

บทที่ 4

การพัฒนาระบบการตรวจรู้อักขระภาษาไทย

4.1 ผังงานของระบบการตรวจรู้อักขระภาษาไทย

รายละเอียดของระบบตรวจรู้อักขระที่พัฒนาขึ้นได้แสดงอยู่ในรูปที่ 4.1 โดยที่ส่วนของชุดคำสั่งที่ใช้จัดเตรียมคำในภาษาไทย (Set THAI Word Module) ส่วนของชุดคำสั่งที่ใช้เลือกอักขระที่จะปรับปรุงแก้ไขรูปแบบ (Select Character to Modify Module) และส่วนของชุดคำสั่งที่ใช้เปลี่ยนแปลงแก้ไขรูปแบบของอักขระ (Modify Character Module) จะมีรายละเอียดแสดงไว้ในภาคผนวก ข

4.2 ลักษณะของจอภาพที่แสดงรายการต่างๆ (Screen Menu)

4.2.1 จอภาพรายการหลัก (Main Menu Screen)

เป็นจอภาพที่แสดงถึงรายการต่างๆที่ใช้ในการตรวจรู้อักขระ (รูปที่ 4.2)

4.2.2 จอภาพแสดงการจัดคำในภาษาไทย

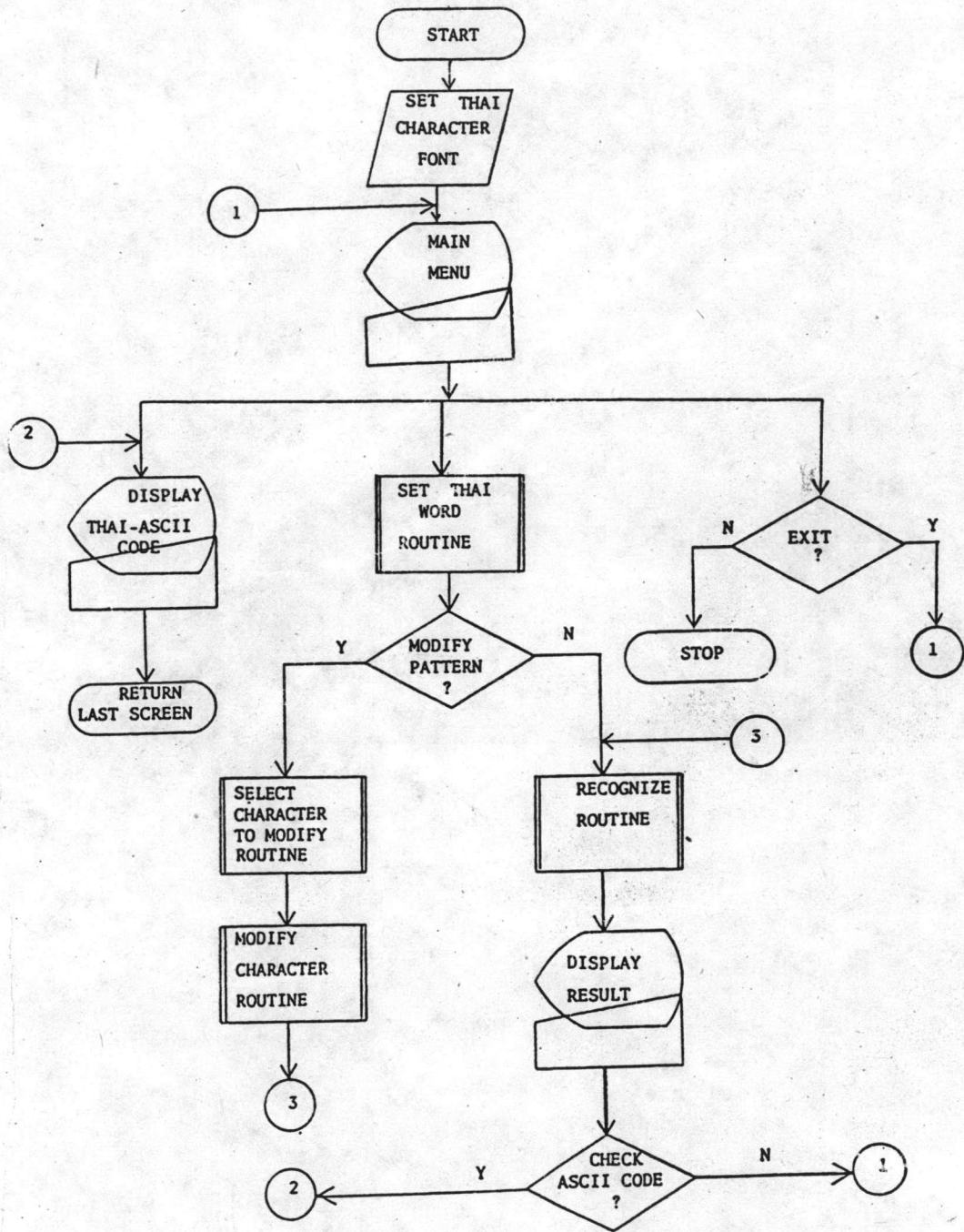
(THAI Word 4 Levels Screen)

เป็นจอภาพที่ช่วยในการจัดคำภาษาไทยทั้ง 4 ระดับเพื่อการตรวจรู้อักขระ (รูปที่ 4.3)

4.2.3 จอภาพแสดงการปรับปรุงแก้ไขอักขระ

(Modify Pattern Screen)

เป็นจอภาพที่ใช้ในการเลือกอักขระภาษาไทยแต่ละรูปมาปรับปรุงแก้ไขรูปแบบของอักขระ (รูปที่ 4.4)



รูปที่ 4.1 แสดงผังงานของระบบตรวจรู้อักขระที่พัฒนาขึ้น

4.2.4 จอภาพแสดงรายการตารางรหัสแอสกีของอักขระภาษาไทย
(THAI ASCII Code Table)

เป็นจอภาพที่แสดงรหัสแอสกีของอักขระภาษาไทยแต่ละรูป
(รูปที่ 4.5) โดยยึดถือตามรหัสที่สำนักงานมาตรฐานอุตสาหกรรม กระทรวง
อุตสาหกรรมกำหนดขึ้น

4.3.5 จอภาพแสดงผลการตรวจรู้อักขระภาษาไทย
(Recognize Result)

เป็นจอภาพที่แสดงผลของการตรวจรู้อักขระ โดยจะแสดง
ผลลัพธ์เป็นรหัสแอสกีในระดับเดียวกัน (รูปที่ 4.6)

4.3 การสับเปลี่ยนข้อมูลภายในหน่วยความจำหลัก (8), (9)
(Memory Transfers)

การใช้หน่วยความจำหลักเป็นสื่อข้อมูลที่ใช้ในการประมวลผลเป็นวิธี
การที่มีประสิทธิภาพมากกว่าการใช้หน่วยความจำสำรอง (Secondary
Storage) เก็บข้อมูลเพื่อการประมวลผล ทั้งนี้เพราะเวลาที่ใช้ในการอ่านและ
พิมพ์ (I/O) ข้อมูลระหว่างหน่วยความจำหลักกับสื่อข้อมูลเป็นช่วง เวลาที่มากเมื่อ
เทียบกับการประมวลผลของ เครื่องคอมพิวเตอร์ ฉะนั้นการสับเปลี่ยนข้อมูลใน
หน่วยความจำหลักจึง เป็นวิธีที่ช่วยเพิ่มประสิทธิภาพของระบบการตรวจรู้อักขระให้
ดียิ่งขึ้น

หลักการสับเปลี่ยนข้อมูลในหน่วยความจำหลัก กระทำได้ด้วยการบรรจุ
ข้อมูลต่างๆ ที่ต้องใช้ในการประมวลผลลงในแฟ้มข้อมูลก่อน เมื่อจะเริ่มต้นประมวล
ผลก็ให้นำแฟ้มข้อมูลดังกล่าวบรรจุลงในหน่วยความจำหลัก ในขณะที่ประมวลผลอยู่
เมื่อเครื่องคอมพิวเตอร์ต้องการประมวลผลข้อมูลใดก็เพียงแต่สับเปลี่ยนของข้อมูล
เหล่านั้นในหน่วยความจำหลัก เครื่องคอมพิวเตอร์ก็จะได้ข้อมูลที่ต้องการประมวล
ผลทันที แต่การสับเปลี่ยนข้อมูลในหน่วยความจำหลักต้องคำนึงถึงตำแหน่งที่บรรจุ
ข้อมูลต่างๆ ในหน่วยความจำหลักเช่น ชุดคำสั่ง ตัวแปลชุดคำสั่ง (Compiler)
ชุดคำสั่งควบคุมระบบ (Operating System) เป็นต้น เพราะหากสับเปลี่ยน
ข้อมูลผิดตำแหน่งอาจทำให้เครื่องคอมพิวเตอร์หยุดการทำงานได้

```
*****  
* THAI CHARACTER RECOGNITION *  
*****  
Copyright 1986  
Function Key Assignment
```

```
<F1> : Help Menu  
<F2> : Set THAI WORD (4 Levels)  
<F3> : Modify THAI character pattern  
<F4> : THAI-ASCII code table  
<F5> : Recognize THAI WORD  
<F6> : Select Character font  
<F7> : Return Last Screen  
<F8> : Exit to MS-DOS
```

Select <F1> - <F8>

Master Thesis
Dept. of Computer Science

Faculty of Engineering
Chulalongkorn University

รูปที่ 4.2 แสดงจอภาพรายการหลัก

ก ข ช ค ต ส ง จ
 ฉ ช ฌ ฉ ญ ฎ ฏ ฐ
 ท ฒ ฒ ด ต ถ ท ธ
 น บ ป ผ ฟ พ ฟ ภ
 ม ย ร ฤ ล ภา ว ค
 ษ ส ห ฬ อ ฮ ะ
 ๑ ำ ๑ ๑ ๑ ๑ ๑
 ๒ ๒ ๒ ๒ ๒ ๒ ๒

๑ ๒ ๓ + ๔

THAI WORD
 (4 levels)

USING ↑ ↓ → ←
 <RETURN>
 &
 <A> - <X>

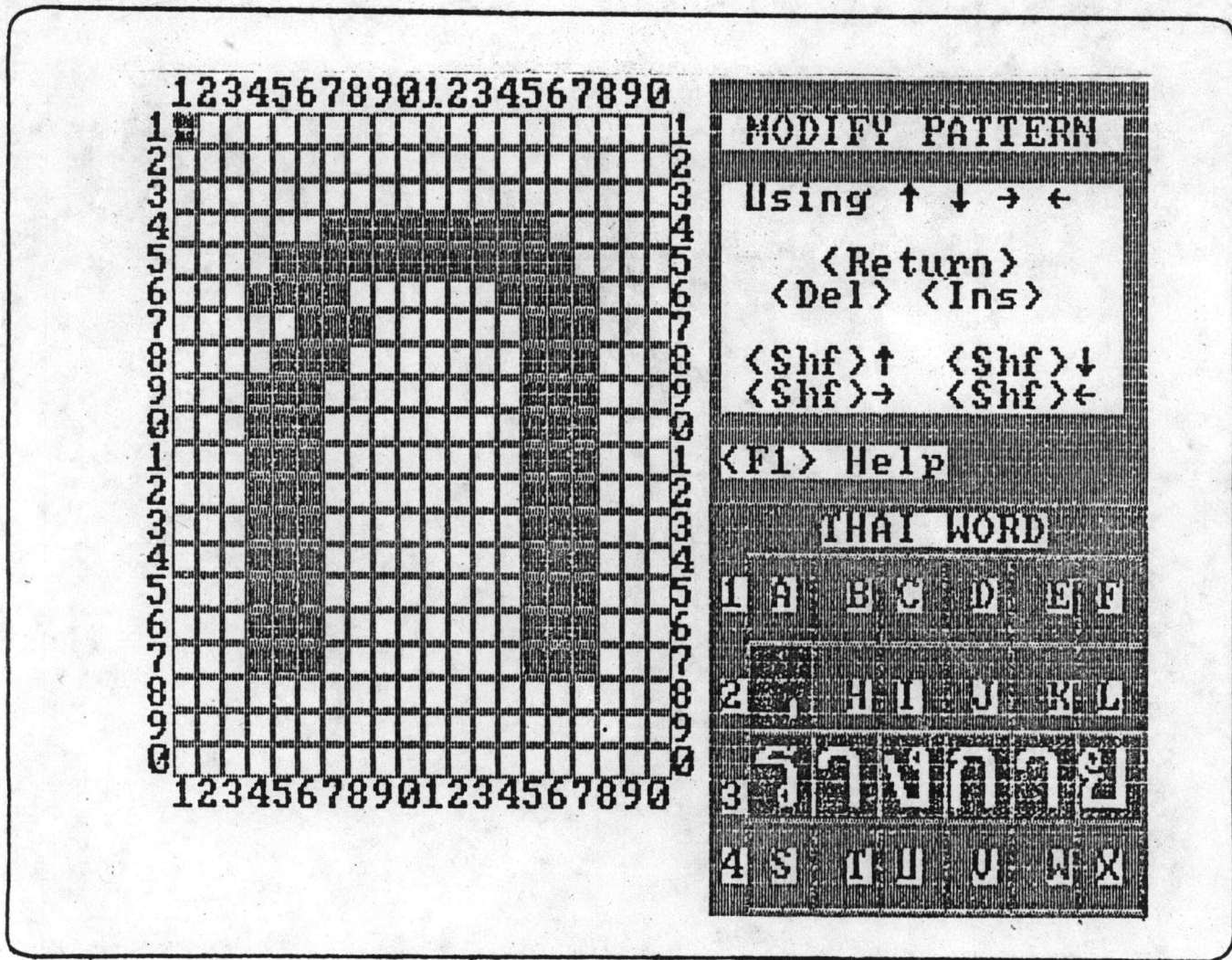
<F1> Help

THAI WORD

1	A	B	D	E	F
2	Y	H	J	K	L
3	M	N	P	Q	R
4	T	U	V	W	X



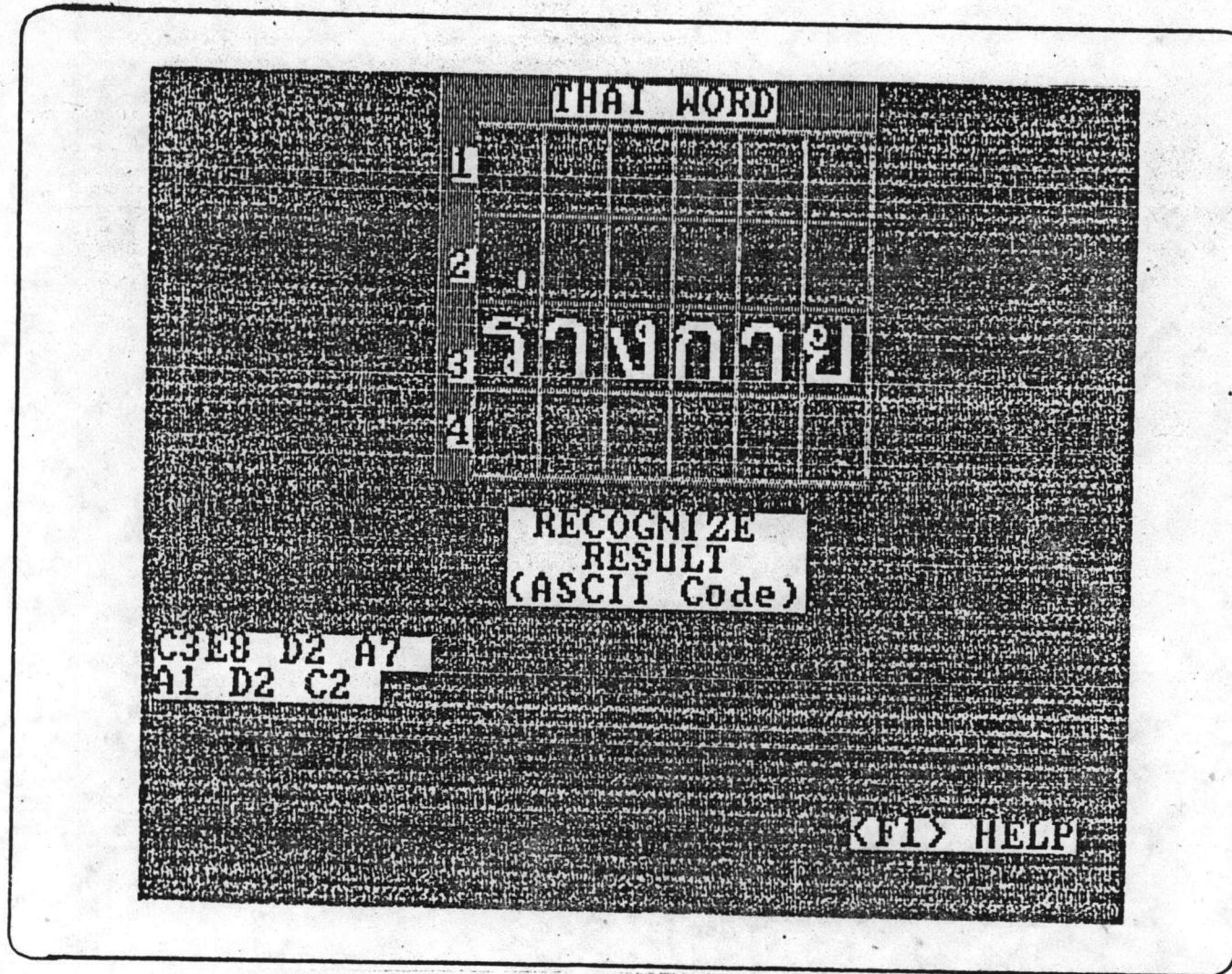
รูปที่ 4.3 แสดงจอภาพรายการจัดคำภาษาไทยทั้ง 4 ระดับ



รูปที่ 4.4 แสดงจอภาพรายการปรับรูปแก้ไขรูปแบบของอักขระ

*** THAI-ASCII CODE TABLE ***																
Hex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		ก	ข	ช	ค	ศ	ฆ	ง	จ	ฉ	ซ	ฌ	ณ	ญ	ฎ	ฏ
B	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ	ฟ
C	ภ	ม	ย	ร	ฤ	ล	ฬ	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ	ฯ
D	๕	๖	๗	๘	๙	๐	๑	๒	๓	๔						
E	๕	๖	๗	๘	๙	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	
* Industrial Standard *																
<RETURN>																

รูปที่ 4.5 แสดงจอภาพรายการตารางรหัสแอสกีของอักขระภาษาไทย



รูปที่ 4.6 แสดงจอภาพที่แสดงผลของอักขระที่สามารถตรวจรู้ได้

(ตำแหน่งของข้อมูลที่บรรจุในหน่วยความจำหลักได้แสดงไว้ในตารางที่ 4.1)
ส่วนแฟ้มข้อมูลและตำแหน่งที่บรรจุแฟ้มข้อมูลในหน่วยความจำหลักมีรายละเอียด
ดังนี้

ตำแหน่ง	ข้อมูลที่บรรจุ
&H38000	แฟ้มข้อมูลที่บรรจุรายการหลัก
&H40000	แฟ้มข้อมูลที่บรรจุรายการที่ใช้แก้ไขรูปแบบอักขระ
&H48000	แฟ้มข้อมูลที่บรรจุรายการที่แสดงรหัสแอสกีของอักขระภาษา ไทย
&H50000	แฟ้มข้อมูลที่บรรจุรายการแสดงแบบพิมพ์และการจัดอักขระ ภาษาไทย
&H58000	ตำแหน่งที่เก็บข้อมูลชั่วคราวเพื่อการสับเปลี่ยนข้อมูล
&HB8000	ส่วนแสดงผลของจอภาพ

เมื่อคอมพิวเตอร์ต้องการประมวลผลข้อมูลใดแทนที่คอมพิวเตอร์จะอ่าน
ข้อมูลจากหน่วยความจำสำรองเข้ามาประมวลผลคอมพิวเตอร์ก็จะใช้วิธีสับเปลี่ยน
ข้อมูลในหน่วยความจำหลักแทนตัวอย่างเช่นการให้จอภาพของเครื่องคอมพิวเตอร์
แสดงผลข้อมูลต่างๆ อย่างฉับพลัน (Switch Page) คอมพิวเตอร์ก็จะประมวล
ผลชุดคำสั่งที่ทำหน้าที่สับเปลี่ยนข้อมูล (รูป 4.7) จากตำแหน่งที่ข้อมูลบรรจุอยู่มา
ยังตำแหน่งที่ &HB8000 ในหน่วยความจำหลักซึ่งเป็นตำแหน่งที่จอภาพจะไปดึง
ข้อมูลมาแสดงผล ข้อมูลที่ต้องการให้แสดงผลก็จะปรากฏบนจอภาพอย่างฉับพลัน

นอกจากการใช้เทคนิคการสับเปลี่ยนข้อมูลในหน่วยความจำแล้วการ
วิจัยนี้ยังได้เพิ่มประสิทธิภาพการคำนวณในระบบการตรวจรู้อักขระนี้ด้วยการนำ
ไมโครโพรเซสเซอร์ หมายเลข 8087 มาช่วยในการประมวลผล จากการนำ
ไมโครโพรเซสเซอร์ หมายเลข 8087 มาช่วยในการประมวลผลนี้สามารถลด
เวลาที่ใช้ในการตรวจรู้อักขระแต่ละรูปลงได้ประมาณ 3 เท่าตัว

Decimal	Hex	Function	Comments
0 16K 32K 48K	00000 04000 08000 0C000	16-64 KB RAM installed on System Board	Typically comes with system
64K 96K 128K 160K 192K 624K	10000 38000 40000 48000 50000 9C000	Main menu screen file Modify character screen file Thai-ASCII code table screen file Set Thai word screen file	
640K	A0000	Reserved	
656K 672K 688K	A4000 A8000 AC000		This space can be used by both the IBM display boards and other manu- facturers' boards. The IBM boards occupy the portions of this area as show.
704K	B0000	Monochrome Display buffer	
720K	B4000		
736K	B8000	Color/Graphics display buffer	
752K	BC000		
768K	C0000		
. .	. .	192 K B memory expansion area	This space requires special software to utilize
944K	EC000		
960K	F0000	Reserved	
976K 992K 1008K	F4000 F8000 FC000	48 K B System ROM	The last 40 K of this ROM comes with the PC.

ตารางที่ 4.1 ตารางแสดงตำแหน่งของข้อมูลที่บรรจุในหน่วยความจำหลัก

```

CODESEG      PUBLIC  MOVE
MOVE         SEGMENT BYTE PUBLIC 'CODE'
PROC        FAR
ASSUME      CS:CODESG
PUSH        BP
MOV         BP,SP
CLD
MOV         BX,[BP+06]
CMP        WORD PTR [BX],+00
JZ         L010D
L010D:      STD
MOV         BX,[BP+14]
MOV         SI,[BX]
MOV         BX,[BP+12]
PUSH        ES
MOV         ES,[BX]
MOV         BX,[BP+10]
MOV         DI,[BX]
MOV         BX,[BP+08]
MOV         CX,[BX]
MOV         BX,[BP+16]
PUSH        DS
MOV         DS,[BX]
REPZ       MOVSB
POP         DS
POP         ES
POP         BP
RET         12
MOVE        ENDP
CODESEG     ENDS

```

รูปที่ 4.7 แสดงชุดคำสั่งที่ใช้สับเปลี่ยนข้อมูลในหน่วยความจำหลัก

4.4 การตรวจรู้อักขระภาษาไทย

การตรวจรู้อักขระนี้แบ่งเป็น 3 ส่วนคือ

4.4.1 การแปลงอักขระต้นแบบให้เป็นโครงร่างของอักขระ

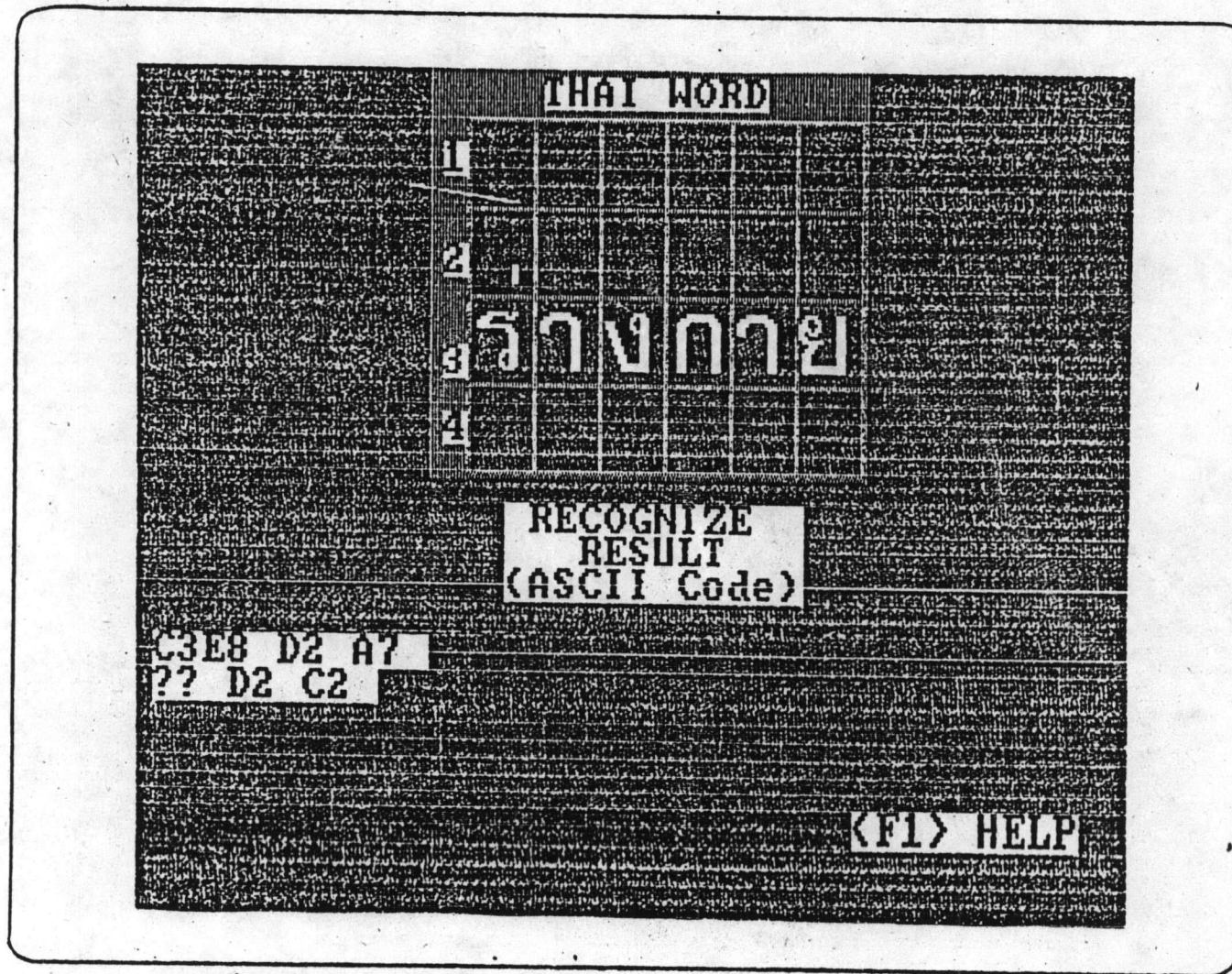
4.4.2 การแทนโครงร่างของอักขระด้วยรหัสรวม

4.4.3 การค้นหารหัสรวมกับรหัสต้นแบบ

รายละเอียดของการตรวจรู้อักขระได้กล่าวมาแล้วในบทที่ 3

4.5 การแสดงผลของการตรวจรู้อักขระ

ตามที่ได้กล่าวมาแล้วในหัวข้อที่ 3.6 ว่า ผลลัพธ์ที่ได้จากการตรวจรู้อักขระภาษาไทยจะแสดงอยู่ในรูปของรหัสแอสกีเป็นแถวในระดับเดียวกัน โดยจะแสดงผลของการตรวจรู้อักขระบนจอภาพเดียวกับอักขระที่ใช้ในการตรวจรู้ (รูปที่ 4.6) ส่วนผลลัพธ์ของอักขระที่ไม่สามารถตรวจรู้ได้จะใช้สัญลักษณ์ "??" แทนรหัสแอสกีของอักขระนั้น (รูปที่ 4.8)



รูปที่ 4.8 แสดงจอภาพแสดงผลของอักขระที่ไม่สามารถตรวจรู้