บทที่ 4

การดำเนินการวิจัย

4.1 คัดเลือกค่าลักษณะเด่น

ในการพัฒนาระบบรู้จำเสียงพูดภาษาต่างๆ จะต้องมีการปรับค่าพารามิเตอร์บางอย่างให้เหมาะสมกับภาษานั้นๆ ค่าลักษณะเด่นเป็นสิ่งที่แสดงถึงลักษณะพิเศษของเสียงแต่ละเสียงซึ่งแต่ละภาษาก็มีความแตกต่างกัน จึงต้องมีการคัดเลือกค่าลักษณะเด่นที่เหมาะสมกับการรู้จำภาษาไทย

ได้ออกแบบการทดลอง เริ่มจากการสร้างแบบจำลองเสียงด้วย

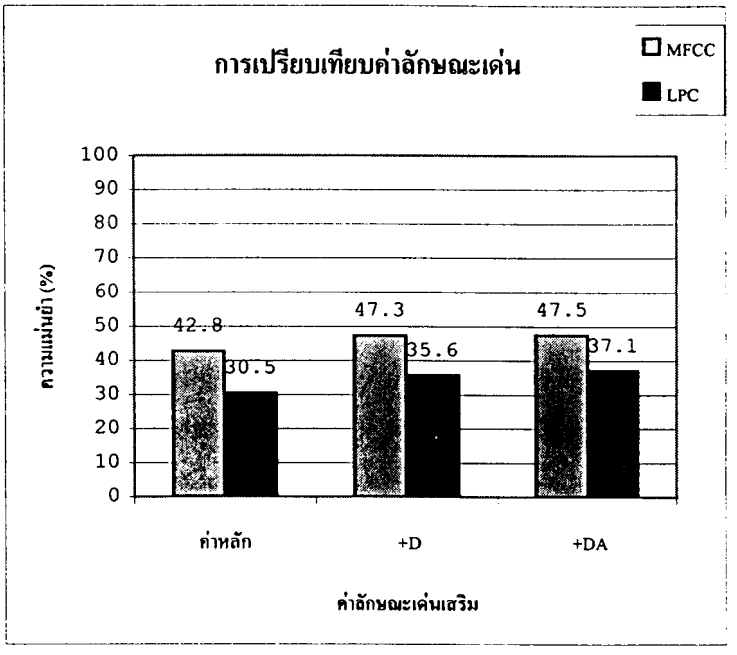
- ข้อมูลฝึกฝน
- ชุดหน่วยเสียง $C_i+V_t+C_f$
- โครงสร้างแบบจำลองเสียง 5 สเตท 3 มิกซเจอร์
- ค่าลักษณะเด่นที่นำมาเปรียบเทียบแบ่งเป็น 2 กลุ่มคือกลุ่มของ MFCC และ LPC แต่ละกลุ่มมีรายละเอียดปลีกย่อยกลุ่มละ 3 แบบคือ
    1. ค่าลักษณะเด่นเพียงอย่างเดียว (MFCC/LPC)
    2. เพิ่มค่าอนุพันธ์อันดับที่ 1 (MFCC+D/LPC+D)
    3. เพิ่มค่าอนุพันธ์อันดับที่ 1 และค่าอนุพันธ์อันดับที่ 2 (MFCC+DA/LPC+DA)

ซึ่งจะได้แบบจำลองเสียงรวมทั้งสิ้น 6 แบบตามค่าลักษณะเด่นที่แตกต่างกัน จากนั้นจึงทำการหาค่าความแม่นยำของแบบจำลองเสียงแต่ละแบบโดยวิธีการจำแนกหน่วยเสียงด้วยข้อมูลทดสอบ

การทดลองเริ่มจากการกำหนด การคำนวณค่าลักษณะที่ความยาวเฟรม 25 mS ระยะระหว่างเฟรม 10 mS โครงสร้าง HMM แบบ Diagonal covariance, Plain model, 1 Stream 36 element feature vector ด้วยขั้นตอนการ initialize 20 ครั้ง single re-estimate 20 ครั้ง และ embedded re-estimate 40 ครั้ง ส่วนการทดสอบแบบจำลองเสียงในส่วนการเปรียบเทียบผลได้ทำลบผลการรู้จำที่เป็นเสียงเงียบออกไปเพื่อทำการเปรียบเทียบเฉพาะข้อมูลที่มีเสียงเท่านั้น

สำหรับค่าอื่นๆ ที่ไม่ได้กล่าวถึงเป็นค่าปริยายของโปรแกรมรู้จำเสียงพูด HTK และวิธีการปรับตั้ง (Setup) ระบบรู้จำเสียงพูดทั้งการสร้างแบบจำลองเสียง และ การทดสอบระบบรู้จำ มีรายละเอียดแสดงอยู่ใน ภาคผนวก

ผลการทดลอง



ภาพประกอบ 4.1 กราฟเปรียบเทียบความแม่นยำของค่าลักษณะเด่นแบบต่างๆ

วิเคราะห์ผลการทดลอง

จากผลการทดลองในภาพประกอบ 4.1 พบว่าค่าคุณลักษณะเด่นที่ทำให้แบบจำลองเสียงมีประสิทธิภาพสูงสุดคือ MFCC+DA หากเปรียบเทียบกับค่าลักษณะเด่น MFCC+D พบว่ามีความแม่นยำน้อยกว่าเพียง 0.1% ซึ่งไม่มีนัยสำคัญ สรุปได้ว่าค่าลักษณะเด่นทั้งสองแบบที่มีความเหมาะสมที่จะนำมาใช้ในการสร้างแบบจำลองเสียง จึงกำหนดให้ค่าลักษณะเด่นแบบ MFCC+DA เป็นค่าที่เหมาะสมในการสร้างแบบจำลองเสียงพูด

4.2 คัดเลือกโครงสร้าง HMM

การคัดเลือกโครงสร้าง HMM ที่เหมาะสมจะต้องสร้างแบบจำลองเสียงที่มีโครงสร้าง HMM ที่แตกต่างกัน แล้วทำมาหาค่าความแม่นยำโดยวิธีการจำแนกหน่วยเสียง เพื่อคัดเลือกโครงสร้าง HMM ที่เหมาะสมสำหรับแบบจำลองเสียงพูดภาษาไทย

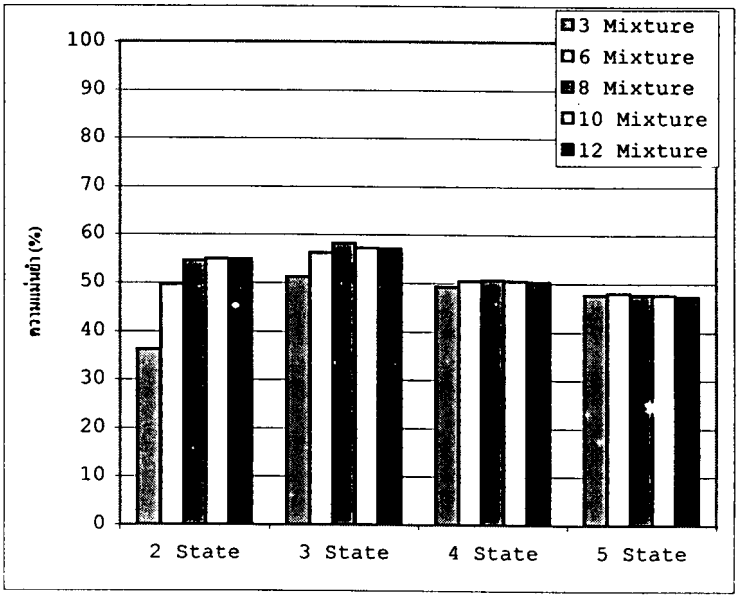
การทดลองจะเริ่มจากการสร้างแบบจำลองเสียงด้วย

- ข้อมูลฝึกฝน
- ชุดหน่วย $C_i+V_t+C_f$

- ค่าลักษณะเด่นแบบ MFCC+DA
- ลักษณะของแบบจำลองที่นำมาคัดเลือกมีจำนวนสเตทตั้งแต่ 2-5 และจำนวนมิกซเจอร์ตั้งแต่ 3-12

ซึ่งจะได้แบบจำลองเสียงจำนวนหนึ่งตามโครงสร้าง HMM ที่แตกต่างกัน หลังจากนั้นจึงนำแบบจำลองเสียงทั้งหมดมาหาค่าความแม่นยำ โดยวิธีการจำแนกหน่วยเสียงด้วยข้อมูลทดสอบ (รูปแบบการจำแนกตามชุดหน่วยเสียงของแบบจำลองเสียงที่ทดสอบ)

ผลการทดลอง



ภาพประกอบ 4.2 กราฟเปรียบเทียบความแม่นยำของชุดโครงสร้าง HMM

วิเคราะห์ผลการทดลอง

ผลการทดลองจากภาพประกอบ 4.2 พบว่าลักษณะของแบบจำลองเสียงแบบ 3 สเตท 8 มิกซเจอร์ มีความแม่นยำสูงสุดคือ 58.3% ซึ่งระดับความแม่นยำของสเตทที่ 3 มีค่าสูงกว่าเมื่อเทียบกับสเตทอื่นที่จำนวนมิกซเจอร์เท่ากัน ดังนั้นงานวิจัยนี้จึงได้เลือกโครงสร้าง HMM แบบ 3 สเตท 8 มิกซเจอร์เป็นโครงสร้างจะใช้ในการสร้างแบบจำลองเสียงพูดต้นแบบ

4.3 คัดเลือกชุดหน่วยเสียง

ชุดหน่วยเสียงเป็นส่วนประกอบสำคัญในการพัฒนาแบบจำลองเสียง ชุดหน่วยเสียงจึงจำเป็นต้องมีหน่วยเสียงที่ครอบคลุมเสียงทั้งหมดในภาษาไทยได้แก่ เสียงพยัญชนะ สระ ตัวสะกด วรรณยุกต์ ซึ่งการทดลองจะเริ่มจากการสร้างแบบจำลองเสียงด้วย

- ข้อมูลฝึกฝน
- โครงสร้างHMM 3 สเตท 8 มิกซเจอร์
- ค่าลักษณะเด่นแบบ MFCC+DA
- ชุดหน่วยเสียงมี 2 ชุดคือ ชุดหน่วยเสียงพยัญชนะ สระวรรณยุกต์ ตัวสะกด ($C_i+V_t+C_f$) และ ชุดหน่วยเสียงพยัญชนะ สระวรรณยุกต์ ตัวสะกดวรรณยุกต์ ($C_i+V_t+C_{ft}$)

จากนั้นจะได้แบบจำลองเสียง 2 แบบ ตามชุดหน่วยเสียงที่ต่างกัน หลังจากนั้นจึงหาค่าความแม่นยำของแบบจำลองเสียงทั้งสอง โดยวิธีการจำแนกหน่วยเสียง ด้วยข้อมูลทดสอบ เนื่องจากค่าความแม่นยำที่ได้เป็นความสามารถในการจำแนกหน่วยเสียงต่างชุดกันไม่เหมาะสมในการเปรียบเทียบกัน จึงได้กำหนดรูปแบบการจำแนกสำหรับชุดหน่วยเสียง $C_i+V_t+C_f$ ให้ส่วนตัวสะกดวรรณยุกต์ มีค่าเทียบเท่ากับตัวสะกด ในชุดหน่วยเสียง $C_i+V_t+C_{ft}$ ดังตาราง 4.1

ชุดหน่วยเสียง $C_i+V_t+C_f$	ชุดหน่วยเสียง $C_i+V_t+C_{ft}$
-น	-น (สามัญ), -น (เอก) , -น (โท) , -น (ตรี) , -น (จัตวา)
-ง	-ง (สามัญ) , -ง (เอก) , -ง (โท) , -ง (ตรี) , -ง (จัตวา)

ตาราง 4.1 วิธีการปรับชุดหน่วยเสียง $C_i+V_t+C_{ft}$ ให้ตรงกับชุดหน่วยเสียง $C_i+V_t+C_f$

ผลการทดลอง

ชุดหน่วยเสียง	ค่าความแม่นยำ (%)
$C_i+V_t+C_f$	58.3
$C_i+V_t+C_{ft}$	54.5

ตาราง 4.2 แสดงค่าความแม่นยำของชุดหน่วยเสียง (%)

วิเคราะห์ผลการทดลอง

ผลการทดลองจากตาราง 4.2 ชุดหน่วยเสียง $C_i+V_t+C_f$ มีความแม่นยำสูงกว่าชุดหน่วยเสียง $C_i+V_t+C_f$ 4% ซึ่งความแตกต่างนี้ถือว่ามีค่าสูง และก่อนการทดลองนี้ได้ตั้งสมมติฐานว่าตัวสะกดมีส่วนช่วยในการจำแนกวรรณยุกต์ได้แต่ผลการทดลองทำให้ทราบว่า การกำหนดตัวสะกดเพื่อช่วยในการจำแนกวรรณยุกต์นั้นทำให้ความแม่นยำโดยรวมลดลง ดังนั้นจึงกำหนดให้ชุดหน่วยเสียง $C_i+V_t+C_f$ เป็นชุดหน่วยเสียงที่เหมาะสมในการสร้างแบบจำลองเสียงพูด

เมื่อพิจารณากลุ่มตัวสะกด พบว่าการสร้างแบบจำลองเสียงโดยกำหนดหน่วยเสียงตัวสะกดแบบแยกตามกลุ่มวรรณยุกต์นั้น จะมีประสิทธิภาพด้อยกว่า การกำหนดหน่วยเสียงตัวสะกดแบบไม่ขึ้นกับวรรณยุกต์

4.4 สร้างและหาค่าความแม่นยำของแบบจำลองเสียงต้นแบบ

เมื่อคัดเลือกสิ่งที่เหมาะสมสำหรับการสร้างแบบจำลองเสียงได้แล้ว ต่อมาจึงเป็นการสร้างแบบจำลองเสียงต้นแบบโดยใช้ปัจจัยแวดล้อมที่ได้คัดเลือกแล้ว ได้แก่

- ข้อมูลฝึกฝน
- โครงสร้างHMM 3 สเตท 8 มิกซเจอร์
- ค่าลักษณะเด่นแบบ MFCC+DA
- ชุดหน่วยเสียง $C_i+V_t+C_f$

ได้แบบจำลองเสียงต้นแบบ เพื่อนำมาหาค่าความแม่นยำด้วยการจำแนกหน่วยเสียงโดยแบ่งการจำแนกออกเป็นกลุ่มๆได้แก่ กลุ่มพยัญชนะ กลุ่มสระ กลุ่มตัวสะกด และรวมทุกกลุ่ม เพื่อแสดงให้เห็นถึงจุดอ่อนและจุดแข็งของแบบจำลอง และสามารถนำไปปรับปรุงแก้ไขได้ตรงจุด ในการจำแนกหน่วยเสียงตามกลุ่มย่อยนั้นได้จัดกลุ่มข้อมูลทดสอบเป็น 3 กลุ่มตามกลุ่มหน่วยเสียง แล้วจึงนำข้อมูลแต่ละกลุ่มเข้ากระบวนการจำแนกเพื่อให้ได้ค่าความแม่นยำในการจำแนกหน่วยเสียงของแต่ละกลุ่มออกมา

นอกจากนี้ยังมีการพิจารณาผลการจำแนกรายหน่วยเสียงในแต่ละกลุ่ม โดยการนับจำนวนหน่วยเสียงที่จำแนกได้อย่างถูกต้องและไม่ถูกต้อง ทำให้ทราบว่าหน่วยเสียงใดที่ระบบจำแนกผิดพลาดบ่อยครั้ง หน่วยเสียงใดที่ระบบสามารถจำแนกได้อย่างแม่นยำ ซึ่งผลการจำแนกรายหน่วยเสียงนี้อยู่ในตารางแสดงความสับสนในการจำแนกหน่วยเสียง (Confusion Matrix) ภายในตารางหน่วยเสียงในแนวตั้งคือข้อมูลที่ระบบรู้จำได้ หน่วยเสียงในแนวนอนคือข้อมูลที่ถูกต้อง ค่าในตารางคือจำนวนหน่วยที่ระบบจำแนกได้ ตัวอย่างค่า 38 ในแถว l คอลัมน์ r ของตาราง 4.4 หมายความ

ว่า ระบบจำแนกได้ว่าเป็นหน่วยเสียง l หรือ ล แต่หน่วยเสียงที่ถูกต้องคือหน่วยเสียง r หรือ ร มีจำนวนการจำแนกที่เป็นเช่นนี้ 38 หน่วยเสียง ซึ่งก็คือระบบจำแนกผิดจาก ร เป็น ล จำนวน 38 หน่วยเสียง

ผลการทดลอง

กลุ่มหน่วยเสียง	ค่าความแม่นยำ(%)
พยัญชนะ	75.3
สระ	39.6
ตัวสะกด	77.6
รวม	58.3

ตาราง 4.3 แสดงค่าความแม่นยำการจำแนกกลุ่มหน่วยเสียง

	l	m	n	ng	p	ph	phl	phr	pl	pr	r	s	t	th	thr	tr	w	z
l	70	0	5	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1
m	5	74	6	3	0	0	0	0	0	0	0	0	0	0	0	0	2	1
n	3	5	82	9	2	0	0	0	0	0	2	1	0	0	0	0	1	0
ng	0	1	2	9	0	0	0	0	0	0	0	0	0	0	0	0	2	1
p	0	0	0	0	16	0	0	0	0	1	1	0	3	0	0	0	1	1
ph	0	0	0	1	0	33	0	0	0	0	0	0	0	18	0	0	0	1
phl	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0
phr	0	0	0	0	0	4	0	3	0	0	1	1	0	0	0	0	0	0
pl	1	0	0	0	1	0	0	0	0	2	1	0	0	0	0	0	0	0
pr	2	0	0	0	0	1	0	0	1	4	6	0	0	0	0	0	0	0
r	38	0	6	0	0	1	0	0	0	6	39	0	2	0	0	1	1	0
s	1	1	0	0	0	0	0	0	0	0	0	120	0	0	0	0	0	0
t	1	0	0	0	3	1	0	0	0	0	0	1	22	1	0	0	0	0
th	0	0	0	0	0	4	0	1	0	0	0	3	1	94	0	0	0	0
thr	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tr	3	0	0	0	0	0	0	0	0	6	6	0	3	0	0	0	0	0
w	2	3	0	1	0	1	0	0	0	0	0	0	0	0	0	0	34	1
z	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	5	24

ตาราง 4.4 แสดงความสับสนในการจำแนกหน่วยเสียงกลุ่มพยัญชนะ

	a0	a1	a2	a3	a4	aa0	aa1	aa2	aa3	aa4
a0	58	13	9	15	0	16	7	4	2	2
a1	8	61	2	27	1	2	12	2	2	5
a2	15	1	11	4	0	1	3	1	0	3
a3	10	21	1	19	1	6	9	2	2	1
a4	4	4	0	2	2	1	0	0	1	1
aa0	5	5	0	8	3	79	29	6	2	20
aa1	3	6	0	2	1	15	13	5	1	6
aa2	5	2	0	4	0	32	21	13	3	9
aa3	0	2	0	0	0	5	1	2	1	2
aa4	2	3	0	0	0	9	7	3	0	24

ตาราง 4.5 แสดงความสับสนในการจำแนกหน่วยเสียงสระ อะ และ สระ อา

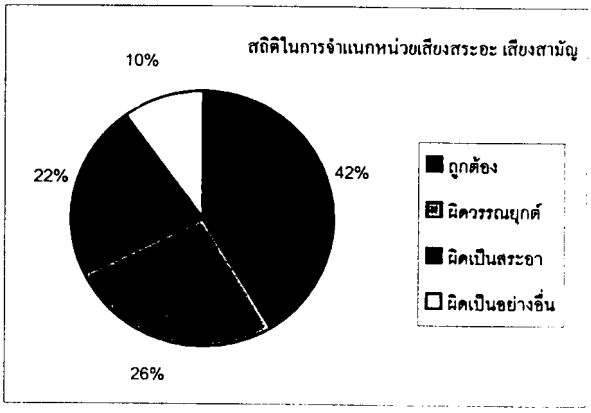
	ch_	f_	j_	k_	l_	m_	n_	ng_	p_	s_	t_	w_
ch_	0	0	0	0	0	0	0	0	0	0	1	0
f_	0	1	0	0	0	0	0	0	0	0	1	0
j_	0	0	168	1	0	1	4	7	0	0	0	8
k_	0	3	1	40	0	2	0	1	1	0	7	5
l_	0	0	0	0	0	0	0	0	0	0	0	0
m_	0	0	0	1	0	100	17	12	0	0	1	2
n_	0	0	6	6	0	12	157	48	2	0	18	3
ng_	0	2	1	5	0	4	15	133	0	0	0	6
p_	0	9	1	12	0	0	1	0	36	0	5	5
s_	0	0	0	0	0	0	0	0	0	4	1	0
t_	0	6	3	20	0	1	3	0	3	0	64	1
w_	0	0	0	1	0	0	1	1	0	0	0	45

ตาราง 4.6 แสดงความสับสนในการจำแนกหน่วยเสียงตัวสะกด

วิเคราะห์ผลการทดลอง

ผลการทดลอง จากตาราง 4.3 เมื่อเปรียบเทียบความแม่นยำของกลุ่มหน่วยเสียงทั้งสาม พบว่า กลุ่มสระมีประสิทธิภาพต่ำที่สุด ทำให้ความแม่นยำรวมมีค่าเพียง 58.3%

เมื่อพิจารณาความสับสนในการจำแนกสระจากตาราง 4.5 พบว่าการจำแนกวรรณยุกต์ในสระเดียวกันมีความผิดพลาดค่อนข้างสูง ตัวอย่างเช่น สระ อะ (a0, a1, a2, a3, a4) ที่เป็นเสียงสามัญมีจำนวน 140 หน่วย จำแนกได้ถูกต้อง 58 หน่วย จำแนกผิดวรรณยุกต์ 37 หน่วย จำแนกเป็นสระอา 31 หน่วย จำแนกผิดเป็นสระอื่น 14 หน่วย ดังภาพประกอบ 4.3



ภาพประกอบ 4.3 แผนภาพแสดงสถิติในการจำแนก สระ อะ เสียงสามัญ

สำหรับกลุ่มพยัญชนะและตัวสะกด ในตาราง 4.4 และตาราง 4.6 การจำแนกที่ผิดพลาดจะพบในหน่วยเสียงที่มีการออกเสียงคล้ายคลึงกันซึ่งมีจำนวนไม่มาก เช่น การจำแนกพยัญชนะ พ เป็นพยัญชนะ ท การจำแนกพยัญชนะ ร เป็นพยัญชนะ ล การจำแนกตัวสะกด น เป็นตัวสะกด ง การจำแนกตัวสะกด ด เป็นตัวสะกด ก

4.5 วิเคราะห์ความสามารถในการรู้จำหน่วยเสียง พยางค์ และคำศัพท์

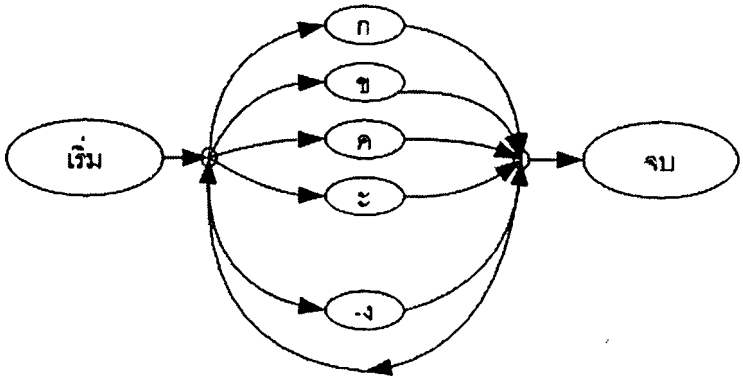
ขั้นตอนนี้เป็นการหาค่าความแม่นยำและค่าความถูกต้องของการรู้จำในมุมมองที่แตกต่างกันตั้งแต่ การรู้จำหน่วยเสียง การรู้จำพยางค์ การรู้จำคำศัพท์ ของแบบจำลองเสียงต้นแบบ ในการทดลองนี้ได้นำแบบจำลองเสียงต้นแบบมาหาค่าความแม่นยำด้วยวิธีการรู้จำเสียงพูด

โดยใช้ข้อมูลต่อไปนี้

- ข้อมูลทดสอบ
- แบบจำลองเสียงต้นแบบ
- รูปแบบการรู้จำ 4 แบบได้แก่
    1. การรู้จำหน่วยเสียง
    2. การรู้จำหน่วยเสียงตามลำดับของพยางค์
    3. การรู้จำพยางค์
    4. การรู้จำคำศัพท์ (เป็นคำศัพท์บางส่วน)

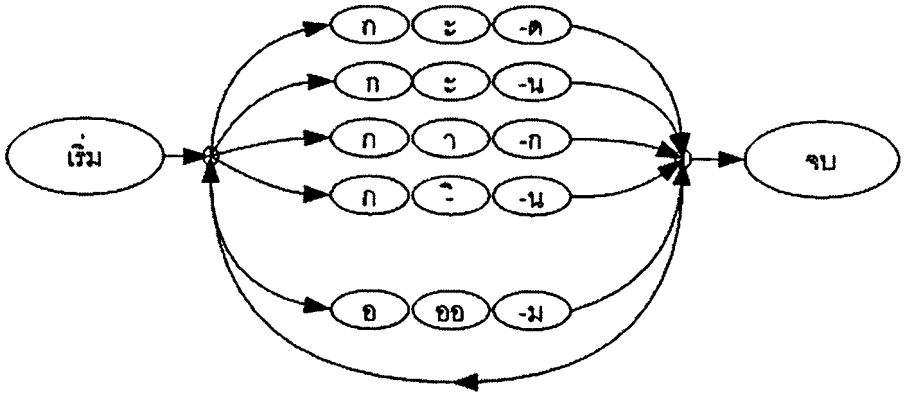
การรู้จำแต่ละแบบมีรายละเอียดดังต่อไปนี้

1. การรู้จำหน่วยเสียง กำหนดผลการรู้จำให้ทุกหน่วยเสียงมีโอกาสเกิดขึ้นอย่างไม่จำกัดในแต่ละประโยค ดังภาพประกอบ 4.4 โดยใช้เทคนิคการกำหนดพจนานุกรมเพื่อการรู้จำหน่วยเสียง



ภาพประกอบ 4.4 รูปแบบการรู้จำหน่วยเสียง

2. การรู้จำหน่วยเสียงตามลำดับของพยางค์ (หน่วยเสียง 2) กำหนดผลการรู้จำให้ลำดับของหน่วยเสียงที่เกิดขึ้นตรงตามกฎการสร้างพยางค์คือหน่วยเสียงที่เกิดขึ้นจะต้องประกอบด้วย พยัญชนะ สระ ตัวสะกด ตามลำดับดังแสดงในภาพประกอบ 4.5 โดยไม่จำกัดจำนวนชุดพยางค์ และใช้พจนานุกรมหน่วยเสียง

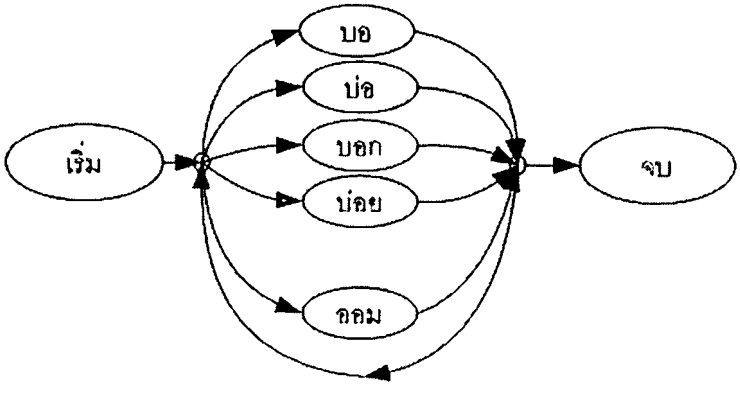


ภาพประกอบ 4.5 รูปแบบการรู้จำหน่วยเสียงตามลำดับของพยางค์

3. การรู้จำพยางค์ กำหนดให้ผลการรู้จำอยู่ในรูปแบบของพยางค์ ด้วยพจนานุกรมพยางค์ ดังตาราง 4.7 โดยมีรูปแบบการรู้จำไม่จำกัดพยางค์ดังภาพประกอบ 4.6 วิธีนี้ได้นำมาจากเทคนิคการปรับปรุงพจนานุกรมและไวยากรณ์เพื่อการรู้จำพยางค์

รูปเสียง	ลำดับหน่วยเสียง
บอ	บ ออ
บ่อ	บ อ่อ
บอก	บ อ่อ –ก
บ่อย	บ อ่อ –ย
ใบ	บ อะ –ย
บัน	บ อะ –น

ตาราง 4.7 ตัวอย่างพจนานุกรมพยางค์

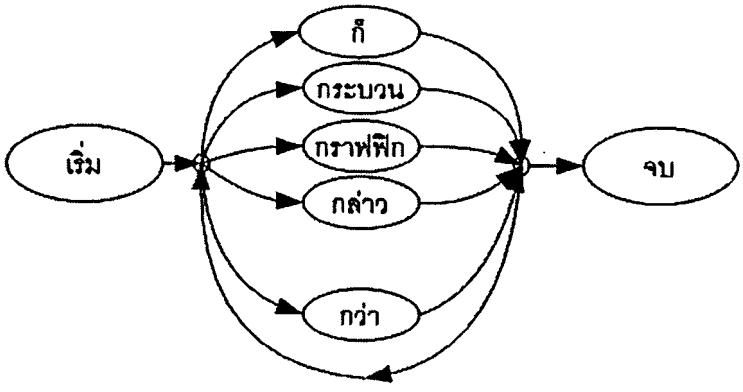


ภาพประกอบ 4.6 รูปแบบการรู้จำพยางค์

4. การรู้จำคำศัพท์ จำเป็นต้องใช้พจนานุกรมคำศัพท์ดังตาราง 4.8 โดยใช้รูปแบบการรู้จำแบบไม่จำกัดคำศัพท์ดังภาพประกอบ 4.7

รูปเสียง	ลำดับหน่วยเสียง
ก็	ก อ้อ
กระบวน	กร อะ บ อัว –น
กราฟฟิก	กร อ๊า -ฟ ฟ อิ –ก
กล่าว	กล อ่า –ว
กว่า	กว อ่า

ตาราง 4.8 ตัวอย่างพจนานุกรมคำศัพท์



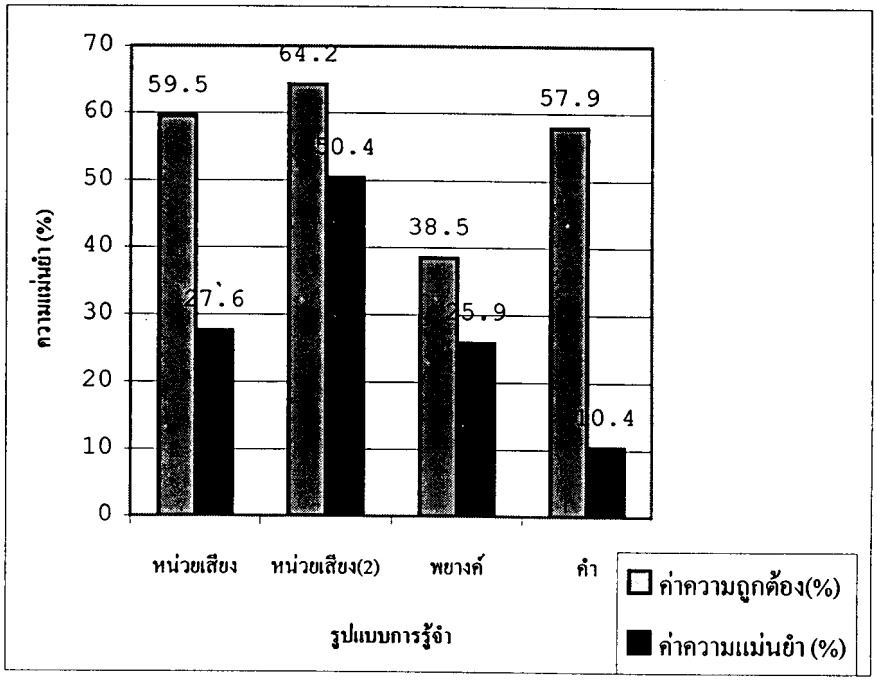
ภาพประกอบ 4.7 รูปแบบการรู้จำคำศัพท์

ในการทดลอง ข้อมูลภายในพจนานุกรมไม่สามารถกำหนดให้เป็นอักษรภาษาไทยดังแสดงในตาราง 4.8 ได้เนื่องจากข้อจำกัดของชุดโปรแกรมที่นำมาใช้งานจึงจำเป็นต้องใช้ชุดหน่วยเสียงที่กำหนดขึ้นมาเป็นคำศัพท์และคำอ่านในพจนานุกรม ดังแสดงในตาราง 4.9 ดังนั้นการแสดงข้อมูลบางอย่างอาจเป็นตัวอักษรภาษาอังกฤษเพื่อให้สอดคล้องกับข้อมูลจริงในการทดลอง

รูปเสียง	ลำดับหน่วยเสียง
k_@@_2	k @@2
kr_a_z^_1_b_uua_n^_0	kr a1 b uua0 n_
kr_aa_p^_3_f_i_k^_1	kr aa3 p_ f i1 k_
kl_aa_w^_1	kl aa1 w_
kw_aa_z^_1	kw aa1

ตาราง 4.9 ตัวอย่างพจนานุกรมคำศัพท์ที่ใช้ในการทดลอง

ผลการทดลอง



ภาพประกอบ 4.8 ความแม่นยำในการรู้จำของแบบจำลองเสียงต้นแบบ

วิเคราะห์ผลการทดลอง

ผลการทดลองจากภาพประกอบ 4.8 ความแม่นยำในการรู้จำหน่วยเสียงยังต่ำมาก ค่าความถูกต้องและความแม่นยำที่แตกต่างกันมากแสดงถึงความผิดพลาดในการรู้จำเกินที่ค่อนข้างสูง (เพราะค่าความถูกต้องไม่ได้นำการรู้จำเกินมาคิด) ทำให้วิเคราะห์ได้ว่าหน่วยเสียงบางหน่วยอาจมีความคล้ายคลึงกับเสียงเงียบ (silence) ดังนั้นเมื่อไม่มีเสียงใดๆเข้าระบบก็อาจรู้จำเป็นหน่วยเสียงเหล่านั้นได้

สำหรับหน่วยเสียง(2) มีความแม่นยำสูงขึ้นเมื่อเทียบกับหน่วยเสียงปกติ สรุปได้ว่าการกำหนดลำดับของหน่วยเสียงให้เหมาะสมจะช่วยให้ประสิทธิภาพในการรู้จำสูงขึ้นได้

เนื่องจากการรู้จำหน่วยเสียงที่ยังไม่ดีพอ ประกอบกับพยางค์ที่รู้จำได้มีจำนวนมาก (รวบรวมจากข้อมูลเสียงที่มีอยู่ ไม่ใช่พยางค์ทั้งหมดในภาษาไทย) ทำให้การรู้จำพยางค์ รวมไปถึงการรู้จำคำศัพท์มีความแม่นยำลดลง

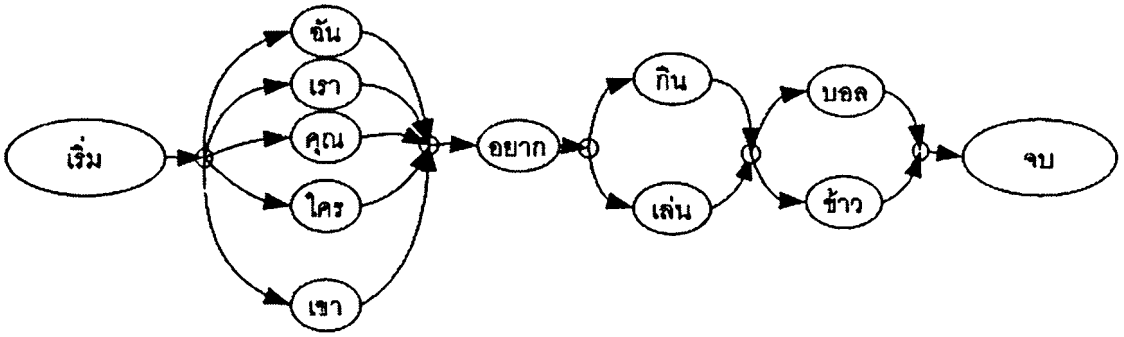
สรุปได้ว่าความซับซ้อนหรือขั้นตอนในการรู้จำมากเท่าไรความแม่นยำก็จะลดลงมากเท่านั้นสังเกตจากผลการรู้จำหน่วยเสียง พยางค์ คำ ตามลำดับ

4.6 การพัฒนาไวยากรณ์ภาษาเบื้องต้น

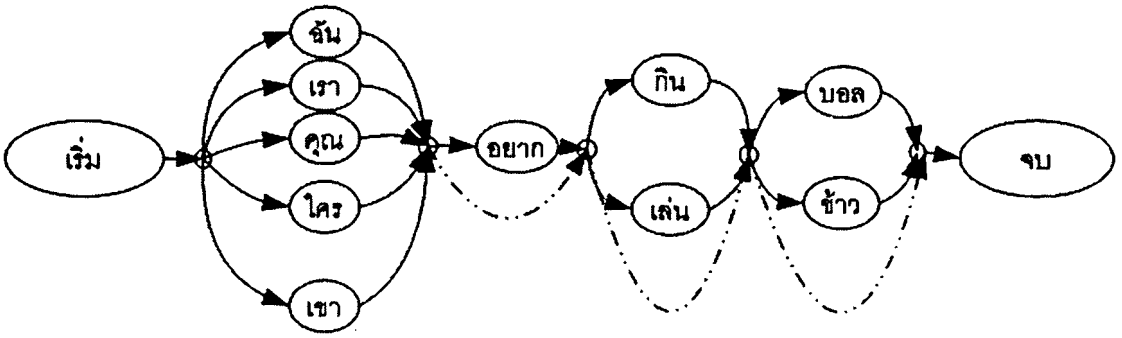
ในการหาค่าความแม่นยำของการรู้จำหน่วยเสียง พยางค์ และคำ ทำให้ทราบถึงศักยภาพของแบบจำลองเสียงว่าเหมาะกับการรู้จำคำศัพท์จำนวนมากน้อยเพียงใด ขั้นตอนนี้จึงเป็นการกำหนดรูปแบบการรู้จำหรือไวยากรณ์ภาษาสำหรับการนำไปใช้ในงานต่างๆ

ในการทดลอง เริ่มจากการออกแบบไวยากรณ์ภาษารูปแบบต่างๆแล้วนำไปทดสอบร่วมกับแบบจำลองเสียงต้นแบบเพื่อให้ได้ค่าความแม่นยำออกมาโดยวิธีการรู้จำเสียงพูด ได้กำหนดข้อมูลสำหรับการรู้จำดังต่อไปนี้

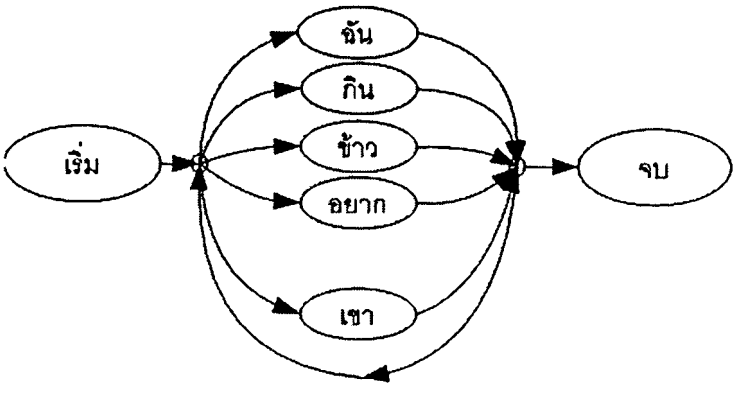
- เสียงจากผู้พูดชาย 1 คนพูด 5 ประโยคที่กำหนดประโยคละ 2 ครั้ง บันทึกเสียงภายในห้องบันทึกเสียงที่ไม่มีสัญญาณรบกวน
- แบบจำลองเสียงต้นแบบ
- เนื่องจากประสิทธิภาพการรู้จำที่ยังต่ำ จึงต้องกำหนดขอบเขตของการรู้จำให้แคบเพื่อให้การรู้จำมีประสิทธิภาพที่สุด กำหนดจำนวนคำศัพท์ที่รู้จำได้ประมาณ 30 คำและไวยากรณ์ 3 แบบดังนี้
  1. ประโยคบอกเล่าพื้นฐาน 4 คำ ผลการรู้จำต้องเรียงลำดับ ประธาน ขยายกริยา กริยา กรรม ดังภาพประกอบ 4.9
  2. ประโยคบอกเล่าแบบซับซ้อน 4 คำ ผลการรู้จำต้องเรียงลำดับ ประธาน ขยายกริยา กริยา กรรม โดยส่วนขยายกริยา กริยา กรรมจะมีหรือไม่ก็ได้ ดังภาพประกอบ 4.10
  3. คำศัพท์ทั่วไป ผลการรู้จำเป็นคำศัพท์ได้ทุกคำโดยไม่จำกัดจำนวน ดังภาพประกอบ 4.11



ภาพประกอบ 4.9 ไวยากรณ์สำหรับการรู้จำประโยคบอกเล่าพื้นฐาน



ภาพประกอบ 4.10 ไวยากรณ์สำหรับการรู้จำประโยคบอกเล่าซับซ้อน



ภาพประกอบ 4.11 ไวยากรณ์สำหรับการรู้จำรู้จำคำศัพท์ทั่วไป

กำหนดชุดคำศัพท์ โดยแยกตามกลุ่มตามลำดับของคำในประโยค ดังนี้

ประธาน	= ฉัน เขา เรา เธอ ผม คุณ ใคร
ขยายกริยา	= อยาก กำลัง ชอบ
กริยา	= เดิน นั่ง นอน ไปเที่ยว กิน ดู เล่น ตี
กรรม	= ข้าว บอล ทีวี การ์ตูน กีฬา น้ำ รถ

ชุดประโยคที่ใช้ในการรู้จำ มีดังนี้

- ผมจะกินข้าว
- ฉันกำลังนั่งรถ
- เขาชอบเล่นบอล
- เราอยากดูทีวี
- ใครไม่ชอบกินน้ำ

ผลการทดลอง

```
sil ph_u0_m c_a1 sil t_ii0 kh_aa2_w sil = ผมจะกินข้าว
sil ph_u0_m j_aa1_k t_ii0 sil kh_aa2_w sil = ผมอยากกินข้าว
sil ch_a4_n c_a1 d_uu0 kh_aa2_w sil = ฉันจะดูข้าว
sil ch_a4_n c_a1 t_ii0 kh_aa2_w sil = ฉันจะกินข้าว
sil kh_a4_w ch_@@2_p l_e2_n b_@@0_l sil = เขาชอบเล่นบอล
sil kh_a4_w m_a2_j_ch_@@2_p l_e2_n b_@@0_l sil = เขาไม่ชอบเล่นบอล
sil r_a0_w j_aa1_k d_uu0 sil th_ii0_w_ii0 sil = เราอยากดูทีวี
sil r_a0_w j_aa1_k d_uu0 sil th_ii0_w_ii0 sil = เราอยากดูทีวี
sil khr_a0_j m_a2_j_ch_@@2_p t_ii0 sil kh_aa2_w sil = ใครไม่ชอบกินข้าว
sil khr_a0_j m_a2_j_ch_@@2_p k_i0_n r_o3_t = ใครไม่ชอบกินรถ
```

ภาพประกอบ 4.12 ผลการรู้จำของไวยากรณ์ประโยคบอกเล่าพื้นฐาน

ไวยากรณ์ภาษา	ค่าความแม่นยำ(%)
ประโยคบอกเล่าพื้นฐาน	67.5
ประโยคบอกเล่าซับซ้อน	57.5
คำศัพท์ทั่วไป	47.5

ตาราง 4.10 แสดงค่าความแม่นยำของการรู้จำไวยากรณ์ภาษารูปแบบต่างๆ

วิเคราะห์ผลการทดลอง

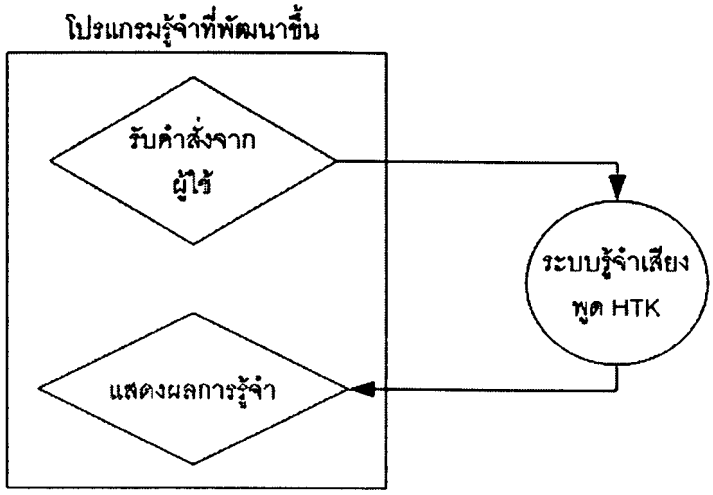
จากผลการทดลองในตารางที่ 4.10 พบว่าความซับซ้อนของไวยากรณ์ยิ่งมากยิ่งทำให้ความแม่นยำในการรู้จำลดลง และคำศัพท์บางคำไม่สามารถรู้จำได้อย่างถูกต้อง เช่นคำว่า น้ำ จากภาพประกอบ 4.12 จึงได้ทำการทดลองเพิ่มเติมเพื่อพิจารณาผลการรู้จำหน่วยเสียงพบว่า ไม่ตรงกับลำดับหน่วยเสียงในพจนานุกรม เช่นคำว่า น้ำ ควรมีลำดับหน่วยเสียงเป็น น อ๊ะ –ม แต่ผลการรู้จำหน่วยเสียงคือ น อ้า –ม แสดงให้เห็นว่า ข้อมูลหน่วยเสียงภายในแบบจำลองยังไม่ตรงกับความเป็นจริง นอกจากนี้ยังพบว่าการรู้จำไม่มีความเที่ยงตรงจากการพูดคำเดียวกัน 2 ครั้งแต่ให้ผลการรู้จำที่ต่างกัน หมายความว่าข้อมูลหน่วยเสียงภายในแบบจำลองมีความซ้ำซ้อนกันมากเมื่อลักษณะเสียงเปลี่ยนไปเล็กน้อยทำให้ผลการรู้จำเปลี่ยนไปด้วย

4.7 การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับรู้จำเสียงพูด

โดยปกติชุดโปรแกรมรู้จำเสียงพูด HTK มีการทำงานอยู่บนแบตช์โหมด (Batch Mode เป็นการทำงานแบบรับงานไปทำเป็นชุด โดยทำงานที่ได้รับมาตามลำดับจนหมดงานในชุดนั้น) โดยรับคำสั่งจากผู้ใช้จากคอมมานด์ไลน์ ซึ่งไม่สะดวกในการใช้งาน หากมีการแสดงผลที่เป็นข้อความภาษาไทยที่ผู้ใช้อ่านง่ายและ มีปุ่มคอยรับคำสั่งต่างๆ แบบกราฟฟิกโหมดจะทำให้ระบบรู้จำเสียงพูดมีความสะดวกในการใช้งานมากขึ้น จึงได้ทำการพัฒนาโปรแกรมรู้จำเสียงพูดสำหรับใช้งานจริง มีการรับเสียงพูดจากไมโครโฟนและแสดงผลออกทางหน้าจอ

4.7.1 การออกแบบ

งานวิจัยนี้ได้ออกแบบและพัฒนาโปรแกรมคอมพิวเตอร์ที่ทำหน้าที่เป็นส่วนติดต่อกับผู้ใช้และควบคุมการทำงานของชุดโปรแกรม HTK โปรแกรมนี้ประกอบด้วยส่วนรับคำสั่งจากผู้ใช้ คอยการกดปุ่มว่าต้องการ เริ่มการรู้จำเสียงพูด หรือหยุดการรู้จำเสียงพูด และส่วนแสดงผลมีหน้าจอที่แสดงผลการรู้จำเป็นข้อความภาษาไทย หรือสัญลักษณ์ต่างๆ ตามที่ผู้ใช้กำหนดได้ ดังภาพประกอบ 4.13



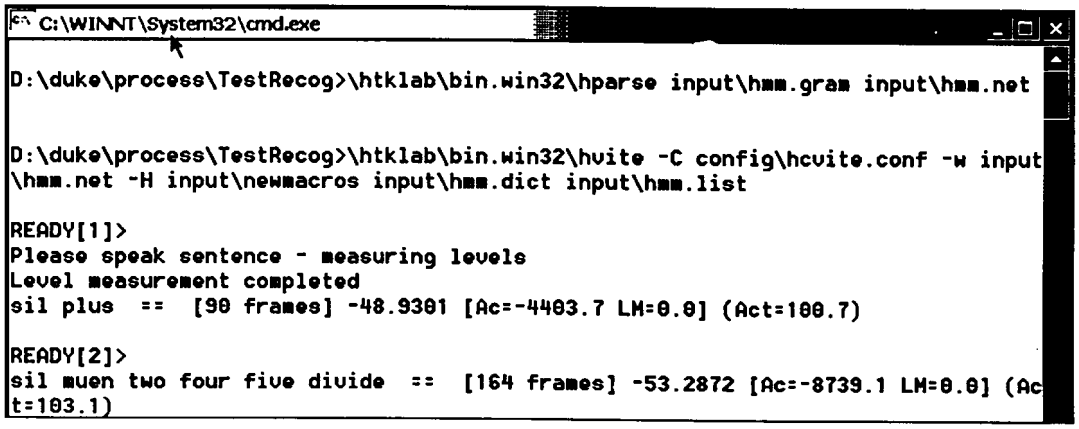
ภาพประกอบ 4.13 ลักษณะการทำงานของโปรแกรมรู้จำ

4.7.2 การพัฒนา

ได้ทำการพัฒนาโปรแกรมตามแบบที่วางไว้ โดยใช้ภาษาจาวา ใช้คอมไพเลอร์เวอร์ชัน 1.4.2 (J2SDK-1.4.2) โปรแกรมที่ใช้พัฒนาคือ เน็ตบีน 3.3.1 (NetBeans 3.3.1) ส่วนสำคัญที่เป็นหน้าที่หลักของโปรแกรมนี้คือ การสั่งงานแบทช์ไฟล์และคอยนำผลที่แสดงบนแบตช์โหมดมาปรับรูปแบบให้เหมาะสมสำหรับการแสดงผลบนหน้าจอโปรแกรมรู้จำที่เป็นแบบกราฟฟิก

โดยปกติการรู้จำเสียงพูดด้วยชุดโปรแกรม HTK จะใช้วิธีการพิมพ์คำสั่งลงไปและผลที่ได้จะแสดงในรูปแบบของข้อความดังภาพประกอบ 4.14 ซึ่งคำสั่งที่ใช้ในการรู้จำคือ

hvite -C hcvite.conf -w hmm.net -H newmacros hmm.dict hmm.list

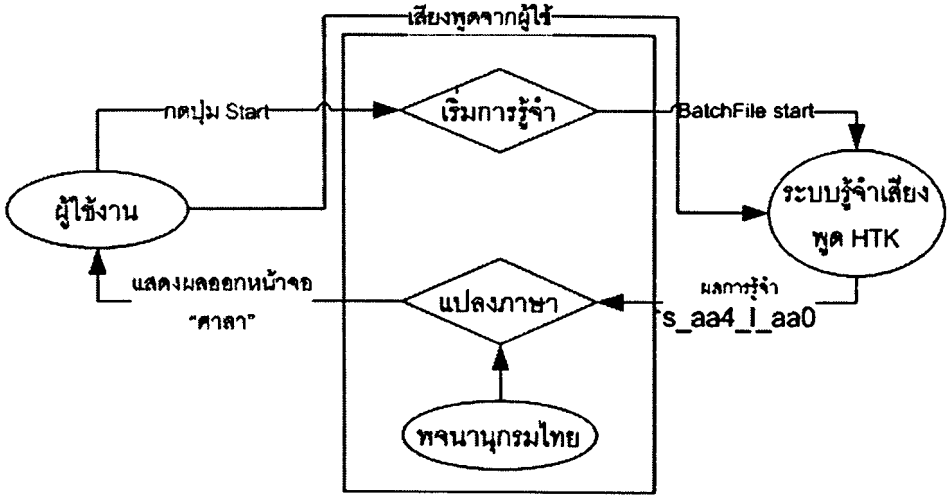


ภาพประกอบ 4.14 วิธีการใช้งานชุดโปรแกรม HTK

จะเห็นได้ว่าคำสั่งที่ใช้ในการรู้จำเสียงพูดประกอบด้วยไฟล์ข้อมูลหลายไฟล์ที่จำเป็นต้องใช้ในการรู้จำ ซึ่งไฟล์แต่ละไฟล์มีความสำคัญดังนี้

- hvite คือโปรแกรมหนึ่งในชุดโปรแกรม HTK ทำหน้าที่รู้จำเสียงพูด
- hcvite.conf คือโหมดการทำงานของ hvite คอยกำหนดวิธีการทำงานและค่าตัวแปรต่างๆ
- hmm.net คือ ไวยากรณ์ภาษาเป็นส่วนสำคัญของการกำหนดรูปแบบการรู้จำ
- newmacros คือ แบบจำลองเสียงพูดที่เก็บข้อมูลของหน่วยเสียงทั้งหมดไว้
- hmm.dict คือ พจนานุกรมที่เก็บคำศัพท์และลำดับหน่วยเสียงในการรู้จำ
- hmm.list คือ ชุดหน่วยเสียงที่ใช้ในการรู้จำ

รายละเอียดการทำงานของโปรแกรมรู้จำนี้เริ่มเมื่อผู้ใช้กดปุ่มเริ่มการรู้จำเสียงพูด (Start) โปรแกรมรู้จำจะไปสั่งการแบทช์ไฟล์ที่เรียกใช้ชุดโปรแกรม HTK ให้เริ่มการรู้จำเสียงพูดทันที ผลจากขั้นตอนนี้ระบบรู้จำจะเริ่มการบันทึกเสียงและรู้จำเสียงที่เข้ามาในระบบ จากนั้นจะแสดงผลการรู้จำเป็นคำศัพท์ในรูปแบบของตัวอักษรภาษาอังกฤษ ในขณะเดียวกันโปรแกรมรู้จำได้นำข้อมูลนี้มาแปลงภาษาด้วยพจนานุกรมไทยและนำไปแสดงผลบนหน้าจอโปรแกรมรู้จำตามขั้นตอนในภาพประกอบ 4.15



ภาพประกอบ 4.15 ขั้นตอนการทำงานของโปรแกรมรู้จำ

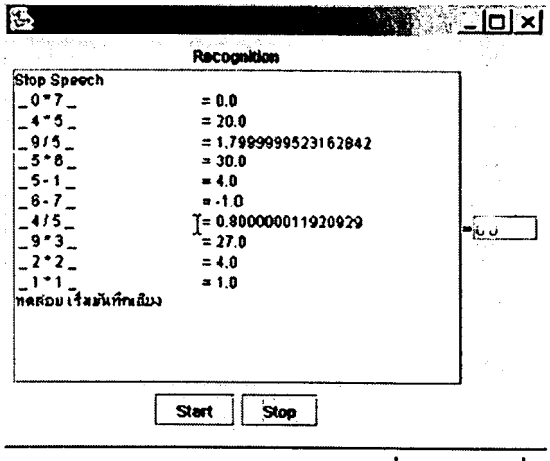
พจนานุกรมไทย เป็นพจนานุกรมสัญลักษณ์ที่กำหนดโดยผู้ใช้ว่าต้องการให้คำศัพท์แต่ละคำนั้นแสดงผลในรูปแบบใด เช่น ตัวเลข คำศัพท์ หรือสัญลักษณ์อื่นๆดังตาราง 4.11 เนื่องจากพจนานุกรมของระบบรู้จำนั้นแสดงผลเป็นกลุ่มอักษรภาษาอังกฤษ ดังภาพประกอบ 4.14 ส่วนพจนานุกรมไทยจะทำหน้าที่แปลงข้อมูลให้เหมาะสมกับการแสดงผล

คำศัพท์	คำแสดง
b_@@0_l	บอล
c_a1	จะ
ch_@@2_p	ชอบ
ch_a4_n	ฉัน
d_uu0	ดู
j_aa1_k	อยาก
k_a0_m_l_a0_ng	กำลัง
k_aa0_t_uu0_n	การ์ตูน

ตาราง 4.11 พจนานุกรมไทยสำหรับโปรแกรมรู้จำ

ตัวอย่างการนำโปรแกรมรู้จำไปใช้แทนเครื่องคิดเลข โดยให้ผู้ใช้พูดตัวเลขและเครื่องหมายทางคณิตศาสตร์ที่เป็นสมการที่ถูกต้องเช่น 2+2 ระบบจะรู้จำเสียงเหล่านั้นเป็นตัวเลขและเครื่องหมายและแสดงผลเป็นสมการพร้อมค่าที่ได้จากคำนวณดังภาพประกอบ 4.16 ในการทดสอบโปรแกรมรู้จำได้ใช้ข้อมูลที่เป็นปัจจัยแวดล้อมดังต่อไปนี้

- เสียงพูดจากผู้พูดชาย 1 คนบันทึกเสียงในห้องบันทึกเสียง
- แบบจำลองเสียงต้นแบบ
- พจนานุกรมตัวเลข
- ไวยากรณ์สมการคณิตศาสตร์ ตัวอย่างเช่น 2+2 , 4-2, ...
- พจนานุกรมไทย-สมการคณิตศาสตร์ (สำหรับการแสดงผล) เช่น เลข 2 รูปเสียงคือ s_@@4_ng สิ่งที่จะแสดงผลที่กำหนดในพจนานุกรมไทยคือ 2



ภาพประกอบ 4.16 หน้าจอโปรแกรมรู้จำที่ใช้แทนเครื่องคิดเลข

ผลการพัฒนาและทดสอบประสิทธิภาพ ทำให้ได้โปรแกรมรู้จำที่อำนวยความสะดวกในการรับคำสั่งและแสดงผลได้อย่างถูกต้อง อีกทั้งความแม่นยำของการรู้จำไม่ได้ลดลงแต่อย่างใด เนื่องจากไม่ได้เปลี่ยนแปลงวิธีการรู้จำเสียงพูด เพียงแต่ปรับปรุงวิธีการสั่งการ และการแสดงผลเท่านั้น

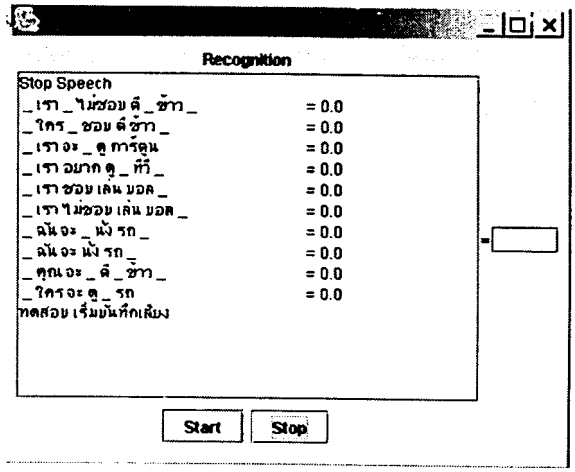
4.8 ทดสอบการรู้จำในสภาพแวดล้อมจริง

เมื่อการพัฒนาระบบรู้จำทั้งระบบเสร็จสิ้นก็จะต้องมีการทดสอบการทำงาน ได้ออกแบบการทดลองเพื่อหาประสิทธิภาพของระบบรู้จำเสียงพูดที่พัฒนาขึ้น โดยใช้ข้อมูลดังต่อไปนี้

- แบบจำลองเสียงต้นแบบ
- พจนานุกรม 24 คำ
- ไวยากรณ์ประโยคบอกเล่าพื้นฐาน
- พจนานุกรมไทย 24 คำ
- เสียงพูดจากผู้พูดชาย 1 คน บันทึกด้วยเครื่องคอมพิวเตอร์แบบพกพา(ต่างจากที่ใช้ในการทดลอง) ในสภาพแวดล้อมภายในบ้าน
- ข้อความที่พูด 5 ข้อความ จากชุดประโยคของการทดลองในขั้นตอนที่ 4.6

เริ่มต้นด้วยการให้ผู้พูดพูดข้อความที่กำหนด ข้อความละ 2 ครั้ง แล้วดูผลการรู้จำจากหน้าจอโปรแกรมรู้จำ และทำการเปรียบเทียบผลการรู้จำกับชุดประโยคได้เป็นค่าความแม่นยำออกมา แล้วการวัดผลการรู้จำในระดับคำศัพท์ โดยเปรียบเทียบเสียงที่พูดกับข้อความที่แสดงออกทางหน้าจอคอมพิวเตอร์

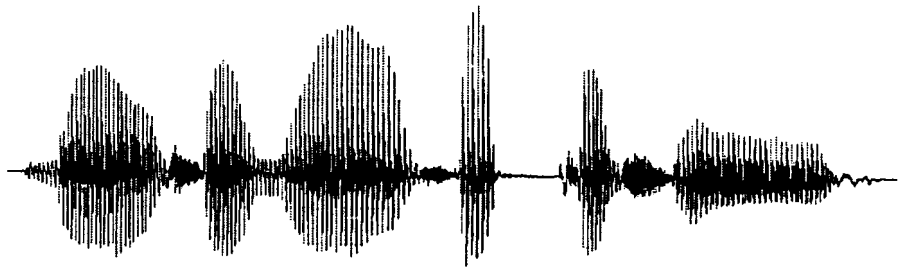
ผลการทดลอง ได้ค่าความแม่นยำ 55%



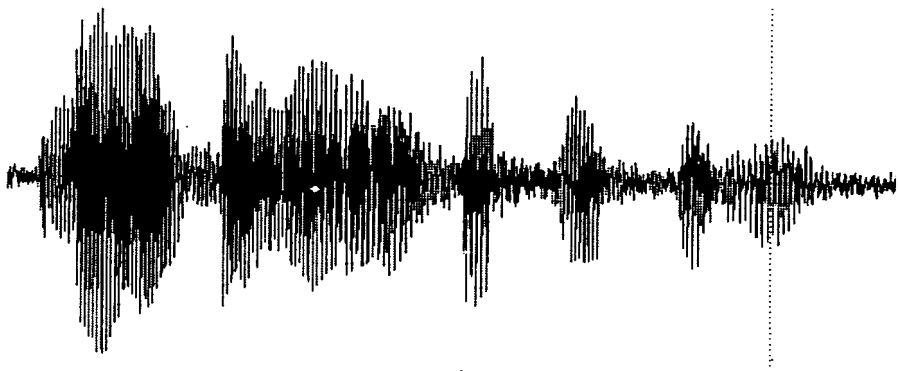
ภาพประกอบ 4.17 ผลการรู้จำจากระบบรู้จำเสียงพูดต้นแบบ

วิเคราะห์ผลการทดลอง

จากภาพประกอบ 4.17 พบปัญหาเช่นเดียวกับการทดลองในขั้นตอนที่ 4.6 แต่การทดลองนี้ใช้เครื่องคอมพิวเตอร์และสภาพแวดล้อมที่เปลี่ยนไป 2 สิ่งนี้จึงเป็นส่วนสำคัญที่ทำให้ความแม่นยำลดลงเมื่อเทียบกับผลการทดลองในขั้นตอนที่ 4.6 จึงได้ทำการเปรียบเทียบข้อมูลเสียงจากทั้งสองการทดลองดังภาพประกอบ 4.18 และ 4.19 พบว่ามีสัญญาณรบกวนปรากฏในข้อมูลเสียงของการทดลองปัจจุบัน ทำให้สามารถสรุปสาเหตุของความแม่นยำที่แตกต่างกันได้ว่า เป็นเพราะสภาพแวดล้อมและอุปกรณ์ที่ใช้ประกอบการบันทึกเสียงที่แตกต่างกัน



ภาพประกอบ 4.18 ลักษณะเสียงที่บันทึกภายในห้องบันทึกเสียง



ภาพประกอบ 4.19 ลักษณะเสียงที่บันทึกภายนอกห้องบันทึกเสียง