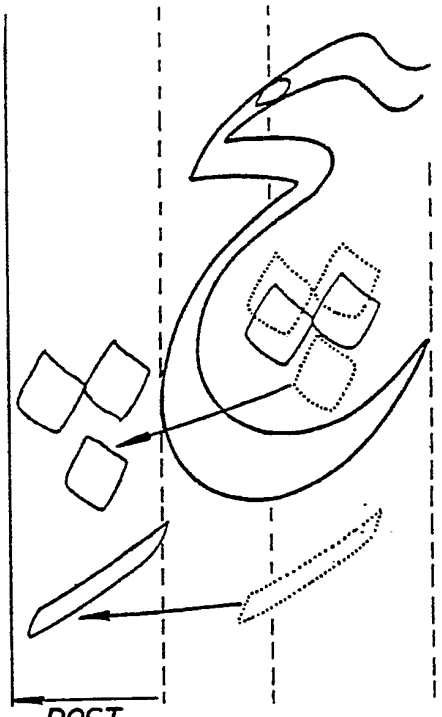


FIG. 28C.

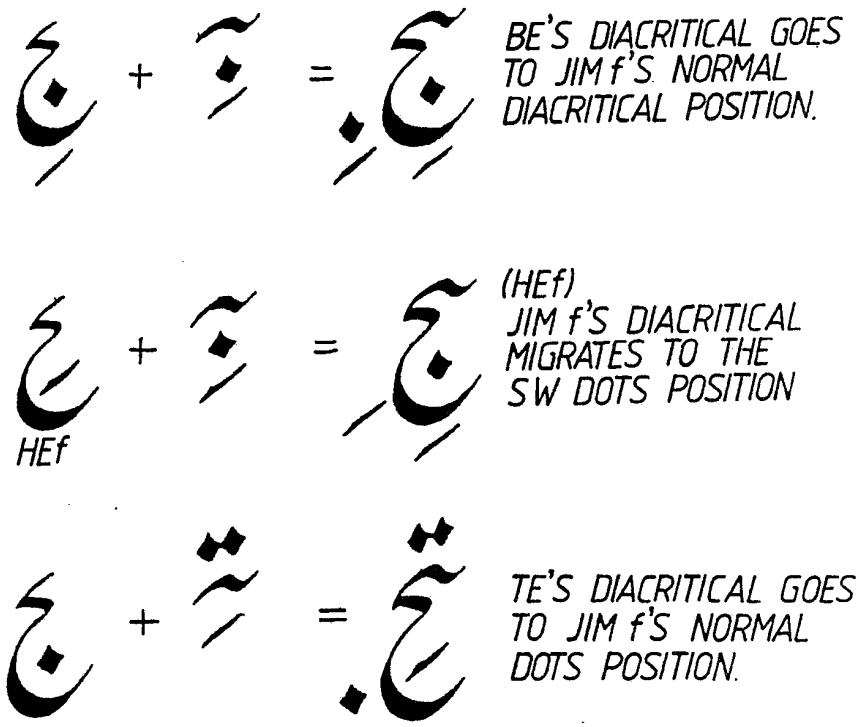


FIG. 28D.

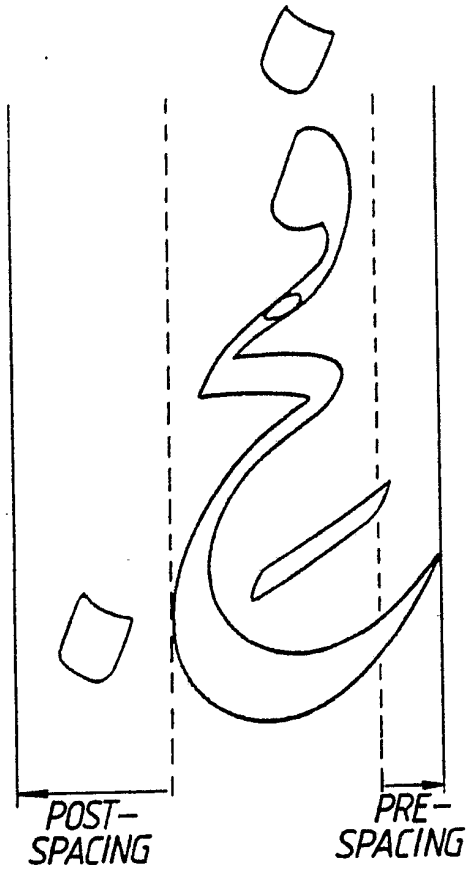


FIG. 28E.

FIG. 28F.

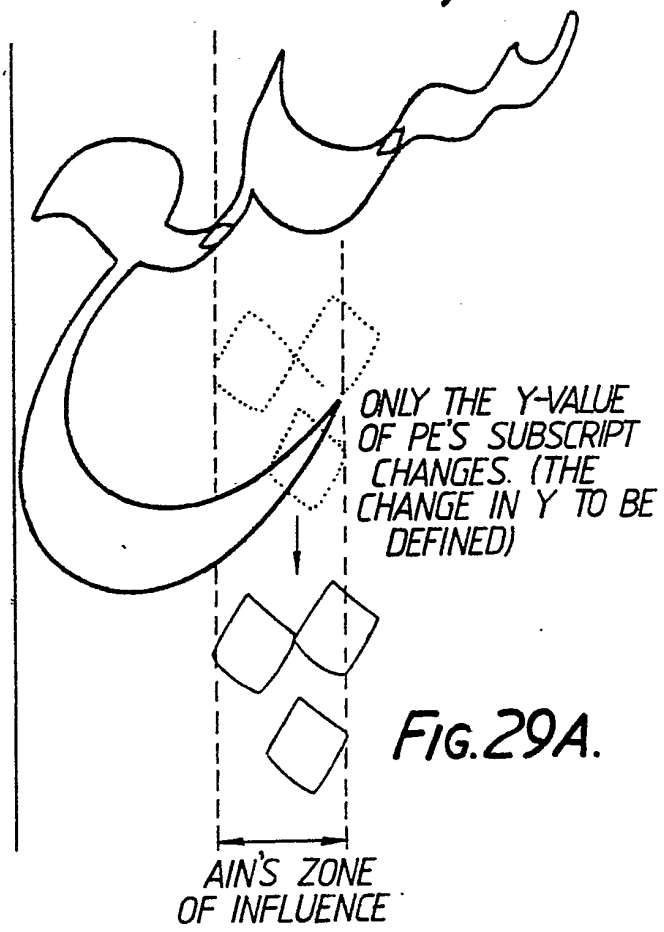


FIG. 29A.

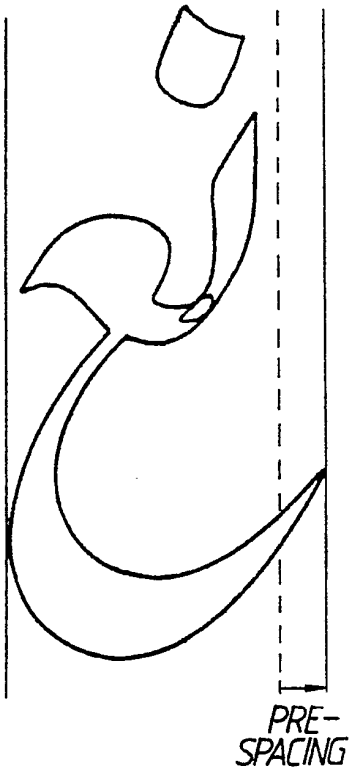
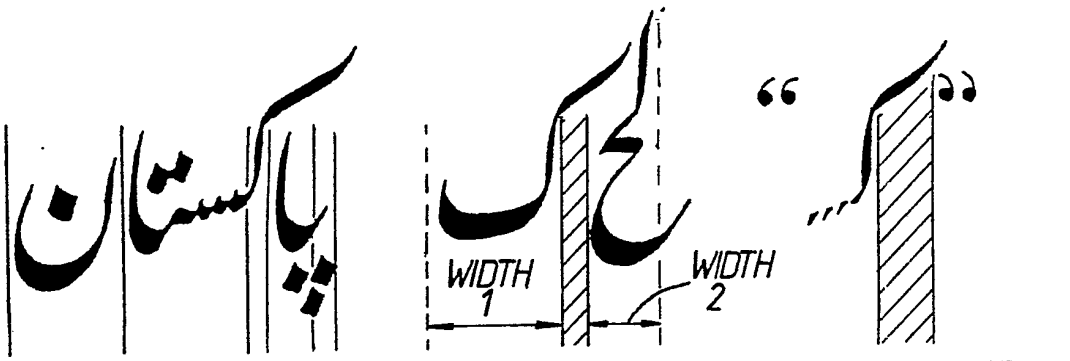


FIG. 29B.



THIS MAY BE 'NORMAL' SPACING (KAF KERNS)

SPACE ADDED BETWEEN KAFu & JIM f

SPACE ADDED BETWEEN KAFi & QUOTES

FIG. 30.

2208556

Improvements relating to Printing

This invention relates to printing and is concerned with the printing of non-Roman scripts, particularly those in Nastaliq, which is the written form of Urdu. The complexity of this script has generally meant having it done by calligraphers, which is highly labour intensive and slow. For the rapid dissemination of information, it is highly desirable that some means for typesetting it should be devised, making it legible to those familiar with the calligraphic script.

The problem is particularly acute with Nastaliq, which has very complicated rules on how a word is written. Calligraphy demands that each letter should flow into the next one without a noticeable break. For this reason, each letter may have various different forms depending on the shape and thickness of the other letter or letters to which it is to be joined.

Another difficulty is that as the writing proceeds from right to left, each word descends. Thus, not only do letters have to be placed in the right relationship horizontally, but also vertically. Depending on their position in the word, they must be at a certain height in relation to a notional base line.

A further characteristic is that letters can take different forms depending on the other letters to which they are joined. To take an English example, in the words BAT, BET, BIT, BOTH and BUT, the B is followed by a different letter in each. If the Nastaliq rules were observed, the B might be written in five different ways, each determined by the following letter. This can apply throughout the word.

Other features of Nastaliq are dots and diacriticals. Dots are put above or below the letters and their position has to be carefully related to them. If a dot is to be placed above a letter, for example, then it must be adjusted in accordance with the height of that letter above the notional base line. Furthermore, its position can be affected by the letters on either side which might have an overlapping portion, dot or diacritical which would interfere with that dot in its normal position. It therefore has to be shifted to a different position, depending on the make-up of the adjacent letters. The same problems apply to the diacriticals.

One approach to typesetting such a complex script has been simply to have a font comprising all the complete words in the language, comprising 30 - 40 thousand items. That is a heavy-handed method, without any flexibility. It is the aim of this invention to

provide a system in which the letters can be stored in their individual forms and combined together calligraphically with any dots and diacriticals in the correct relationship. It has been found that the letters can be broken down into a manageable number of different shapes or characters and their combination can be governed by rules of calligraphy programmed into a computer.

According to the present invention there is provided printing apparatus including a keyboard or other input device, a font store with a plurality of digital coded representations of shapes of characters of an alphabet, program storing means storing encoded logic and data related to character combinations, means for storing sequences of character signals resulting from operation of the input device, means subject to the program storing means operable to scan such sequences to select from the font store digitized representations of character shapes each positioned and appropriate to conjoin with an adjacent shape, and to output the resultant encoded selections of conjoined characters.

There may be, for at least some characters, two or more possible shapes determinable according to their position in a conjoined group. But the input device will have a single selector element for each such character to produce the associated signal for the

sequence storing means.

At least some character shapes may terminate at opposite ends at different levels, and the digital coded representations will then include parameters indicative of the level difference. Likewise, at least some character widths may differ from those of others, and the digital coded representations will include parameters indicative of width.

The scanning and selecting means may be operable, in a normal routine, to select each character shape by reference to the immediately preceding character shape but for certain combinations of characters there may be sub-routines whereby the normal routine is modified. By 'preceding' is meant immediately before in the selection process: a word may be built up from its end in which case 'preceding' does not apply to the character actually read before the one in question.

In such a sub-routine, one character may influence the shape of at least one preceding one, causing consequential re-selection of other character shapes if necessary to maintain the appropriate conjunctions.

In another such sub-routine, if a certain character shape when conjoined would add a level difference bringing the total level difference of a group of character shapes selected previous to that point above a predetermined limit, an alternative character shape is

selected, causing consequential re-selection of the character shapes if necessary to maintain the appropriate conjunctions.

The font store may also have a plurality of digital coded representations of symbols associable with, but not physically linked to, said characters and augmenting the alphabet, the program storing means storing encoded logic and data related to the combination of these symbols with the characters, the character sequence storing means being adapted to store symbol signals resulting from operation of the input device, and the scanning and selecting means being arranged to include in its output encoded representations of these symbols, each positioned in the appropriate relationship to the conjoined characters.

A set of these symbols may be associable with one group of characters, each character of that group forming the basis of sub-groups whose members have different symbols associated therewith and may include the character alone. In this case, the input device may have a single selector element for each said member, and the sequence storing means may be arranged to record each said member separately as a character signal and a symbol signal.

Another set of these symbols may be independently selectable on the input device, each selection causing

the symbol to be associated with the adjacent character selected.

When, according to the normal association of each symbol with its character, there would be interference between symbols and/or between symbols and adjacent characters, the scanning and selecting means will output encoded representations of the symbols shifted to non-interfering positions.

For a better understanding of the invention, one embodiment will now be described, by way of example, with reference to the accompanying drawings, in which:

Figures 1 to 22 show in table form the basic letters of the Nastaliq alphabet in their various different shapes or characters,

Figure 23 shows a combination of letters and the vertical shift between them arising from the inclined script,

Figure 24 illustrates the dots which may be applied to the basic letters,

Figure 25 illustrates some of the diacriticals which may be applied to the basic letters,

Figure 26 shows a dot adjustment feature when there is mutual interference.

Figure 27 shows the letter BARI YE in its final form alone and in combination, to explain associated dot displacement,

Figure 28 shows the letter JIM in its final form, alone and in combination, to explain associated dot and diacritical displacement,

Figure 29 shows the letter AIN in its final form in combination with others, and

Figure 30 illustrates the spacing of certain letters.

The keyboard or other input device and the various computer components whose functions are outlined above will not be described in detail since their arrangement and programming will be evident from the following.

The basic letters of the Nastaliq alphabet can be listed as follows:

1. ALIF
2. BE
3. JIM
4. DAL
5. RE
6. SIN
7. SAD
8. TOE
9. AIN
10. FE
11. QAF
12. KAF
13. GAF
14. LAM
15. MIM
16. NUN
17. VAO
18. HAYC
19. HAYA
20. HAMZA
21. CHHOTI YE
22. BARI YE

Furthermore, there are a number of extra letters

which take the form of some of these basic letters but with an addition or modification of the dots (to be described later) applied to them. These will be referred to as derivative letters. An example is the basic letter BE which has derivatives including PE, TE, SE and YE.

As explained previously, each letter may have a different shape or character depending on its position in a word, namely whether it is the initial letter (i), a medial letter (m), a final letter (f), or whether it is unconnected (u), standing on its own. There may further be various different forms of a letter in the initial, medial and final positions, although there is only one unconnected form of each. Figures 1 to 22 illustrate the alphabet with all the various different characters, and in each left hand column they are identified by the letter itself, followed by i, m, f or u indicating its position, and finally by a number indicating the actual shape for that group. For the purpose of digitizing, this order is changed, in this example, each character being identified by a four figure number with a first number corresponding to the position (i=1, m=2, f=3 and u=4), the next two numbers indicating the position of the letter in the alphabet, and the final number identifying the shape as before. As there are only 242 shapes or characters, this could

be done with less digits, simply numbering them 1-242, but the coding used here makes relating characters to letters much easier.

When there is a derivative, an extra digit could identify which one, the first four being determined as for the basic.

Thus the number 2077, for example, is an unambiguous identification of SAD when in a medial position and taking its seventh shape or character. 20262 indicates the second letter in the BE group (counting BE as the first) and represents medial PE taking the sixth shape or character. Each such number can readily be stored and handled by a computer, and ultimately used to retrieve the appropriate character from the font when required.

Referring back to Figures 1 to 22, it will be noted that some letters apparently have no initial or medial forms. This does not necessarily mean that they cannot exist in that part of a word. When they are required there, they take the form of a derivative of one of the basic letters. For example, NUN in its medial form has a single dot over the medial BE character.

In Nastaliq, a word is not necessarily a single continuous string of characters; there can be a break, so that a word is in fact a combination of part-words. For convenience, references in this specification to

'words' will mean complete or part-words.

It will be assumed that a compositor will have a keyboard whose main group of keys are those corresponding to the twenty-two basic letters of the alphabet, and all their derivatives. It is of course impractical to put every version of a particular letter on the associated key, and so a single representative character is used to identify it, conveniently the unconnected version. Thus the compositor will simply type words using the basic syllabary. The computer will translate the key actions into the correct selection and organisation of the various characters.

First, since there is mutual interdependence of the characters, there will normally be no selection of characters until a word is completed, this being signified by actuation of a spacing bar or key. Thus there will be temporary storage of the keyed inputs, and when the word is complete it is composed backwards from the final character.

Many letters have only one final form, and so that selection is made instantly. However, others have a number of alternative shapes, and which one is used has to be determined from the immediately preceding medial, or initial if it is a two letter word. This is ascertained from a stored table listing all the acceptable combinations from the font of medial-final,

initial-final, medial-medial and initial-medial characters. This will be referred to as the generic selection table (GST). For the most part, there will be only one acceptable pairing in each case, and so working back from the final character, all the other characters will be uniquely determined. However, there are exceptions, and these will be embodied in further tables and routines, some of which are described below.

Account also has to be taken of the script having an incline down from right to left. Referring to Figure 23, to position characters correctly, a base line labelled O is taken, and when a word is being assembled the height of each character in relation to that line is matched to that of each adjacent character so that there are smooth transitions throughout. There will preferably be small overlaps, as indicated by the shaded areas, to disguise any minor inaccuracies. The incline occurs since practically every character starts at the right hand end at a higher level than that at which it terminates at the left hand end. This difference will be referred to as the base jump value and will normally be positive, although there could be small negative values. In any event it will be stored as an essential parameter of each character, and as a word is assembled from the left hand end, each character will be shifted upwards by the base jump of the preceding character.

Initial characters will not need base jump values, since nothing is joined on to their right hand ends and they take their vertical positions from the final and medial characters. The final character of each word will of course be located on the base line O, and the height h of any character junction will be the accumulated sum of the base jumps to that point.

Each character will also have a given width and this too will be an essential parameter to store and take account of when a word is being made up, in order to preserve continuity.

As mentioned above, there are exceptions to the main routine of composition using the GST for which extra routines have to be provided. Representative examples are:

1) BE-SIN
2) BE-BE
3) KAF/GAF m3, i3
4) KAF/GAF m4
5) MIM m3

For example, for the first pair BE-SIN there is a special rule when there are the following finals preceded by medial BE and when the latter is preceded by medial or initial SIN:

BE (f1, f2 only)

SAD

TOE

AIN

FE

QAF

VAO

Here the medial BE will be BE m4, while SIN will be determined by the normal process from BEm4 in the GST.

Similarly, BEm4 is selected when the following medials are preceded by medial BE and the latter is preceded by medial or initial SIN:

SAD

TOE

AIN

FE

KAF (m2 only)

GAF (m2 only)

The BE-BE combination follows the GST when there are just two BEs, BEf1 being preceded by BEm4 and BEf2 by BEm1. If it is a word of just two BEs then it is BEf1, BEi2. However, when two or more are strung together preceding a final BE, variations occur depending on whether there is an odd or even number of medial BEs. With an even number, or where there is a single medial BE preceded by an initial BE, the medial BE preceding

the final BE is Bm4. This means that the final BE is BEf1, from the GST, while the preceding BEs follow a selection pattern alternating BEm2 and BEm4 towards the initial letter. In the other case, except for the combination BE-BE-SIN (see above), the medial BE preceding the final BE is Bm1, and this dictates that the final character is BEf2. Preceding BEm1 is BEm4 alternating with BEm2 as before.

The third exception relates to KAFm3 and GAFm3, and KAFi3 and GAFi3, being shortened alternatives of KAFm1 and GAFm1, and KAFi1 and GAFi1, respectively. Should any of those taller versions normally have been selected according to the GST, then if the total base jump height h to that point has reached a critical value, a substitution is made, bringing in the shorter character. This could also affect the next preceding characters, which may have to be re-selected according to the GST.

In a similar way KAFm4 and GAFm4 are alternatives for KAFm2 and GAFm2 and the latter are replaced by the former if the total base jump height h to that point is above a set value, but only if the character is the first medial following the initial. This substitution will also require a check on the GST to ensure the correct initial character is used.

The fifth exception concerns the effect of the following initials on medial MIMm1:

BE

JIM

SIN

SAD

TOE

AIN

FE

MIM

HAYA

If this relationship occurs, MIMm3 is substituted for MIMm1, and this in turn may require a re-adjustment of the initial.

It should be emphasised that these rules have been formulated on the basis of a particular style of calligraphy and its dissection into the characters shown in Figures 1 to 22. It is possible that another analysis of a slightly different, but equally intelligible, script would produce a different breakdown of characters, and modifications to the exceptions. However, the composition of a word will generally require the following procedures, all carried out by a computer programmed accordingly.

1) To wait for a complete word to be entered from the keyboard,

2) To select characters directly if there are no

alternatives or to make reference to the GST if there is ambiguity,

3) If the GST resolves the ambiguity, to cause the characters so identified to be selected,

4) If there are any combinations requiring special rules to be applied, to recognise them and run routines to resolve which characters should be selected, and to select them accordingly.

5) To reckon up the collective base jumps and make substitutions in the special cases if required,

6) To assign a height to each finally selected character in accordance with the base jumps, and to assign horizontal positions in accordance with the character widths.

The basic word is then complete, at least within the computer. However, there are further symbols to be considered, in the form of dots and diacriticals, which will now be discussed.

There are nine dots, five of which are placed above various characters, and four of which are placed below, and they take the forms shown in Figure 24. In fact, a dot as referred to herein may be two or three dots in a group, or a configuration not normally regarded as a dot, but it is convenient to class them all as dots. Some of the basic letters have a dot as an essential part, and their derivatives remove or substitute a

different dot while retaining the character or shape of the basic. Other derivatives add a dot to an undotted basic. For example, BE and its derivatives always require dots while JIM and its derivatives have them for some but not for all in the group. A third kind, such as LAM, never take them. There can only be one dot per character, above or below, and they may thus be identifiable as positive or negative.

As mentioned above, the keyboard has keys for all the basic letters and their derivatives, and so there are no separate dot keys. However, the dots will be stored as information separate from that on the characters or shapes of Figures 1 to 22. When a key for a dotted letter is operated, the character and dot information is extracted and processed separately, the character being determined as described above and the dot being held in reserve to be finally positioned as described below once the string of characters is settled.

There are considerably more diacriticals, about twenty-five, and just a small selection is illustrated in Figure 25. Again, they may be above or below the characters. They are equivalent to vowels, and each has a corresponding key on the keyboard, to be operated immediately before the key for the letter with which that diacritical is to be associated.

All letter shapes can take diacriticals, although BARI YE which serves only as a long vowel does not normally have them on its own account. However, a diacritical from a neighbouring character may be floated below it, as described later. There can be two placements per character, one above and one below.

A problem with these extra symbols is that their lateral extent may be greater than that of the character with which they are associated, and so if the adjacent characters also have dots and diacriticals, there could be interference. The solution to this is to shift them laterally, leaving the one or ones furthest to the left in their initially assigned positions.

To carry this out, each character (and here is meant a character, A say, with subscript dots or diacriticals) has a "knock-on" value, which represents the shift to the right necessary to be applied to any subscript of the next preceding character B in order to maintain mutual clearance. This is made cumulative so that, if that next character B also has a knock-on value, the two values are added and applied to any subscript dots or diacriticals of a third character C next to the right of B. This can continue throughout the word. If a character with a subscript has no knock-on value (the dot or diacritical being sufficiently compact in relation to the main portion of the character) this

simply carries over any existing total knock-on value to the next character to the right, that subscript being "shunted" accordingly. However, if a character has no subscript, the knock-on effect will be lost, since subscripts from characters to the left have room to be set out without encroaching on the character next to the right.

There is a further problem should there be a knock-on value subsisting at the initial character. There is no further character for the dot or diacritical to float under. However, this is solved, once the problem is signalled, by providing an imaginary character, referred to as the pre-space, and shifting the homeless subscript into it at a height that it would occupy if there really was a character there. This is illustrated in Figure 26, where the original dot positions, before the knock-on is applied, are shown in broken lines.

There are exceptions to this dot and diacritical placement, involving the following letters:

BARI YE f

JIM f

AIN f

These three letters in their final form (there is only one version of each) all have a substantial tail sweeping under and to the right, finishing at a point

substantially upstream, in the direction opposite to that of writing, in relation to the junction with the immediately preceding medial, or initial.

Referring to Figure 27A, this shows BARI YE and its total lateral extent is composed of its basic width on the left (from its starting point to its left hand extremity) and a zone of influence, defined by the length of the tail to the right of that starting point.

A preceding medial or initial may have a subscript with which this tail would interfere, and so there are rules dictating where such subscripts should be shifted to if this is the case.

If any preceding medial or initial characters have subscripts falling within the zone of influence, entirely or by more than a substantial proportion of their width (say 50%, although it may differ from that), then those subscripts are shifted underneath the BARI YE character, being set below the tail by the same distance that they would otherwise be below their own character. If there is only one such subscript to have this shift, then it stays vertically below its own character, as illustrated in Figure 27B.

This Figure also illustrates the rule that, with a short word where the zone of influence extends to the right of the initial character, then a pre-space must be provided.

However, when two or more subscripts have to be displaced to below the BARI YE tail, they are distributed evenly over the whole length (the basic width plus the zone of influence). This is shown in Figure 27C. This overrides any would-be knock-on shift, and the knock-on values are cancelled.

JIMf also has a basic width and zone of influence, as shown in Figure 28A, and its group of normal variations with dots and diacriticals is shown in Figure 28B.

The subscripts of a preceding character could compete with the dot or diacritical within the large bight of JIMf if it was more than 50% within the zone of influence. This is resolved by that dot or diacritical giving way and being replaced by that of the preceding character. The displaced one shifts in line with the slope of the word to a position outside the bight, into a 'post-space', analagous to the pre-space. At the same time, if there is a diacritical below the tail, as in CHEf, then that is moved to the left in sympathy, remaining vertically below the shifted dot. However, it moves horizontally, rather than with the word slope, since without the intervening tail there would otherwise be an excessive gap. This is illustrated in Figure 28C, and further combinations following these rules are shown in Figure 28D.

It is possible that both a pre-space and a post-space will be required as illustrated in Figure 28E, which shows a combination of FEi and JIMf.

If the character preceding JIMf has a knock-on value, then that value is not cancelled. The subscript concerned will cause the shift to the post-space, as described, of the dot and/or diacritical of JIMf, and at the same time shift the subscript of the next preceding letter to the right. This is illustrated in Figure 28F, where YE has a knock-on value.

The third character with a tail, AINf, obeys the basic rule for BARI YE, namely that if any subscripted character comes entirely, or by more than 50%, within the zone of influence, then the subscript is displaced vertically to below the tail, as shown in Figure 29A. Also, similarly to JIMf, when combined with a narrow initial character, it will require a pre-space as shown in Figure 29B.

Similar techniques will apply to the adjustment of superscripts where there is potential conflict. In particular, there will be knock-on values, zones of influence (around KAF and GAF) and pre-spacing. Post-spacing is unlikely to be required.

There are also the unconnected characters which nevertheless require to be spaced a proper distance from adjacent characters, and these include, apart from the

letter characters of Figures 1 to 22, signs such as brackets and punctuation marks. Also, of course, there is the spacing at each end of a set of connected characters

The rules governing this will be held in a further table, and this will be subdivided into the rules determining the spacing between:

1) Any pair of unconnected characters,

2) Any unconnected character and a final character of a preceding combination of characters,

3) Any unconnected character and an initial character of a succeeding combination of characters, and

4) Any final character and any initial character of two discrete combinations of characters.

Most spacings will be normal, corresponding to a standard gap, and so the table need only list exceptions in terms of what has to be added to or subtracted from the normal spacing. Figure 30 illustrates various character combinations in some of which the spacing has to be adjusted.

This spacing is not the same as word spacing: a word may be more than one set of connected characters in succession, or the combination of an unattached character before or after a connected set, and these are known as compound words. The keyboard operator would normally signal a break in such a word by pressing a

key separate from the word spacing key or bar.

Word spacing may be selected through a wide range, and the computer may be programmed to justify, if required. If the word spacing is chosen to be small, then the spacing within compound words will be also, reducing to zero at the extreme. However, as the word spacing is increased, although the compound word spacing is increased as well, it is not done so at a proportional rate, but rather less than that. It is a matter of choice whether a compound word is broken at the end of a line.

Whenever a backward tail or kern, such as in KAF or GAF, begins a line, or wherever a forward tail or kern ends a link, as in RE, then pre- or post-spacing may be required.

CLAIMS

1. Printing apparatus including a keyboard or other input device, a font store with a plurality of digital coded representations of shapes of characters of an alphabet, program storing means storing encoded logic and data related to character combinations, means for storing sequences of character signals resulting from operation of the input device, means subject to the program storing means operable to scan such sequences to select from the font store digitized representations of character shapes each positioned and appropriate to conjoin with an adjacent shape, and to output the resultant encoded selections of conjoined characters.

2. Apparatus as claimed in Claim 1, wherein there are, for at least some characters, two or more possible shapes determinable according to their position in a conjoined group, the input device having a single selector element for each such character to produce the associated signal for the sequence storing means.

3. Apparatus as claimed in Claim 1 or 2, wherein at least some character shapes terminate at opposite ends at different levels, and the digital coded representations include parameters indicative of the level difference.

4. Apparatus as claimed in Claim 1, 2 or 3, wherein at least some character widths differ from those of others, and the digital coded representations include parameters indicative of width.

5. Apparatus as claimed in any preceding claim, wherein the scanning and selecting means are operable, in a normal routine, to select each character shape by reference to the immediately preceding character shape, but wherein for certain combinations of characters there are sub-routines whereby the normal routine is modified.

6. Apparatus as claimed in Claim 5, wherein in such a sub-routine one character influences the selection of the shape of at least one preceding one, causing consequential re-selection of other character shapes if necessary to maintain the appropriate conjunctions.

7. Apparatus as claimed in Claim 5 or 6, as appendent to Claim 3, wherein in such a sub-routine, if a certain character shape when conjoined would add a level difference bringing the total level difference of a group of character shapes selected previous to that point above a pre-determined limit, an alternative character shape is selected, causing consequential re-selection of other character shapes if necessary to maintain the appropriate conjunctions.

8. Apparatus as claimed in any preceding claim, wherein the font store also has a plurality of digital

coded representations of symbols associable with, but not physically linked to, said characters and augmenting the alphabet, the program storing means storing encoded logic and data related to the combination of these symbols with the characters, the character sequence storing means being adapted to store symbol signals resulting from operation of the input device, and the scanning and selecting means being arranged to include in its output encoded representations of these symbols each positioned in the appropriate relationship to the conjoined characters.

9. Apparatus as claimed in Claim 8, wherein a set of said symbols are associable with one group of characters, each character of that group forming the basis of a sub-group whose members have different symbols associated therewith and may include the character alone, wherein the input device has a single selector element for each said member, and wherein the sequence storing means is arranged to record each said member separately as a character signal and a symbol signal.

10. Apparatus as claimed in Claim 8 or 9, wherein a set of said symbols is independently selectable on the input device, each selection causing the symbol to be associated with the adjacent character selected.

11. Apparatus as claimed in Claim 8, 9 or 10,

wherein when, according to the normal association of each symbol with its character, there would be interference between symbols and/or between symbols and adjacent characters, the scanning and selecting means output encoded representations of the symbols shifted to non-interfering positions.

12. Printing apparatus substantially as hereinbefore described with reference to the accompanying drawings.

Published 1988 at The Patent Office, State House, 66 71 High Holborn, London WC1R 4TP. Further copies may be obtained from The Patent Office, Sales Branch, St Mary Cray, Orpington, Kent BR5 3RD. Printed by Multiplex techniques ltd, St Mary Cray, Kent. Con. 1/87.