บทที่ 4

การศึกษาลักษณะของคำและการพัฒนาขั้นตอนในการจดจำ

4.1 การศึกษาลักษณะของคำ

การทดลองศึกษาลักษณะของคำเพื่อเป็นแนวทางในการพัฒนาขั้นตอนในการจดจำ ได้เตรียมการไว้ 2 ขั้นตอน คือ

4.1.1 การทดลองเพื่อศึกษาลักษณะโดยทั่วไปของคำพูด

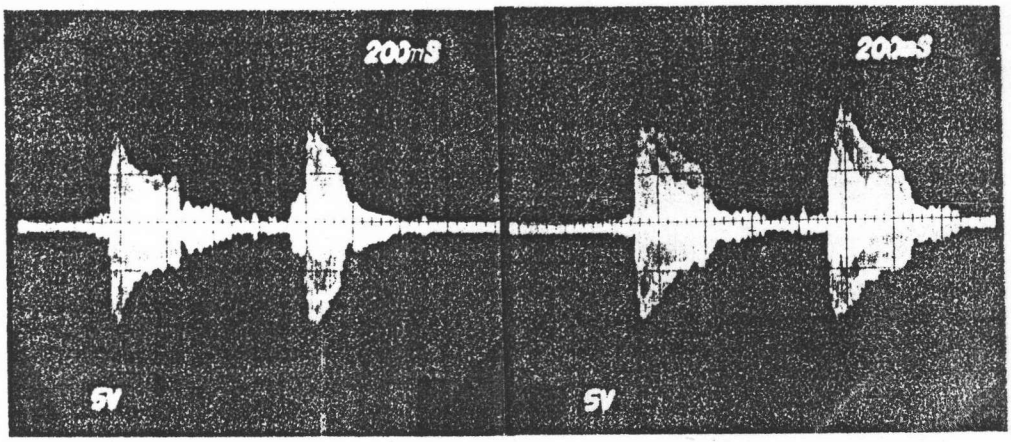
วัตถุประสงค์ เพื่อศึกษาลักษณะรูปสัญญาณ

การทดลอง ทำการพูดคำตัวอย่าง จาก 0 – 9 หลาย ๆ ครั้ง ผ่านไมโครโฟน และจับภาพสัญญาณขาออกจากวงจรกรองผ่านความถี่ต่ำ (3.2.2) โดยใช้สตอเรจออสซิโลสโคป (**storage oscilloscope**) ลักษณะของรูปสัญญาณเป็นดังรูปที่ 4.1

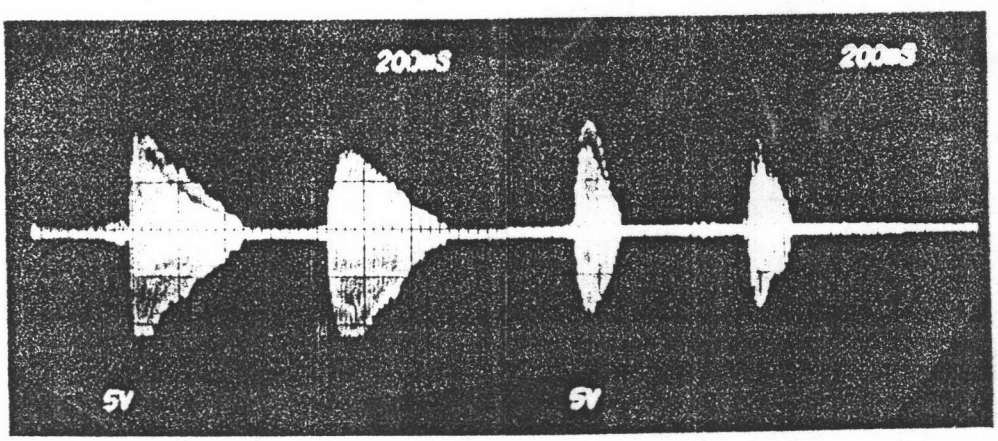
ในการออกเสียงคำแต่ละคำ จะมีความยาวแตกต่างกัน ตามเสียงสระ ค่าระยะเวลาในการออกเสียงคำ เป็นดังตารางที่ 4.1

คำพูด	ช่วงเวลาของคำพูด (มิลลิวินาที)
0	500 – 600
1	400 – 500
2	500 – 600
3	500 – 600
4	400 – 500
5	500 – 600
6	200 – 300
7	200 – 250
8	400 – 500
9	500 – 600

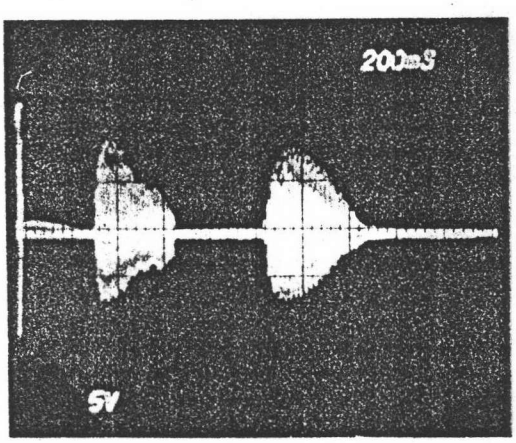
ตารางที่ 4.1 แสดงช่วงเวลาในการออกเสียงเมื่อนับ 0 – 9



"ศูนย์" "หนึ่ง" "สอง" "สาม"



"สี่" "ห้า" "หก" "เจ็ด"



"แปด" "เก้า"

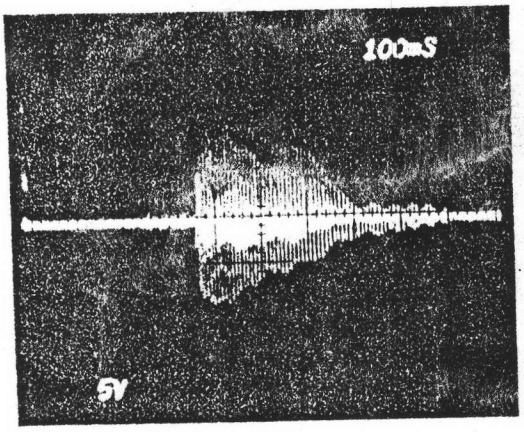
รูปที่ 4.1 แสดงรูปสัญญาณของคำพูดศูนย์ถึงเก้า

จากค่าช่วงเวลาในการออกเสียงสามารถแบ่งคำเป็น 2 พวกคือ คำที่มีเสียงสั้น จะมีช่วงเวลาในการออกเสียง 200 – 300 msec และคำที่มีเสียงยาวอยู่ในช่วง 300 – 600 msec ดังแสดงในตารางที่ 4.2

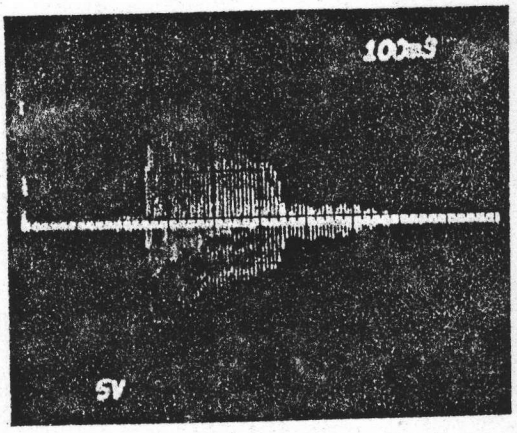
ประเภทของคำ	ช่วงเวลา (m sec)	คำพูด
เสียงสั้น	200 – 300	6, 7
เสียงยาว	300 – 600	0,1,2,3,4,5,8,9

ตารางที่ 4.2 แสดงกลุ่มของคำพูดแยกตามระยะเวลาการออกเสียง

การออกเสียงคำ ๆ เดียวกัน และการออกเสียงคำที่มีวิธีการออกเสียงคล้าย ๆ กัน จะให้รูปสัญญาณที่มีลักษณะคล้ายกัน เช่น คำพูด "2" และ "3" เป็นต้น ส่วนคำที่มีวิธีการออกเสียงแตกต่างกัน จะให้รูปสัญญาณที่มีลักษณะแตกต่างกัน เช่น "1" และ "9" ดังแสดงในรูปที่ 4.2(ก) และ 4.2(ข)

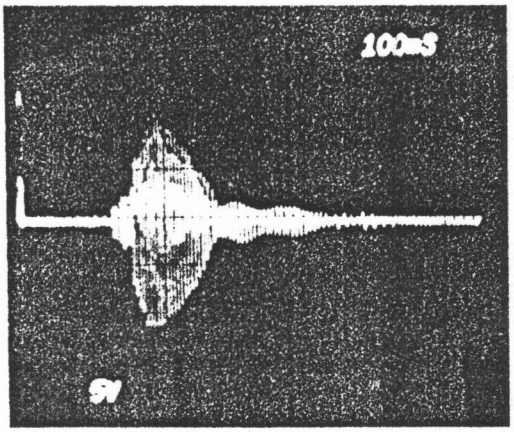


"สอง"

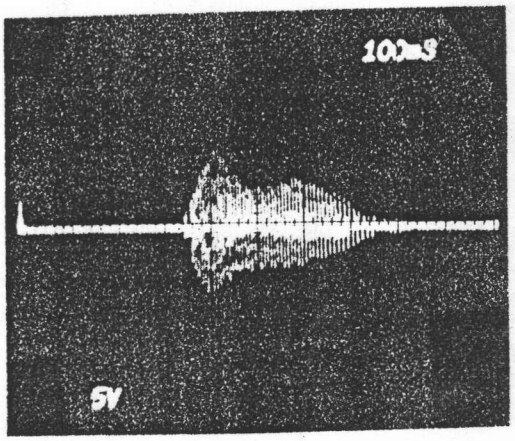


"สาม"

รูปที่ 4.2(ก) แสดงภาพสัญญาณที่มีลักษณะคล้ายคลึงกัน



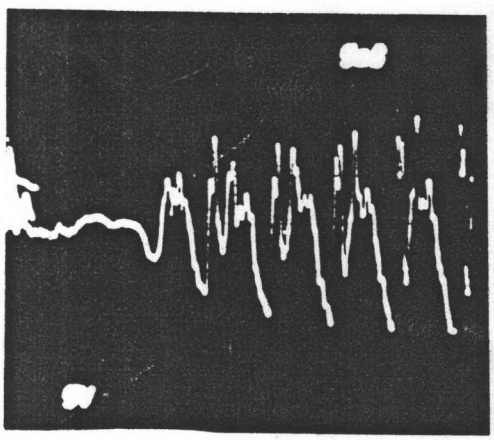
"หนึ่ง"



"เก้า"

รูปที่ 4.2(ข) แสดงภาพสัญญาณที่มีลักษณะแตกต่างกัน

เมื่อทำการขยายภาพสัญญาณจะเห็นได้ว่า สัญญาณเสียงเป็นสัญญาณที่มีคาบเวลา (periodie signal) แต่ไม่เป็นรูปซายน์ธรรมชาติ (natural sinusoidal wave form) ซึ่งแสดงว่าสัญญาณเสียงประกอบด้วยสัญญาณความถี่พันดาเมนทัล (fundamental frequency) และความถี่ยาร์โมนิค (harmonic frequency) ค่าคาบเวลาจะเปลี่ยนไปตลอดช่วงของคำซึ่งจะมีขึ้นกับระดับเสียงที่ใช้ในการออกเสียงคำนั้น ๆ ลักษณะของภาพขยายรูปสัญญาณเป็นดังรูปที่ 4.3



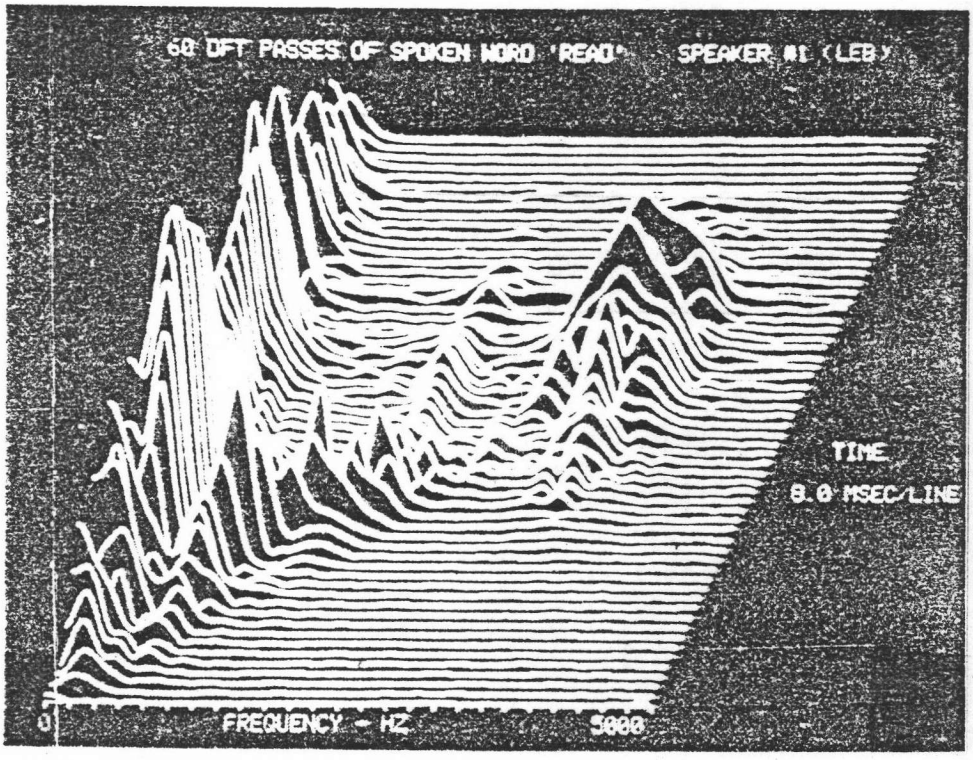
รูปที่ 4.3 แสดงภาพขยายของสัญญาณเสียง "9" ในช่วงต้นของคำ

4.1.2 การทดลองเพื่อศึกษาลักษณะเชิงความถี่

วัตถุประสงค์ เพื่อศึกษาลักษณะการกระจายพลังงานเสียงในบางช่วงความถี่

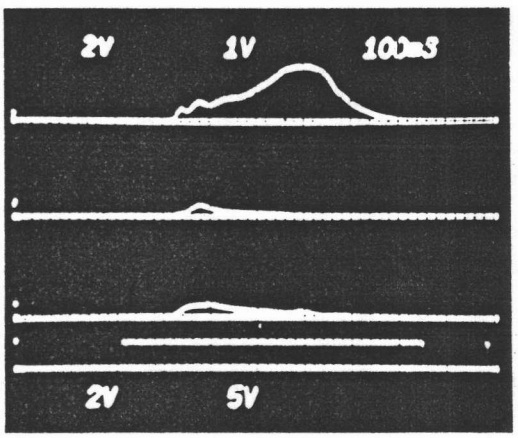
การทดลอง ทำการพูดคำพูด "0" ถึง "9" ผ่านไมโครโฟน และจับภาพสัญญาณขาออกจากวงจรกรองผ่านความถี่ (3.2.4) โดยใช้สตอเรจออสซิโลสโคป

ภาพสัญญาณที่ปรากฏจะเป็นภาพเอ็นเวลล็อป (envelope) ของฮาร์โมนิคที่ปรากฏในแต่ละแถบความถี่ ลักษณะของรูปสัญญาณนี้ มีความคล้ายคลึงกันสำหรับคำพูดคำเดียวกัน และมีลักษณะแตกต่างกันสำหรับคำพูดคนละคำ เนื่องจากการเคลื่อนไหวที่แตกต่างกันของอวัยวะที่ใช้ในการพูด จะทำให้เกิดอะคูสติกเรโซแนนซ์ (acoustic resonance) ต่างกันไป ซึ่งจะปรากฏในรูปขององค์ประกอบฮาร์โมนิคของสัญญาณเสียง ตัวอย่างของการกระจายพลังงานเสียงเป็นดังรูปที่ 4.4

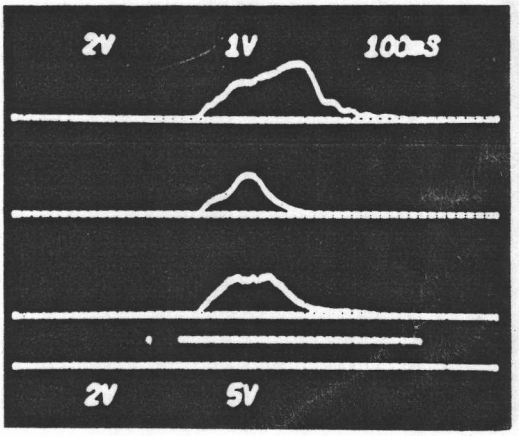


รูปที่ 4.4 แสดงการกระจายของพลังงานเสียงไปตามแถบความถี่ต่าง ๆ ในการพูดคำว่า "READ" จากการทดลองของ Tufts, Levinson และ Rao

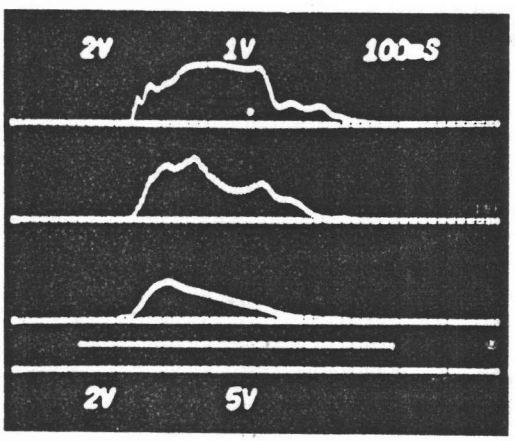
ลักษณะของรูปสัญญาณที่จับได้ในแต่ละแถบความถี่ จากการทดลองเป็นดังรูปที่ 4.5



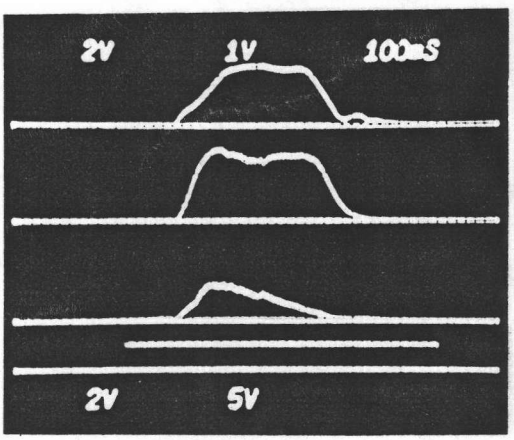
"ศูนย์"



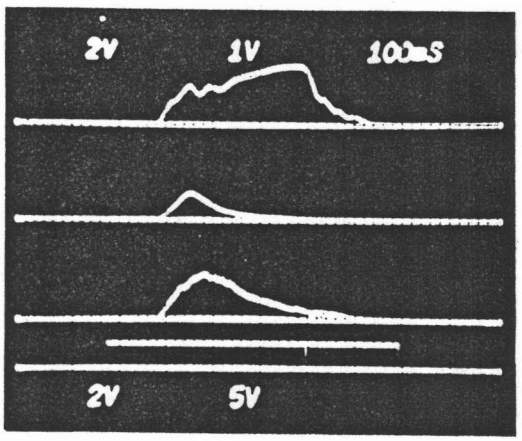
"หนึ่ง"



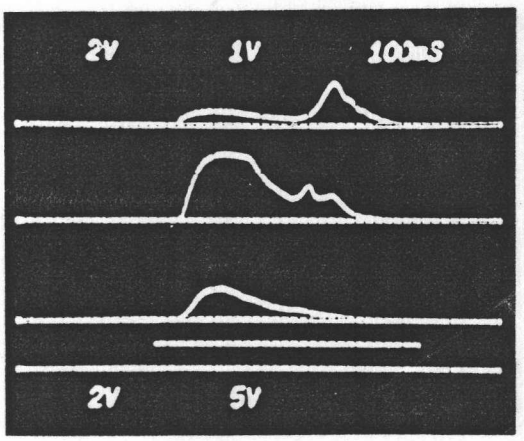
"สอง"



"สาม"

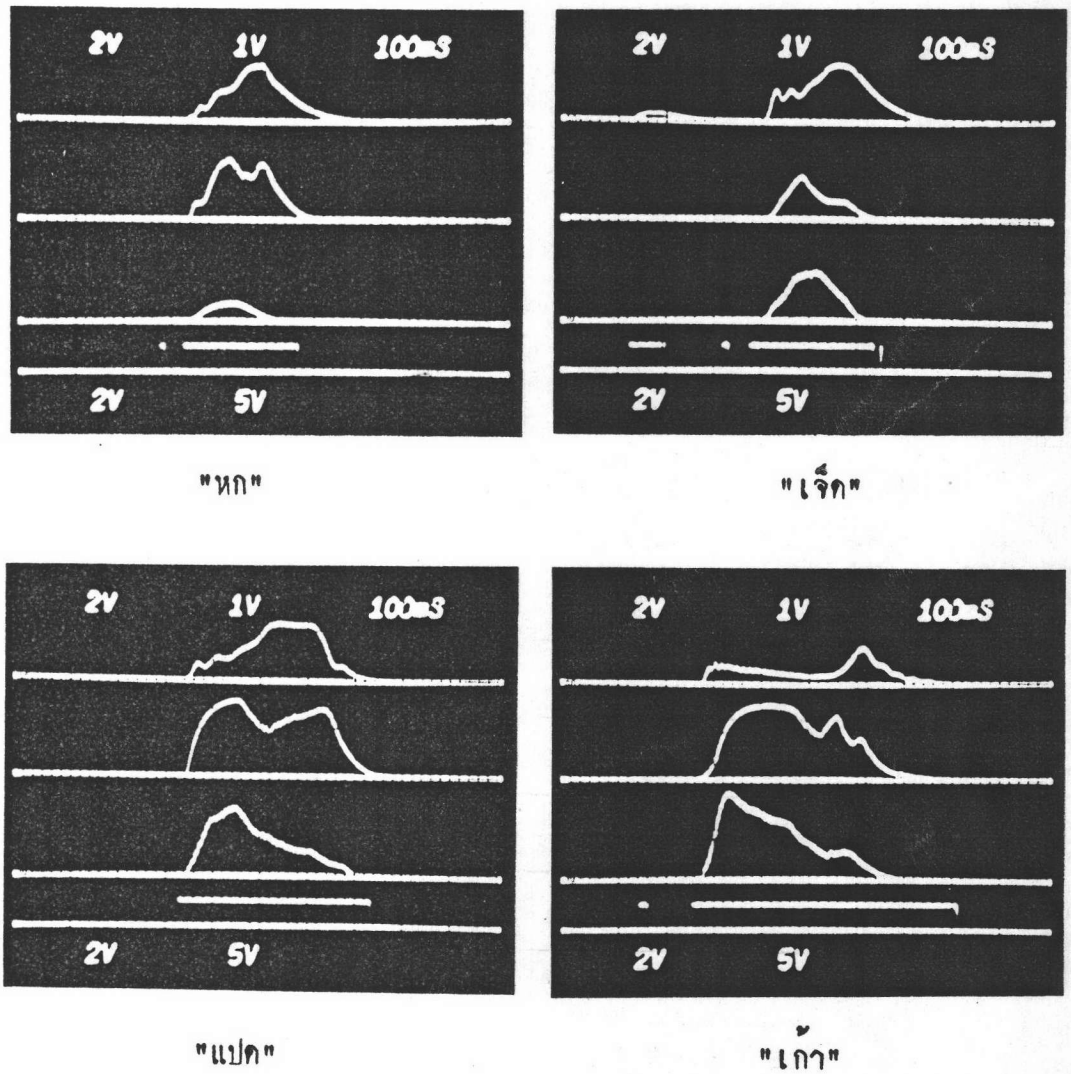


"สี่"



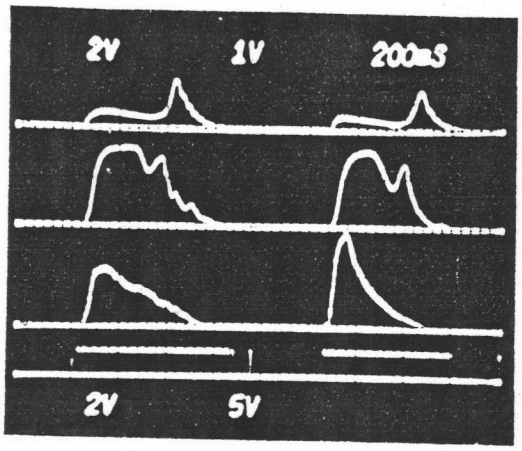
"ห้า"

15432920

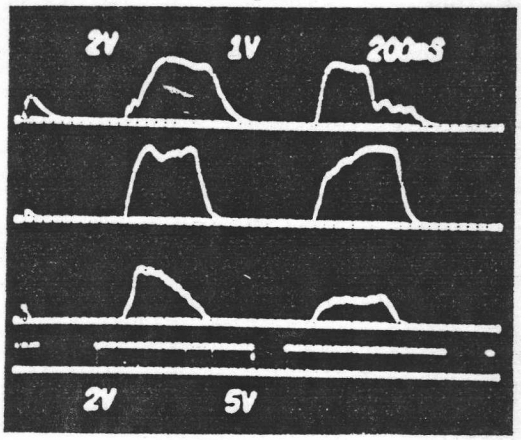


รูปที่ 4.5 แสดงรูปสัญญาณที่ปรากฏในแต่ละแถบความถี่ของคำพูดศูนย์ถึงเก้า

ลักษณะของรูปสัญญาณนี้จะแปรเปลี่ยนไปบ้างเนื่องจากในการพูดแต่ละครั้งจะมีลักษณะการออกเสียงที่แตกต่างกันเล็กน้อย ซึ่งมีผลต่อออะคูสติกเรโซแนนซ์ ความแตกต่างนี้แสดงให้เป็นดังรูปที่ 4.6



"เก้า"



"แปด"

รูปที่ 4.6 แสดงความแตกต่างของรูปสัญญาณในการพูดคำ ๆ เดียวกัน 2 ครั้ง

ในการวิเคราะห์เชิงปริมาณพลังงานเสียงที่ปรากฏในแต่ละแถบความถี่ โดยการหาค่าผลรวมของค่าสัญญาณที่อ่านได้จากวงจรเปลี่ยนสัญญาณเสียงต่อหน่วยเวลา ดังแสดงในตารางที่ 4.4 ข้อมูลที่ใช้แสดงไว้ในภาคผนวก

Pattern	Duration			Average voice signal								
				BPF 3 (1,000 Hz.)			BPF 2 (800 Hz.)			BPF 1 (200 Hz.)		
	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean
0	44	37	40	3	0	0	10	2	4	26	4	8
1	41	36	37	13	3	6	9	1	2	36	16	25
2	45	37	41	3	0	1	24	10	18	34	2	20
3	44	39	41	3	1	1	44	30	39	30	10	18
4	42	38	40	7	0	2	1	0	0	32	14	19
5	41	30	34	15	2	9	54	18	36	8	1	4
6	20	14	16	5	0	2	73	35	47	39	5	24
7	16	14	15	30	6	18	12	3	7	22	4	10
8	29	26	26	36	8	16	108	89	100	28	10	17
9	55	40	44	36	20	28	66	41	55	6	2	3

ตารางที่ 4.3 แสดงค่าเฉลี่ยของสัญญาณเสียงต่อหน่วยเวลา
ของคำพูด "0" – "9"

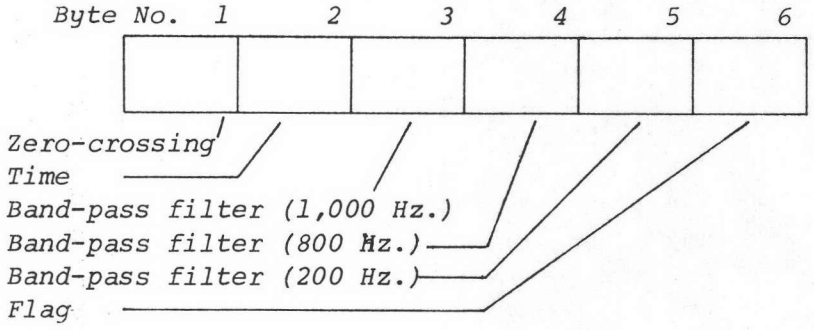
4.2 การพัฒนาขั้นตอนในการจดจำ

จากตารางที่ 4.4 จะพบว่าค่าเฉลี่ยของสัญญาณเสียงของคำแต่ละคำ เมื่อพิจารณาจากวงจรกรองแถบความถี่ทั้งสามมาประกอบกัน มีความแตกต่างกันมากพอสมควร ซึ่งสามารถนำมาใช้เป็นตัวประกอบการตัดสินใจในการแยกกลุ่มของคำ การพัฒนาขั้นตอนทำในลักษณะให้ระบบเก็บตัวอย่างเสียงที่ผู้พูดต้องการจะพูด โดยทำการวิเคราะห์หาค่าเฉลี่ยของสัญญาณเสียงจากข้อมูลของแต่ละวงจรกรองแถบความถี่ แล้วเก็บค่าที่วิเคราะห์ได้พร้อมความหมาย หลังจากเสร็จสิ้นการเก็บตัวอย่าง ระบบจะสร้างตารางสถิติของค่าเฉลี่ยของสัญญาณเสียงไว้เป็นตารางอ้างอิง เมื่อระบบรับสัญญาณเสียงเข้ามาใหม่ จะวิเคราะห์สัญญาณเสียงตามวิธีการเดิม แล้วจะเปรียบเทียบผลวิเคราะห์กับตารางสถิติเพื่อหาความหมายที่เป็นไปได้มากที่สุด

โครงสร้างของข้อมูลของระบบที่ใช้ เป็นดังนี้

4.2.1 ระเบียนข้อมูล ใช้เก็บข้อมูลเสียงที่ระบบทำการสุ่มตัวอย่างเข้ามา แต่ละระเบียนแทนการสุ่มตัวอย่าง 1 ครั้ง การสุ่มตัวอย่างทำด้วยอัตรา 67.4 ครั้งต่อวินาที ขอบเขตของระเบียนข้อมูลเป็นดังรูปที่ 4.7

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

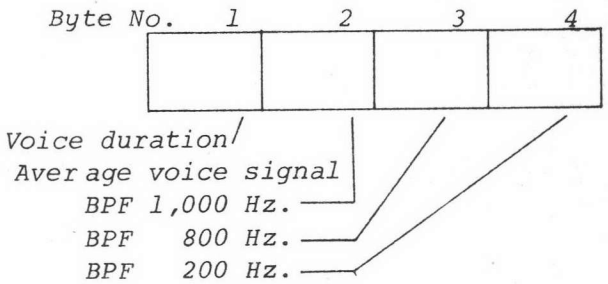


ไบท์ที่	ข้อมูล
1	ค่าที่อ่านจากวงจรตรวจจับการผ่านศูนย์
2	เวลา อ่านจากวงจรนับเวลา
3	ค่าสัญญาณจากวงจรกรองแถบความถี่ 1000 H_z
4	ค่าสัญญาณจากวงจรกรองแถบความถี่ 800 H_z
5	ค่าสัญญาณจากวงจรกรองแถบความถี่ 200 H_z
6	Flag 00 = ยังมีระเบียนต่อมา, FF = ระเบียนสุดท้าย

รูปที่ 4.7 แสดงระเบียนข้อมูลเสียง

หน่วยความจำที่ใช้เก็บข้อมูลเสียงเริ่มตั้งแต่แอดเดรส (address) $319E_{16}$ ความยาว 1,024 ไบท์

4.2.2 ระเบียนข้อมูลวิเคราะห์ ใช้เก็บผลการวิเคราะห์ค่าเฉลี่ยของสัญญาณเสียงจากแต่ละวงจรกรองแถบความถี่ และเก็บค่าระยะเวลาของการออกเสียงคำด้วย การหาค่าเฉลี่ยทำโดยหาผลรวมของสัญญาณเสียงจากแต่ละวงจรกรองแถบความถี่หารด้วยระยะเวลาของการออกเสียงคำ ๆ นั้น ระเบียนข้อมูลเป็นดังรูปที่ 4.8



$$\text{Average voice signal} = \frac{\Sigma(\text{voice signal})}{\text{Duration}}$$

รูปที่ 4.8 แสดงการจัดระเบียนข้อมูลวิเคราะห์

4.2.3 แฟ้มข้อมูลตัวอย่างเสียง ใช้เก็บผลการวิเคราะห์ (จาก 4.2.2) พร้อมความหมาย ความหมายจะได้จากข้อมูลที่ผู้พูดกดแป้นพิมพ์ แฟ้มข้อมูลแบ่งเป็น 2 ส่วน คือ

4.2.3.1 ส่วนควบคุมแฟ้มข้อมูล ยาว 6 ไบท์ ประกอบด้วย

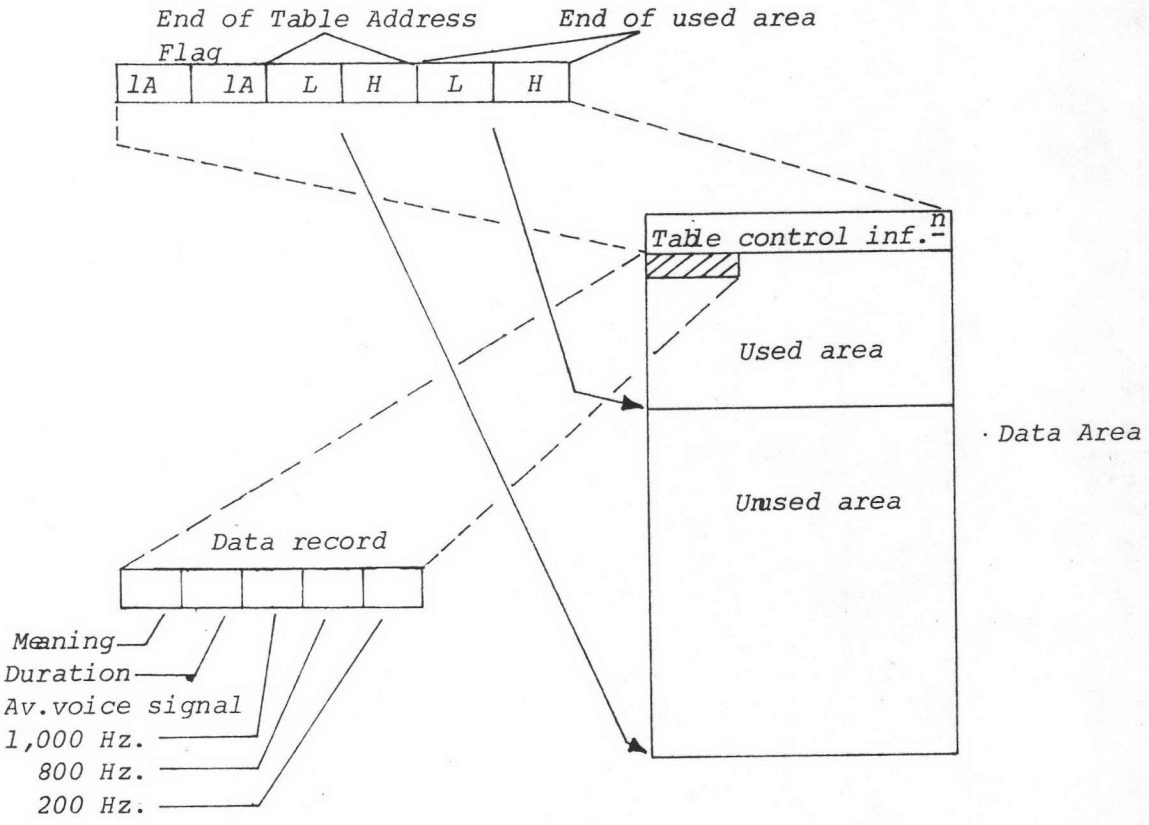
4.2.3.1.1 แฟล็ก (Flag) บอกว่าแฟ้มข้อมูลถูกฟอร์แมท (Format) เรียบร้อยแล้ว

4.2.3.1.2 แอดเดรสบอกการสิ้นสุดของแฟ้มข้อมูล

4.2.3.1.3 แอดเดรสบอกเขตสิ้นสุดที่แฟ้มข้อมูลถูกใช้ไป

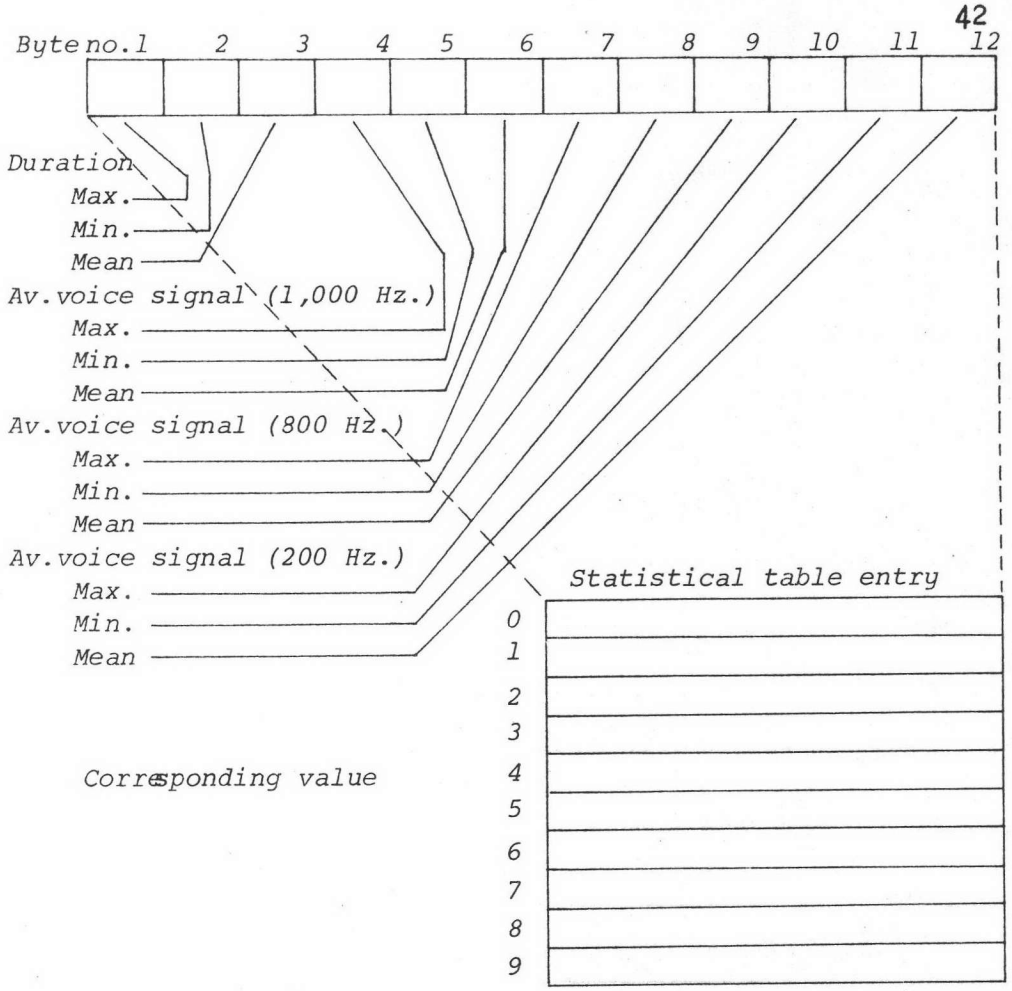
4.2.3.2 ส่วนเก็บระเบียนข้อมูล ยาว 5 ไบท์ ประกอบด้วยความหมายของเสียงตัวอย่าง และค่าจากระเบียนข้อมูลวิเคราะห์

ลักษณะของแฟ้มข้อมูลเป็นดังรูป 4.9 การนำระเบียนข้อมูลเข้าสู่แฟ้มข้อมูล ระบบจะหาระเบียนที่มีความหมายเดียวกัน ถ้าพบจะนำระเบียนใหม่แทรกไว้ข้างหน้า ระเบียนแรกที่มีความหมายตรงกัน ถ้าไม่พบจะแทรกไว้ที่ระเบียนแรกในส่วนข้อมูลของแฟ้มข้อมูล ลักษณะของแฟ้มข้อมูลจริงเป็นดังภาคผนวก



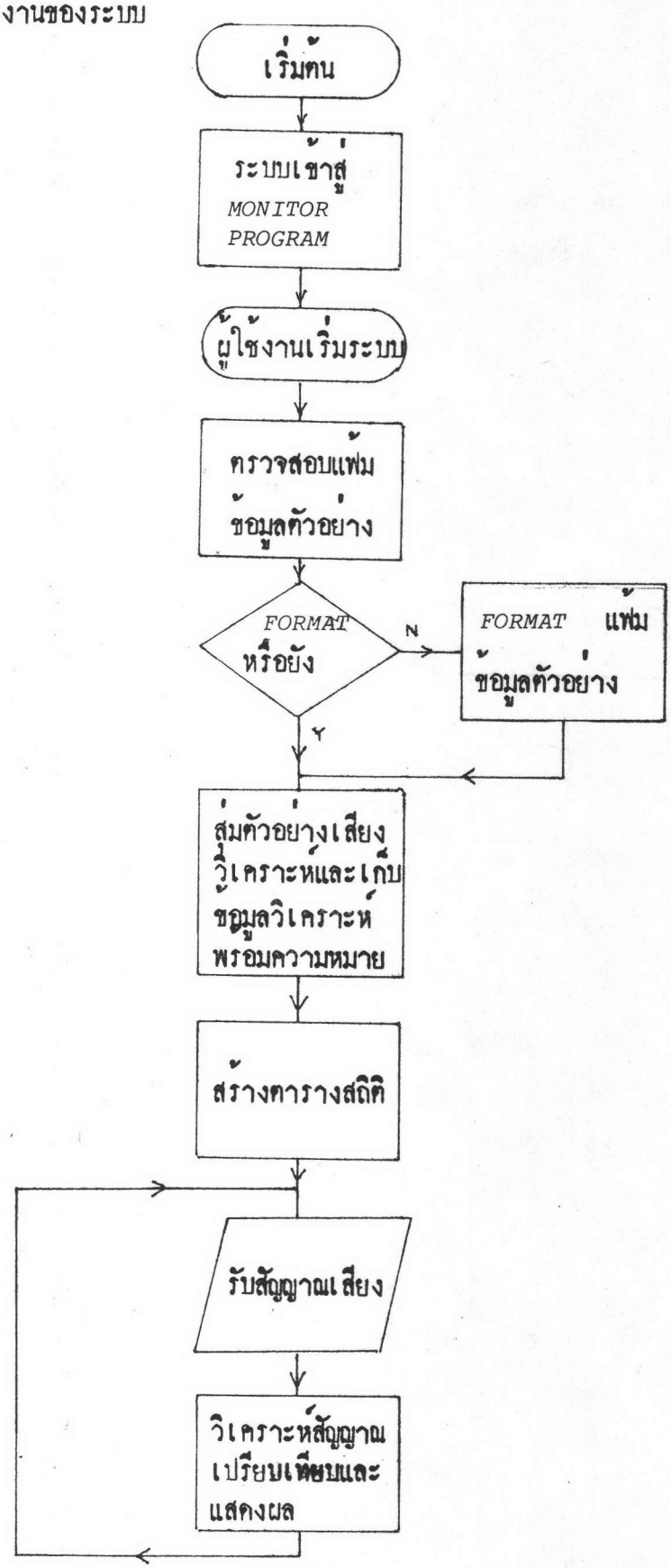
รูปที่ 4.9 แสดงแฟ้มข้อมูลตัวอย่างเสียงที่ถูกเก็บในหน่วยความจำ

4.2.4 แฟ้มข้อมูลตารางสถิติ ใช้เก็บข้อมูลสถิติจากแฟ้มข้อมูลตัวอย่างเสียง ระบบจะทำการคำนวณค่าสูงสุด ค่าต่ำสุด และค่าเฉลี่ยของค่าเฉลี่ยสัญญาณเสียงจากแต่ละแถบความถี่และระยะเวลาของการเกิดเสียงสำหรับแต่ละความหมาย ลักษณะของแฟ้มข้อมูลตารางสถิติเป็นดังรูปที่ 4.10



รูปที่ 4.10 แสดงตารางสถิติค่าเฉลี่ยของสัญญาณเสียงที่ปรากฏในแต่ละแถบความถี่ที่เก็บภายในหน่วยความจำ

4.3 ผังงานของระบบ



การทำงาน

เมื่อเริ่มเปิดเครื่อง ระบบจะเข้าสู่โปรแกรมมอนิเตอร์ (Monitor) ซึ่งจะทำหน้าที่ในการตรวจสอบหน่วยความจำ นำข้อมูลเข้าสู่และแสดงผลข้อมูลที่อยู่ภายในหน่วยความจำ ณ แอดเดรสที่กำหนด รวมทั้งสั่งให้ระบบทำงานตามโปรแกรมโดยเริ่มจากแอดเดรสใด ๆ ก็ได้ ระบบจะแสดงผล 18 00 X X ซึ่งเป็นแอดเดรสต่ำสุดของหน่วยความจำแบบ RAM ที่ติดตั้งอยู่กับระบบ

โปรแกรมของระบบจดจำจะเริ่มต้นที่ แอดเดรส 1000_{16} การเริ่มต้นระบบผู้ใช้งานจะต้องกดปุ่มดังนี้

ขั้นที่	ปุ่มที่กด	ค่าที่หน่วยแสดงผล
1	1 0	18 00 10
2	H	10 00

ขณะนี้เครื่องจะพร้อมทำงานตามโปรแกรมเริ่มต้นที่แอดเดรส 1000_{16}

3	E x c	L r n-

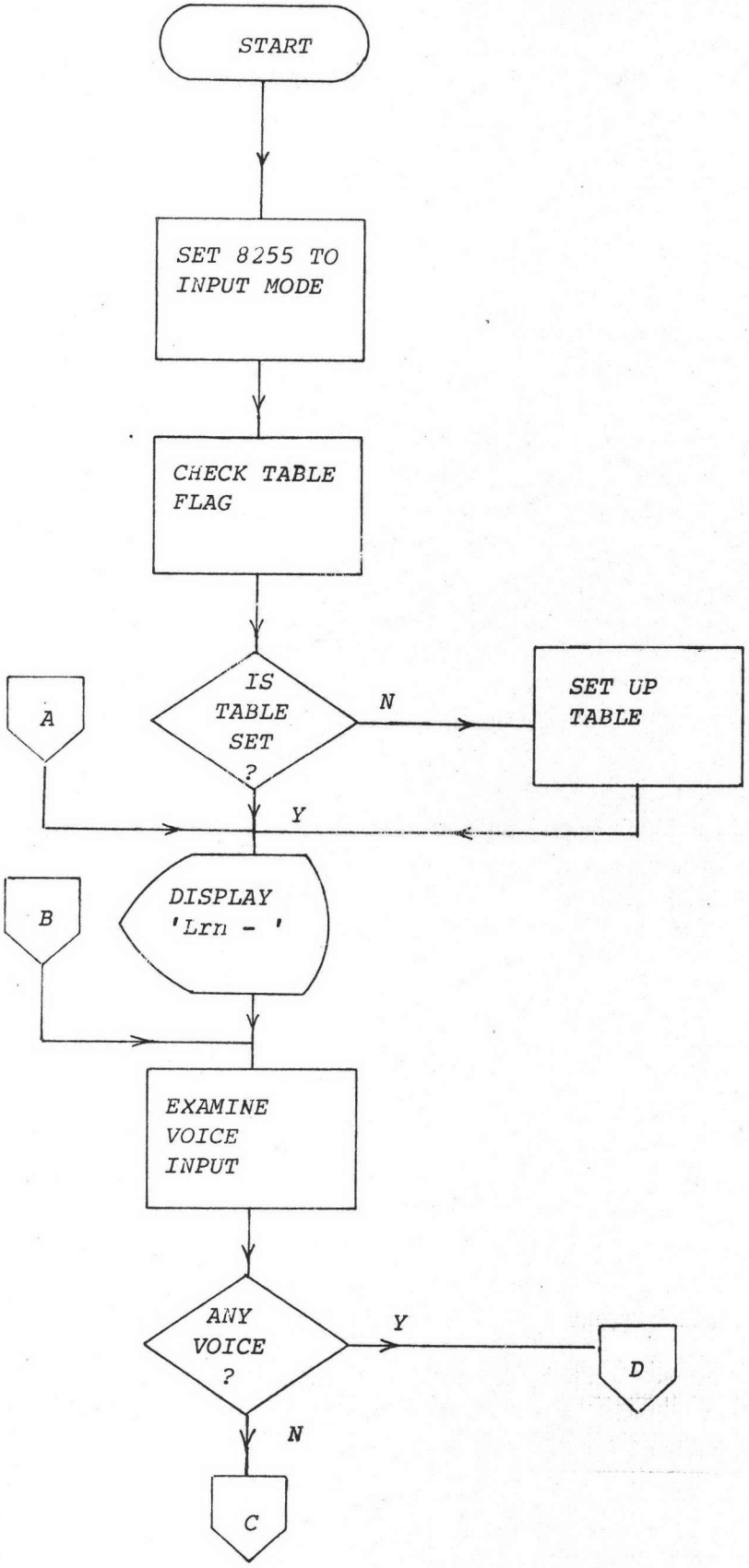
การกดปุ่ม E x จะทำให้เครื่องเริ่มทำงานตามโปรแกรมของระบบจดจำ ซึ่งจะแสดง L r n - แสดงว่าเครื่องอยู่ในภาวะเรียนรู้ ขณะนี้ระบบจะรอรับสัญญาณเสียงพูดจากไมโครโฟน เมื่อระบบรับเสียงพูด LED บนหน้าปัทม์ทั้ง 8 จะกระพริบ และจะหยุดกระพริบเมื่อสิ้นเสียง ระบบจะทำการวิเคราะห์หาค่าเฉลี่ยของสัญญาณเสียงและรอรับความหมาย 0 – 9 จากแป้นพิมพ์ ให้ผู้ใช้งานกดความหมายหนึ่ง แล้วกดปุ่ม RTN ระบบจะรับเอาความหมายและเก็บข้อมูลวิเคราะห์พร้อมความหมายไว้ในแฟ้มข้อมูลตัวอย่างเสียง เมื่อผู้ใช้งานให้ตัวอย่างเสียงเรียบร้อยแล้วให้กดปุ่มดังนี้

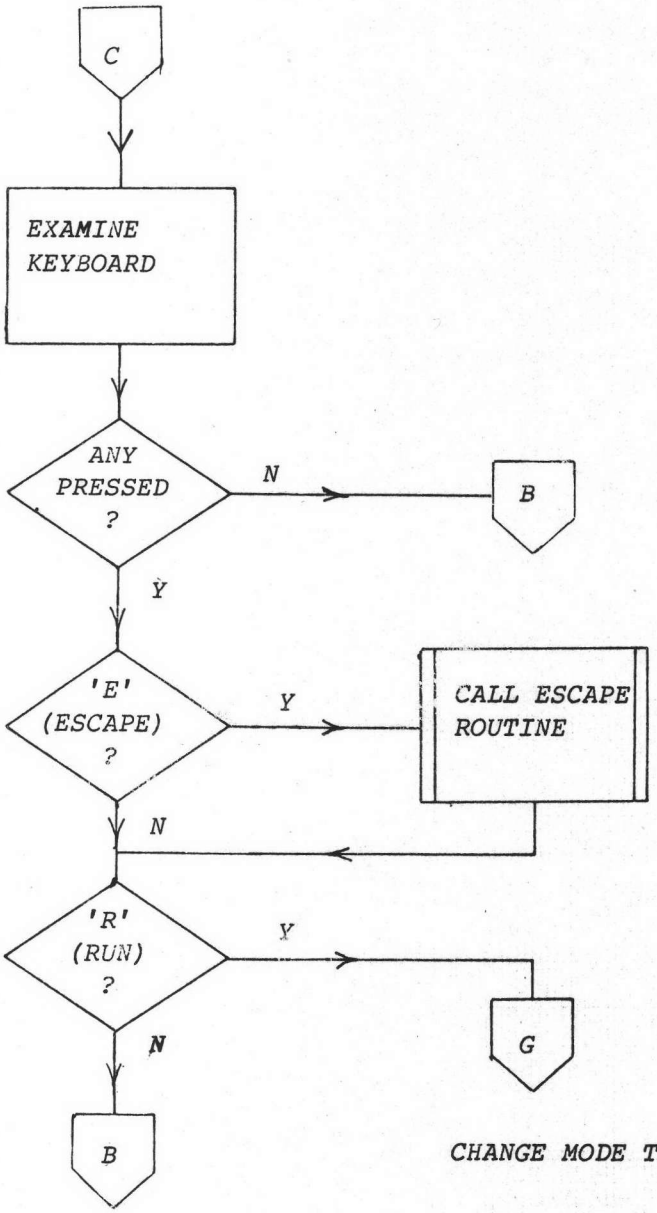
ขั้นที่	ปุ่มที่กด	ค่าที่หน่วยแสดงผล
4	R	–

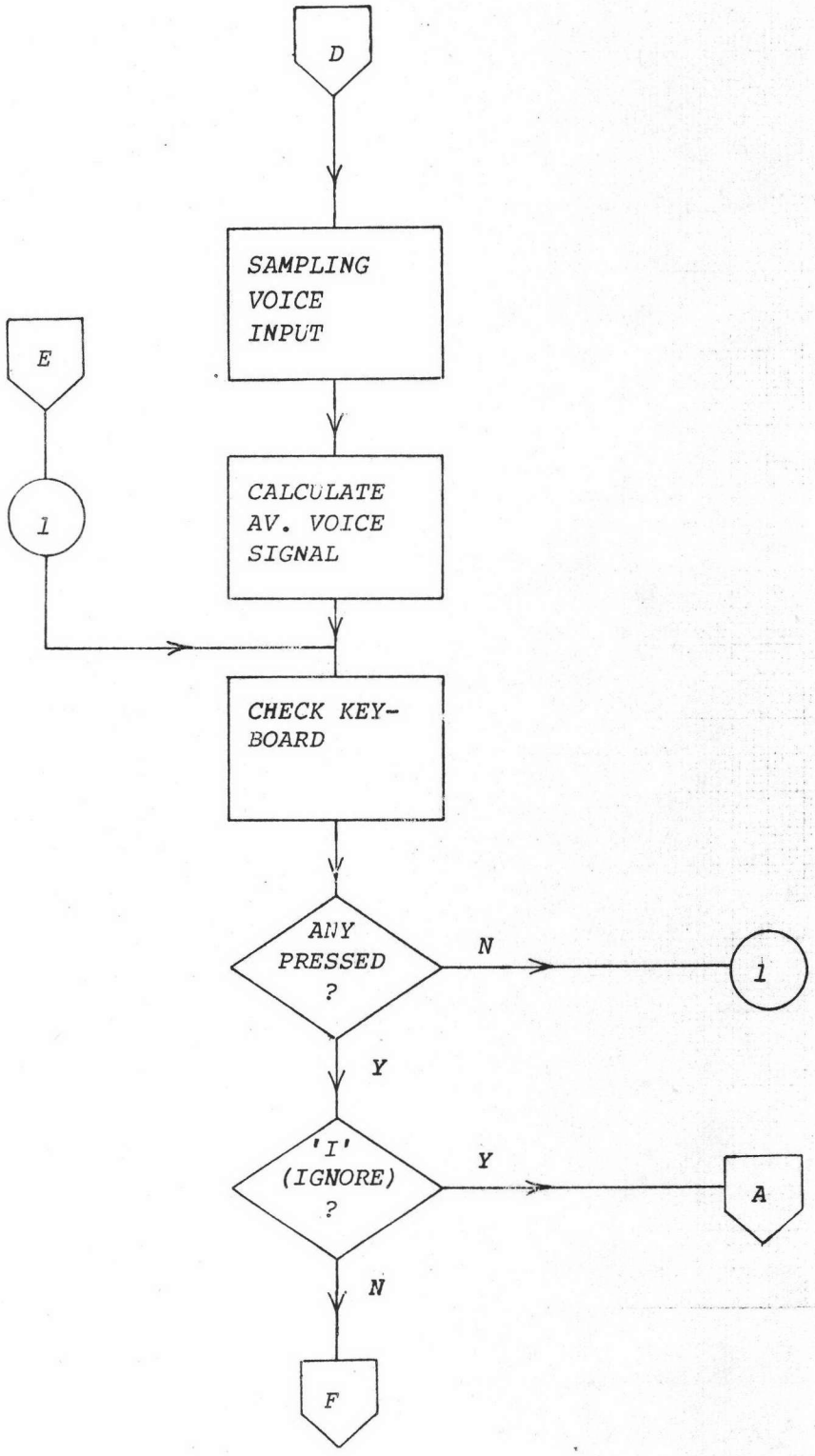
เมื่อกดปุ่ม R ระบบจะนำข้อมูลซึ่งเก็บไว้ในแฟ้มข้อมูลตัวอย่างเสียงมาสร้างตารางสถิติ จากนั้นระบบจะอยู่ในสภาพพร้อมรับสัญญาณเสียง เมื่อผู้ใช้งานพูดคำพูดระบบ

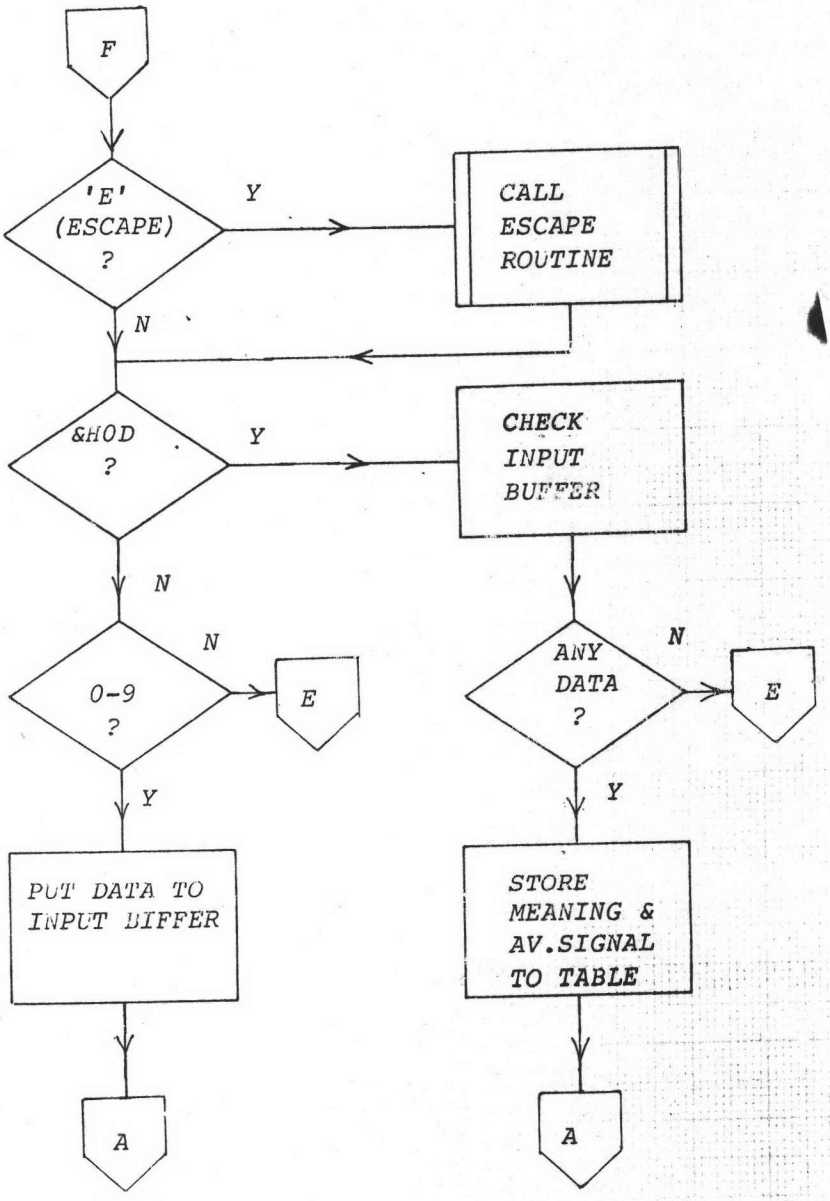
จะวิเคราะห์สัญญาณเสียง เปรียบเทียบกับตารางสถิติ แล้วแสดงผลลัพท์ที่หน่วยแสดงผล หลักที่ 6 เมื่อผู้ใช้งานกดแป้นพิมพ์ แป้นใดแป้นหนึ่ง ตัวเลขที่แสดงผลจะหายไป ซึ่งแสดงว่าระบบพร้อมที่จะรับสัญญาณเสียงใหม่

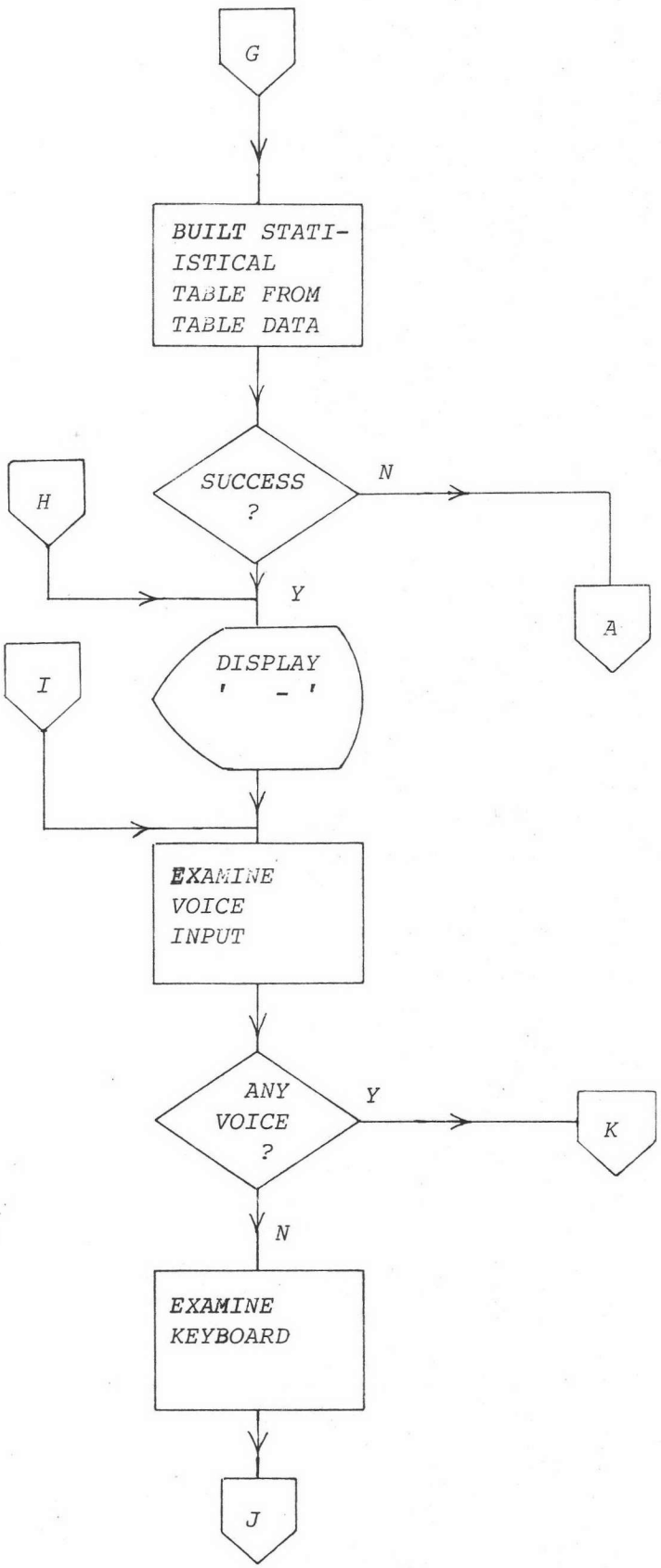
4.4 ผังงานของระบบจดจำ

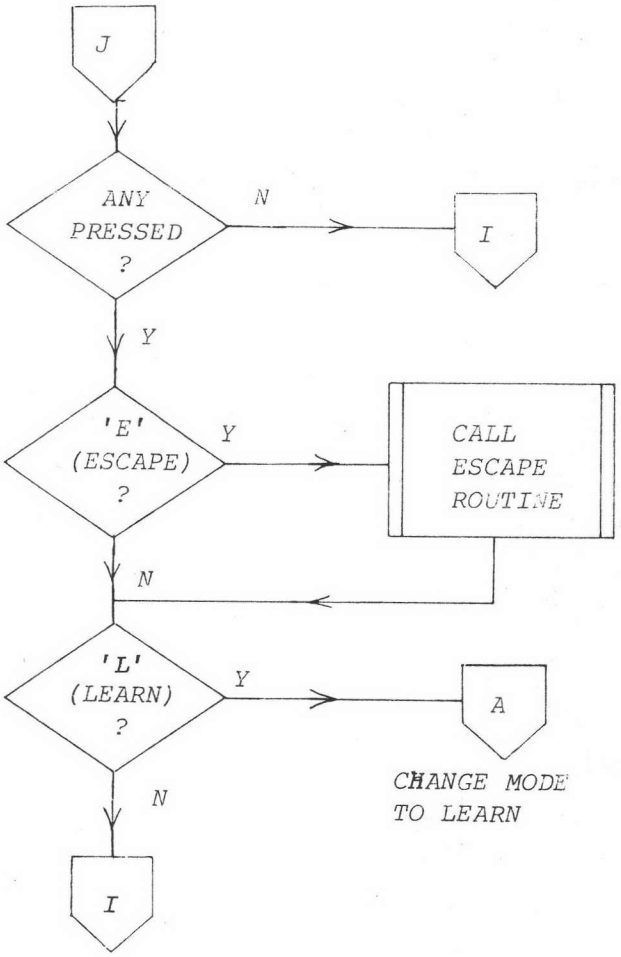


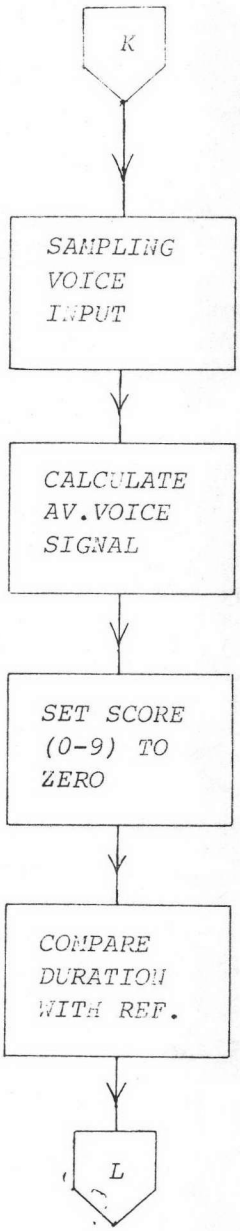

C
EXAMINE KEYBOARD
ANY PRESSED ?
N
B
Y
'E' (ESCAPE) ?
Y
CALL ESCAPE ROUTINE
N
'R' (RUN) ?
Y
G
N
B
CHANGE MODE TO DECODE


D
SAMPLING VOICE INPUT
E
1
CALCULATE AV. VOICE SIGNAL
CHECK KEY-BOARD
ANY PRESSED ?
N
1
Y
'I' (IGNORE) ?
Y
A
N
F

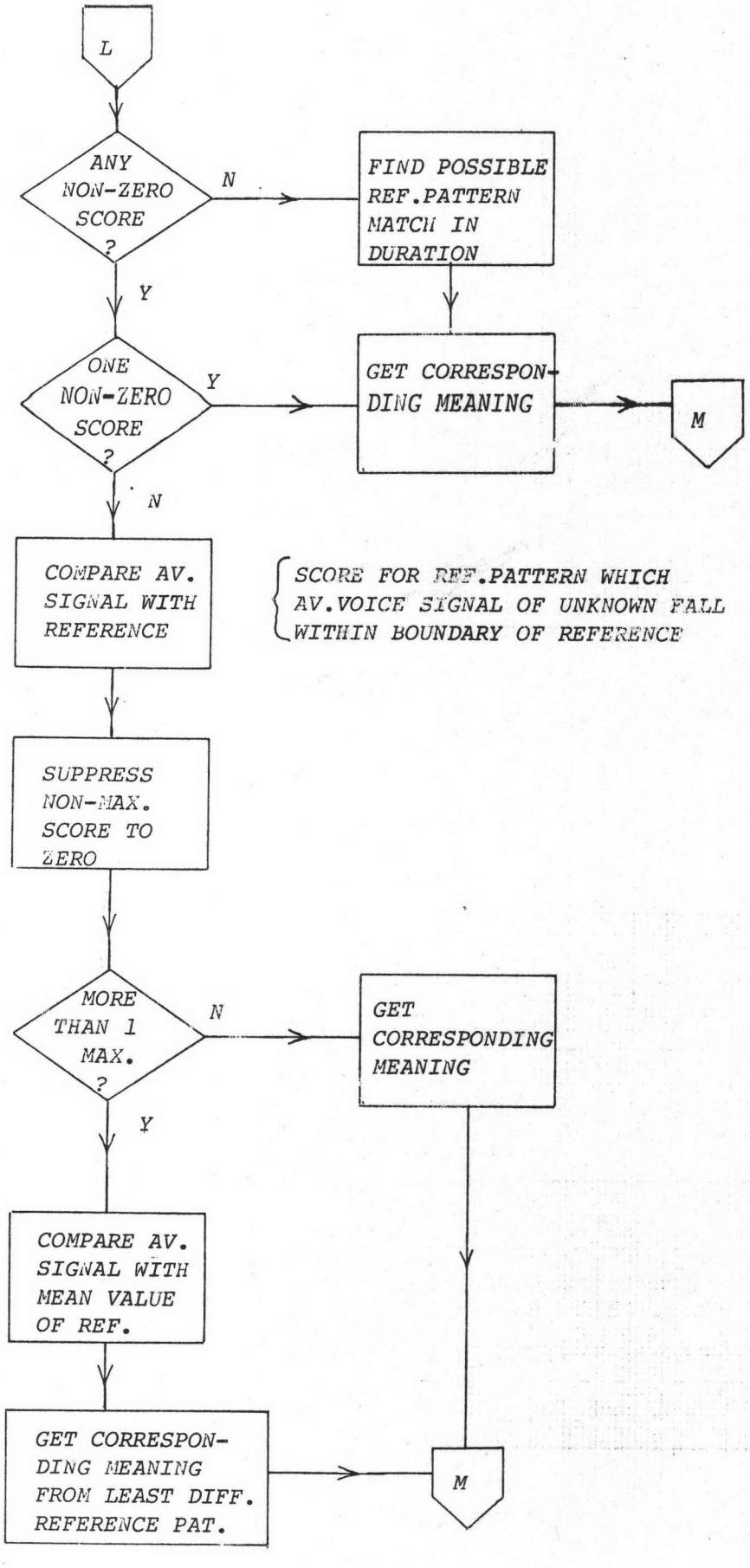

F
'E' (ESCAPE) ?
Y
CALL ESCAPE ROUTINE
N
&HOD ?
Y
CHECK INPUT BUFFER
N
0-9 ?
N
E
Y
PUT DATA TO INPUT BIFFER
A
ANY DATA ?
N
E
Y
STORE MEANING & AV.SIGNAL TO TABLE
A

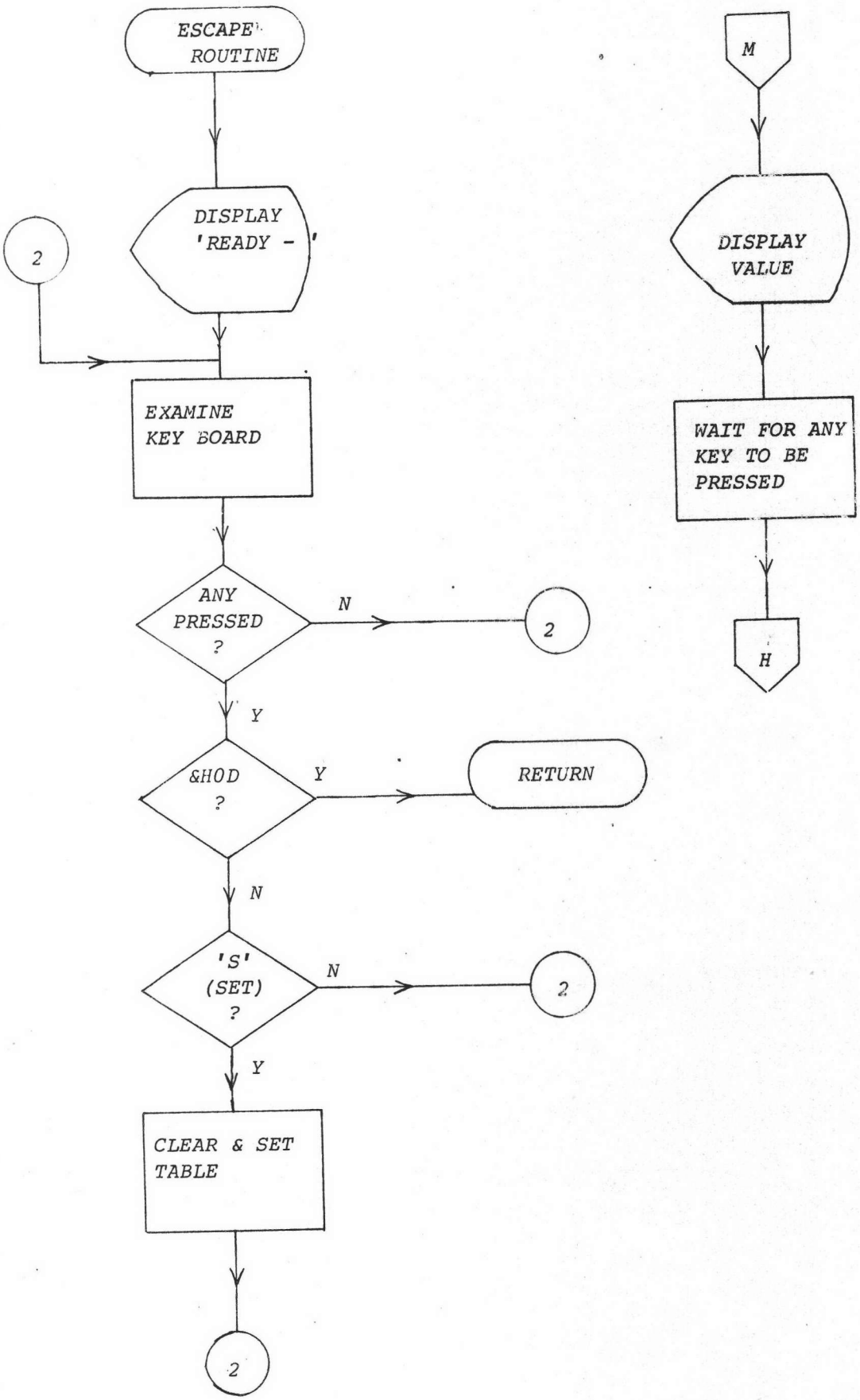

G
BUILT STATI-
ISTICAL
TABLE FROM
TABLE DATA
SUCCESS
?
N
H
Y
A
DISPLAY
' _ '
I
EXAMINE
VOICE
INPUT
ANY
VOICE
?
Y
K
N
EXAMINE
KEYBOARD
J





IF VALUE FALL WITHIN n th REFERENCE BOUNDARY THEN SCORE(n)=SCORE(n)+1


L
ANY NON-ZERO SCORE ?
N
FIND POSSIBLE REF.PATTERN MATCH IN DURATION
Y
ONE NON-ZERO SCORE ?
Y
GET CORRESPONDING MEANING
M
N
COMPARE AV. SIGNAL WITH REFERENCE
SCORE FOR REF.PATTERN WHICH AV.VOICE SIGNAL OF UNKNOWN FALL WITHIN BOUNDARY OF REFERENCE
SUPPRESS NON-MAX. SCORE TO ZERO
MORE THAN 1 MAX. ?
N
GET CORRESPONDING MEANING
Y
COMPARE AV. SIGNAL WITH MEAN VALUE OF REF.
GET CORRESPONDING MEANING FROM LEAST DIFF. REFERENCE PAT.
M


ESCAPE ROUTINE
DISPLAY 'READY -'
2
EXAMINE KEY BOARD
ANY PRESSED ?
N
2
Y
&HOD ?
Y
RETURN
N
'S' (SET) ?
N
2
Y
CLEAR & SET TABLE
2
M
DISPLAY VALUE
WAIT FOR ANY KEY TO BE PRESSED
H