

한글문자 모아쓰기 Display의 한방안 (A method for the Display of Hangeul in its traditional Combined Form)

安 秀 桔
(Ann, Sou Guil)

要 約

한글의 幾何學的 構造上의 複雜度로부터 한글 各 字素의 character generator用 diode matrix의 size를 推算하고 이 字素 matrix 3枚(初聲, 中聲, 終聲)의 OR動作으로서 모아쓰기 한글의 모든 組合이 可能함을 보였고 水平母音用初聲과 垂直母音用初聲의 字板 選擇과, 終聲이 있을 때와 없을때의 中聲길이의 變更 등이 比較的 簡單한 論理回路로서 switch될 수 있음을 보였다.

새로 15bit 內部 code를 提案하였는데 이는 86個의 sparse diode matrix로서 모든 組合을 다 카버하고 그러면서도 decoding機構의 規模를 줄여준다.

Abstract

The required minimum size of character diode matrix of Korean letters is estimated from the topological complexity of letter structure.

The OR combination of three letter boards (diode matrix) gives all possible Hangeul whole letters in proper traditional combined form with minimum required discernibility.

Two forms of first consonants (centre located ones for horizontal vowels and leftward displaced ones for vertical and composed vowels) are switched by only 1 bit of the vowel code. The vowel pattern length is modified by again the last four bits of the code.

A new 15bit sized inner code is proposed which permits considerably small sized decoding mechanism.

1. 序 論

한글에 關한 研究는 그동안 꾸준한 發達을 하여 왔고 言語學的인 見地에서뿐만 아니라 自然科學의 關心으로도 그 努力이 높히 評價될만 하다. 音聲學的인 見地에서의 研究⁽¹⁾⁽²⁾로부터 한글文字의 形態學的⁽³⁾⁽⁴⁾⁽⁵⁾⁽⁶⁾⁽⁷⁾인 研究들은 電算機에 對한 한글入力を 可能하게 할 것이며 line printer에 依한 한글出力은 한국電算機市場에 關聯이 있는 모든 會社에서 注力하고 있으며 各各의 會社가 여러가지方法을 開發하여 實用化의 段階에 들어선것으로 볼 수 있다.⁽⁹⁾⁽¹⁰⁾⁽¹¹⁾

Braun tube에 依한 display도 꾸준한 發達을 하여 왔고 역시 一部會社에서 商品化하고 있다.

通信理論의인 見地에서 極히 重要性을 띄고있는 Entropy에 關한 研究⁽⁸⁾를 high light로 하여 本格的인 段階에 이르렀으나, 한글에 關한 研究는 maker側の 努力以外的 面에서 關心이 줄어들고 있는 느낌이

다. 이는 지금 段階로서 研究가 實用化研究段階가 되어 費用이 越等히 많이 든다는 事實과 反對로 學問的인 關心에 依해서 研究를 振興시킬 機關이 없거나 있어도 研究費를 投入할만한 財源이 없는 탓으로 보인다. 제나라 國語의 諸般研究에 對해서 더욱히 日常生活에 便宜를 가지울 한글의 電算機入出力에 關해서 좀 더 積極的인 研究및마침이 要望된다.

2. 한글의 電算機 內部code

한글의 內部code는 政府나 其他 公共機關이 定해 준 바 없이 maker마다 獨自的인 code를 使用하고 있다. CDC에서는 6bit code⁽⁹⁾이고 IBM과 univac에서는 8bit⁽¹⁰⁾⁽¹¹⁾이기 때문에 모아쓰기를 爲해서는 이것의 3 倍程度의 bit數가 必要하여 8bit가지고 足한 Alphanumeric의 境遇에 比하면 memory의 浪費가 심하다.

한글 모아쓰기의 內部 code는 1970年度에 박안기氏가 提案한 Full key⁽¹²⁾ 방식以外에 組織的인 方法으로 code를 計算할 수 있는 組織的 code⁽¹³⁾가 提案되어 必要한 bits數가 16bits면 足하다는 利點을 가져왔으나

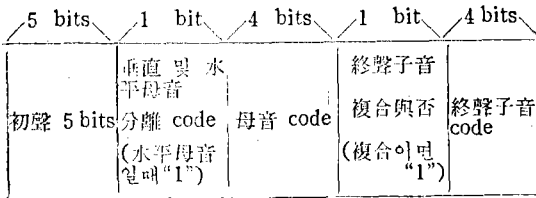
* 正會員, 서울大學校 工科大學 電子工學科
College of Engineering, Seoul National University.
接受日字: 1975年 3月 5日

decoding을 爲해서는 16bits 全體에 對해서 그 state를 살피기 위해서 入力端子가 많아야 하므로 三聲分離形 內部 code를 提案하였다.⁽¹⁴⁾ 이는 初聲, 中聲, 終聲을 各各 別途로 decoding 하기 때문에 decoder의 input 數를 16個로 부터 5個로 줄일 수 있다. 現行한글에 있어서 初聲은 重子音(ㄱ, ㅋ, ㆁ, ㆁ, ㆁ)을 獨立된 字母로 取扱해서 19個, 中聲은 垂直母音만의 複母音(ㅏ, ㅑ, ㅓ, ㅕ) 및 垂直, 水平母音混合複母音(ㅘ, ㅙ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ)를 모두 獨立된 字母로 取扱해서 21個이기 때문에 5 bits까지도 可能한 code數 32보다 적어서 5bits로서 收容이 可能하다. 終聲에 關해서는 單子音 14個以外에 重子音이나 複子音의 境遇가 13個(ㄴ, ㄷ, ㄹ, ㄴ, ㄷ, ㄹ, ㄴ, ㄷ, ㄹ, ㄴ, ㄷ, ㄹ, ㄴ, ㄷ, ㄹ) 있기 때문에 單子音 또는 複子音인가 單子音인가를 分類하는데 1bit 그리고 나머지를 위해서 4bits로 하면 다 收容될 수 있다. 따라서 初聲, 中聲, 終聲에 各各 5bits를 配當한 15bits의 內部 code로서 모든 한글을 code化할 수 있으며 그 構成은 第1

圖와 같다. 이것을 한글細密分離型 code라고 부른다면 이 code에서는 前記 maker들이 쓰고있는 6bits 또는 8bits code의 境遇에 비해 memory길이도 적어지고 또한 子音 code의 memory에서의 占有位置가 確定되어서 有利하다. 한글 原分離表順序를 尊重한다면 可能한 하나의 code方式은 第2圖와 같다. 初聲과 終聲에서 單子音의 境遇는 같은 code를 쓰고 있다.

2 縱記方式的 橫記方式에로의 變遷이 가져오는 筆書能率問題

水平母音과 垂直母音의 分離에 母音의 長枝部分이 基準이 되었던것과 같이 筆記上에 있어서 모든 母音을 短枝와 長枝部分으로 區分한다면 筆記方向이 上으로부터 下로가 爲主로 되었던 元來한 制定時代의 筆記方法의 效率과 지금과 같이 左에서 右로 쓰는 筆記方法의 境遇와에는 人間工學的인 見地에서 差가 있다. 다시말해서 한글은 制定當時筆記方向이 上으로부터 下였다는 事實이 水平母音과 垂直母音의 配當에 影響을 주었을 可能性이 있다. 文教部에서 1955년에 調査⁽¹⁵⁾한 바에 依하면 中聲母音을 10個 字素로 나눈 頻度는 第3圖와 같다.



第1圖

					初聲		中聲		終聲	
bs	b ₁	b ₂	b ₃	b ₄	0	1	0	1	0	1
	0	0	0	1	ㄱ	ㄴ	ㅏ	ㅑ	ㅇ	ㅣ
	0	0	1	0	ㅋ	ㄷ	ㅓ	ㅕ	ㄴ	ㅗ
	0	0	1	1	ㆁ	ㅞ	ㅙ	ㅛ	ㄴ	ㅜ
	0	1	0	0	ㄴ	ㅓ	ㅕ	ㅛ	ㄴ	ㅜ
	0	1	0	1	ㄷ	ㅞ	ㅙ	ㅛ	ㄴ	ㅜ
	0	1	1	0	ㅞ	ㅙ	ㅛ	ㅜ	ㅛ	ㅜ
	0	1	1	1	ㅙ	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ
	1	0	0	0	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ
	1	0	0	1	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ
	1	0	1	0	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ
	1	0	1	1	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ
	1	1	0	0	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ
	1	1	0	1	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ
	1	1	1	0	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ
	1	1	1	1	ㅛ	ㅜ	ㅛ	ㅜ	ㅛ	ㅜ
	0	0	0	0						

第2圖

차	례	발	침	반	도	%
1		ㅏ		56,644		25.66
2		ㅑ		52,507		24.04
3		ㅓ		30,190		12.82
4		ㅕ		28,092		12.86
5		ㅛ		24,684		11.30
6		ㅜ		14,488		6.63
7		ㅠ		8,223		3.76
8		ㅡ		1,850		0.85
9		ㅣ		1,643		0.75
10		ㅞ		706		0.32
계				2,181,429		99.99

第3圖

各各의 字素의 頻度와 長枝數를 合해서 水平長枝와 垂直長枝의 頻度를 살피 보면 各各

水平長枝頻度비중=31.96%

垂直長枝頻度비중=68.03%이다.

즉 한글母音은 決定的으로 垂直方向으로 쓰는 것이 글자의 흐름의 方向에 無理가 적음을 알 수 있다. 따

라서 人間工學的인 見地에서 母音字素의 水平化를 考慮해 볼 必要가 있다. 特히 “ㄱ”은 全體母音의 $\frac{1}{4}$ 을 超過하고 있으니 이 字素 하나만이라도 水平形으로 改定한다면 能率을 向上시킬수 있다. 但 이 25.66%는 “ㄱ” “내” “나” 등에 들어있는 “ㄱ”도 包含하고 있다.

4. Character generator matrix의 크기와 必要個數

1973年の 現況解説⁽¹⁶⁾⁽¹⁷⁾에서 枝路數와 折點(시작점, 끝점, 굴절점, 교점)數에 依해서 複雜度를 計量하는 方法을 提案하여 가장 複雜한 韓文字素(그것이 matrix의 크기를 決定하는 것이기 때문에)가 Alphabet의(절점數로 보아) 1.23倍, (枝路數로 보아) 1.22倍가 됨을 指摘하였는데 모아쓰기 韓글의 境遇는 橫方向으로는 그 字素가 들어가는 境遇(垂直母音의 境遇)가 最大이고 縱方向은 3字素(水平母音의 境遇)이나 美的見地에서 水平母音을 縱方向으로는 省略하여야 하기 때문에 半으로 치면 모아쓰기 matrix의 크기는 다음과 같이 推산할 수 있다.

橫方向 matrix크기 = 初聲字音(Alphanumeric경우의 5×1.22) + 垂直母音(이 경우 5이던 充分) = 11

縱方向 matrix크기 = 初聲字音(Alphanumeric경우의 7×1.22) + 水平母音을 위한 $7 \times 1.22 \times \frac{1}{2}$

+ 終聲字音(初聲에 準하여 7×1.22) = 21

여기서 한 matrix에 여러字素를 넣은내에서 오는 可能節減率을 2字素인 橫方向에 12%程度, 3字素인 縱方向에서는 24%程度로 잡고 그만큼 matrix의 크기를 줄여 준다면 大略 10×16 이 된다. 이들은 現在入手可能한 logic element(ROM RAM等)의 크기가 10bits 또는 16bits임에 비추어 浪費가 없어서 알맞게 보인다. 實際 10×16 bits로 character generator를 만들어 보면 “ㄸ”等 斜線이 많은 境遇에도 充分히 攄게 보이고 다른 글자에 對해서는 여유가 있는듯 하다.

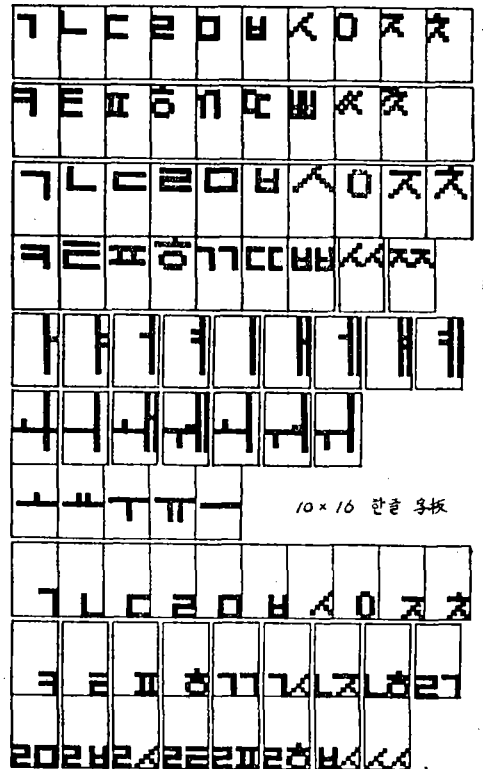
必要한 character generator의 個數에 對해서는 初聲 19個가 水平母音에 對해서는 中心線에 자리잡게 할 수 있으나 垂直母音에 對해서는 左側으로 옮겨져야 하기 때문에 必要 matrix는 38個이나 display를 위한 scanning機構는 19個를 위한 것으로 充分하다. 같은 初聲字素 두가지는 同時에 scanning되는데 쓰지않는 字板편의 shift register는 第6番에 bit에 依해서 Reset로 hold되어 있다. 이들 19雙의 matrix의 選定은 처음 다섯 bits에 따라서 2個의 16channel demultiplexer 74154에 依해서 行하여지고 이들은 中聲, 終聲에 對해서도 같은 方法으로 行하여진다. 따라서 初聲 demult-

iplexer가 2個 中聲 및 終聲 demultiplexer도 各各 2個씩이 必要하다.

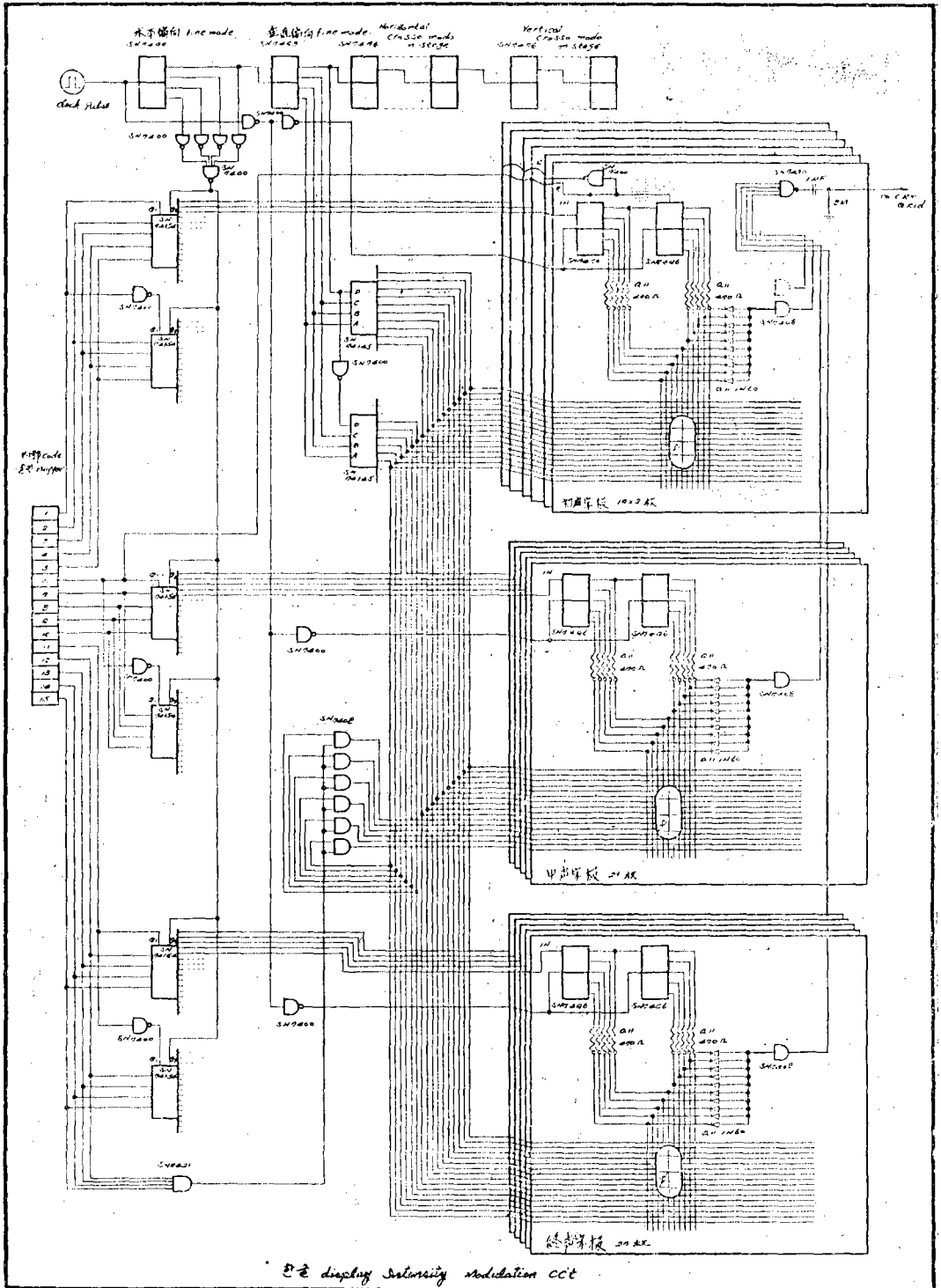
水平方向에 對한 scanning은 shift register에 10個中 1個만이 “1”인 pulse train을 入力에 導入함으로써 行하여지고 出力이 交代로 “1”이 되는데 470Ω 抵抗의 遠端도 電壓이 올간다. 그러나 diode가 들어있는 交點에서는 電位가 오르지않고 이것이 CRT에 光點으로 나타나게 된다.

中聲用 matrix의 必要個數는 中聲字素의 個數와 같은 21個이다. 但 垂直母音은 終聲이 없을 때를 위해서 11行, 12行, 13行까지 길게 뻗어있으나 다시 終聲有無에 따라서 gating된다. 따라서 終聲이 있을때에는 11行以下는 母音이 내려가지 않으며 그以下는 終聲이 全的으로 차지하게, 終聲이 없을 때에는 終聲 code는 $\times 0000$ 이다(\times 는 don't care條件). 따라서 0000가 檢出되면 11行以下를 enable한다. 그밖에 경우에는 11行以下의 母音字板이 inhibit되고 終聲이 全的으로 차지하게 된다.

終聲用 matrix는 27個로서 終聲이 없을 때 code의 第11位 以下 bit는 모두 “0”이다. 따라서 全體必要字板은 86個($19 \times 2 + 21 + 27$)이며 vertical scanning bit의

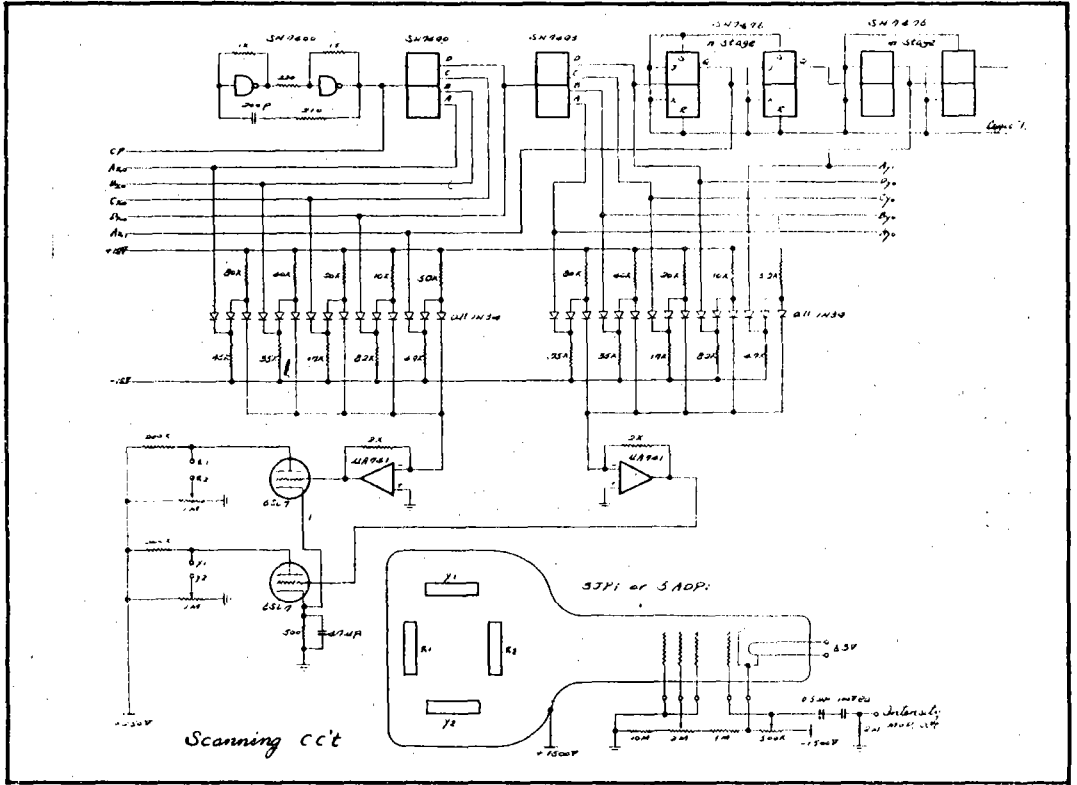


第 4 圖



色電 display salinity modulation CCt

第5圖



第5圖

數는 初聲이 雙으로 되어 있어 交代로 높고 있기 때문에 67(19+21+27)이다. 이中 3個의 matrix가 同時에 scan하게 된다. 이 數量은 많은듯하나 字母에 따라서 中心位置를 옮길 必要가 없으면서도 OR機能(positive logic)으로서 三聲部를 合成시킬 수 있기 때문에 機構가 比較的簡單하면서도 받아들일만한 美觀은 된다.

普通妥當性이 있어 보이는 한글pattern의 例를 第4圖에 보였다.

5. 回路와 構成

實際回路를 第5圖에 보인다. CRT表面에 2"行 2"列의 모아쓰기한글을 display한다면 먼저 거칠게 2"行 2"列의 交點上에 光點이 머물게 하되 이들은 實際 머물러있지 않고 그 近處에서 10列 16行으로 微小한 scanning을 하게 된다. 따라서 前者를 G.M. (Grosso modo) scanning이라 하고 後者를 F.M. (Fine modo) scanning이라고 하면, 그리고 flickering을 없애기 위해 1秒 20회의 display를 한다면 G.M.scan速度는 2"×2"×20이 될것이고 따라서 F.M.scan을 위해서는 2"×2"×20×10×16의 repetition rate를 갖는 clock pulse

를 必要로 한다. 이 clock pulse는 먼저 10進法(F.M. Horizontal scan을 위해)으로 divide되고 BCD出力이 D-A converter를 구동하여 matrix의 처음行의 10個의 點이 나오되 matrix의 pattern에 따라 明點이 되기도 하고 暗點이 되기도 한다. 이 scanning에 의해서 X方向 FMscan이 完成하면서 나오는 하나의(10分之 1로 나누어진) pulse는 Y軸의 F.M. scan을 1 step시켜 다음 X軸 scan은 1行만큼 앞은 位置에서 되풀이하게 된다. 10×16의 F.M.scan이 完成되면 하나의 pulse가 나오면서 G.M. X軸 scan(一字移動)을 行하되 이와 同時에 output buffer나 그다음 display할 한글의 code를 收容한 다음 buffer가 이어지거나 shift register의 內容이 shift되거나 한다. 繼續된 display를 위해서 2"×2"의 다시 말해서 display하게 될 文字의 數만큼의 Buffer 또는 2"×2" bit의 shift register가 15個 있어야 한다. 3雙의 16channel demultiplexer 74154가 한글內部 code의 各 5bits의 字素 code를 받아 이 code에 依해 指定된 字板(character generator matrix) 3個(各各初聲, 中聲, 終聲을 形成하는)가 同時에 scanning되어서 字板上에 字形을 나타내기 위해 diode

encoding되어 있는 部分만은 scan될때에 low potential로 나타나고 이들은 内部 code에 依해 選擇된 3板의 字板에서 서로 OR되어 diode가 있는 곳은 (그 3字板中 이라면 어느것에 있을 경우에나) 光點이 생기고 심지어는 같은곳에 겹쳐있어도 OR作用에 아무런 支障이 없다. 이미 說明한 바와 같이 第 6bit는 두벌씩있는 初聲字板(한벌은 左側上部를 차지하고 또한벌은 上部中央에 자리하고 있음)을 選擇하는데 쓰이고 第12bit부터 第15bit까지는 終聲에 없을 때 Binary code 0000가 收容되어 있어서 中聲 pattern의 下部길이를 定해준다.

4bit binary counter 7493의 出力에서 最終의 state가 나올 때는 한자가 完全히 display된 것이기 때문에 Buffer의 adress를 advance시켜서 다음 Buffer의 內容에 따라서 display할 字板群(初聲, 中聲, 終聲)을 指定하게 된다. 이 advance機構는 回路圖에 나타내지 않았으나 hardware的인 方法以外에 software的인 方法이 있다. 第 6番 bit가 1일때는 中聲이 水平母音이기 때문에 終聲 scauning用 pulse를 第3 register에 도입하여 右邊에 있는 終聲을 한가운데로 옮기게 한다.

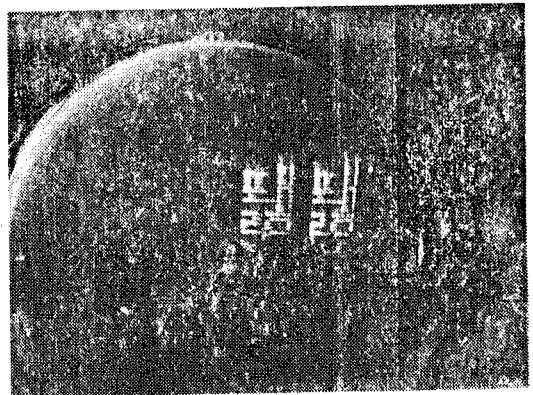
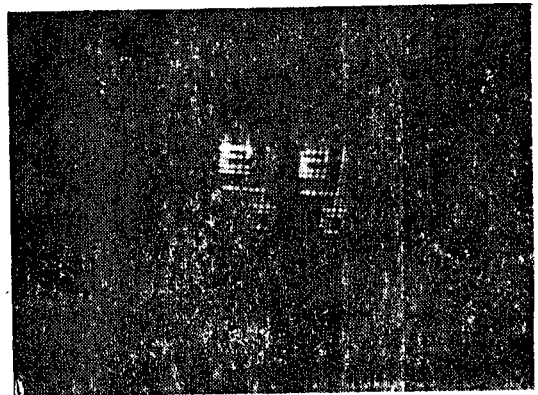
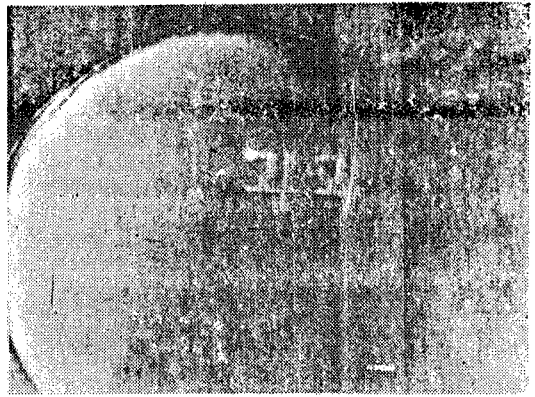
6. 結果와 檢討

Display된 몇개의 한글을 사진찍은 것을 第6圖에 보였다. 86個의 字板이 있으면 現在使用되고 있는 모든 모아쓰기한글을 다 display할 수 있고 美的見地에서 滿足할만한 pattern을 찾아낼 수 있는 程度의 豊饒함이 있고 그러한 反面 diode의 實地占有面積은 一字素面積에 限定되어서 全體字板 크기에 比하면 一部에 極限되기 때문에 實際必要한 diode의 數는 뜻밖에도 8223과 같은 FROM을 쓸수도 있어서 크기를 越等히 줄일 수 있는데 이 경우에 있어서도 semimanual한 bit by bit programming을 한다면 fusing해야할 bit가 적다는 것은 有利하다. 但 diode matrix에 比해서 access가 좋지않고 出力 bit가 8bits이기 때문에 10bits를 위해서는 浪費가 있다.

試製品은 2行 2列로서 4字의 display를 行하고 있는데 CRT에 對한 誘導가 全無해서 光點을 極히 적게할 수 있다면 同時에 display하는 字數를 크게할 수는 있을 것이다(Astigmatism까지 校正해서). 그러나 그 경우에 있어 Clock pulse rate를 相當히 키워야하고 그 結果로 scanning에 쓰이는 D-A converter의 conversion speed가 더 커야 하는데 이들은 現 state of art로서 充分히 解決될 수 있다.

7. 結 論

Alphanumeric character generator의 size 5×7에



第 6 圖

익숙한 大部分의 研究者의 觀念에서 보았을때 한글도 역시 過히 크지않는 size로 限定하는 것을 期待하지만 實際 한글의 三聲部基本의 音節表記方法으로는 必然코 Pattern이 大型化하기 마련이고 따라서 字素결집數와 모아쓰기구성 字素數를 檢討하여 10×16으로 하였는데 이는 頻도가 높지않는 複雜한 字素의 경우에는 아직도 最小必要 size이다. 但 頻度關係때문에 一般의 으로 남아도는 느낌을 준다.

한글은 하나하나의 字素도 複雜性에 있어서 單純한 경우와 複雜한 경우와 사이에 相當한 差가 있는데 matrix를 위해서는 既述한바와 같이 그중 가장 복잡한 것이 size를 決定하기 때문에 特別히 頻度가 적으면 시 複雜한 字素의 再調整을 생각하여 曠되하다. Alphabet의 경우도 特別히 筆記體를 위해서는 最少努力의 原則에 따라 簡單하게 쓸 수 있는 方向으로 도래하여 왔을 것은 事實일것이고 보면 前記 복잡度의 fluctuation이 적고 最少努力의 方向으로 옮겨지는 유연성이 기대된다.

size에 比하면 matrix用 diode數(또는 FROM等에 있어서는 fusing spot數)는 많지 않다. 必要한 基板의 數는 12×15cm의 22pin萬能基板으로 50枚程度이고 專用板을 開發한다면 30枚程度로 줄일 수 있다. 現在 50枚中 43枚가 matrix基板이다.

256bit bipolar Read only Memory를 쓰면 初, 中, 終 全體에 86個 package를 使用하면 되고 10×16代身 8×16 matrix를 쓴다면 半減할 수 있다. 이 程度의 規模로서 모든 한글 모아쓰기의 組合이 可能하기 때문에 字板. 字素 pattern決定에 專門家(美術이나 書道關係)들의 努力을 거치면 字板을 規格化할 수 있고 ROM製作도 本格化할 수 있을 것으로 보인다. 三聲部의 matrix의 面積配分에 가장많은 時間을 使用하여 決定한 것이지만 必要하다면 이들의 再配分도 考慮할 수 있다. 但 現在 本研究에서 使用한 試作品의 規模로서 充分하다는 것은 結論지을 수 있고 8×16bit로 칩는다 면 勿論 規模를 줄일 수 있다.

8. 感 謝

本研究의 資金은 1974年度 文敎部 教授研究造成費에서 나왔으며 設計와 實現에 있어서는 Digital technique에 特別히 圓熟의 境地에 있는 電氣通信研究所 電子傳送係 金析箕係長 team에 힘입은바 큼을 밝혀두는 바이다.

9. 參 考 文 獻

1. 崔嶺泰 “周波數分析에 의한 한글 音聲特性” 1972年

度秋季電子學術研究發表論文要約集 1972年 9月

2. 申龍澈, 崔嶺泰 “한글子音의 周波數分析의 연구” 大韓電子工學會誌 第10卷 第3號 1973年 6月
3. 강린구, 이형세 “한글자체의 특징추출의 한方式” 大韓電子工學會誌 第6卷 第2號 1969年 9月
4. 李柱根, “한글文字의 電子計算組織에 適應하기 위한 特徵抽出에 관한 연구(I)” 대한전자공학회지 第6卷 第4號 1969년 12월
5. 李柱根, 이광우 “한글문자의 인식에 관한 연구(II)” 대한전자공학회지 第7卷 第3號 1970년 11월
6. 柳時正 “한글入力の 配턴化에 대한 考察” 성균관대학교 경제개발대학원 석사학위 신청논문 1973년 1월
7. 李柱根 “한글문자의 인식에 관한 연구(N)” 대한전자공학회지 第9卷 第4號
8. 李柱根, 崔興文 “韓國語音節의 Entrophy에 관한 연구” 대한전자공학회지 第11卷 第3號 1974年 6月
9. 성기수 “EDPS한글화에 관한 연구” 1971년 1월
10. Univac “한글 Data처리에 대한 설명서” KD-7201 1968년 12월
11. IBM KOREA Inc. “Programmers Guide for using Hangeul” Form No. Hangeul 1-1
12. 박안기 “중앙전자계산소 1970년도 연구개발사업보고서” 1970년 12월
13. 文明鎬 “한글의 機械處理에 대한 考察” 성균관대학교 경제개발대학원석사학위 청구논문 1971년 12월
14. 崔道坤 “EDPS에 있어서 한글內部 codes와 Display에 관한 연구” 성균관대학교 경제개발대학원석사학위 청구논문
15. 문교부 “우리말에 쓰인 글자의 갖기조사” 1955년
16. 安秀桔 “電子計算機의 한글入出力에 관한 研究現況과 한글반풀어쓰기 提案” 대한전자공학회지 제10권 제2호 1973년 4월
17. 安秀桔 “電子計算機의 한글入出力에 관한 研究現況과 한글반풀어쓰기 提案(II)” 대한전자공학회지 제10권 제3호 1973년 6월