

การชี้เฉพาะคำสำคัญเสียงพูดภาษาไทยบนพื้นฐานของการตรวจสอบหน่วยเริ่ม
และหน่วยตามของพยางค์



นายเฉลิมวุฒิ ไวกนะ

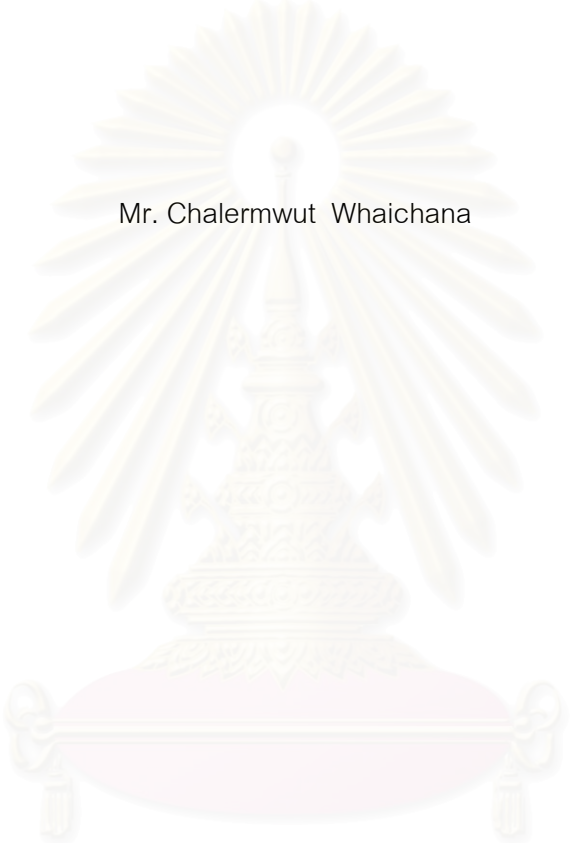
วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-17-5138-9

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

THAI SPEECH KEYWORD SPOTTING BASED ON SUBSYLLABLE ONSETS
AND RHYMES VERIFICATION

Mr. Chalermwut Whaichana



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2005
ISBN 974-17-5138-9

หัวข้อวิทยานิพนธ์

การชี้เฉพาะคำสำคัญเสียงพูดภาษาไทยบนพื้นฐานของการ
ตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์

โดย

นายเฉลิมวุฒิ ไชชนะ

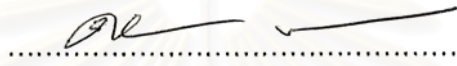
สาขาวิชา

วิศวกรรมไฟฟ้า


อาจารย์ที่ปรึกษา


รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล

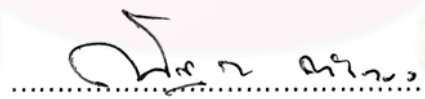
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยอนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต


..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(อาจารย์สุวิทย์ นาคพิระยุทธ)


..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล)


..... กรรมการ
(อาจารย์ ดร. ณัฐกร ทับทอง)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เฉลิมวุฒิ ไวชนะ : การชี้เฉพาะคำสำคัญเสียงพูดภาษาไทยบนพื้นฐานของการ
ตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์. (THAI SPEECH KEYWORD
SPOTTING BASED ON SUBSYLLABLE ONSETS AND RHYMES VERIFICATION)

อ. ที่ปรึกษา: รศ. ดร.สมชาย จิตะพันธ์กุล, 78 หน้า. ISBN: 974-17-5138-9

วิทยานิพนธ์นี้ มีวัตถุประสงค์เพื่อพัฒนากรรณวิธีในการชี้เฉพาะคำสำคัญเสียงพูด
ภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยการประยุกต์ใช้หลักการของแบบจำลองฮิดเดนมาร์คอฟ ซึ่งเป็น
แบบจำลองของเสียงระดับย่อยของพยางค์ ใช้วิธีการตรวจสอบหน่วยเริ่มและหน่วยตามของ
พยางค์ มีจำนวนคำสำคัญ 70 คำแบ่งออกเป็น 3 ชุดตามจำนวนคือ 20 คำ 40 คำและ 70 คำ
เสียงพูดที่นำมาเป็นต้นแบบและเป็นแบบทดสอบประกอบด้วยเสียงผู้ชาย 30 คนและเสียงผู้หญิง
20 คน

ผลการทดสอบค่า *FOM* ของระบบการชี้เฉพาะแบบไม่ขึ้นกับผู้พูดมีอัตราเฉลี่ยร้อยละ
54.45 โดยค่า *FOM* เฉพาะชุดคำสำคัญ 20 คำมีค่าเฉลี่ยร้อยละ 64.59 ค่า *FOM* เฉพาะชุดคำ
สำคัญ 40 คำมีค่าร้อยละ 51.11 และค่า *FOM* เฉพาะชุดคำสำคัญ 70 คำมีค่าร้อยละ 47.65
ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา _____ วิศวกรรมไฟฟ้า _____

สาขาวิชา _____ วิศวกรรมไฟฟ้า _____

ปีการศึกษา _____ 2548 _____

ลายมือชื่อนิสิต _____

ลายมือชื่ออาจารย์ที่ปรึกษา _____

447 06939 21 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: KEYWORD SPOTTING / HIDDEN MARKOV MODEL / SUBSYLLABLE /
ONSET / RHYME / FIGURE OF MERIT

CHALERMWUT WHAICHANA: THAI SPEECH KEYWORD SPOTTING BASED ON
SUBSYLLABLE ONSETS AND RHYMES VERIFICATION. THESIS ADVISOR:
ASSOC. PROF. SOMCHAI JITAPUNKUL, 78 pp. ISBN: 974-17-5138-9

The objective of this thesis is to develop an algorithm for Speaker Independent Thai Speech Keyword Spotting. The methodology is to use the Hidden Markov Model to create the subsyllable models based on subsyllable onsets and rhymes verification algorithm. The 70 keywords are subdivided into three sets comprising 20, 40 and 70 keywords. The separated speech training set and testing set are composed of 30 males and 20 females.

The Figure of Merit (FOM) of this speaker – independent keyword spotting system is 54.45 percent. The FOM of the three sets of amount of keywords are 64.59 percent for 20 keywords set, 51.11 percent for 40 keywords set, and 47.65 percent for 70 keywords set.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department Electrical Engineering

Field of study Electrical Engineering

Academic year 2005

Student's signature 

Advisor's signature 

กิตติกรรมประกาศ

ในการดำเนินงานวิจัยในงานวิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี ข้าพเจ้าขอกราบ
ขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำแนะนำ
ในการดำเนินการวิจัย ตลอดจนได้จัดหาอุปกรณ์และทุนการวิจัยเพื่อดำเนินการวิจัยได้อย่าง
สะดวกและครบถ้วน

ขอขอบคุณ คุณเอกฤทธิ์ มณีน้อย ที่ได้กรุณาสละเวลาให้คำแนะนำและคำปรึกษาที่
สำคัญอย่างยิ่งต่อการดำเนินการวิจัยครั้งนี้

ขอขอบคุณทุกท่านที่ได้สละเวลาเพื่อการบันทึกเสียง และเพื่อนนิสิตปริญญาโททุกท่านที่
ได้ให้ความช่วยเหลือและเป็นกำลังใจให้เสมอมา

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดา ที่ได้มอบความอบอุ่น การสนับสนุนในทุก
ด้านและเป็นแรงใจกำลังใจตลอดเวลาที่ได้ศึกษาวิจัยจนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไป
ด้วยดี



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ญ
สารบัญคำศัพท์.....	ฎ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
แนวทางที่นำเสนอ.....	6
วัตถุประสงค์.....	7
ขอบเขตวิทยานิพนธ์.....	7
ขั้นตอนและวิธีดำเนินงาน.....	7
ประโยชน์ที่คาดว่าจะได้รับ.....	8
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	9
การประมวลผลสัญญาณเบื้องต้น.....	10
การสกัดลักษณะสำคัญ.....	12
การทดสอบความคล้ายคลึงกันของรูปแบบ.....	15
ระบบแบบจำลองฮิดเดน มาร์คอฟ.....	15
ขั้นตอนวิธีการตัดสินใจ.....	29
ประเภทของแบบจำลองฮิดเดน มาร์คอฟ.....	30
คุณลักษณะด้านกลศาสตร์ของเสียงพูดภาษาไทย.....	32
การตรวจหากลุ่มคำ.....	35
แบบจำลองภาษา.....	38
บทที่ 3 วิธีดำเนินการวิจัย.....	39
การกำหนดวิธีการสร้างชุดคำสำคัญ.....	39
การเก็บตัวอย่างข้อมูลเสียงพูด.....	39
รายละเอียดขั้นตอนในการชี้เฉพาะคำสำคัญภาษาไทย.....	43

บทที่ 4 ผลการวิจัย.....	48
ความแม่นยำของระบบชี้เฉพาะคำสำคัญ.....	48
ค่า FOM ของระบบชี้เฉพาะคำสำคัญ.....	48
ROC Curve.....	49
ผลการชี้เฉพาะคำสำคัญภาษาไทย.....	49
วิเคราะห์ผลการชี้เฉพาะคำสำคัญภาษาไทย.....	58
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	60
สรุปผลการวิจัย.....	60
ข้อเสนอแนะสำหรับงานวิจัยในอนาคต.....	61
รายการอ้างอิง.....	62
ภาคผนวก.....	65
ประวัติผู้เขียนวิทยานิพนธ์.....	78

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

	หน้า
ตารางที่ 2.1 รายละเอียดของขั้นตอนในการกำเนิดลำดับค่าสังเกต.....	17
ตารางที่ 2.2 รายละเอียดของปัญหาพื้นฐานสามประการของแบบจำลองฮิตเดน มาร์คอฟ.....	18
ตารางที่ 2.3 รายละเอียดกระบวนการไปหน้า.....	21
ตารางที่ 2.4 รายละเอียดกระบวนการย้อนกลับ.....	23
ตารางที่ 2.5 รายละเอียดขั้นตอนวิธีการ Viterbi.....	25
ตารางที่ 2.6 รายละเอียดกระบวนการประมาณค่าซ้ำของ Baum – Walch.....	26
ตารางที่ 2.7 รายละเอียดการประมาณค่าซ้ำของค่าความน่าจะเป็นของค่าสังเกต แบบเกาส์เซียน.....	29
ตารางที่ 2.8 รายละเอียดของพัญชนะเสียงเดี่ยว.....	33
ตารางที่ 2.9 รายละเอียดของพัญชนะเสียงคู่.....	34
ตารางที่ 2.10 รายละเอียดของเสียงตัวสะกด.....	34
ตารางที่ 2.11 รายละเอียดของสระเสียงเดี่ยว.....	34
ตารางที่ 2.12 รายละเอียดของสระเสียงผสม.....	34
ตารางที่ 3.1 ตัวอย่างของผลที่ได้จากการทดสอบระบบ.....	46
ตารางที่ 4.1 ผลอัตราความแม่นยำเฉลี่ยของชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน.....	50
ตารางที่ 4.2 ผลอัตราความแม่นยำเฉลี่ยของชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน.....	50
ตารางที่ 4.3 รายละเอียดผลการทดสอบค่า FOM ของแบบจำลองเสียงระดับย่อย ของพยางค์ประเภท Monophone.....	51
ตารางที่ 4.4 รายละเอียดผลการทดสอบค่า FOM ของแบบจำลองเสียงระดับย่อย ของพยางค์ประเภท Triphone.....	51
ตารางที่ 4.5 รายละเอียดผลการทดสอบค่า FOM ของแบบจำลองเสียงระดับย่อย ของพยางค์ประเภท Onset – Rhyme.....	52
ตารางที่ 4.6 การเปรียบเทียบค่า FOM ของชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน.....	52
ตารางที่ 4.7 การเปรียบเทียบค่า FOM ของชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน.....	52
ตารางที่ 4.8 รายละเอียดของความผิดพลาดของแบบจำลอง Monophone.....	56
ตารางที่ 4.9 รายละเอียดของความผิดพลาดของแบบจำลอง Triphone.....	56
ตารางที่ 4.10 รายละเอียดของความผิดพลาดของแบบจำลอง Onset – Rhyme.....	57

ตารางที่ 4.11	รายละเอียดความผิดพลาดที่ปรากฏของแบบจำลอง Monophone.....	57
ตารางที่ 4.12	รายละเอียดความผิดพลาดที่ปรากฏของแบบจำลอง Triphone.....	57
ตารางที่ 4.13	รายละเอียดความผิดพลาดที่ปรากฏของแบบจำลอง Onset – Rhyme.....	58
ตารางที่ ก. 1	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Monophone ของชุดคำสำคัญลำดับที่ 1 – 28.....	68
ตารางที่ ก. 2	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Monophone ของชุดคำสำคัญลำดับที่ 29 – 55.....	69
ตารางที่ ก. 3	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Monophone ของชุดคำสำคัญลำดับที่ 56 – 70.....	70
ตารางที่ ก. 4	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 1 – 17.....	71
ตารางที่ ก. 5	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 18 – 41.....	72
ตารางที่ ก. 6	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 42 – 57.....	73
ตารางที่ ก. 7	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 58 – 70.....	74
ตารางที่ ก. 8	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Onset - Rhyme ของชุดคำสำคัญลำดับที่ 1 – 27.....	75
ตารางที่ ก. 9	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Onset - Rhyme ของชุดคำสำคัญลำดับที่ 28-54.....	76
ตารางที่ ก. 10	รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ ประเภท Onset-Rhyme ของชุดคำสำคัญลำดับที่ 55-70.....	77

สารบัญภาพ

	หน้า
รูปที่ 2.1 รายละเอียดของแบบจำลองรูปแบบการชี้เฉพาะคำสำคัญภาษาไทย.....	9
รูปที่ 2.2 ฟังก์ชันกรอบชนิด Hamming Window.....	11
รูปที่ 2.3 Mel Scale ของความถี่สัญญาณเสียง.....	13
รูปที่ 2.4 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล.....	14
รูปที่ 2.5 รายละเอียดลำดับกระบวนการในการคำนวณค่าตัวแปรไปหน้า $\alpha_t(i)$	21
รูปที่ 2.6 รายละเอียดลำดับกระบวนการในการคำนวณค่าตัวแปรย้อนกลับ $\beta_t(i)$	23
รูปที่ 2.7 รายละเอียดลำดับเหตุการณ์ในการคำนวณเหตุการณ์ร่วมในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t + 1$	27
รูปที่ 2.8 ค่าความน่าจะเป็นของค่าสังเกตที่มีการกระจายแบบเกาส์เซียน 2 มิติ.....	28
รูปที่ 2.9 แบบจำลองแบบเออร์กอดิกที่มี 3 สถานะ.....	30
รูปที่ 2.10 แบบจำลองแบบซ้าย - ขวาที่มี 5 สถานะ.....	31
รูปที่ 2.11 วัฏวะที่เกี่ยวข้องกับการสร้างเสียงพูด.....	33
รูปที่ 2.12 รายละเอียดขั้นตอน Bottom – up Detector.....	37
รูปที่ 2.13 รายละเอียดขั้นตอน Top - Down Verifier.....	37
รูปที่ 2.14 รายละเอียดของแบบจำลองภาษา.....	38
รูปที่ 3.1 รายละเอียดของแบบจำลองรูปแบบการชี้เฉพาะคำสำคัญภาษาไทย.....	40
รูปที่ 3.2 รายละเอียดขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลอง ฮิดเดนมาร์คอฟ.....	41
รูปที่ 3.3 รายละเอียดขั้นตอนการชี้เฉพาะคำสำคัญภาษาไทย.....	42
รูปที่ 4.1 กราฟ Log - likelihood ของระบบการชี้เฉพาะคำสำคัญภาษาไทย.....	50
รูปที่ 4.2 ROC Curve ของชุดคำสำคัญจำนวน 20 คำของเสียงของผู้พูดที่อยู่ในชุดฝึกฝน.....	53
รูปที่ 4.3 ROC Curve ของชุดคำสำคัญจำนวน 40 คำของเสียงของผู้พูดที่อยู่ในชุดฝึกฝน.....	53
รูปที่ 4.4 ROC Curve ของชุดคำสำคัญจำนวน 70 คำของเสียงของผู้พูดที่อยู่ในชุดฝึกฝน.....	54
รูปที่ 4.5 ROC Curve ของชุดคำสำคัญจำนวน 20 คำของเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน.....	54
รูปที่ 4.6 ROC Curve ของชุดคำสำคัญจำนวน 40 คำของเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน.....	55
รูปที่ 4.7 ROC Curve ของชุดคำสำคัญจำนวน 70 คำของเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน.....	55

สารบัญคำศัพท์

ความแม่นยำ	accuracy
ขั้นตอนวิธีการ	algorithm
ชุดรหัส	codebook
เสียงพูดต่อเนื่อง	continuous speech
แบบจำลองฮิดเดนมาร์คอฟ	hidden markov model
คำสำคัญ	keyword
แบบจำลองภาษา	language model
สัมประสิทธิ์เซปสตรอลบนความถี่เมล	mel frequency cepstral coefficient
หน่วยเริ่มของพยางค์	onset
เสียงระดับย่อยของพยางค์	phoneme
กรรมวิธีประมวลผลเบื้องต้น	preprocessing
การรู้จำ	recognition
หน่วยตามของพยางค์	rhyme
ค่าคะแนน	score
เสียงพูด	speech
การชี้เฉพาะ	spotting
สถานะ	state
จุดเริ่มเปลี่ยน	threshold
การฝึกฝน	training
การควอนไทซ์แบบเวกเตอร์	vector quantization
การตรวจสอบ	verification

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้การใช้เทคโนโลยีการประมวลผลสัญญาณเสียงพูด (Speech Processing) ได้เริ่มมีการนำมาใช้งานอย่างแพร่หลาย และในอนาคตคาดว่าจะเข้ามามีบทบาทต่อชีวิตประจำวันมากขึ้น เพื่อความสะดวกในการใช้งาน ผู้ใช้งานจะใช้เพียงคำพูดเพื่อการสั่งงานด้วยเสียงเท่านั้น ทำให้ผู้ใช้สามารถประหยัดเวลาและสะดวกสบายกับขั้นตอนในการใช้งาน ดังนั้นการพัฒนาเพื่อให้อุปกรณ์มีความสามารถเข้าใจในภาษาพูดของมนุษย์ หรือการพัฒนาการประมวลผลสัญญาณเสียงพูด มีความจำเป็นอย่างยิ่งที่จะต้องสร้างระบบการรู้จำเสียงพูด (Speech Recognition) ให้มีประสิทธิภาพมากยิ่งขึ้น ยกตัวอย่างเช่น การพัฒนาระบบการรู้จำเสียงพูด ทั้งการรู้จำเสียงพูดแบบต่อเนื่อง (continuous speech recognition) และการชี้เฉพาะคำสำคัญ (keyword spotting) เป็นต้น อย่างไรก็ตาม การพัฒนาการประมวลผลเสียงพูดนี้ คอมพิวเตอร์ยังคงไม่สามารถเข้าใจความหมายของคำพูดหรือประโยคที่พูดทั้งหมดได้ โดยเฉพาะอย่างยิ่งในคำพูดภาษาไทย จึงถูกจำกัดไว้เพื่อพัฒนาในด้านการประยุกต์เฉพาะทาง เช่น การพิมพ์ตามคำพูด (Dictation) การโต้ตอบเพื่อการจองตั๋วเครื่องบิน การสั่งงานด้วยเสียง การยืนยันตัวบุคคล (Identification) เป็นต้น ซึ่งแต่ละการประยุกต์จะมีขั้นตอนรายละเอียดแตกต่างกันออกไป

ในการประยุกต์ใช้ระบบการรู้จำคำพูดเพื่อการสั่งงานด้วยเสียง หรือ การค้นหาคำสำคัญในเอกสารเสียง การชี้เฉพาะคำสำคัญมักจะถูกนำมาใช้เนื่องจากมีจุดเด่นในด้านของความยืดหยุ่นในการใช้งานและใช้เวลาในการประมวลผลน้อยเพราะคำที่ต้องการรู้จำมีน้อยกว่าเมื่อเทียบกับระบบการรู้จำคำพูดต่อเนื่อง การรู้จำคำพูดต่อเนื่องมีความจำเป็นต้องสร้างแบบจำลองให้ครอบคลุมกับการออกเสียงเป็นภาษาทั้งหมด เช่น การสั่งเครื่องเสียงให้เล่นซีดีแผ่นที่ 5 เพลงที่ 1 ด้วยวิธีการออกคำสั่งด้วยเสียงพูดนั้น ต้องมีการกำหนดคำสั่งไว้ในเครื่องก่อน โดยหากเรากำหนดให้มีคำสำคัญ คือคำว่า “แผ่นที่” “เพลงที่” และ ตัวเลข การที่ผู้ใช้ให้คำสั่งว่า “เล่นซีดีแผ่นที่ 5 เพลงที่ 1” หรือ “เล่นเพลงที่ 1 จากซีดีแผ่นที่ 5” เครื่องก็จะสามารถรู้จำคำสั่งที่ป้อนเข้ามาได้ทั้งสองรูปแบบ เนื่องจากเราได้กำหนดไว้ว่าหลังจากคำว่า “แผ่นที่” จะเป็นตัวเลขของแผ่นที่ต้องการจะเล่น และหลังจากคำว่า “เพลงที่” จะเป็นลำดับของเพลงที่ต้องการจะเล่น ดังนั้นการสลับตำแหน่งของ “เพลงที่” และ “แผ่นที่” จึงไม่มีผลต่อความสามารถในการรู้จำด้วยวิธีชี้เฉพาะคำสำคัญ อีกทั้งการชี้เฉพาะคำสำคัญยังมีความเหมาะสมในการใช้งานกับการประยุกต์ที่ต้องการรู้จำแค่เพียงคำสำคัญอีกด้วย เช่น “ผมอยากฟังเพลงที่ 1 จากซีดีแผ่นที่ 5 ครับ” เครื่องก็จะสามารถรู้จำ

คำสั่งที่ป้อนเข้ามาได้เนื่องจากคำอื่น ๆ นอกจากคำว่า “แผ่นที่” “เพลงที่” และ ตัวเลข จะไม่ถูกนำมาพิจารณาเป็นคำสั่ง นอกเหนือจากการประยุกต์ใช้การชี้เฉพาะคำสั่งสำคัญกับการสั่งงานด้วยเสียงแล้ว การชี้เฉพาะคำสั่งสำคัญยังสามารถนำไปใช้ได้กับงานอื่น ๆ อีกมากมาย เช่น การค้นหาเอกสารเกี่ยวกับเสียงตามคำสั่ง ระบบโต้ตอบทางโทรศัพท์ เป็นต้น

เทคนิคที่นิยมนำมาใช้สำหรับการชี้เฉพาะคำสั่งสำคัญ สามารถแบ่งเป็นวิธีต่าง ๆ ได้ดังนี้

1. N-best Recognizer

การรู้จำเสียงพูดต่อเนื่องที่มีคำศัพท์ขนาดใหญ่ใช้สร้าง N – best Recognizer ของเสียงพูด เทคนิคนี้เหมาะสมกับระบบที่มีจำนวนคำสั่งสำคัญมาก ใช้ทรัพยากรด้านหน่วยความจำไม่มากเมื่อเทียบกับ Phone Lattice แต่จะทำงานช้ากว่ามาก หากต้องการให้ทำงานได้ด้วยความเร็วเท่า ๆ กันจึงต้องใช้เครื่องคำนวณที่มีประสิทธิภาพสูงกว่ามาก จึงมักใช้ในงานที่ไม่ต้องการผลลัพธ์ในทันทีแต่ต้องการความถูกต้องสูง เช่น งานจัดการเอกสารเกี่ยวกับเสียง (Audio Document) (Knill, K.M. and Young, S.J., 1999) เป็นต้น

2. Acoustic Model

เป็นการนำแบบจำลองต่าง ๆ มาใช้ในการจำลองคุณลักษณะของเสียงเพื่อสร้างเป็นต้นแบบในการประมวลผลเปรียบเทียบกับเสียงที่ป้อนเข้ามา เพื่อระบุเสียง (หรือ องค์ประกอบที่ต้องใช้ในการออกเสียง เช่น เสียงพยัญชนะต้นที่ใช้ประกอบขึ้นเป็นเสียง ‘ก’ ‘ข’ ‘ค’ ... เป็นต้น) ที่ถูกต้อง ในปัจจุบันมีแบบจำลองที่นิยมใช้ได้แก่ HMM, ANN และ Phoneme based ซึ่งข้อดีของเทคนิคนี้คือไม่มีการจำกัดคำสั่งที่ใช้ในการค้นหา อีกทั้งมีความถูกต้องในการรู้จำดี แต่มีข้อเสียตรงที่ใช้เวลาในการคำนวณมากเมื่อมีการชี้เฉพาะคำใหม่ (Bo - Ren Bai, Chiu - Yu Tseng and Lin - Shan lee, 1997; Chung - Hsien Wu and Yeou - Jiunn Chen, 2001; Dharanipragada, S. and Roukos, S., 1998; Kawabata, T. and Kohda, M., 1987; Medress, M.F., Deller, T.C., Kloker, D.R. et al., 1978; Rohlicek, J.R., Jeanrenaud, P., Gish, H. et al., 1993)

3. Phone Lattice

มีลักษณะเป็นโครงข่ายของเสียงย่อยของพยางค์ที่นำมาต่อกันเพื่อให้เกิดเป็นพยางค์ มีความยืดหยุ่นในด้านความสามารถในการค้นหาแต่ละคำสั่งและงานของเอกสารเกี่ยวกับเสียง มีข้อเสียคือใช้ทรัพยากรมาก ความน่าเชื่อถือบน Phone Recognition ยังไม่มากนัก แต่มีความรวดเร็วในการประมวลผลสูง (Foote, J.T., 1999; Foote, J.T., Young, S.J., Jones, G.J.F. et al., 1997; James, D.A. and Young, S.J., 1994)

สำหรับงานวิจัยที่เกี่ยวข้องกับการชี้เฉพาะคำสำคัญที่ได้ถูกนำเสนอมาแล้วนั้น เทคนิคที่ใช้กันมากที่สุดคือ Acoustic Model โดยมีข้อดีคือ ให้ความแม่นยำในการรู้จำ และมีความรวดเร็วในการประมวลผล เหมาะสำหรับการใช้กับการประยุกต์ที่มีคำสำคัญไม่มาก การเพิ่มจำนวนคำสำคัญในเทคนิคนี้สามารถทำได้ไม่จำกัด เพราะว่าเทคนิคนี้ไม่มีโครงสร้างของการเข้าคู่ของคำสำคัญ และใช้แบบจำลองเป็นหน่วยย่อยของพยางค์ จึงมีจำนวนแบบจำลองตามองค์ประกอบของคำสำคัญ ถ้าระบบการชี้เฉพาะมีการเพิ่มคำสำคัญที่มีเสียงของหน่วยย่อยของพยางค์นอกเหนือจากแบบจำลองในระบบที่มีอยู่ก็จะเสียเวลาในการฝึกฝนค่าพารามิเตอร์ในแบบจำลองของเสียงที่มีอยู่ในคำสำคัญคำใหม่

จากงานวิจัยของ Takeshi และ Masaki (Kawabata, T. and Kohda, M., 1987) มีขั้นตอนของระบบการชี้เฉพาะคำสำคัญเสียงพูดภาษาญี่ปุ่น ซึ่งสามารถแบ่งออกเป็น 2 ขั้นตอนคือ

1. Bottom – up Phoneme Detector
2. Top – down Phoneme Verifier

การทดลองของงานวิจัยนี้กำหนดให้ผู้ชาย 3 คนพูด 13 ประโยคภาษาญี่ปุ่นเฉพาะทาง ซึ่งใน 13 ประโยคนั้นบรรจุคำที่แตกต่างกัน 234 คำและกำหนดให้มีคำสำคัญ 7 คำซึ่งจะต้องถูกตรวจพบ 65 ครั้ง ซึ่งวิธี Bottom – up Phoneme Detector ใช้ Context – independent feature ในการรู้จำ มีข้อดีคือใช้เวลาในการประมวลผลน้อย แต่ผลที่ได้มีความน่าเชื่อถือต่ำ จึงใช้วิธี Top – down Phoneme Verifier เพื่อเพิ่มความถูกต้องจากการใช้ Context – dependent feature ในการรู้จำ วิธีนี้มีข้อดีคือสามารถแยกแยะเสียงย่อยของพยางค์ที่มีความคล้ายคลึงกันได้ แต่ข้อเสียของวิธีนี้คือเมื่อใช้ในการรู้จำคำพูดที่เป็นคำต่อเนื่องจะให้ความน่าเชื่อถือต่ำ เมื่อดูผลการทดลองพบว่ามี ความถูกต้อง 90.8 %

งานวิจัยของ J.R. Rohlicek , P. Jeanrenaud , K. Ng , H. Gish , B. Musicus และ M. Siu (1993) ได้เสนอการชี้เฉพาะคำสำคัญเสียงพูดภาษาอังกฤษ โดยมีขั้นตอนต่าง ๆ แบ่งเป็น 3 ส่วนตามลำดับ คือ

1. HMM Acoustic Models
2. Overall HMM Structure
3. Scoring Method

งานวิจัยนี้กำหนดให้คนพูด 70 คน โดยแบ่งเป็น 48 คนสำหรับนำมาฝึกฝนและ 22 คนสำหรับนำมาทดสอบ มีคำสำคัญ 20 คำ ซึ่งวิธี HMM Acoustic Models ได้ใช้แบบจำลอง 2 แบบคือ Whole – word และ Phoneme – based เพื่อใช้ในการรู้จำ ในส่วนของ Overall HMM Structure จะมีแบ่งประเภทในการประมวลผลเป็นแบบ

1. Keyword “pieces” Alternate Model
2. Phoneme Alternate Model
3. Large Vocabulary Model
4. Large Vocabulary + Language Model

ขั้นตอนต่อมาใช้วิธี Scoring Method เป็นการคำนวณคะแนนโดยการใช้ทฤษฎีของ Baum – Welch ซึ่งทั้งระบบนี้จะใช้เวลาในการประมวลผลค่อนข้างมาก ผลการทดลองพบว่ามีความถูกต้องประมาณ 74.6 %

สำหรับงานวิจัยของ Bo – Ren Bai , Chiu – Yu Tseng และ Lin – Shan Lee (1997) เป็นการชี้เฉพาะคำสำคัญของเสียงพูดภาษาจีนกลาง (Mandarin) ในงานวิจัยนี้มีคำสำคัญที่ใช้เป็นจำนวนมาก โดยขั้นตอนต่าง ๆ ที่ใช้แบ่งออกเป็น 4 ขั้นตอน คือ

1. Syllable Boundaries Estimation
2. Acoustic Recognition
3. Keyword Matching
4. Re – scoring

ผลการทดลองของงานวิจัยนี้ซึ่งมีคำสำคัญถึง 2,611 คำ ซึ่งวิธี Syllable Boundaries Estimation เป็นการหาขอบเขตของพยางค์โดยดูจากค่าพลังงานของเสียง ในส่วนนี้จะทำการแยกแยะชนิดของขอบเขตซึ่งมีอยู่ 3 ชนิดคือ

1. Hard Syllable Boundaries
2. Soft Syllable Boundaries
3. Phrase Boundaries

ซึ่งการแยกแยะชนิดของขอบเขตเหล่านี้จะช่วยเพิ่มความเร็วในการคำนวณ ขึ้นต่อมาจะใช้วิธี Acoustic Recognition ซึ่งในงานวิจัยนี้ใช้ Context Independent, Initial และ Final Models ในส่วนถัดมาเป็นวิธี Keyword Matching และส่วนสุดท้ายคือวิธี Re – scoring ในส่วนนี้ใช้เพื่อเพิ่มความถูกต้องให้ระบบ ในงานวิจัยนี้มีความถูกต้อง 85.79 %

ในการเลือกวิธีการที่เฉพาะคำสำคัญไปใช้ในงานประยุกต์ต่าง ๆ จำเป็นต้องพิจารณาเงื่อนไขของการประยุกต์ใช้เป็นสำคัญ เนื่องจากแต่ละวิธีจะมีข้อดีและข้อเสียแตกต่างกันไป เช่น การนำเทคนิคการที่เฉพาะคำสำคัญไปใช้กับการสอบถามข้อมูลเกี่ยวกับการท่องเที่ยวในเขตพื้นที่ที่เป็นชายทะเล จะเห็นว่าปริมาณคำสำคัญที่ใช้นั้นจะมีปริมาณน้อย และการประยุกต์นี้ต้องการความเร็วในการประมวลผลสูง โดยอาจยอมให้เกิดความผิดพลาดขึ้นได้บ้าง ดังนั้นเราจึงเลือกใช้เทคนิค Acoustic Model เนื่องจากเป็นเทคนิคที่เหมาะสมกับการที่เฉพาะคำสำคัญในการประยุกต์ใช้งานแบบนี้ เช่น เทคนิคที่ใช้ในงานวิจัยของ Takeshi (1987) เพราะทำให้ผลการทดลองที่ดีและไม่จำเป็นต้องใช้เทคนิคอื่นมาช่วยเพิ่มความถูกต้องอันจะมีผลให้ต้องใช้เวลาในการคำนวณมากขึ้น เนื่องจากมีคำสำคัญไม่มาก

เพื่อให้การทดสอบระบบที่จะพัฒนาขึ้นทำได้สะดวกและง่าย งานวิจัยนี้กำหนดขอบเขตของการทดสอบไว้ในเรื่องการท่องเที่ยวในเขตพื้นที่ที่เป็นทะเลในประเทศไทย ดังนั้นจะมีการกำหนดคำสำคัญเป็นชื่อของจังหวัดในประเทศไทยและชื่อสถานที่ท่องเที่ยวในประเทศไทย จึงมีแนวคิดที่จะใช้วิธี Bottom – up Phoneme Detectors เพื่อใช้ในการตรวจสอบขั้นต้นในการที่เฉพาะคำสำคัญ เพราะวิธีนี้ใช้เวลาในการประมวลผลน้อย และในขั้นตอนนี้ไม่ต้องการความแม่นยำมากและใช้ Top – down Phoneme Verifiers เพื่อตรวจสอบคำสำคัญ ซึ่งในงานวิจัยของ Takeshi (1987) ใช้ Context – dependent features เป็น Triphone เพื่อทำการตรวจสอบคำสำคัญ เนื่องจากแบบจำลองเสียง Triphone ในแต่ละคำนั้น จะมีองค์ประกอบของแบบจำลองค่อนข้างมาก ซึ่งเป็นผลทำให้เสียเวลาในการคำนวณมากกว่าเมื่อเทียบกับแบบจำลองเสียง Onset และ Rhyme ซึ่งงานวิจัยในการรู้จำเสียงพูดภาษาไทยนั้น ใช้แบบจำลองเสียง Onset และ Rhyme ในระบบการรู้จำเสียงพูดต่อเนื่องพบว่าให้ผลการทดลองที่ดี มีความถูกต้องมากกว่า 90 % (วิศรุต อาชุนทร, 2545; Maneenoi, E., 2002) จึงมีความเห็นว่าใช้แบบจำลองเสียง Onset และ Rhyme ในขั้นตอนการตรวจสอบคำสำคัญ เพื่อนำผลมาเปรียบเทียบกับความเร็วในการประมวลผลและความแม่นยำของระบบการที่เฉพาะคำสำคัญ และนำวิธีเหล่านี้มาสร้างระบบการที่เฉพาะคำสำคัญเสียงพูดภาษาไทยบนพื้นฐานของการตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์

การค้นคว้าวิจัยในด้านการรู้จำเสียงพูดภาษาไทยที่มีในประเทศไทยนั้น ได้มีการศึกษาค้นคว้าวิจัยมาอย่างต่อเนื่องดังจะเห็นได้จากผลงานวิจัยต่าง ๆ เช่น การตรวจรู้จำเสียงพูดภาษาไทย (ทวิ ประทุมทาน, 2530) ระบบการรับรู้เสียงพูดแบบต่างบุคคล (ไพศาล ธรรมโพธิทอง, 2533)

การรู้จำเสียงพูดตัวเลขเป็นภาษาไทย แบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดน มาร์คอฟ โมเดลและ
 เวกเตอร์ควอนไทซ์เซชัน (เสาวลักษณ์ อารีพงศา, 2538) การรู้จำเสียงพูดสระภาษาไทยโดด ๆ ไม่
 ขึ้นกับผู้พูดโดยการวัดสเปกตรัมดิสแตนซ์และใช้ไดนามิกไทม์วาร์ปิง (ระพีพัฒน์ เพ็ญศิริ, 2538)
 การรู้จำคำพูดภาษาไทยโดยใช้ลักษณะบ่งความต่างของหน่วยเสียง (ณัฐกร ทับทอง, 2538) ระบบ
 การรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (วิศรุต อาชุนทร,
 2539) การรู้จำหน่วยเสียงสระภาษาไทยโดยใช้โครงข่ายประสาทเทียม (เอกฤทธิ์ มณีน้อย, 2541)
 ระบบบ่งชี้ผู้พูดแบบระบบเปิดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบหลายชุดรหัส (พงศ์ไท ทา
 สะครู, 2542) ระบบการรู้จำผู้พูด : การบ่งชี้ผู้พูดแบบขึ้นกับบทพูด (ศวิต กาสุริยะ, 2542) การรู้จำ
 เสียงพูดภาษาไทยอย่างคงทนโดยใช้สัมประสิทธิ์เมลฟรีเควินซีเซปสตรอลของค่าอัตราสัมพันธ์
 ของเสียงพูดที่มีสัญญาณรบกวน (เจษฎา กานต์ประชา, 2545) และการศึกษาหน่วยเริ่มของ
 พยางค์เชิงกลศาสตร์ : พื้นฐานสำหรับการรู้จำเสียงพูดต่อเนื่องภาษาไทย (วิศรุต อาชุนทร,
 2545) เป็นต้น โดยงานวิจัยเหล่านี้ล้วนเป็นการวิจัยเกี่ยวกับการรู้จำเสียงพูดภาษาไทย ดังนั้น
 งานวิจัยนี้จึงทำการวิจัยเกี่ยวกับการชี้เฉพาะคำสำคัญภาษาไทยเพื่อเพิ่มความยืดหยุ่นของระบบ
 การรู้จำและยังเป็นแนวทางในการประยุกต์ใช้งานในชีวิตประจำวันที่เกี่ยวข้องกับการรู้จำเสียงพูด
 ซึ่งระบบการชี้เฉพาะคำสำคัญภาษาไทยนี้ยังไม่ได้มีการทำวิจัยมาก่อน งานวิจัยนี้อาศัยหลักการที่
 ได้มาจากการวิจัยของ Takeshi (1987) ซึ่งได้แนวความคิดของระบบชี้เฉพาะคำสำคัญ โดยการ
 นำขั้นตอนการตรวจสอบคำสำคัญมาประยุกต์ใช้ และในงานวิจัยระบบบ่งชี้ผู้พูดแบบระบบเปิด
 โดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบหลายชุดรหัส (พงศ์ไท ทาสะครู, 2542) ซึ่งได้นำการ
 วิเคราะห์การสกัดลักษณะสำคัญของเสียงภาษาไทยมาประยุกต์ใช้ และงานวิจัยการศึกษาหน่วย
 เริ่มของพยางค์เชิงกลศาสตร์ : พื้นฐานสำหรับการรู้จำเสียงพูดต่อเนื่องภาษาไทย (วิศรุต อาชุน
 ทร, 2545) ซึ่งได้นำแบบจำลองฮิดเดน มาร์คอฟของหน่วยเริ่มและหน่วยตามของพยางค์ (Onset
 – Rhyme) มาประยุกต์ใช้ในการชี้เฉพาะคำสำคัญภาษาไทย เพื่อให้สามารถชี้เฉพาะคำสำคัญ
 ภาษาไทยจากเสียงพูดต่อเนื่องภาษาไทยได้ ด้วยการปรับเปลี่ยนโครงสร้างระบบให้สามารถชี้
 เฉพาะคำสำคัญบนความหลากหลายของข้อมูลเสียงพูดได้

1.2 แนวทางที่นำเสนอ

ในวิทยานิพนธ์นี้เสนอ การรู้จำโดยใช้แบบจำลองฮิดเดน มาร์คอฟ โดยใช้แบบจำลองของ
 เสียงระดับย่อยของพยางค์ประเภท Monophone เพื่อใช้ในการรู้จำ และใช้แบบจำลองของหน่วย
 เริ่มของพยางค์ (Onset) และหน่วยตามของพยางค์ (Rhyme) ใช้ตรวจสอบคำสำคัญในเสียงพูด
 ต่อเนื่องภาษาไทย เพื่อเพิ่มความถูกต้อง และความเร็วในการประมวลผล ของระบบชี้เฉพาะคำ
 สำคัญในภาษาไทย

1.3 วัตถุประสงค์

1. เพื่อพัฒนาระบบการรู้จำเสียงพูดภาษาไทย โดยการที่เฉพาะคำสำคัญบนพื้นฐานของการตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์
2. เพื่อสร้างระบบการรู้จำเสียงพูดภาษาไทยบนพื้นฐานของการตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์
3. เพื่อให้เครื่องคอมพิวเตอร์สามารถรู้จำคำที่เฉพาะภาษาไทยได้
4. เพื่อเป็นแนวทางสำหรับการศึกษาวิธีการรู้จำวิธีการอื่น ๆ

1.4 ขอบเขตวิทยานิพนธ์

1. สร้างระบบการรู้จำเสียงพูดภาษาไทยบนพื้นฐานของการตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์ซึ่งคำสำคัญเป็นชื่อจังหวัดในประเทศไทยและชื่อสถานที่ท่องเที่ยวในประเทศไทย
2. เปรียบเทียบประสิทธิภาพของระบบเมื่อกำหนดให้จำนวนคำสำคัญในระบบแตกต่างกันไม่เกิน 100 คำ
3. เปรียบเทียบประสิทธิภาพของระบบเมื่อใช้ Triphone และ Onset – Rhyme ในส่วนของขั้นตอนการตรวจสอบคำสำคัญ

1.5 ขั้นตอนและวิธีดำเนินงาน

1. ศึกษา ค้นคว้าและเก็บรวบรวมข้อมูลรายละเอียดดังต่อไปนี้
 - คุณลักษณะของเสียงพูด
 - การวิเคราะห์และวัดค่าลักษณะสำคัญ
 - การประยุกต์ใช้การรู้จำคำสำคัญภาษาไทยในการรู้จำเสียงพูด
2. การวิเคราะห์ระบบการรู้จำเสียงพูดภาษาไทย
 - ออกแบบบทพูดเพื่อนำมาใช้เก็บตัวอย่างเสียง
 - เก็บข้อมูลเสียงพูด
 - พัฒนาโปรแกรมในการวิเคราะห์การรู้จำคำสำคัญภาษาไทย
3. สร้างระบบการรู้จำเสียงพูดภาษาไทย
 - พัฒนาโปรแกรมในการฝึกฝนแบบจำลองของคำสำคัญ
 - พัฒนาโปรแกรมในการวิเคราะห์ค่าความน่าจะเป็นของผลลัพธ์ที่ได้
4. ทำการทดสอบ แก้ไขและประเมินผลระบบ
5. สรุปรวบรวมผลการวิจัยทั้งหมด พร้อมทั้งจัดทำเอกสารเกี่ยวกับวิทยานิพนธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงวิธีการชี้เฉพาะคำสำคัญภาษาไทย
2. ทราบถึงกรรมวิธีที่เหมาะสมในการชี้เฉพาะคำสำคัญภาษาไทย
3. เป็นแนวทางในการพัฒนาระบบการชี้เฉพาะคำสำคัญภาษาไทยจำนวนคำศัพท์มาก

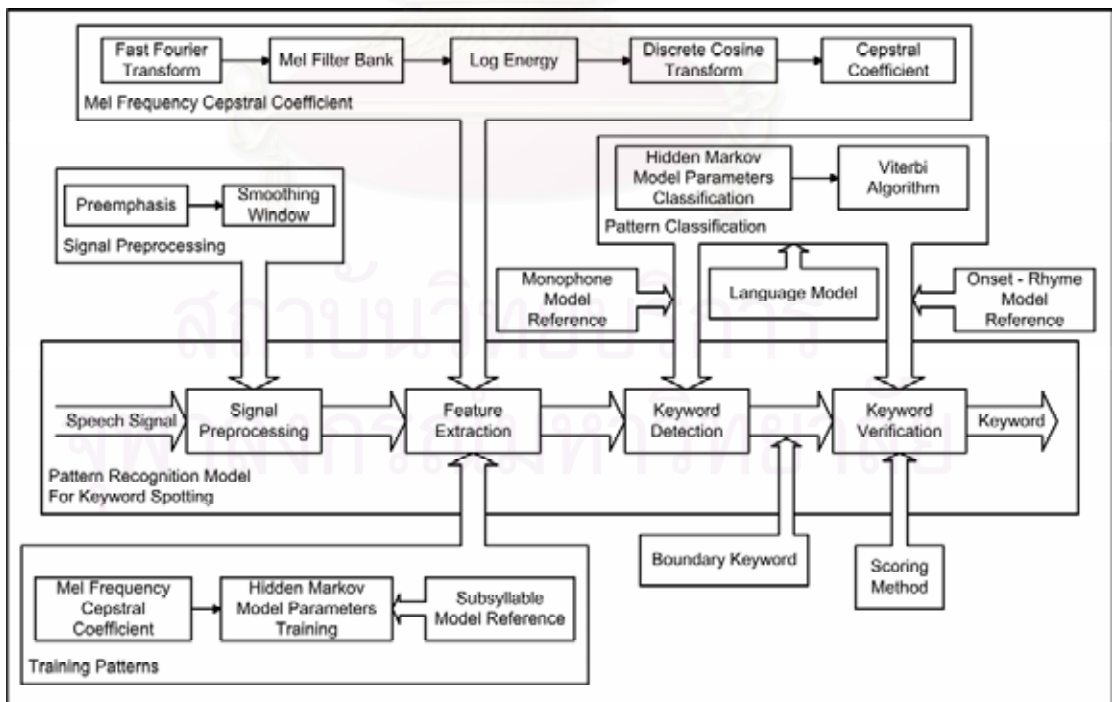


สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

ความรู้พื้นฐานและทฤษฎีที่เกี่ยวข้อง

รายละเอียดในบทที่ 2 นี้ จะกล่าวถึงทฤษฎีต่าง ๆ ที่เกี่ยวข้องและอ้างอิงถึงในงานวิจัยนี้ โดยแบ่งออกเป็น 4 ขั้นตอนดังรูปที่ 2.1 โดยขั้นตอนแรกเริ่มจากทฤษฎีที่เกี่ยวข้องกับการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) ขั้นตอนที่สองเป็นขั้นตอนการวิเคราะห์และวัดค่าลักษณะสำคัญ (Feature Measurement) ที่เกี่ยวข้องกับทฤษฎีของสัมประสิทธิ์เมลฟรีควีนซีเซปสตรอล ขั้นตอนที่สามเป็นการจำแนกรูปแบบ (Pattern Classification) เกี่ยวข้องกับทฤษฎีของการเปรียบเทียบรูปแบบที่นำมาทดสอบกับต้นแบบที่มีอยู่ โดยใช้หลักการของแบบจำลองฮิดเดนมาร์คอฟ และขั้นตอนสุดท้ายเป็นขั้นตอนการตัดสินใจ (Decision Algorithm) ในการเลือกต้นแบบที่ตรงกับรูปแบบที่นำมาทดสอบมากที่สุดโดยอาศัยวิธีการของ Viterbi (Viterbi Algorithm) และยังมีความรู้ทางคุณลักษณะด้านกลศาสตร์ของเสียงพูดภาษาไทย การตรวจหา กลุ่มคำและประเภทแบบจำลองของเสียงระดับย่อย ซึ่งในหัวข้อนี้ได้รวมถึงทฤษฎี Bottom - up Detector และ Top - Down Verifier (Kawabata, T. and Kohda, M., 1987) และหลักการของแบบจำลองภาษา (Language Model)



รูปที่ 2.1 รายละเอียดของแบบจำลองรูปแบบการชี้เฉพาะคำสำคัญภาษาไทย

ก. การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนกรรมวิธีในการจัดเตรียมข้อมูล จากข้อมูลดิบของเสียงพูดที่ได้จากการบันทึกเสียงนำมาผ่านกรรมวิธีประมวลผลสัญญาณเชิงเลข เพื่อใช้ในการประมวลผลในขั้นตอนต่อไป เนื่องจากสัญญาณเสียงพูดโดยรวมจะแปรเปลี่ยนตามเวลาและไม่เสถียร (Nonstationary) ดังนั้นในการประยุกต์ใช้งานกรรมวิธีประมวลผลสัญญาณเชิงเลขกับสัญญาณเสียงพูด จึงต้องแบ่งสัญญาณเสียงพูดในแต่ละกรอบเสียงพูดมีความเสถียรและไม่แปรเปลี่ยนตามเวลา (Stationary) จากนั้นจึงสามารถทำการประมวลผลสัญญาณเชิงเลขกับสัญญาณเสียงพูดในแต่ละกรอบเสียงพูดได้

กรรมวิธีในการประมวลผลสัญญาณเบื้องต้นแบ่งได้เป็น 2 กรรมวิธีย่อย ได้แก่ กรรมวิธีเน้นล่วงหน้า (Preemphasis) และกรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

1. กรรมวิธีเน้นล่วงหน้า (Preemphasis)

ขั้นตอนการเน้นล่วงหน้าเป็นการบีบอัดช่วงพิสัยพลวัต (Dynamic Range) ของสัญญาณเสียงพูด โดยการให้ความลาดเอียงในเชิงความถี่แบบราบลงซึ่งจะส่งผลให้อัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าสูงขึ้น ในทางปฏิบัติแล้วจะนำสัญญาณผ่านตัวกรองเชิงเลขอันดับหนึ่ง (First Order Digital Filter) ที่มีฟังก์ชันถ่ายโอนดังแสดงในสมการที่ (2-1) และ (2-2) (Furui, S., 1985) เมื่อ a เป็นสัมประสิทธิ์ตัวกรอง, $\tilde{s}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรรมวิธีเน้นล่วงหน้าที่ n และ $s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n และ $s(n-1)$ เป็นค่าสัญญาณเสียงพูดขาเข้าค่าก่อนหน้าที่ $n-1$ ดังสมการ

$$H(z) = 1 - az^{-1} \dots\dots\dots (2-1)$$

$$\tilde{s}(n) = s(n) - as(n-1) \dots\dots\dots (2-2)$$

โดยกำหนดให้ค่าสัมประสิทธิ์ของตัวกรอง a มีค่าเข้าใกล้ 1 เมื่อใช้ร่วมกับการวิเคราะห์หาค่าลักษณะสำคัญของเสียง จะกำหนดให้ค่า $a = 0.95$ เนื่องจากเป็นค่าที่ให้ผลดีที่สุดสำหรับการวิเคราะห์ (วิศรุต อาชูปุตร, 2539; เสาวลักษณ์ อารีพงศา, 2538; เอกฤทธิ์ มณีน้อย, 2541)

2. กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

ขั้นตอนกรรมวิธีวางกรอบขนาดสัญญาณจัดเป็นขั้นตอนในการเตรียมข้อมูลในแต่ละกรอบข้อมูลเสียงพูดเพื่อการวิเคราะห์ โดยการคูณแต่ละค่าของสัญญาณในกรอบข้อมูลเสียงพูดด้วยค่าฟังก์ชันกรอบ (Window Function) ผลของการวางกรอบขนาดสัญญาณมี 2 ประการ โดยประการแรกเป็นการลดทอนแอมพลิจูดอย่างช้า ๆ ที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อป้องกันการเปลี่ยนแปลงอย่างกะทันหันที่จุดปลาย และประการที่สอง เป็นการสร้างค่า

การประสานสำหรับการผลแปลงฟูรีเยร์ของฟังก์ชันกรอบและแถบสเปกตรัมของเสียงพูด สำหรับการวิเคราะห์เสียงพูดในงานวิจัยนี้จะกำหนดให้ใช้ค่าฟังก์ชันกรอบชนิด Hamming Window ซึ่งค่าฟังก์ชันดังแสดงในรูปที่ 2.2 ซึ่งกรรมวิธีการวางกรอบสัญญาณจะเป็นไปตามสมการที่ (2-3) และ (2-4) (Furui, S., 1985)

$$\tilde{x}_l(n) = x_l(n) \cdot w(n) \dots\dots\dots(2-3)$$

$$w(n) = 0.54 - 0.46 \cos\left[\frac{2\pi n}{N-1}\right] \dots\dots\dots(2-4)$$

เมื่อ $l = 0, 1, \dots, L-1$ และ $n = 0, 1, \dots, N-1$

กำหนดให้ $x(n)$ คือค่าสัญญาณเสียงพูดของข้อมูลที่ n

$\tilde{x}(n)$ คือค่าสัญญาณเสียงพูดที่ผ่านกรรมวิธีวางกรอบ

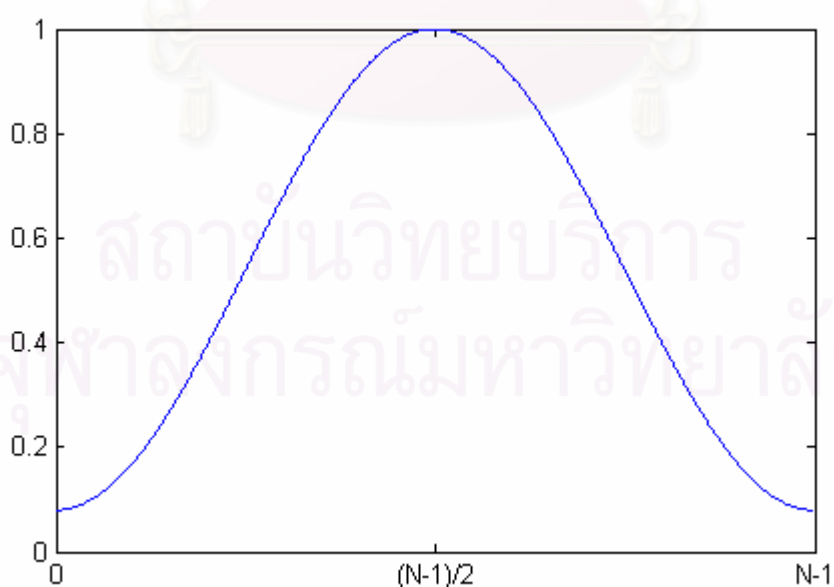
$w(n)$ คือฟังก์ชันกรอบชนิด Hamming Window

N คือจำนวนของข้อมูลในแต่ละกรอบเสียงสัญญาณพูด

n คือลำดับข้อมูลในกรอบเสียงสัญญาณพูดที่ l

L คือจำนวนของกรอบสัญญาณเสียงพูด

l คือลำดับของกรอบสัญญาณเสียงพูด



รูปที่ 2.2 ฟังก์ชันกรอบชนิด Hamming Window

ข. การสกัดลักษณะสำคัญ (Feature Extraction)

การวิเคราะห์และวัดค่าลักษณะสำคัญเป็นการวิเคราะห์สัญญาณเสียงพูด เพื่อเก็บรวบรวมลักษณะสำคัญของเสียงพูดแต่ละเสียง สำหรับการฝึกฝนระบบให้รับรู้ถึงความแตกต่างของเสียงพูดแต่ละเสียงและเพื่อใช้ในการเปรียบเทียบแยแ่กความแตกต่างของเสียงพูดแต่ละเสียงออกจากกัน

จากการศึกษาพบว่างานวิจัยหลายงาน (Knill, K.M. and Young, S.J., 1996; Nitta, T., Tanaka, S., Masai, Y., et al., 1996) ได้ใช้สัมประสิทธิ์การประมาณเชิงเส้น หรือ LPC เป็นลักษณะสำคัญ นอกจากนี้ก็มี สัมประสิทธิ์ Cepstral และ สัมประสิทธิ์ MFCC (Mel Frequency Cepstral Coefficient) โดยใช้งานร่วมกับแบบจำลองฮิดเดนมาร์คอฟ และให้อัตราการรู้จำที่ดีกับงานได้รู้จำคำพูด (Ahkuputa, V., Jitapunkul, S., Maneenoi, E. et al., 1998) ซึ่งเป็นงานที่มีความคล้ายคลึงกับงานบ่งชี้ผู้พูด ณ ระดับหนึ่ง ในงานวิจัยระบบรู้จำคำพูดภาษาไทยที่ผ่านมาจะเห็นว่า ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลให้ผลดีที่สุด เมื่อเปรียบเทียบกับสัมประสิทธิ์การประมาณเชิงเส้นและสัมประสิทธิ์เซปสตรอล (พงศ์ไท ทาสระคุ, 2542) ดังนั้นการวัดค่าลักษณะสำคัญในงานวิจัยนี้จะใช้วิธีการหาสัมประสิทธิ์เซปสตรอลบนความถี่เมล

สัมประสิทธิ์เซปสตรอลบนความถี่เมล (Mel Frequency Cepstral Coefficients, MFCC)

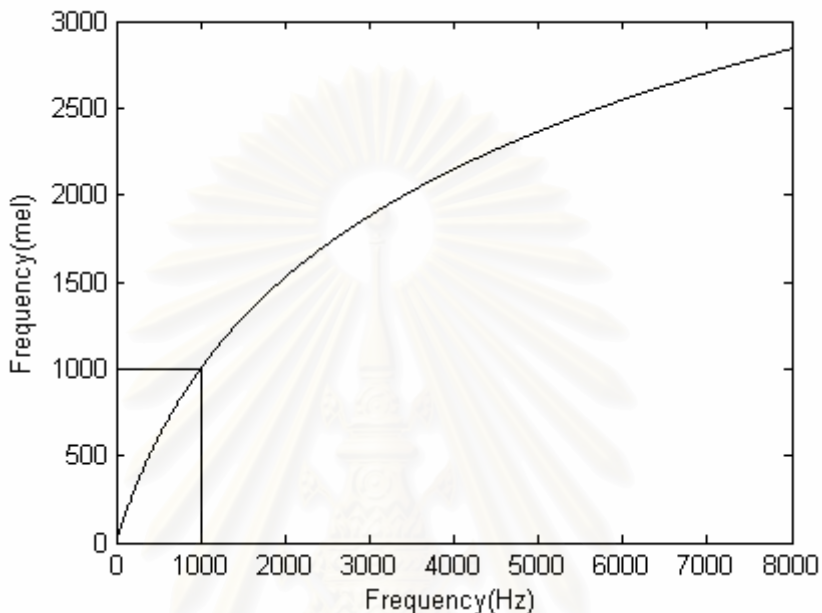
สัมประสิทธิ์เซปสตรอลบนความถี่เมลคำนวณได้จากการวิเคราะห์โดยไม่ใช้พารามิเตอร์ (Non-Parameter Analysis) คือ ไม่มีการสร้างแบบจำลองสัญญาณเพื่อปรับค่าพารามิเตอร์ในแบบจำลองให้ได้เป็นสัญญาณที่เราต้องการ แต่เป็นการนำสัญญาณผ่านวงจรกรองแบบผ่านแถบความถี่ (Band Pass Filter) หลายวงจร โดยแต่ละวงจรกรองมีช่วงความถี่ที่ผ่านได้แตกต่างกัน และวงจรกรองที่ใช้เป็นชุดวงจรกรองแบบดิจิทัล (Digital Filter Bank) ที่ใช้วิเคราะห์สเปกตรัมของเสียงสัญญาณพูด โดยเลียนแบบตามการได้ยินของมนุษย์ สามารถทำได้หลายวิธีขึ้นอยู่กับกรรมวิธีที่นำมาใช้ได้แก่ (Tolba, H. and O'Shaughnessy, D., 1998)

- Uniform Spacing (Fourier Transform)
- Exponential Spacing (Wavelet Transform)
- Perceptually-Derived Spacing (Mel Scale หรือ Bark Scale)

ในงานวิจัยนี้ได้เลือกใช้กรรมวิธี Perceptually-Derived Spacing หรือเรียกอีกชื่อหนึ่งว่า Mel Scale เป็นการเปลี่ยนแกนความถี่สัญญาณเสียง (Acoustic Frequency) ไปเป็นแกนความถี่ของการได้ยิน (Perceptual Frequency) จึงจะกล่าวถึงแต่กรรมวิธีเท่านั้น Mel Scale ใช้หลักการ Mapping จากการประมาณความถี่ช่วงที่เป็นเชิงเส้น (Linear Frequency Scale) เป็น

ความถี่ที่ไม่เชิงเส้น (Nonlinear frequency Scale) ตามลักษณะการได้ยินของมนุษย์ดังแสดงในสมการที่ (2-5) (Tolba, H. and O'Shaughnessy, D., 1998)

$$mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \dots\dots\dots(2-5)$$



รูปที่ 2.3 Mel Scale ของความถี่สัญญาณเสียง

เมื่อ f คือช่วงความถี่ที่ยังเป็นเชิงเส้น จากรูปที่ 2.3 เห็นได้ว่าลักษณะการได้ยินของมนุษย์ในช่วงที่เป็นเชิงเส้น (Linear Scale) คือประมาณช่วงความถี่ตั้งแต่ 0 ถึง 1000 Hz และช่วงที่ไม่เป็นเชิงเส้น (Logarithmic Scale) คือช่วงความถี่ตั้งแต่ 1000 Hz ขึ้นไป แบบดิวัดท์ของวงจรรองสามารถคำนวณได้จากสมการที่ (2-6) ดังนี้

$$BW_{critical} = 25 + 75 \left[1 + 1.4 \left(\frac{f}{1000} \right)^2 \right]^{0.69} \dots\dots\dots(2-6)$$

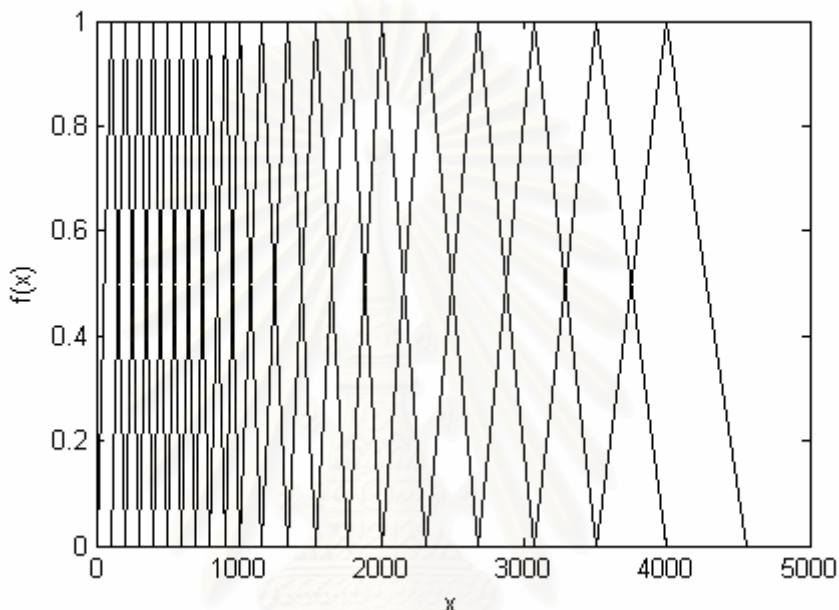
วิธีการคำนวณแบบจำลองของชุดวงจรรอง (Filter Bank Model) ที่ง่ายที่สุดและให้ผลที่มีประสิทธิภาพดีคือการแปลงแบบฟูริเยร์ (Fourier Transform) ของสัญญาณ และหาข้อมูลออกของชุดวงจรรองของแต่ละวงจรรองจากความถี่ที่ใส่เข้าไป การคำนวณหาข้อมูลออกของชุดวงจรรองสามารถคำนวณได้ตามสมการที่ (2-7)

$$X_N = \frac{1}{N_{FB}} \sum_{k=1}^{N_{FB}} \omega_{FB}(k) X(f + \delta f(f, k)) \dots\dots\dots(2-7)$$

เมื่อ N_{FB} คือจำนวนของตัวอย่างที่ใช้ในการหาค่าเฉลี่ย

ω_{FB} คือฟังก์ชันการถ่วงน้ำหนัก

$\delta f(f, k)$ คือฟังก์ชันที่อธิบายถึงความถี่ที่อยู่ใกล้เคียงกับความถี่ f เพื่อนำมาใช้คำนวณหาค่าเฉลี่ย



รูปที่ 2.4 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล

ในการคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลใช้ค่าข้อมูลออกที่ได้จากวงจรกรองแถบผ่าน N วงจรเมื่อใส่ค่าล็อกของขนาดสเปกตรัมของสัญญาณเสียงเข้าไป เพื่อประมาณความถี่ตอบสนองของ Basilar Membrane ใน Cochlea ที่อยู่ในหูชั้นในวงจรกรองที่นำมาใช้นี้เป็นวงจรกรองแบบสามเหลี่ยมดังแสดงในรูปที่ 2.4 เพราะฉะนั้นการคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล (C_n) สามารถหาได้จากสมการดังต่อไปนี้ (Tolba, H. and O'Shaughnessy, D., 1998)

$$C_n = \sum_{k=1}^N X_k \cos\left(\frac{\pi n}{N}(k - 0.5)\right), \quad n = 1, 2, \dots, M \dots\dots\dots(2-8)$$

เมื่อ M คือจำนวนของสัมประสิทธิ์เซปสตรอล

N คืออันดับในการวิเคราะห์

X_k คือค่าล็อกของพลังงานที่ค่าข้อมูลออกของวงจรกรอง k วงจร เมื่อ $k = 1, 2, \dots, N$

ค. การทดสอบความคล้ายคลึงกันของรูปแบบ (Pattern Similarity Testing)

การทดสอบความคล้ายคลึงกันของรูปแบบ (Pattern Similarity Testing) หรือการจำแนกรูปแบบ (Pattern Classification) เกี่ยวข้องกับการเปรียบเทียบรูปแบบเสียงพูดระหว่างคำพูดหรือวลีที่ไม่ทราบรูปแบบกับรูปแบบที่ได้จัดเก็บไว้แล้ว ซึ่งก็คือชุดรูปร่างต้นแบบ (Templates) หรือแบบจำลอง (Models) ของเสียงพูด ภายหลังจากการเปรียบเทียบแต่ละครั้งจะได้ค่าความไม่คล้ายคลึงกัน (Dissimilarity Scores) หรือค่าระยะทาง (Distance Scores) เพื่อใช้ในขั้นตอนวิธีการตัดสินใจในการเลือกรูปแบบที่เหมาะสมใกล้เคียงที่สุดต่อไป

ขั้นตอนวิธีการในการจำแนกรูปแบบที่ถูกนำมาใช้มากที่สุด สามารถแบ่งออกได้เป็น 4 วิธีการ (Rabiner, L.R., 1989) ได้แก่ การเข้าคู่ต้นแบบ (Template Matching) ระบบตามกฎเกณฑ์ (Rule-Based System) ระบบแบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Model, HMM) และเครือข่ายประสาท (Neural Network) แต่เนื่องจากงานวิจัยนี้ใช้เทคนิคระบบแบบจำลองฮิดเดนมาร์คอฟ จึงจะกล่าวถึงแต่เฉพาะรายละเอียดและทฤษฎีของวิธีการจำแนกรูปแบบด้วยระบบแบบจำลองฮิดเดนมาร์คอฟเท่านั้น

ง. ระบบแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model)

ระบบแบบจำลองฮิดเดนมาร์คอฟนี้ถือเป็นขั้นตอนวิธีการจำแนกรูปแบบที่ดีที่สุดวิธีการหนึ่งที่มีอยู่ในขณะนี้โดยอาศัยวิธีการทางสถิติ (Rabiner, L.R., 1989) ขั้นตอนวิธีการนี้มีข้อได้เปรียบที่สำคัญเหนือวิธีการเข้าคู่ต้นแบบก็คือสามารถเก็บข้อมูลรายละเอียดในทางสถิติเกี่ยวกับเสียงพูดไว้ได้มากกว่าวิธีการเข้าคู่ต้นแบบ โดยเก็บข้อมูลการกระจายที่สมบูรณ์ของลักษณะสำคัญที่มีอยู่ในข้อมูลฝึกฝน จึงสามารถจำแนกความแตกต่างระหว่างเสียงพูดได้ดีมากยิ่งขึ้น อีกทั้งขั้นตอนวิธีการนี้ยังอาศัยการโปรแกรมแบบพลวัต (Dynamic Programming) ทำให้มีความรวดเร็วในการประมวลผลมากยิ่งขึ้น

แบบจำลองฮิดเดนมาร์คอฟ เป็นแบบจำลองที่ใช้กระบวนการทางด้านการสถิติมาสร้างแบบจำลองของสัญญาณ แบบจำลองนี้ได้รับการพัฒนาเป็นเวลานานแล้ว เริ่มแรกที่มีการเผยแพร่ผลงานเมื่อประมาณปี 1960-1970 โดย Baum และผู้ร่วมงาน จากนั้นก็ได้มีการนำความรู้พื้นฐานของแบบจำลองฮิดเดนมาร์คอฟไปประยุกต์ใช้กับงานด้านต่างๆ โดยเฉพาะ งานด้านการรู้จำเสียงพูด ได้มีการพัฒนาแบบจำลองฮิดเดนมาร์คอฟมาประยุกต์ใช้ในการรู้จำเสียงพูดด้วย เนื่องจากผลของการนำแบบจำลองนี้มาใช้ให้อัตราการรู้จำที่ดีเมื่อเทียบกับการใช้ระบบรู้จำแบบอื่น ซึ่งมีงานวิจัยจำนวนมากที่นิยมใช้แบบจำลองของฮิดเดนมาร์คอฟในการรู้จำเสียงพูด (เสาวลักษณ์ อารีพงศา, 2538; วิศรุต อาชูปุตระ, 2539; พงศ์ไท ทาสรักษา, 2542; ศวิต กาศุริยะ, 2542)

1. องค์ประกอบของแบบจำลองฮิดเดนมาร์คคอฟ

องค์ประกอบของแบบจำลองฮิดเดนมาร์คคอฟ ประกอบไปด้วยพารามิเตอร์ต่าง ๆ ดังนี้

- 1) N คือจำนวนสถานะที่อยู่ภายในแบบจำลอง ซึ่งโดยทั่วไปแล้วแต่ละสถานะจะเชื่อมโยงถึงกันด้วยวิธีการที่ทำให้สถานะใด ๆ สามารถเข้าถึงสถานะอื่น ๆ ได้ แต่ละสถานะแสดงได้ด้วย $S = \{S_1, S_2, \dots, S_N\}$ โดยมีสถานะที่เวลา t แสดงได้ด้วย q_t
- 2) M คือจำนวนสัญลักษณ์ของค่าสังเกตต่อสถานะ ซึ่งสัญลักษณ์ของค่าสังเกตจะสัมพันธ์กับผลลัพธ์ขาออกทางกายภาพของระบบที่ถูกจำลอง แต่ละสัญลักษณ์สามารถแสดงได้ด้วย $V = \{v_1, v_2, \dots, v_M\}$
- 3) การกระจายของความน่าจะเป็นในการเปลี่ยนแปลงสถานะ (State Transition Probability Distribution) $A = \{a_{ij}\}$ เมื่อ

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N \dots\dots\dots(2-9)$$

ในกรณีเฉพาะที่สถานะใด ๆ สามารถเข้าถึงสถานะอื่นได้ภายในขั้นตอนเดียวจะกำหนดให้ $a_{ij} > 0, \forall i, j$ ส่วนในกรณีอื่นนอกเหนือจากนี้จะกำหนดให้ $a_{ij} = 0$ สำหรับ (i, j) เพียงคู่เดียวหรือมากกว่า

- 4) การกระจายความน่าจะเป็นของค่าสังเกต (Observation Probability Distribution) มี 2 ชนิด คือ

- การกระจายของความน่าจะเป็นของสัญลักษณ์ของค่าสังเกต (Observation Symbol Probability Distribution) $B = \{b_j(k)\}$ ในสถานะที่ j เมื่อ

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j], \quad \begin{matrix} 1 \leq j \leq N \\ 1 \leq k \leq M \end{matrix} \dots\dots\dots(2-10)$$

การกระจายความน่าจะเป็นของค่าสังเกตชนิดนี้มีลักษณะการกระจายแบบไม่ต่อเนื่อง (Discrete Distribution) อาศัยหลักการควอนไทซ์แบบเวกเตอร์และใช้สัญลักษณ์แทนลำดับค่าสังเกต

- การกระจายความน่าจะเป็นของค่าสังเกตแบบเกาส์มิกเจอร์ (Gaussian mixture distribution) $B = \{b_j(\mathbf{O}_t)\}$ ในสถานะที่ j เมื่อ

$$b_j(\mathbf{O}_t) = \sum_{m=1}^M c_{jm} N(\mathbf{O}_t; \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}) \dots\dots\dots(2-11)$$

การกระจายความน่าจะเป็นของค่าสังเกตชนิดนี้มีลักษณะการกระจายแบบต่อเนื่อง (Continuous Distribution) เป็นการกระจายแบบเกาส์ ประกอบด้วยการกระจายแบบเกาส์ที่มีค่าพารามิเตอร์แตกต่างกันตามจำนวนมิถุนอร์

5) การกระจายของสถานะเริ่มต้น (Initial State Distribution) $\pi = \{\pi_i\}$ เมื่อ

$$\pi_i = P[q_t = S_i], \quad 1 \leq i \leq N \quad \text{.....(2-12)}$$

โดยการกำหนดค่าที่เหมาะสมให้กับองค์ประกอบ N, M, A, B, π ของแบบจำลองฮิดเดนมาร์คอฟ ซึ่งใช้ในการกำเนิดลำดับค่าสังเกต เมื่อแต่ละค่าสังเกต O_t เป็นสัญลักษณ์ที่ได้จาก V และ T เป็นจำนวนค่าสังเกตทั้งหมดที่มีในลำดับ ซึ่งมีขั้นตอนวิธีการดังนี้

$$O = \{O_1, O_2, \dots, O_T\} \quad \text{.....(2-13)}$$

ตารางที่ 2.1 รายละเอียดของขั้นตอนในการกำเนิดลำดับค่าสังเกต

ขั้นตอนที่ 1	เลือกสถานะเริ่มต้น $q_1 = S_i$ ที่สัมพันธ์กับการกระจายของสถานะเริ่มต้น π
ขั้นตอนที่ 2	กำหนดให้ $t = 1$
ขั้นตอนที่ 3	เลือก $O_t = v_k$ ที่สัมพันธ์กับการกระจายของความน่าจะเป็นของสัญลักษณ์เมื่ออยู่ในสถานะ S_i เช่น $b_i(k)$
ขั้นตอนที่ 4	เคลื่อนย้ายไปยังสถานะใหม่ $q_{t+1} = S_j$ ที่สัมพันธ์กับการกระจายของความน่าจะเป็นในการเปลี่ยนแปลงสถานะสำหรับสถานะ S_i เช่น a_{ij}
ขั้นตอนที่ 5	กำหนดให้ $t = t + 1$ แล้วกลับไปทำซ้ำขั้นตอนที่ 3 ใหม่ถ้า $t < T$ นอกเหนือจากนี้ให้ยุติกระบวนการ

ขั้นตอนดังกล่าวนี้สามารถได้ทั้งการกำเนิดค่าสังเกต และเป็นแบบจำลองเพื่อบอกถึงความเหมาะสมในการกำเนิดลำดับค่าสังเกตด้วยแบบจำลองฮิดเดนมาร์คอฟ ดังนั้นการกำหนดคุณสมบัติเฉพาะของแบบจำลองฮิดเดนมาร์คอฟต้องการคุณสมบัติเฉพาะของพารามิเตอร์ของแบบจำลองสองค่า (N และ M) คุณสมบัติเฉพาะของสัญลักษณ์ของค่าสังเกต และคุณสมบัติเฉพาะของการวัดค่าความน่าจะเป็นได้แก่ A, B, π โดยเขียนอยู่ในรูปแบบย่อเพื่อบ่งบอกชุดของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองดังนี้

$$\lambda = (A, B, \pi) \quad \text{.....(2-14)}$$

2. ปัญหาพื้นฐานสามประการของแบบจำลองฮิดเดนมาร์คคอฟ

ในการประยุกต์ใช้งานแบบจำลองฮิดเดนมาร์คคอฟในทางปฏิบัตินั้น ก็คือการแก้ไข ปัญหาพื้นฐานทั้งสามประการ โดยมีรายละเอียดดังแสดงในตารางที่ 2.2

ตารางที่ 2.2 รายละเอียดของปัญหาพื้นฐานสามประการของแบบจำลองฮิดเดนมาร์คคอฟ

ปัญหาพื้นฐานข้อที่ 1	เมื่อกำหนดลำดับค่าสังเกต $O = \{O_1, O_2, \dots, O_T\}$ และแบบจำลอง $\lambda = (A, B, \pi)$ จะทำการหาค่าความน่าจะเป็นของลำดับค่าสังเกต $P(O \lambda)$ ตามแบบจำลองที่กำหนดให้ได้อย่างไร
ปัญหาพื้นฐานข้อที่ 2	เมื่อกำหนดลำดับค่าสังเกต $O = \{O_1, O_2, \dots, O_T\}$ และแบบจำลอง λ จะทำการเลือกลำดับสถานะที่สัมพันธ์กับ $Q = q_1 q_2 \dots q_T$ ซึ่งมีความเหมาะสมที่สุดกับแบบจำลองที่กำหนดให้ได้อย่างไร
ปัญหาพื้นฐานข้อที่ 3	จะทำการปรับค่าพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ ใดๆ เพื่อให้ค่าความน่าจะเป็นของลำดับค่าสังเกต $P(O \lambda)$ มีค่ามากที่สุด

ปัญหาพื้นฐานข้อแรก คือปัญหาในการประเมินค่า ซึ่งก็คือการหาค่าความน่าจะเป็นของลำดับค่าสังเกตที่สร้างจากแบบจำลอง เมื่อกำหนดแบบจำลองและลำดับค่าสังเกตมาให้ หรืออีกนัยหนึ่งก็คือการแสดงว่าแบบจำลองที่กำหนดให้เข้าคู่กันได้กับลำดับค่าสังเกตที่กำหนดให้ได้ดีเพียงใด ตัวอย่างเช่น ถ้าในกรณีที่พิจารณาเลือกกระหว่างแบบจำลองหลายแบบ ผลลัพธ์ของปัญหาพื้นฐานข้อแรกจะช่วยในการเลือกแบบจำลองที่เข้าคู่กันได้ดีที่สุดกับค่าสังเกต

ปัญหาพื้นฐานข้อที่สอง คือความพยายามในการเปิดเผยส่วนที่แบบจำลองปิดบังไว้ โดยเฉพาะกรณีของแบบจำลองที่ด้วยประสิทธิภาพจะไม่มีลำดับสถานะที่ถูกต้อง ดังนั้นในทางปฏิบัติจึงใช้กฎเกณฑ์ของความเหมาะสมที่สุดในการแก้ไขปัญหานี้ซึ่งมีด้วยกันหลายประเภท ดังนั้นการเลือกกฎเกณฑ์จึงเท่ากับเป็นการเปิดเผยลำดับสถานะที่ถูกปกปิดโดยแบบจำลอง

ปัญหาพื้นฐานข้อที่สาม คือการทำให้พารามิเตอร์ของแบบจำลองมีประสิทธิภาพมากที่สุด เพื่อที่จะอธิบายลำดับค่าสังเกตได้ดีที่สุด ลำดับค่าสังเกตที่ใช้ในการปรับพารามิเตอร์ของแบบจำลองเรียกว่า “ลำดับฝึกฝน” (Training Sequences) เนื่องจากถูกใช้ในการฝึกฝนแบบจำลองฮิดเดนมาร์คคอฟ ปัญหาในการฝึกฝนนี้จะช่วยให้ปรับแต่งพารามิเตอร์ของแบบจำลองให้เหมาะสมมากที่สุดกับข้อมูลฝึกฝนที่สังเกต

ในการประยุกต์ใช้งานแบบจำลองฮิดเดนมาร์คคอฟกับการรู้จำคำพูดนั้น เริ่มต้นจากการแบ่งคำศัพท์ออกเป็นหน่วยย่อยของพยางค์ตามลักษณะกลศาสตร์ของภาษาไทย แล้วจึง

ออกแบบสร้างแบบจำลองฮิดเดนมาร์คอฟ N สถานะสำหรับแต่ละหน่วยย่อยของพยางค์ของชุดคำศัพท์ W คำ สัญญาณเสียงพูดของแต่ละหน่วยย่อยของพยางค์จะถูกแทนที่ด้วยลำดับเวลาของเวกเตอร์รหัสเชิงสเปกตรัม การเข้ารหัสจะอาศัยชุดรหัสเชิงสเปกตรัมที่ประกอบด้วยเวกเตอร์เชิงสเปกตรัม M เวกเตอร์ที่เป็นเอกลักษณ์ จึงทำให้แต่ละค่าสังเกตจะเป็นดรรชนีของเวกเตอร์เชิงสเปกตรัมที่ใกล้เคียงกับสัญญาณเสียงพูดต้นฉบับมากที่สุด ดังนั้นแต่ละหน่วยย่อยของพยางค์จะมีลำดับการฝึกฝนที่ประกอบด้วยจำนวนลำดับของดรรชนีชุดรหัสของหน่วยย่อยของพยางค์ ขั้นตอนแรกเริ่มจากการสร้างแบบจำลองของหน่วยย่อยของพยางค์ของแต่ละคำโดยการแก้ปัญหาพื้นฐานข้อที่ 3 เพื่อประมาณค่าพารามิเตอร์ของแบบจำลองให้เหมาะสมที่สุดสำหรับแต่ละแบบจำลอง ขั้นตอนที่สองเป็นการสร้างความเข้าใจในความหมายทางกายภาพของสถานะของแบบจำลองโดยการแก้ปัญหาพื้นฐานข้อที่ 2 เพื่อแบ่งแยกแต่ละลำดับฝึกฝนของคำไปยังแต่ละสถานะและศึกษาถึงคุณสมบัติของเวกเตอร์เชิงสเปกตรัมที่ทำให้เกิดค่าสังเกตในแต่ละสถานะ โดยในขั้นตอนนี้จะทำการปรับแต่งแบบจำลองเพื่อเพิ่มพูนความสามารถในการจำลองแบบลำดับคำพูด ขั้นตอนสุดท้ายภายหลังจากการออกแบบชุดของแบบจำลองฮิดเดนมาร์คอฟ พร้อมทั้งปรับให้มีประสิทธิภาพที่เหมาะสมแล้ว การรู้จำคำพูดที่ไม่รู้จักมาก่อนจะอาศัยการแก้ปัญหาพื้นฐานข้อที่ 1 เพื่อให้คะแนนแต่ละแบบจำลองของหน่วยย่อยของพยางค์ด้วยลำดับค่าสังเกตที่ใช้ทดสอบและเลือกค่าซึ่งมีแบบจำลองที่ให้คะแนนสูงที่สุด

3. การแก้ไขปัญหามูลฐานสามประการของแบบจำลองฮิดเดนมาร์คอฟ

ในการแก้ไขปัญหามูลฐานสามประการของแบบจำลองฮิดเดนมาร์คอฟ ตามปัญหาที่ได้กำหนดนั้นสามารถอธิบายวิธีแก้ไขในแต่ละปัญหาดังต่อไปนี้

3.1 การแก้ไขปัญหามูลฐานข้อที่ 1

เมื่อกำหนดแบบจำลอง λ การหาค่าความน่าจะเป็น $P(O | \lambda)$ ของลำดับค่าสังเกต $O = \{O_1, O_2, \dots, O_T\}$ จะอาศัยวิธีการหาค่าทุกลำดับสถานะความยาว T ที่เป็นไปได้ทั้งหมดตามลำดับ เมื่อ T เป็นจำนวนค่าสังเกต พิจารณาลำดับของสถานะที่มีค่าจำกัดดังนี้

$$Q = q_1 q_2 \cdots q_T \dots\dots\dots(2-15)$$

เมื่อ q_i เป็นสถานะเริ่มต้น ค่าความน่าจะเป็นของลำดับค่าสังเกต O สำหรับลำดับสถานะสามารถหาได้ดังนี้

$$P(O | Q, \lambda) = \prod_{i=1}^T P(O_i | q_i, \lambda) \dots\dots\dots(2-16)$$

เมื่อกำหนดให้ค่าสังเกตเป็นชนิดที่ไม่ขึ้นแก่กันในทางสถิติจะได้ว่า

$$P(O | Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdots b_{q_T}(O_T) \dots (2-17)$$

ดังนั้นค่าความน่าจะเป็นของลำดับสถานะ Q สามารถเขียนได้เป็น

$$P(Q | \lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T} \dots (2-18)$$

ค่าความน่าจะเป็นระหว่าง O และ Q เป็นเพียงผลคูณของสมการข้างต้นดังนี้

$$P(O, Q | \lambda) = P(O | Q, \lambda) P(Q, \lambda) \dots (2-19)$$

ค่าความน่าจะเป็นของ O เมื่อกำหนดแบบจำลอง $\lambda = (A, B, \pi)$ มาให้ สามารถหาได้โดยผลรวมของค่าความน่าจะเป็นร่วมของลำดับสถานะทั้งหมดที่เป็นไปได้ ดังนี้

$$\begin{aligned} P(O | \lambda) &= \sum_{all Q} P(O | Q, \lambda) P(Q, \lambda) \\ &= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1} a_{q_2} b_{q_2}(O_2) \dots a_{q_{T-1}} a_{q_T} b_{q_T}(O_T) \dots (2-20) \end{aligned}$$

จากสมการที่ (2-20) เริ่มต้นที่เวลา $t=1$ จะเริ่มที่สถานะ q_1 ด้วยความน่าจะเป็น π_{q_1} และให้กำเนิดสัญลักษณ์ O_1 ภายในสถานะเดียวกันด้วยความน่าจะเป็น $b_{q_1}(O_1)$ เมื่อเวลาเปลี่ยนจาก t เป็น $t+1$ และเปลี่ยนไปยังสถานะ q_2 จากสถานะ q_1 ด้วยความน่าจะเป็น $a_{q_1} a_{q_2}$ และให้กำเนิดสัญลักษณ์ O_2 ภายในสถานะเดียวกันด้วยความน่าจะเป็น $b_{q_2}(O_2)$ กระบวนการนี้จะดำเนินต่อไปอย่างต่อเนื่องจนกระทั่งการเปลี่ยนแปลงเวลา T จากสถานะ q_{T-1} ไปยังสถานะ q_T ด้วยค่าความน่าจะเป็น $a_{q_{T-1}} a_{q_T}$ และให้กำเนิดสัญลักษณ์ O_T ภายในสถานะเดียวกันด้วยความน่าจะเป็น $b_{q_T}(O_T)$

ในการคำนวณค่าความน่าจะเป็น $P(O | \lambda)$ ตามสมการที่ (2-20) นั้นจะทำให้เกิดการคำนวณขึ้นด้วยอันดับประมาณ $2T \cdot N^T$ ครั้ง เนื่องจากทุกเวลา $t=1, 2, \dots, T$ จะเกิดสถานะที่เป็นไปได้ N สถานะซึ่งเข้าถึงได้ และในแต่ละลำดับสถานะจะเกิดการขึ้นประมาณ $2T$ ครั้งสำหรับแต่ละพจน์ในผลรวมของสมการที่ (2-20) ซึ่งจำแนกได้เป็นการคูณ $(2T-1) \cdot N^T$ ครั้งและการบวก $N^T - 1$ ครั้ง จึงทำให้เกิดการคำนวณค่าได้โดยใช้สมการนี้เป็นไปไม่ได้ถึงแม้ด้วยค่า N และ T ค่าน้อยๆก็ตาม ดังนั้นจึงมีกระบวนการที่มีประสิทธิภาพในการคำนวณค่าที่เรียกว่า “กระบวนการไปหน้า-ย้อนกลับ” (Forward-Backward Procedure) ซึ่งประกอบด้วยกระบวนการไปหน้าและกระบวนการย้อนกลับร่วมกันดังต่อไปนี้

1) กระบวนการไปหน้า (Forward Procedure)

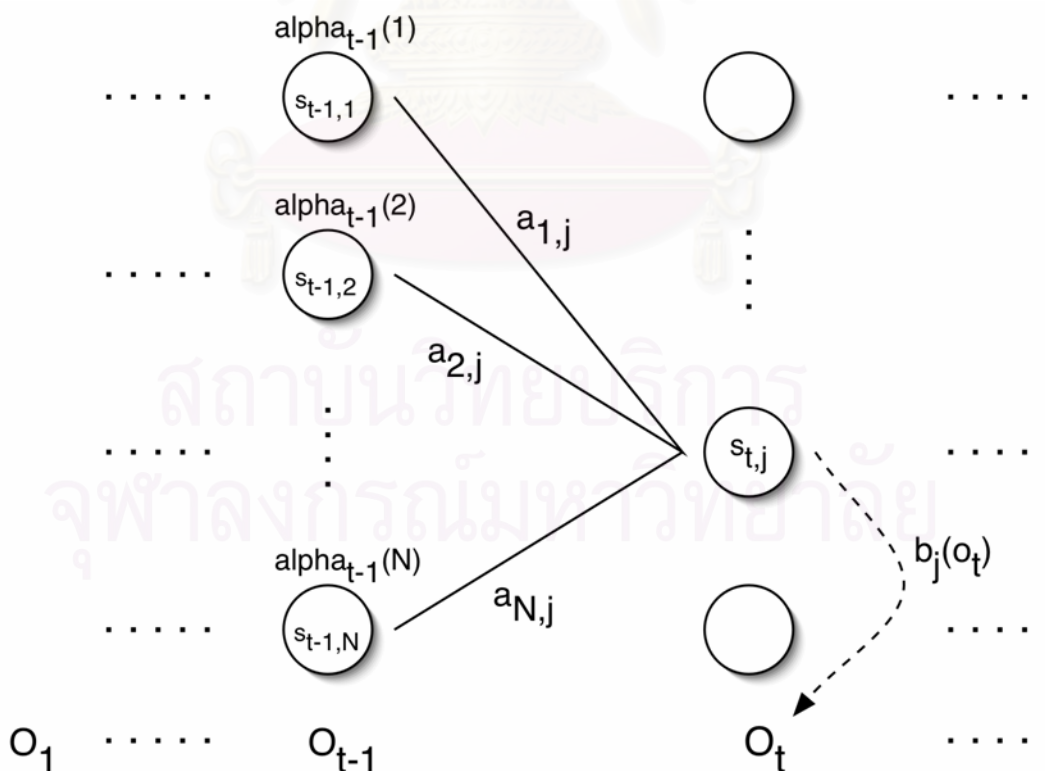
พิจารณาตัวแปรไปหน้า $\alpha_t(i)$ ที่กำหนดไว้แล้วดังนี้

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i | \lambda) \dots\dots\dots(2-21)$$

ค่าความน่าจะเป็นของลำดับค่าสังเกตบางส่วน $O_1 O_2 \cdots O_t$ และสถานะ S_t ที่เวลา t เมื่อกำหนดแบบจำลอง λ จะสามารถหาค่าของ $\alpha_t(i)$ โดยอุปนัยได้ในตารางที่ 2.3 ซึ่งแสดงถึงรายละเอียดของกระบวนการไปหน้าดังนี้

ตารางที่ 2.3 รายละเอียดกระบวนการไปหน้า

ขั้นตอนที่ 1	$\alpha_1(i) = \pi_i b_i(O_1), \dots\dots\dots(2-22)$
กระบวนการเริ่มต้น	$1 \leq i \leq N$
ขั้นตอนที่ 2	$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \dots\dots\dots(2-23)$
กระบวนการอุปนัย	$1 \leq t \leq T-1, \quad 1 \leq i \leq N$
ขั้นตอนที่ 3	$P(O \lambda) = \sum_{i=1}^N \alpha_T(i) \dots\dots\dots(2-24)$
กระบวนการสิ้นสุด	



รูปที่ 2.5 รายละเอียดลำดับกระบวนการในการคำนวณค่าตัวแปรไปหน้า $\alpha_t(i)$

ขั้นตอนกระบวนการเริ่มต้น เป็นการกำหนดค่าเริ่มต้นให้กับค่าความน่าจะเป็นแบบไปหน้าโดยกำหนดให้เป็นความน่าจะเป็นร่วมของสถานะ S_i และค่าสังเกตเริ่มต้น O_1 ขั้นตอนกระบวนการอุปนัยถือเป็นหัวใจสำคัญของกระบวนการไปหน้าดังแสดงในรูปที่ 2.5 ซึ่งแสดงถึงการที่สถานะ S_j ที่สามารถเข้าได้ที่เวลา $t+1$ จาก N สถานะที่เป็นไปได้ $S_i, 1 \leq i \leq N$ ที่เวลา t เนื่องจาก $\alpha_t(i)$ เป็นค่าความน่าจะเป็นของเหตุการณ์ร่วมที่สังเกต $O_1 O_2 \dots O_t$ และเข้าถึงสถานะ S_j ที่เวลา t ผลรวมผลคูณของ N สถานะทั้งหมดที่เป็นไปได้ $S_i, 1 \leq i \leq N$ ที่เวลา t ได้ผลเป็นค่าความน่าจะเป็น S_j ที่เวลา $t+1$ ร่วมกับค่าสังเกตบางส่วนทั้งหมดที่สัมพันธ์กัน

เมื่อเสร็จสิ้นกระบวนการทั้งหมดและได้ค่า S_j โดยค่า $\alpha_{t+1}(j)$ จะได้มาจากค่าสังเกต O_{t+1} ในสถานะ j การคำนวณในสมการที่ (2-23) จะกระทำกับทุกสถานะ $j, 1 \leq j \leq N$ เมื่อกำหนดค่า t ให้ จากนั้นจะวนซ้ำกับทุก $t = 1, 2, \dots, T-1$ ขั้นตอนกระบวนการสิ้นสุด เป็นการคำนวณค่า $P(O | \lambda)$ ด้วยผลรวมของตัวแปรไปหน้าตัวสุดท้าย $\alpha_T(i)$ ซึ่งทำให้ $P(O | \lambda)$ เป็นเพียงผลรวมของ $\alpha_T(i)$ แต่ละตัวมีนิยามดังนี้

$$\alpha_T(i) = P(O_1 O_2 \dots O_T, q_T = S_i | \lambda) \dots \dots \dots (2-25)$$

เมื่อพิจารณาการคำนวณค่า $\alpha_t(j), 1 \leq t \leq T, 1 \leq j \leq N$ นั้น จะเกิดการคำนวณขึ้นด้วยอันดับประมาณ $N^2 T$ ครั้งเมื่อเปรียบเทียบกับ $2TN^T$ ครั้งเมื่อคำนวณโดยตรง ซึ่งจำแนกได้เป็นการคูณ $N(N+1)(T-1) + N$ ครั้งและการบวก $N(N-1)(T-1)$ ครั้ง

การคำนวณค่าความน่าจะเป็นแบบไปหน้าจะอยู่บนพื้นฐานของโครงสร้าง Lattice หรือ Trellis กุญแจสำคัญก็คือเนื่องจากมีเพียง N สถานะดังนั้นลำดับสถานะทั้งหมดที่เป็นไปได้จะรวมเข้ากับปม N ปมในโครงสร้างไม่ว่าลำดับของค่าสังเกตจะยาวเพียงใดก็ตาม ที่เวลา $t=1$ ซึ่งเป็นช่วงเวลาแรกของโครงสร้าง Lattice จะทำการคำนวณค่าของ $\alpha_1(i), 1 \leq i \leq N$ ที่เวลา $t=2, 3, \dots, T$ จะเป็นเพียงการคำนวณค่าของ $\alpha_t(j), 1 \leq j \leq N$ โดยแต่ละครั้งจะคำนวณเพียงค่า $\alpha_{t-1}(i)$ จำนวน N ค่าก่อนหน้านั้นเท่านั้น เนื่องจากแต่ละจุดตาราง N จุดสามารถเข้าถึงได้จาก N จุดตารางเดียวกันของช่องเวลาก่อนหน้านั้น

2) กระบวนการย้อนกลับ (Backward Procedure)

พิจารณาตัวแปรย้อนกลับ $\beta_t(i)$ ที่กำหนดไว้แล้วดังนี้

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | q_t = S_i, \lambda) \dots \dots \dots (2-26)$$

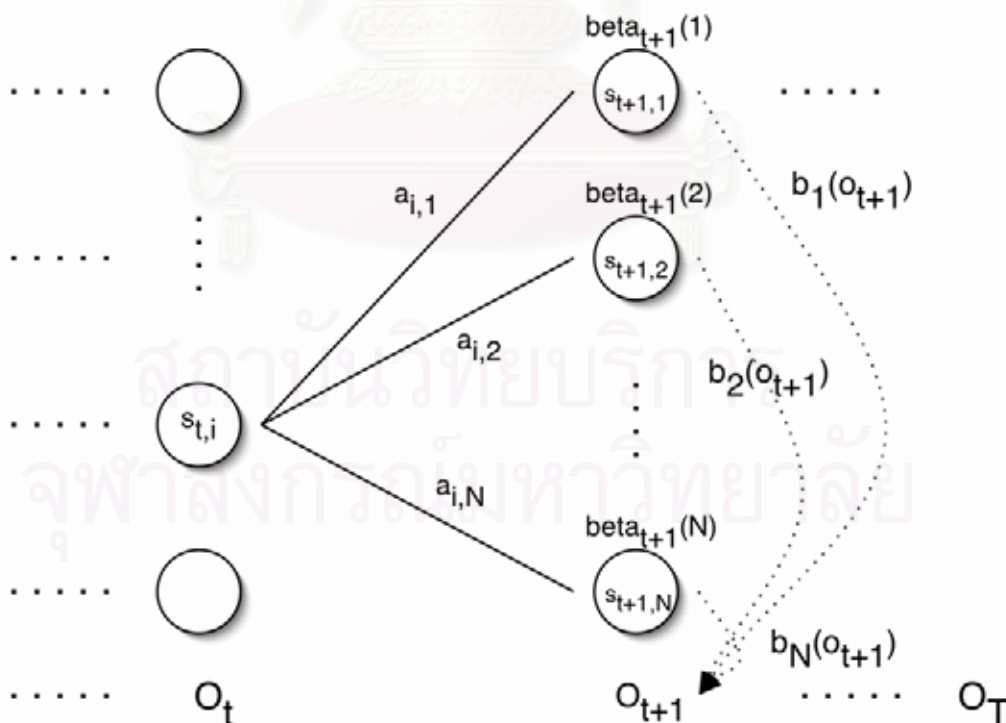
ค่าความน่าจะเป็นของลำดับค่าสังเกตบางส่วนจาก $t+1$ จนถึงสิ้นสุดเมื่อกำหนดสถานะ S_i ที่เวลา t ด้วยแบบจำลอง λ จะสามารถหาค่าของ $\beta_t(i)$ โดยอุปนัยได้ดังตารางที่ 2.4

ขั้นตอนกระบวนการเริ่มต้น เป็นการกำหนดค่าเริ่มต้นให้กับ $\beta_T(i)$ โดยกำหนดให้เป็น 1 ทั้งหมดทุกค่า i ขั้นตอนกระบวนการอุปนัยดังแสดงในรูปที่ 2.6 แสดงถึงการเข้าถึงสถานะ S_j

ที่เวลา t และการให้รายละเอียดของลำดับค่าสังเกตตั้งแต่เวลา $t+1$ เป็นต้นไป จะต้องพิจารณาสถานะ S_j ที่เป็นไปได้ทั้งหมด เพื่อนับรวมการเปลี่ยนแปลงจากสถานะ S_i ไปยัง S_j ของพจน์ a_{ij} รวมไปถึงค่าสังเกต O_{t+1} ในสถานะ j ของพจน์ $b_j(O_{t+1})$ และนับรวมถึงลำดับค่าสังเกตบางส่วนที่ยังเหลืออยู่จากสถานะ j ของพจน์ $\beta_{t+1}(j)$ ด้วย ซึ่งในการคำนวณค่า $\beta_t(i), 1 \leq t \leq T, 1 \leq i \leq N$ นั้น จะเกิดการคำนวณซ้ำด้วยอันดับประมาณ N^2T ครั้ง และสามารถคำนวณโดยอาศัยโครงสร้าง Lattice ได้เช่นเดียวกัน

ตารางที่ 2.4 รายละเอียดกระบวนการย้อนกลับ

ขั้นตอนที่ 1	$\beta_T(i) = 1, \dots\dots\dots(2-27)$
กระบวนการเริ่มต้น	$1 \leq i \leq N$
ขั้นตอนที่ 2	$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \dots\dots\dots(2-28)$
กระบวนการอุปนัย	$t = T - 1, T - 2, \dots, 1, \quad 1 \leq i \leq N$
ขั้นตอนที่ 3	$P(O \lambda) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \dots\dots\dots(2-29)$
กระบวนการสิ้นสุด	



รูปที่ 2.6 รายละเอียดลำดับกระบวนการในการคำนวณค่าตัวแปรย้อนกลับ $\beta_t(i)$

3.2 การแก้ไขปัญหาค่าพื้นฐานข้อที่ 2

ปัญหาพื้นฐานข้อที่ 2 เป็นปัญหาที่เกี่ยวข้องกับการหาลำดับสถานะที่เหมาะสมที่สุดที่สัมพันธ์กับลำดับค่าสังเกตที่กำหนดให้ ความยากในการแก้ปัญหานี้ขึ้นอยู่กับนิยามของลำดับสถานะที่เหมาะสมที่สุดที่เป็นไปได้มากมาย กฎเกณฑ์หนึ่งที่เป็นไปได้คือ การเลือกสถานะ q_t ตัวต่อตัวแล้วคล้ายคลึงกันมากที่สุด กฎเกณฑ์ของความเหมาะสมที่สุดนี้จะทำให้ค่าประมาณของจำนวนสถานะที่ถูกต้องมีค่ามากที่สุด กำหนดให้ตัวแปรที่มีค่าดังนี้

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \dots\dots\dots(2-30)$$

ซึ่งเป็นค่าความน่าจะเป็นของการอยู่ในสถานะ S_i ที่เวลา t เมื่อกำหนดลำดับค่าสังเกต O และแบบจำลอง λ มาให้ สมการที่ (2-30) สามารถเขียนอยู่ในรูปของตัวแปรไปหน้าและตัวแปรย้อนกลับได้ดังสมการที่ (2-31)

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O | \lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)} \dots\dots\dots(2-31)$$

เมื่อ $\alpha_t(i)$ เป็นลำดับค่าสังเกตบางส่วน $O_1 O_2 \dots O_t$ และสถานะ S_i ที่เวลา t ขณะที่ $\beta_t(i)$ เป็นลำดับค่าสังเกตที่คงเหลืออยู่ $O_{t+1} O_{t+2} \dots O_T$ เมื่อกำหนดสถานะ S_i ที่เวลา t ตัวประกอบที่ทำให้เป็นบรรทัดฐานเดียวกันคือ $P(O | \lambda) = \sum_{i=1}^N \alpha_t(i)\beta_t(i)$ ทำให้ $\gamma_t(i)$ กลายเป็นการวัดค่าความน่าจะเป็นดังนี้

$$\sum_{i=1}^N \gamma_t(i) = 1, \quad 1 \leq t \leq T \dots\dots\dots(2-32)$$

โดยการใช้ $\gamma_t(i)$ จะสามารถหาค่าสถานะที่ตัวต่อตัวคล้ายคลึงกันมากที่สุด q_t ที่เวลา t ได้ดังนี้

$$q_t = \arg \max_{1 \leq i \leq N} [\gamma_t(i)], \quad 1 \leq t \leq T \dots\dots\dots(2-33)$$

แม้ว่าสมการที่ (2-33) จะทำให้การประมาณของจำนวนสถานะที่ถูกต้องมีค่ามากที่สุดโดยการเลือกสถานะที่คล้ายคลึงกันมากที่สุดสำหรับแต่ละ t ก็ตาม ก็ยังเกิดปัญหานี้กับผลของลำดับสถานะที่ได้เนื่องจากสมการที่ (2-33) เป็นเพียงการหาสถานะที่คล้ายคลึงกันที่สุดในขณะใดขณะหนึ่งเท่านั้น โดยไม่เกี่ยวข้องกับความน่าจะเป็นในการปรากฏของลำดับสถานะ

วิธีการหนึ่งในการแก้ปัญหา ก็คือการแก้ปัญหากฎเกณฑ์ของความเหมาะสมที่สุด โดยการคำนวณหาค่าลำดับสถานะที่ทำให้ค่าประมาณของคู่สถานะ (q_t, q_t) ที่ถูกต้องมีค่ามากที่สุด กฎเกณฑ์ของความเหมาะสมที่สุดที่นิยมใช้กันมากที่สุดก็คือ การหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียว ขั้นตอนวิธีการในการหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียวนี้อาศัยอยู่บนพื้นฐานของการโปรแกรมแบบพลวัตที่เรียกว่า “ขั้นตอนวิธีการ Viterbi”

ขั้นตอนวิธีการ Viterbi (Viterbi Algorithm)

ขั้นตอนวิธีการ Viterbi เป็นขั้นตอนวิธีในการหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียว $Q = \{q_1, q_2, \dots, q_T\}$ สำหรับลำดับค่าสังเกตที่กำหนดให้ $O = \{O_1, O_2, \dots, O_T\}$ จะกำหนดตัวแปรได้ดังนี้

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1, q_2, \dots, q_t = i, O_1, O_2, \dots, O_T | \lambda] \dots\dots (2-34)$$

โดยที่ $\delta_t(i)$ เป็นค่าความน่าจะเป็นที่มีค่าสูงที่สุดของเส้นทางเดียวที่เวลา t ซึ่งเป็นค่าสังเกต t ค่าแรกและสิ้นสุดในสถานะ S_i ด้วยวิธีการอุปนัยจะได้ว่า

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(O_{t+1}) \dots\dots\dots (2-35)$$

ในการเรียกใช้ค่าลำดับสถานะจำเป็นต้องติดตามค่าอาร์กิวเมนต์ที่ทำให้สมการที่ (2-35) มีค่ามากที่สุดสำหรับแต่ละค่า t และ j โดยอาศัยแถวลำดับ $\psi_t(j)$ ขั้นตอนวิธีการในการหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียว ดังแสดงในตารางที่ 2.5 รายละเอียดของกระบวนการ Viterbi ดังนี้

ตารางที่ 2.5 รายละเอียดขั้นตอนวิธีการ Viterbi

ขั้นตอนที่ 1	$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \dots\dots\dots (2-36)$
กระบวนการเริ่มต้น	$\psi_1(i) = 0$
ขั้นตอนที่ 2	$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad \begin{matrix} 2 \leq t \leq T \\ 1 \leq j \leq N \end{matrix} \dots\dots\dots (2-37)$
กระบวนการวนซ้ำ	$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad \begin{matrix} 2 \leq t \leq T \\ 1 \leq j \leq N \end{matrix} \dots\dots\dots (2-38)$
ขั้นตอนที่ 3	$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \dots\dots\dots (2-39)$
กระบวนการสิ้นสุด	$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \dots\dots\dots (2-40)$
ขั้นตอนที่ 4	$q_t^* = \psi_{t+1}(q_{t+1}^*), \dots\dots\dots (2-41)$
กระบวนการย้อนกลับ	$t = T - 1, T - 2, \dots, 1$
รอยเส้นทาง	

รายละเอียดขั้นตอนวิธีการ Viterbi นั้นคล้ายคลึงกับการคำนวณกระบวนการไปหน้าของสมการที่ (2-22) ถึง (2-24) ยกเว้นกระบวนการย้อนกลับรอยเส้นทางเท่านั้น ความแตกต่างที่สำคัญก็คือการหาค่าที่สูงที่สุดในสมการที่ (2-37) แทนสถานะก่อนหน้าซึ่งใช้แทนผลรวมในสมการที่ (2-23) นั่นเอง

3.3 การแก้ไขปัญหาพื้นฐานข้อที่ 3

ปัญหาพื้นฐานข้อที่สามเกี่ยวข้องกับการค้นหาวิธีการปรับค่าพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ เพื่อให้ค่าความน่าจะเป็นของลำดับค่าสังเกตมีค่ามากที่สุดเมื่อกำหนดค่าพารามิเตอร์ของแบบจำลองมาให้ เนื่องจากไม่มีวิธีการที่แน่นอนในการวิเคราะห์เพื่อแก้ไขปัญหาลำดับค่าสังเกตที่จะให้ค่าความน่าจะเป็นของลำดับค่าสังเกตมีค่ามากที่สุด ถึงแม้ว่าจะกำหนดลำดับค่าสังเกตที่จำกัดให้เป็นข้อมูลฝึกฝนก็ตามก็ยังไม่มีความเหมาะสมที่สุดในการประมาณค่าพารามิเตอร์ของแบบจำลอง แต่จะสามารถเลือก $\lambda = (A, B, \pi)$ ที่ทำให้ $P(O | \lambda)$ มีค่ามากที่สุดโดยใช้กระบวนการวนซ้ำของ Baum-Welch

กระบวนการประมาณค่าซ้ำของ Baum-Welch

(Baum-Welch Reestimation Procedure)

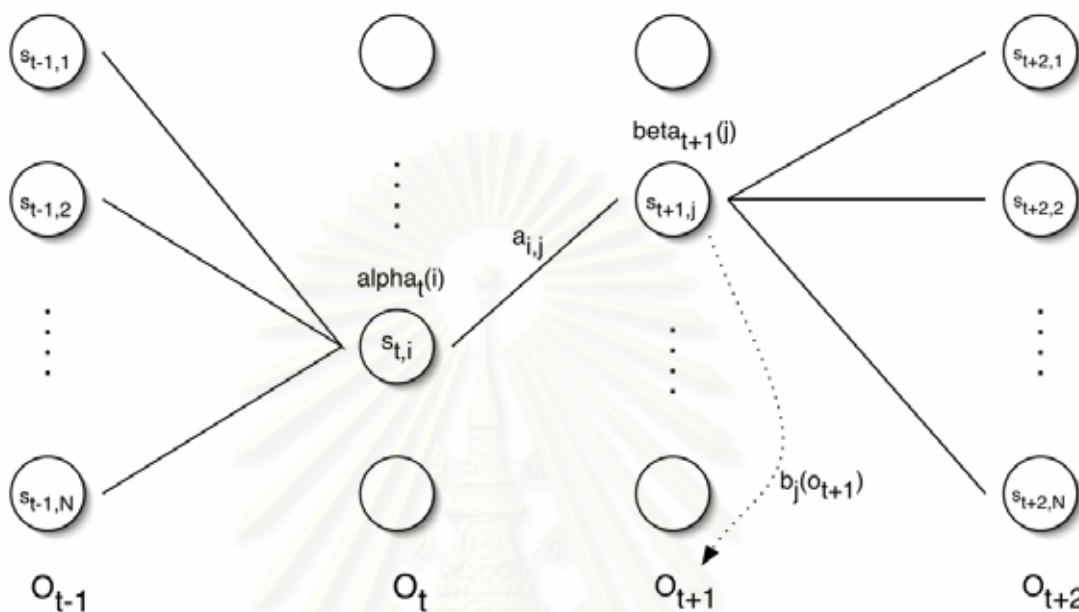
กระบวนการประมาณค่าซ้ำของ Baum-Welch นั้นอยู่บนพื้นฐานของหลักการของความน่าจะเป็นจริงสูงสุด ซึ่งจะช่วยปรับปรุงให้ค่าความน่าจะเป็นของลำดับค่าสังเกตให้มีค่าสูงขึ้น โดยมีขั้นตอนวิธีการดังนี้ (Rabiner, L.R., 1989)

ตารางที่ 2.6 รายละเอียดกระบวนการประมาณค่าซ้ำของ Baum - Welch

การกระจายของสถานะเริ่มต้น	$\hat{\pi}_i = \alpha_1(i)\beta_1(i)$(2-42)
การกระจายของความน่าจะเป็นในการเปลี่ยนแปลงสถานะ	$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)}$(2-43)
การกระจายของความน่าจะเป็นของสัญลักษณ์ของค่าสังเกต	$\hat{b}_j(k) = \frac{\sum_{t=1, O_t=v_k}^{T-1} \alpha_t(j)\beta_t(j)}{\sum_{t=1}^{T-1} \alpha_t(j)\beta_t(j)}$(2-44)

กระบวนการประมาณค่าซ้ำของ Baum-Welch นั้นอยู่บนพื้นฐานของความรู้ความเข้าใจในเรื่องของการประมาณค่าใหม่ของค่าความน่าจะเป็นในการเปลี่ยนแปลง ซึ่งอยู่บนพื้นฐานของจำนวนการเปลี่ยนแปลงจากสถานะ i ไปยังสถานะ j หารด้วยจำนวนการเปลี่ยนแปลงออกจาก

สถานะ i ในทำนองเดียวกันค่าความน่าจะเป็นใหม่ของสัญลักษณ์ขาออกสำหรับสัญลักษณ์ที่ k ที่สถานะ i ได้จากจำนวนครั้งของสัญลักษณ์ที่ออกจากสถานะหารด้วยจำนวนครั้งที่อยู่ในสถานะ ดังแสดงในรูปที่ 2.7



รูปที่ 2.7 รายละเอียดลำดับเหตุการณ์ในการคำนวณเหตุการณ์ร่วมในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t + 1$

เนื่องจากงานวิจัยนี้ได้ใช้ค่าความน่าจะเป็นของค่าสังเกต $b_j(\mathbf{O}_t)$ ที่มีการกระจายแบบเกาส์มิกเจอร์ (Gaussian mixture distribution) ซึ่งค่าความน่าจะเป็นของค่าสังเกตที่มีการกระจายแบบเกาส์มิกเจอร์นี้เป็นค่าความน่าจะเป็นที่มีช่วงต่อเนื่อง (Continuous probability) (Young, S., Kershaw, D., Odell, J. et al., 2000) ดังรูปที่ 2.8 การกระจายแบบเกาส์สามารถช่วยแก้ปัญหาของการกระจายของความน่าจะเป็นของสัญลักษณ์ของค่าสังเกตที่มีช่วงไม่ต่อเนื่อง (discrete probability) ในเรื่องของค่าความน่าจะเป็นของสัญลักษณ์ของค่าสังเกต $b_j(k)$ บางค่าซึ่งสามารถเป็นศูนย์ได้มาจากสาเหตุของการฝึกฝน จะเป็นผลต่อเนื่องทำให้เกิดค่าเกินขอบเขตของคอมพิวเตอร์ที่จะสามารถประมวลผลได้ (overflow) จึงช่วยลดขั้นตอนในการปรับค่าความน่าจะเป็นของสัญลักษณ์ของค่าสังเกต $b_j(k)$ เพื่อไม่ให้มีค่าเป็นศูนย์ และยังช่วยลดขั้นตอนในการควอนไทซ์แบบเวกเตอร์ (Vector Quantization) และขั้นตอนการฝึกฝนชุดรหัส (Codebook Training) เพราะไม่มีความจำเป็นต้องใช้ชุดรหัสในการประมวลผลของแบบจำลองฮิดเดนมาร์คอฟ การคำนวณค่าความน่าจะเป็นของค่าสังเกต $b_j(\mathbf{O}_t)$ ที่มีการกระจายแบบเกาส์มิกเจอร์เป็นการคำนวณแบบเวกเตอร์ สามารถหาค่าความน่าจะเป็นได้จากสมการที่ (2-45) ดังนี้

$$b_j(\mathbf{O}_t) = \sum_{m=1}^M c_{jm} N(\mathbf{O}_t; \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}) \dots\dots\dots (2-45)$$

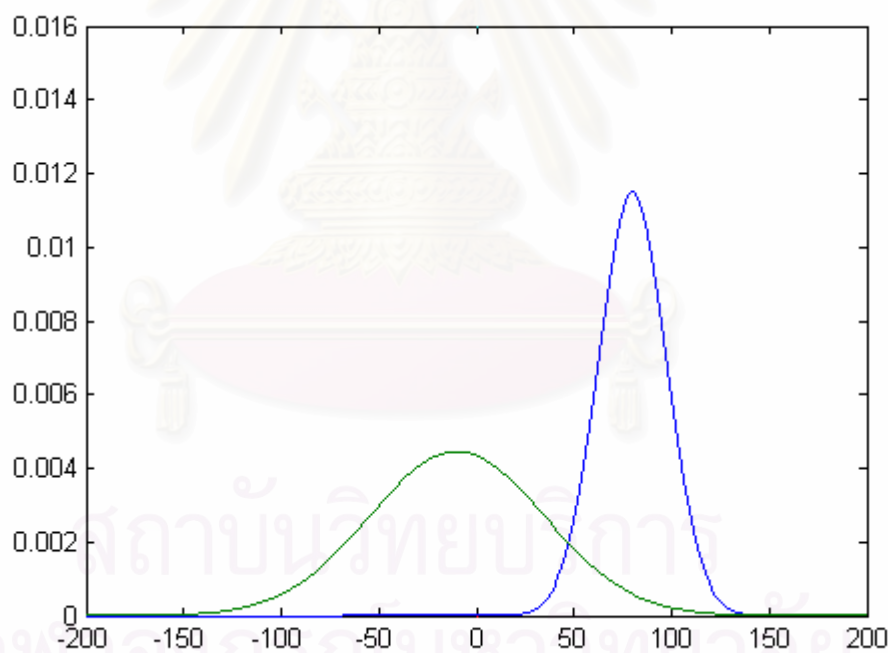
เมื่อ M คือจำนวนของมิกซ์เจอร์

c_{jm} คือค่าถ่วงน้ำหนักของมิกซ์เจอร์ m ที่สถานะ j

$N(\mathbf{O}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ คือการกระจายเกาส์ซึ่งประกอบด้วยเวกเตอร์ค่าเฉลี่ย $\boldsymbol{\mu}$ (Mean vector) และเวกเตอร์ความแปรปรวนร่วม $\boldsymbol{\Sigma}$ (Covariance matrix) ซึ่งมีสมการดังนี้

$$N(\mathbf{O}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{O}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{O}-\boldsymbol{\mu})} \dots\dots\dots (2-46)$$

เมื่อ n คือจำนวนมิติของเวกเตอร์ค่าสังเกต \mathbf{O}_t



รูปที่ 2. 8 ค่าความน่าจะเป็นของค่าสังเกตที่มีการกระจายแบบเกาส์ 2 มิกเจอร์

ในการปรับค่าพารามิเตอร์ของค่าความน่าจะเป็นของค่าสังเกต $b_j(\mathbf{O}_t)$ ซึ่งมีพารามิเตอร์ที่จำเป็นของปรับคือ c_{jm} , $\boldsymbol{\mu}_{jm}$ และ $\boldsymbol{\Sigma}_{jm}$ โดยใช้กระบวนการประมาณค่าซ้ำของ Baum-Welch ซึ่งมีขั้นตอนวิธีการดังนี้

เวกเตอร์ค่าเฉลี่ย	$\hat{\boldsymbol{\mu}}_{jm} = \frac{\sum_{t=1}^T L_{jm}(t) \mathbf{O}_t}{\sum_{t=1}^T L_{jm}(t)} \dots\dots\dots (2-47)$
เวกเตอร์ความแปรปรวนร่วม	$\hat{\boldsymbol{\Sigma}}_{jm} = \frac{\sum_{t=1}^T L_{jm}(t) (\mathbf{O}_t - \boldsymbol{\mu}_{jm})(\mathbf{O}_t - \boldsymbol{\mu}_{jm})'}{\sum_{t=1}^T L_{jm}(t)} \dots\dots (2-48)$
ค่าถ่วงน้ำหนัก	$c_{jm} = \frac{\sum_{t=1}^T L_{jm}(t)}{\sum_{t=1}^T L_j(t)} \dots\dots\dots (2-49)$

เมื่อกำหนดให้ $L_{jm}(t)$ และ $L_j(t)$ มีสมการดังนี้

$$L_{jm}(t) = \frac{\sum_{i=1}^T \alpha_t^-(j) c_{jm} N(\mathbf{O}_t; \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}) \beta_t(j)}{P(\mathbf{O} | \lambda)} \dots\dots\dots (2-50)$$

โดยที่

$$\alpha_t^-(j) = \sum_{i=1}^N \alpha_t(i) a_{ij} \dots\dots\dots (2-51)$$

และ

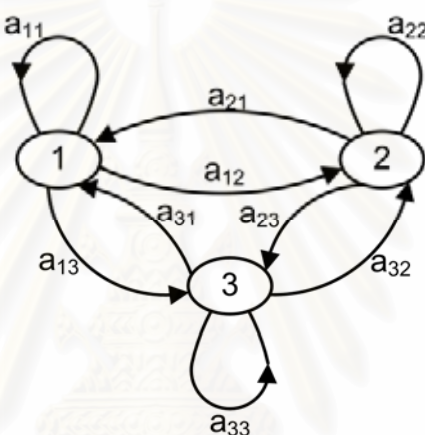
$$L_j(t) = \frac{\alpha_t(j) \beta_t(j)}{P(\mathbf{O} | \lambda)} \dots\dots\dots (2-52)$$

จ. ขั้นตอนวิธีการตัดสินใจ (Decision Algorithm)

ขั้นตอนการตัดสินใจ เป็นขั้นตอนที่เกี่ยวข้องกับกฎเกณฑ์ที่ใช้ในการตัดสินใจเลือกรูปแบบที่มีความคล้ายคลึงกันมากที่สุดระหว่างคำพูดที่ไม่ทราบรูปแบบที่ได้จัดเก็บไว้ล่วงหน้า โดยอาศัยค่าความไม่คล้ายคลึงกันหรือค่าระยะทางที่ได้จากการทดสอบความคล้ายคลึงกันของรูปแบบที่มีค่ามากที่สุด ในขั้นตอนวิธีการแก้ปัญหาพื้นฐานทั้งสามประการของแบบจำลองฮิดเดนมาร์คอฟนั้น การแก้ปัญหาพื้นฐานข้อที่สองจัดเป็นขั้นตอนวิธีการตัดสินใจเลือกรูปแบบที่เหมาะสมที่สุดในการรู้จำ วิธีการหนึ่งที่ถูกนำมาใช้กับแบบจำลองฮิดเดนมาร์คอฟก็คือขั้นตอนวิธีการของ Viterbi (Rabiner, L.R., 1989) โดยมีขั้นตอนวิธีการดังแสดงในการแก้ปัญหาพื้นฐานข้อที่สอง เมื่อแบบจำลองที่ใช้ในงานวิจัยนี้เป็นแบบจำลองของเสียงส่วนย่อยของพยางค์ ในระบบการรู้จำ คำศัพท์ต้องนำแบบจำลองของเสียงส่วนย่อยของพยางค์มาประกอบกันเป็นคำศัพท์จึงมีความจำเป็นต้องสร้างแบบจำลองภาษาเพื่อวิเคราะห์ค่าความน่าจะเป็นของเสียงพูดเมื่อใช้แบบจำลองภาษาของคำศัพท์นั้น ๆ คำศัพท์ที่ให้ค่าความน่าจะเป็นมากที่สุดจะตัดสินใจให้แทนด้วยช่วงเวลาของเสียงพูดที่ถูกประมวลด้วยแบบจำลองภาษาของคำศัพท์คำนี้ ซึ่งรายละเอียดของแบบจำลองภาษาจะอยู่ในหัวข้อที่ 8

จ. ประเภทของแบบจำลองฮิตเดนมาร์คอฟ

ขั้นตอนวิธีการของแบบจำลองฮิตเดนมาร์คอฟส่วนใหญ่เป็นการพิจารณาเพียงกรณีพิเศษของแบบจำลองประเภทเออร์กอดิก ซึ่งเป็นแบบจำลองฮิตเดนมาร์คอฟที่ทุกสถานะต่อเชื่อมถึงกันหมด โดยทุกสถานะของแบบจำลองสามารถเข้าถึงสถานะอื่นได้ในขั้นตอนเดียว ดังนั้นแบบจำลองแบบเออร์กอดิกจึงมีคุณสมบัติที่ทุกสถานะสามารถเข้าถึงได้จากสถานะอื่นด้วย ขั้นตอนที่กำหนดแน่นอนดังแสดงในรูปที่ 2.9 ซึ่งเป็นแบบจำลองที่มีจำนวนสถานะ $N = 3$ สถานะ และมีคุณสมบัติเฉพาะที่สัมพันธ์ a_{ij} ทั้งหมดมีค่าเป็นบวกซึ่งแสดงในสมการที่ (2-53) ดังนี้ (Rabiner, L.R., 1989)

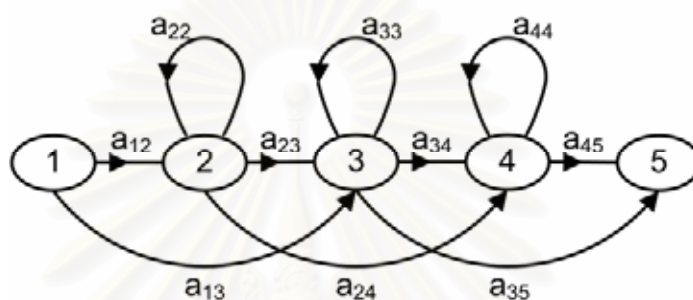


รูปที่ 2.9 แบบจำลองแบบเออร์กอดิกที่มี 3 สถานะ

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \dots\dots\dots (2-53)$$

ในการประยุกต์ใช้งานแบบจำลองฮิตเดนมาร์คอฟกับงานเฉพาะอย่างนั้น ยังมีแบบจำลองประเภทอื่นที่เหมาะสมกับคุณสมบัติที่สังเกตของสัญญาณซึ่งถูกจำลองมากกว่าแบบจำลองประเภทเออร์กอดิกมาตรฐานดังแสดงในรูปที่ 2.10 ซึ่งเป็นแบบจำลองซ้าย - ขวา (Left - Right Model) หรือแบบจำลอง Bakis เนื่องจากลำดับสถานะที่อยู่ภายในที่สัมพันธ์กับแบบจำลองมีคุณสมบัติที่ดัชนีของสถานะจะเพิ่มขึ้นหรือมีค่าเท่าเดิมเมื่อเวลาเพิ่มขึ้นเปรียบเสมือนกับสถานะดำเนินจากซ้ายไปขวา ดังนั้นแบบจำลองซ้าย - ขวาจึงมีคุณสมบัติเหมาะสมในการจำลองแบบสัญญาณที่เปลี่ยนแปลงไปตามเวลา ดังเช่น สัญญาณเสียง เป็นต้น ซึ่งในงานวิจัยนี้ได้ใช้แบบจำลองซ้าย - ขวาที่มีลักษณะของสถานะแรกและสถานะสุดท้ายจะไม่มีกรวนอยู่ใน

สถานะเดิม ถ้าลำดับค่าสังเกตตกอยู่ในสถานะสุดท้ายเป็นอันสิ้นสุดกระบวนการของรู้อำของแบบจำลองนี้ สำหรับแบบจำลองฮิดเดนมาร์คอฟแบบดังกล่าวจะไม่มีค่าความน่าจะเป็นของค่าสังเกตในสถานะแรกและสถานะสุดท้าย (Young, S., Kershaw, D., Odell, J. et al., 2000) ซึ่งมีข้อดีในระบบการรู้อำคำพูดเนื่องจากช่วยลดขั้นตอนในการหาจุดสิ้นสุดของพยางค์ (Endpoint Detection) เพราะเมื่อลำดับค่าสังเกตตกอยู่ในสถานะสุดท้ายเป็นอันสิ้นสุดของการประมวลด้วยแบบจำลอง ก็จะได้จุดสิ้นสุดของเสียงตามแบบจำลอง



รูปที่ 2. 10 แบบจำลองแบบซ้าย - ขวาที่มี 5 สถานะ

คุณสมบัติพื้นฐานของทุกแบบจำลองฮิดเดนมาร์คอฟประเภทซ้าย - ขวานั้น สัมประสิทธิ์ของการเปลี่ยนสถานะจะต้องเป็นไปตามคุณสมบัติที่ไม่อนุญาตให้มีการเปลี่ยนแปลงสถานะไปยังสถานะที่มีดัชนีต่ำกว่าสถานะปัจจุบันตามสมการที่ (2-54) รวมทั้งค่าความน่าจะเป็นเริ่มต้นจะต้องมีคุณสมบัติเป็นไปตามสมการที่ (2-55) ดังนี้

$$a_{ij} = 0, \quad j < i \quad (2-54)$$

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \quad (2-55)$$

เนื่องจากลำดับสถานะจะต้องเริ่มต้นจากสถานะที่ 1 และสิ้นสุดในสถานะที่ N ดังนั้นในแบบจำลองประเภทซ้าย - ขวาจึงต้องเพิ่มเติมเงื่อนไขบังคับให้กับสัมประสิทธิ์ของการเปลี่ยนสถานะ เพื่อไม่ให้เกิดการเปลี่ยนแปลงมากเกินไปดังนี้

$$a_{ij} = 0, \quad j > i + \Delta \quad (2-56)$$

จากรูปที่ 2.12 จะกำหนดให้ค่า $\Delta = 2$ ซึ่งไม่อนุญาตให้มีการข้ามสถานะเกินกว่า 2 สถานะ จึงได้เมตริกซ์ของการเปลี่ยนสถานะเป็นดังสมการที่ (2-58) และสถานะสุดท้ายของ

แบบจำลองประเภทซ้าย – ขวาแบบดังกล่าวนี้อาจมีสัมประสิทธิ์ของการเปลี่ยนสถานะเป็นศูนย์ เพื่อให้สิ้นสุดกระบวนการประมวลผลของแบบจำลองดังสมการที่ (2-57)

$$a_{Ni} = 0, \quad 1 \leq i \leq N \quad (2-57)$$

$$A = \begin{bmatrix} 0 & a_{12} & a_{12} & 0 & 0 \\ 0 & a_{12} & a_{12} & a_{12} & 0 \\ 0 & 0 & a_{12} & a_{12} & a_{12} \\ 0 & 0 & 0 & a_{12} & a_{12} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2-58)$$

ข. ลักษณะด้านกลศาสตร์ศาสตร์ของเสียงพูดภาษาไทย

เสียงพูดภาษาไทย 1 พยางค์จะประกอบด้วย เสียงพยัญชนะ เสียงสระ เสียงตัวสะกด และเสียงวรรณยุกต์ ซึ่งมีลักษณะการประกอบเป็นพยางค์ภาษาไทยดังนี้

$$S = C(C)^T V (V)(C)$$

S = Syllable (พยางค์)

C = Consonant (เสียงพยัญชนะ)

V = Vowel (เสียงสระ)

T = Tone (เสียงวรรณยุกต์)

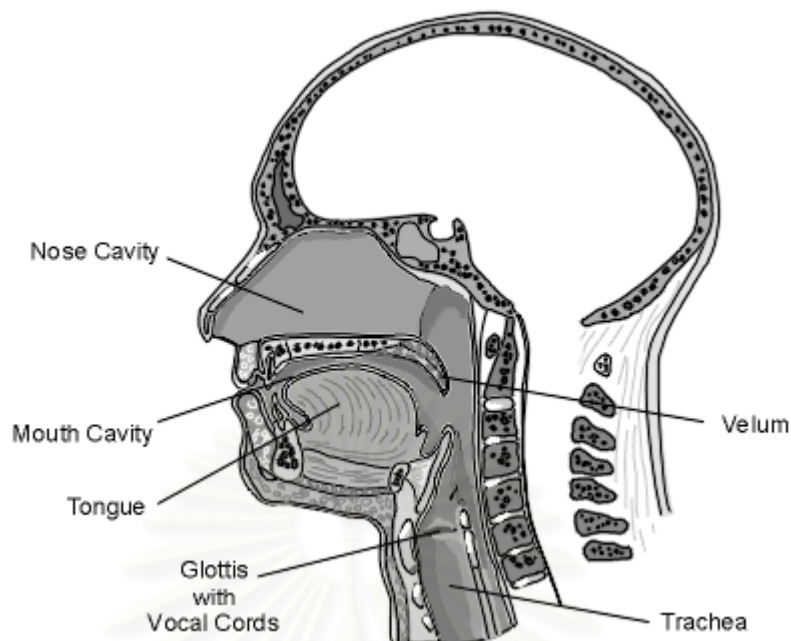
เสียงพยัญชนะภาษาไทยนั้นแบ่งออกเป็น 2 ประเภทคือ

- 1) พยัญชนะเดี่ยว
- 2) พยัญชนะคู่หรือพยัญชนะควบกล้ำ

เสียงสระภาษาไทยแบ่งออกเป็น 2 ประเภทคือ

- 1) สระเดี่ยว
- 2) สระผสม

เสียงที่เกิดขึ้นจะเริ่มต้นจากลมในหลอดลมผ่านเส้นเสียงทำให้เกิดเสียงก้องที่มีความถี่ค่าหนึ่ง ซึ่งความถี่ที่เกิดขึ้นมานี้เรียกว่าความถี่มูลฐาน (Fundamental Frequency) เมื่อเสียงนี้ผ่านช่องปากซึ่งประกอบด้วยอวัยวะต่าง ๆ ภายในช่องปาก อวัยวะเหล่านี้ทำให้เสียงที่เกิดขึ้นแปรเปลี่ยนแตกต่างกันออกไป ริมฝีปากและโพรงจมูกยังมีส่วนในการออกเสียงด้วย รูปที่ 2.11 แสดงอวัยวะที่เกี่ยวข้องกับการสร้างเสียงพูด



รูปที่ 2.11 อวัยวะที่เกี่ยวข้องกับการสร้างเสียงพูด (แหล่งที่มา <http://www.kt.tu-cottbus.de/speech-analysis/tech.html>)

ในภาษาไทยมีพยัญชนะทั้งหมด 44 ตัว แต่จะมีพยัญชนะที่มีเสียงซ้ำ เช่นเสียง “ค” และเสียง “ข” เป็นต้น และพยัญชนะที่ออกเสียงเหมือนกันแต่เสียงวรรณยุกต์ต่างกัน เช่นเสียง “ช” และเสียง “ค” เป็นต้น จึงสามารถแบ่งเสียงของพยัญชนะเสียงเดียวที่แตกต่างกันได้เป็น 21 เสียง เสียงของพยัญชนะต้นเสียงคู่ (ควบกล้ำ) 12 เสียง เสียงของสระเสียงเดียว 18 เสียง เสียงของสระเสียงผสม 6 เสียง และเสียงของตัวสะกด 8 เสียง (วิศรุต อาขุบุตร, 2545; Maneenoi, E., 2002) ดังตารางที่ 2.8 ถึง 2.12

ตารางที่ 2.8 รายละเอียดของพยัญชนะเสียงเดียว

		Place of Articulation					
		Labial	Alveolar	Palatal	Velar	Glottal	
Manner of Articulation	Stops	Voiceless Unaspirated	p (ป)	t (ต)	c (จ)	k (ก)	? (อ)
		Voiceless Aspirated	ph (ฟ/พ)	th (ถ/ท)	ch (ฉ/ช)	kh (ข/ค)	
		Voiced	b (บ)	d (ด)			
	Non – stop	Nasal	m (ม)	n (น)		ng (ง)	
		Fricative	f (ฝ/ฟ)	s (ส/ซ)			h (ห/ฮ)
Trill			r (ร)				
Lateral			l (ล)				
	Approximant	w (ว)		j (ย)			

ตารางที่ 2.9 รายละเอียดของพยัญชนะเสียงคู่

	พยัญชนะต้น					
ตัวควบกล้ำ	p (ป)	t (ต)	k (ก)	ph (พ)	th (ท)	kh (ค/ข)
r(ร)	pr (ปร)	tr (ตร)	kr (กร)	phr(พร)	thr (ทร)	khr (กร)
l(ล)	pl (ปล)		kl (กล)	phl(พล)		khl (คล)
w(ว)			kw (กว)			khw (คว)

ตารางที่ 2.10 รายละเอียดของเสียงตัวสะกด

ตัวสะกด	k [^]	p [^]	t [^]	n [^]	m [^]	ng [^]	j [^]	w [^]
หมวด	แม่กก	แม่กบ	แม่กด	แม่กน	แม่กม	แม่กง	แม่เกย	แม่เกว

ตารางที่ 2.11 รายละเอียดของสระเสียงเดี่ยว

		Tongue Advancement		
		Front	Central	Back
Tongue Height	High	i (อิ)	v (อี)	u (อุ)
		ii (อี)	vv (อี้อ)	uu (อู)
	Medium	e (เอะ)	q (เออะ)	o (โอะ)
		ee (เอ)	qq (เออ)	oo (โอ)
	Low	x (แอะ)	a (อะ)	@ (เออะ)
		xx (แอ)	aa (อา)	@@ (เออ)

ตารางที่ 2.12 รายละเอียดของสระเสียงผสม

Diphthongs	ia (เอี๊ยะ)	va (เอี๊อะ)	ua (อัวะ)
	ia (เอี๊ย)	vva (เอี๊อ)	uua (อัว)

ซ. การตรวจหากลุ่มคำ (Phrase Detector)

ในระบบการชี้เฉพาะคำสำคัญ การวิเคราะห์เสียงนั้นจะกระทำอยู่ในระดับย่อยของพยางค์ (Phoneme) เมื่อเปรียบเทียบกับระดับคำ (Word) และระดับพยางค์ (Syllable) ในคำว่า “ กรุงเทพมหานคร ” ได้ดังนี้

Word	/krung0theep2ma3haa4na3kh@@n0/
Syllable	/krung0/ - /theep2/ - /ma3/ - /haa4/ - /na3/ - /kh@@n0/
Phoneme	/kr/ - /u/ - /ng/ - /th/ - /ee/ - /p/ - /m/ - /a/ - /h/ - /aa/ - /n/ - /a/ - /kh/ - /@@/ - /n/

ในระดับเสียงย่อยของพยางค์ที่นิยมใช้กันจะมี Monophone, Triphone, Onset, Rhyme, Initial และ Final ซึ่งการแบ่งขอบเขตของเสียงย่อยของพยางค์ต่าง ๆ เหล่านี้จะแตกต่างกันออกไป ในแต่ละชนิดจะเหมาะสมกับโครงสร้างของแต่ละภาษาต่าง ๆ กัน ในแต่ละภาษาจะมีคุณลักษณะของการออกเสียงไม่เหมือนกันและจำนวนของเสียงพยัญชนะต้น เสียงสระ และเสียงตัวสะกดไม่เท่ากัน จึงเป็นสาเหตุของการแบ่งขอบเขตของเสียงย่อยของพยางค์นี้ต่างกัน สามารถเปรียบเทียบการแบ่งขอบเขตของเสียงต่าง ๆ ได้ดังนี้

Monophone	/kr/ - /u/ - /ng^/ - /th/ - /ee/ - /p^/ - /m/ - /a/ - /h/ - /aa/ - /n/ - /a/ - /kh/ - /@@/ - /n^/
Triphone	/sil-kr+u/ - /kr-u+ng^/ - /u-ng^+th/ - /ng^-th+ee/ - /th-ee+p^/ - /ee- p^+m/ - /p-m+a/ - /m-a+h/ - /a-h+aa/ - /h-aa+n/ - /aa-n+a/ - /n-a+kh/ - /a-kh+@@/ - /kh-@@+n^/ - /@@-n^+sil/
Onset – Rhyme	/kru/ - /ung^/ - /the/ - /eep^/ - /ma/ - /a/ - /ha/ - /aa/ - /na/ - /a/ - /kh@/ - /@@n^/
Initial – Final	/kr/ - /ung^/ - /th/ - /eep^/ - /m/ - /a/ - /h/ - /aa/ - /n/ - /a/ - /kh/ - /@@n^/

(sil หมายถึง เสียงเงียบ)

ในการวิจัยนี้ได้ใช้เสียงระดับย่อยของพยางค์ (phoneme) 3 ประเภทคือ Monophone Triphone และ Onset – Rhyme ซึ่งเสียงระดับย่อยของพยางค์แต่ละประเภทมีรายละเอียดดังนี้

Monophone

เป็นเสียงระดับย่อยของพยางค์ที่ไม่ขึ้นกับสภาพแวดล้อมของเสียงตอนต้นและตอนท้ายของเสียงระดับย่อยของพยางค์นั้น ๆ หรือไม่นำเสียงก่อนหน้าเสียงระดับย่อยของพยางค์และเสียงต่อจากเสียงระดับย่อยของพยางค์นั้น ๆ มาเกี่ยวข้องด้วย เช่น คำว่า “กรุง” ประกอบเสียงพยัญชนะต้น “กร” สามารถเขียนแทนได้ด้วย /kr/ ในส่วนของเสียงสระ “อุ” สามารถเขียนแทนได้ด้วย /u/ และเสียงตัวสะกด “ง” สามารถเขียนแทนได้ด้วย /ng^/ เป็นต้น

Triphone

เป็นเสียงระดับย่อยของพยางค์ที่ขึ้นอยู่กับสภาพแวดล้อมของเสียงตอนต้นและตอนท้ายของเสียงระดับย่อยของพยางค์นั้น ๆ หรือนำเสียงก่อนหน้าเสียงระดับย่อยของพยางค์และเสียงต่อจากเสียงระดับย่อยของพยางค์นั้น ๆ มาเกี่ยวข้องด้วย เช่น คำว่า “กรุง” ประกอบเสียงพยัญชนะต้น “กร” แต่เนื่องจากว่าก่อนหน้าเสียง “กร” เป็นเสียงเงียบ และต่อจากเสียง “กร” เป็นเสียงสระ “อุ” เพราะฉะนั้นสามารถแทนเสียง “กร” ในเสียงระดับย่อย Triphone สามารถเขียนแทนได้ด้วย /sil - kr + u/ ในส่วนของเสียงสระ “อุ” นั้นมีเสียงก่อนหน้าเป็นเสียง “กร” และต่อจากเสียง “อุ” เป็นเสียง “ง” เพราะฉะนั้นสามารถเขียนแทนเสียง “อุ” ได้เป็น /kr - u + ng^/ และในส่วนของเสียงตัวสะกด “ง” นั้นมีเสียงก่อนหน้าเป็นเสียง “อุ” และต่อจากเสียง “ง” เป็นเสียงเงียบ จึงเขียนแทนเสียง “ง” ได้เป็น /u - ng^ + sil/ เป็นต้น

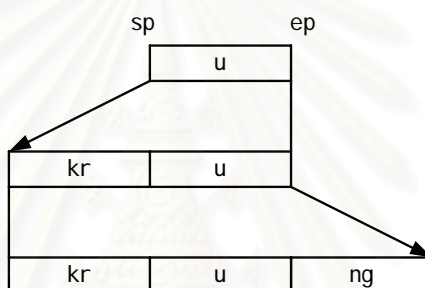
Onset – Rhyme

เป็นเสียงระดับย่อยของพยางค์ที่แบ่งคำพยางค์เดี่ยวออกเป็น 2 ส่วนคือ หน่วยเริ่มของพยางค์ (Onset) และหน่วยตามของพยางค์ (Rhyme) (วิศรุต อาชุนุต, 2545; Maneenoi, E., 2002) ในส่วนของหน่วยเริ่มของพยางค์จะประกอบด้วยเสียงของพยัญชนะต้นและเสียงเริ่มของเสียงสระในคำนั้น เช่น คำว่า “กรุง” สามารถเขียนแทนในส่วนของหน่วยเริ่มของพยางค์ได้เป็น /kr_u/ คำว่า “หิวหิน” สามารถเขียนแทนในส่วนของหน่วยเริ่มของพยางค์ได้เป็น /h_u/ และ /h_i/ ตามลำดับ คำว่า “เกาะ” สามารถเขียนแทนในส่วนของหน่วยเริ่มของพยางค์ได้เป็น /k_@/ เป็นต้น สำหรับในส่วนของหน่วยตามของพยางค์จะประกอบด้วยเสียงของสระและเสียงของตัวสะกด ส่วนคำภาษาไทยที่ไม่มีตัวสะกด หน่วยตามของพยางค์จะประกอบด้วยเสียงของสระเพียงอย่างเดียว เช่น คำว่า “กรุง” สามารถเขียนแทนในส่วนของหน่วยตามของพยางค์ได้เป็น /u_ng^/ คำว่า “หิวหิน” สามารถเขียนแทนในส่วนของหน่วยตามของพยางค์ได้เป็น /uua/ และ /i_n^/ ตามลำดับ คำว่า “เกาะ” สามารถเขียนแทนในส่วนของหน่วยตามของพยางค์ได้เป็น /@/ เป็นต้น

งานวิจัยระบบที่เฉพาะคำสำคัญภาษาไทยนี้ได้อาศัยแนวความคิดจากงานวิจัยของ Takeshi และ Masaki (1987) มาประยุกต์ใช้ซึ่งประกอบด้วยขั้นตอน Bottom – up Detector และขั้นตอน Top – down verifier ซึ่งมีรายละเอียดของขั้นตอนต่าง ๆ ดังนี้

Bottom – up Detector

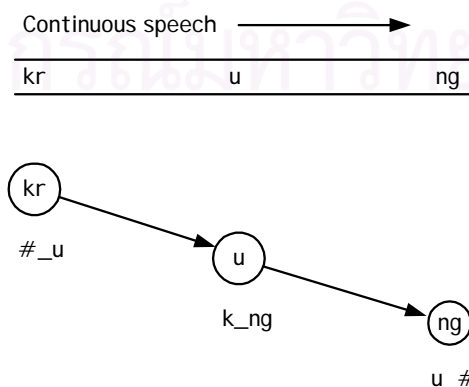
วิธี Bottom – up Detector นั้นใช้ Context – independent phoneme feature ในการตรวจสอบแต่ละส่วนย่อยของพยางค์ ซึ่งวิธีนี้จะใช้เวลาในการประมวลผลน้อย แต่ที่ไม่มีควมน่าเชื่อถือในผลที่ได้มา ในรูปที่ 2.12 เป็นการตรวจหากลุ่มคำโดยเริ่มจากเสียงสระ (vowel phoneme) ก่อนเพราะว่าความแตกต่างกันของเสียงสระนั้นมีน้อยกว่าเสียงพยัญชนะ ทำให้ใช้เวลาในการคำนวณลดลง จากนั้นตรวจสอบเสียงพยัญชนะต้นและเสียงตัวสะกด



รูปที่ 2.12 รายละเอียดขั้นตอน Bottom – up Detector

Top – down Verifier

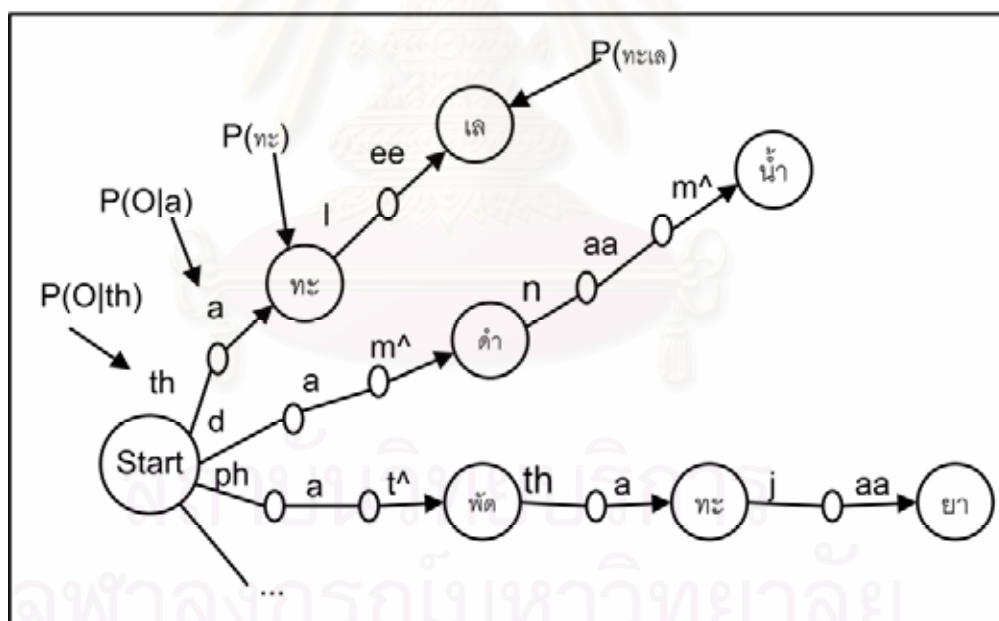
ในขั้นตอนต่อมาเป็นการตรวจสอบกลุ่มคำโดยใช้วิธี Top – down verifier ซึ่งจะพิจารณาผลที่ได้จากวิธี Bottom – up Detector มาตรวจสอบ เพราะวิธีนี้ใช้แบบจำลองของ Context – dependent features ทำให้ผลที่ได้มีความน่าเชื่อถือมากกว่า ในงานวิจัยของ Takeshi (1987) ใช้ Context – dependent features เป็น Triphone ตรวจสอบคำสำคัญดังรูปที่ 2.13



รูปที่ 2.13 รายละเอียดขั้นตอน Top - Down Verifier

ณ. แบบจำลองภาษา (Language Model)

แบบจำลองภาษาเป็นโมเดลของคำต่าง ๆ ที่มาประกอบกันเป็นประโยคที่มีความหมาย และใช้กันในชีวิตประจำวัน เพื่อนำประโยคเสียงพูดมาทดสอบในระบบการรู้จำและให้ค่าความน่าจะเป็นที่ดีที่สุด จะแทนได้ว่าเสียงพูดที่นำมาทดสอบระบบการรู้จำคือประโยคในโมเดลภาษาที่ให้ค่าความน่าจะเป็นที่ดีที่สุด เช่น กำหนดให้ประโยค ซึ่งประกอบด้วยคำต่าง ๆ นำมาต่อกันหลาย ๆ คำ สามารถแทนด้วย $W_n = (w_1 w_2 \dots w_M)$ โดยที่ w_m แทนคำแต่ละคำ โมเดลภาษาจะบอกว่า W_n มีโอกาสเกิดขึ้นให้ค่าเป็นความน่าจะเป็น $P(W_n)$ (Young, S., Kershaw, D., Odell, J. et al., 2000) ซึ่งเมื่อแทนเสียงพูดที่นำมาทดสอบระบบการรู้จำด้วยประโยคทั้งหมดในโมเดลภาษาแล้วจะได้ค่า $P(W_n)$ ที่ให้ค่าความน่าจะเป็นมากที่สุด ซึ่งจะได้ว่าประโยค W_n เป็นประโยคของเสียงพูดที่นำมาทดสอบระบบการรู้จำ สำหรับงานวิจัยระบบการรู้จำเฉพาะคำสำคัญภาษาไทยได้ออกแบบลักษณะของแบบจำลองภาษาให้สามารถใช้งานวิจัยได้โดยกำหนดให้แบบจำลองภาษามีลักษณะเป็นการเชื่อมต่อกันของแบบจำลองของเสียงส่วนย่อยให้เป็นคำสำคัญตามที่กำหนดไว้ในงานวิจัย ดังแสดงในรูปที่ 2.14



รูปที่ 2.14 รายละเอียดของแบบจำลองภาษา

บทที่ 3

ขั้นตอนวิธีในการดำเนินการวิจัย

รายละเอียดในบทที่ 3 นี้กล่าวถึงรายละเอียดของขั้นตอนวิธีในการดำเนินการวิจัย ประกอบไปด้วยรายละเอียดเกี่ยวกับขั้นตอนการฝึกฝนระบบการรู้จำคำสำคัญและขั้นตอนการทดสอบระบบการรู้จำคำสำคัญ นอกจากนี้ยังมีรายละเอียดเกี่ยวกับการกำหนดวิธีการสร้างชุดคำสำคัญภาษาไทยที่ใช้ในการฝึกฝนและและการสร้างชุดประโยคพูดภาษาไทยเพื่อใช้ในการทดสอบ รวมทั้งรายละเอียดในการเก็บตัวอย่างเสียงพูดเพื่อนำมาใช้เป็นตัวอย่างเพื่อฝึกฝนระบบและเป็นตัวอย่างทดสอบ

การกำหนดวิธีการสร้างชุดคำสำคัญ

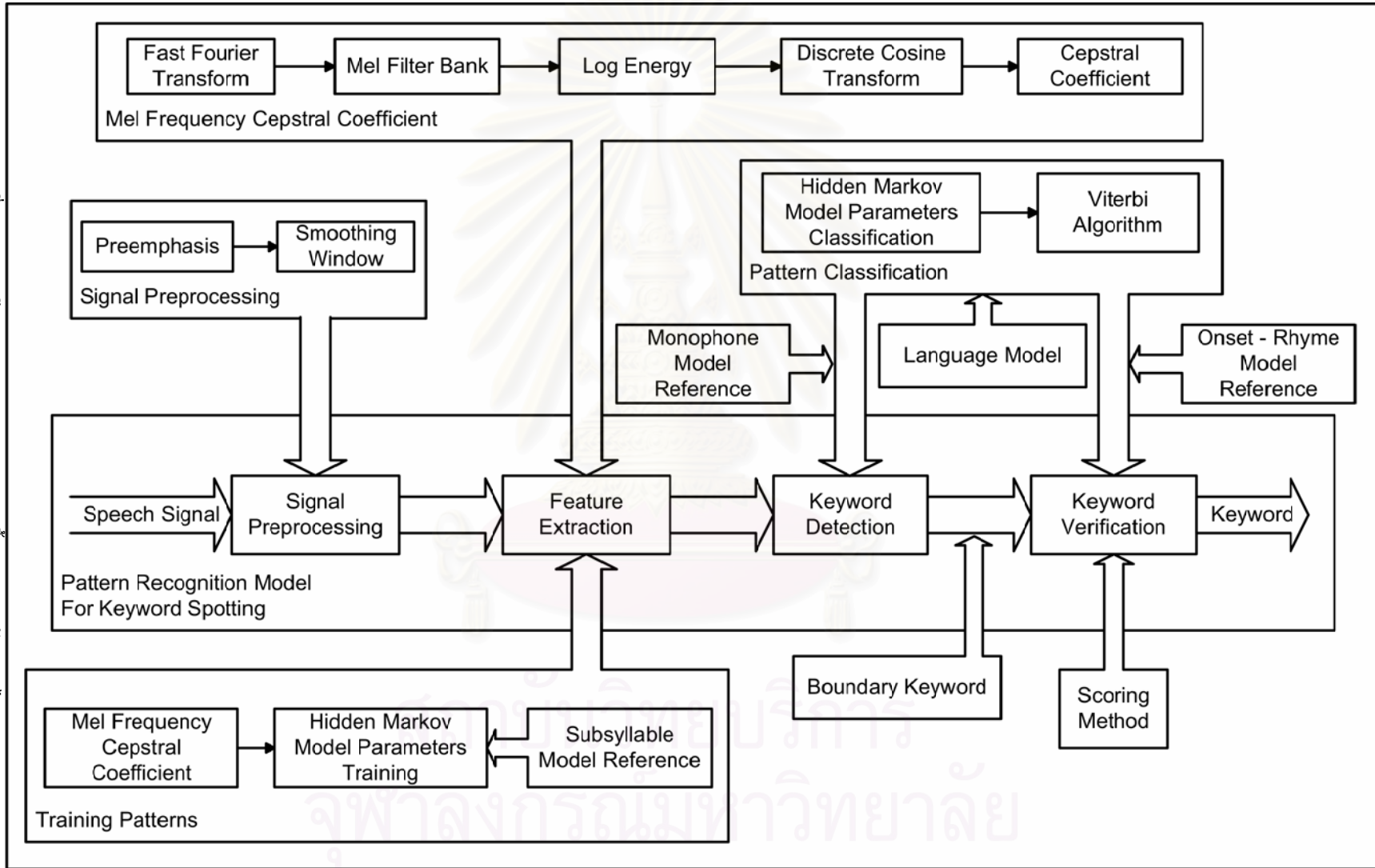
ในการสร้างชุดคำสำคัญสำหรับระบบการรู้จำคำสำคัญภาษาไทยของการวิจัยครั้งนี้ ได้จำกัดจำนวนคำสำคัญ โดยกำหนดให้คำสำคัญเป็นชื่อสถานที่ท่องเที่ยวทางทะเล ชื่อจังหวัดในประเทศไทยและกิจกรรมทางทะเล โดยมีจำนวนคำสำคัญทั้งหมด 70 คำ แบ่งเป็น 3 ชุดตามจำนวนคือ 20 คำ 40 คำและ 70 คำ และชุดประโยคภาษาไทยที่เป็นบทความของแหล่งท่องเที่ยวทางทะเลซึ่งประกอบด้วยข้อความ 6 ย่อหน้า เป็นจำนวนประโยคทั้งหมด 40 ประโยค เพื่อนำมาทดสอบระบบรู้จำคำสำคัญภาษาไทย ดังมีรายละเอียดแสดงในภาคผนวก ก

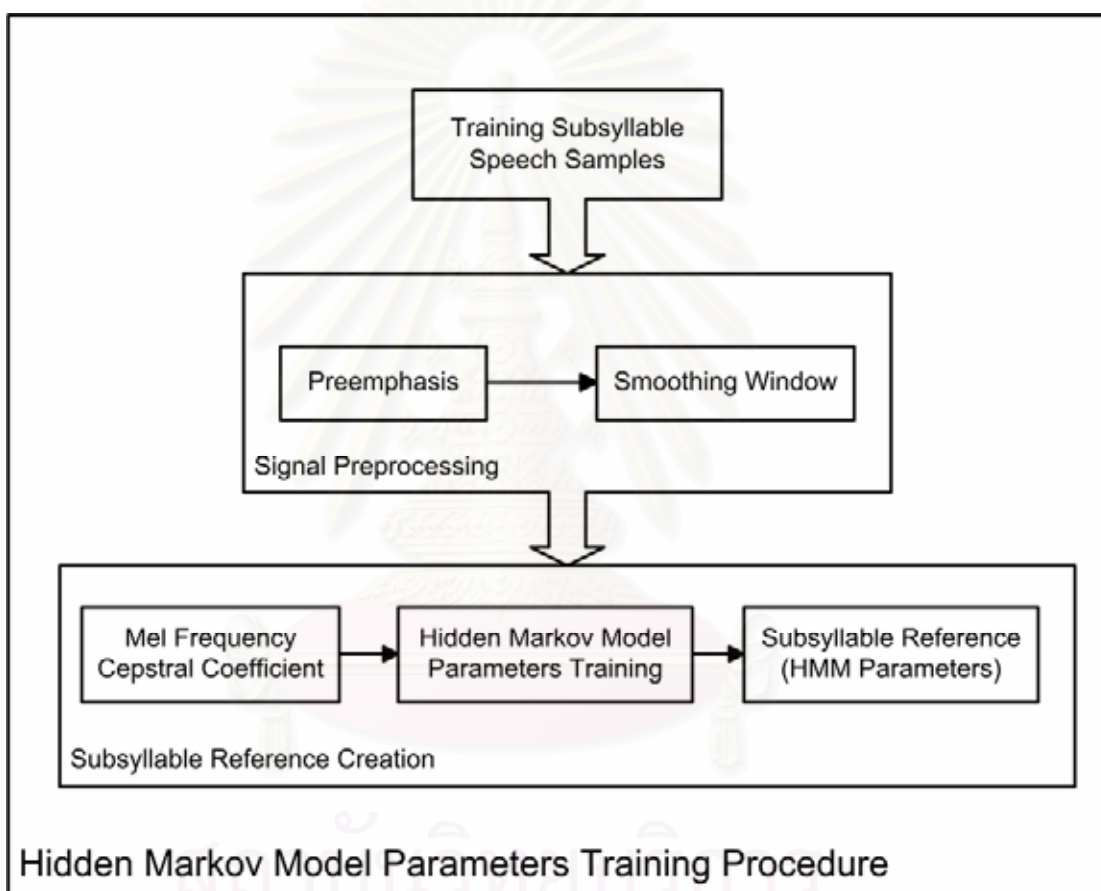
การเก็บตัวอย่างข้อมูลเสียงพูด

การเก็บตัวอย่างข้อมูลเสียงพูดจะอาศัยการเก็บบันทึกข้อมูลไว้ในเครื่องคอมพิวเตอร์ โดยทำการบันทึกเสียง ณ ห้องปฏิบัติการวิจัยประมวลผลสัญญาณดิจิทัล ชั้น 13 ตึกวิศวกรรมศาสตร์ 4 คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ซึ่งได้รับการควบคุมสภาพแวดล้อมขณะทำการบันทึกไม่ให้มีเสียงรบกวนจากภายนอก โดยเสียงพูดที่บันทึกไว้จะจัดเก็บด้วยตัวอย่างขนาด 16 บิตและมีอัตราการสุ่มตัวอย่าง 16 KHz ซึ่งกำหนดให้ผู้พูดบันทึกเสียงพูดเป็นบทความของแหล่งท่องเที่ยวทางทะเลและคำสำคัญทั้งหมด 70 คำ

ในงานวิจัยนี้ทำการเก็บตัวอย่างเสียงพูดจำนวน 50 คน แบ่งเป็นเสียงผู้ชาย 30 คนและเสียงผู้หญิง 20 คน โดยแบ่งออกเป็น 2 ชุด ได้แก่ชุดเสียงพูดเพื่อฝึกฝนระบบ (Training Set) และชุดเสียงพูดเพื่อทดสอบระบบ (Test Set) สำหรับชุดเสียงเพื่อฝึกฝนระบบมีจำนวน 30 คน แบ่งเป็นเสียงผู้ชาย 20 คนและเสียงผู้หญิง 10 คน ชุดเสียงเพื่อทดสอบระบบมีจำนวน 20 คนแบ่งเป็นเสียงผู้ชาย 10 คนและเสียงผู้หญิง 10 คน

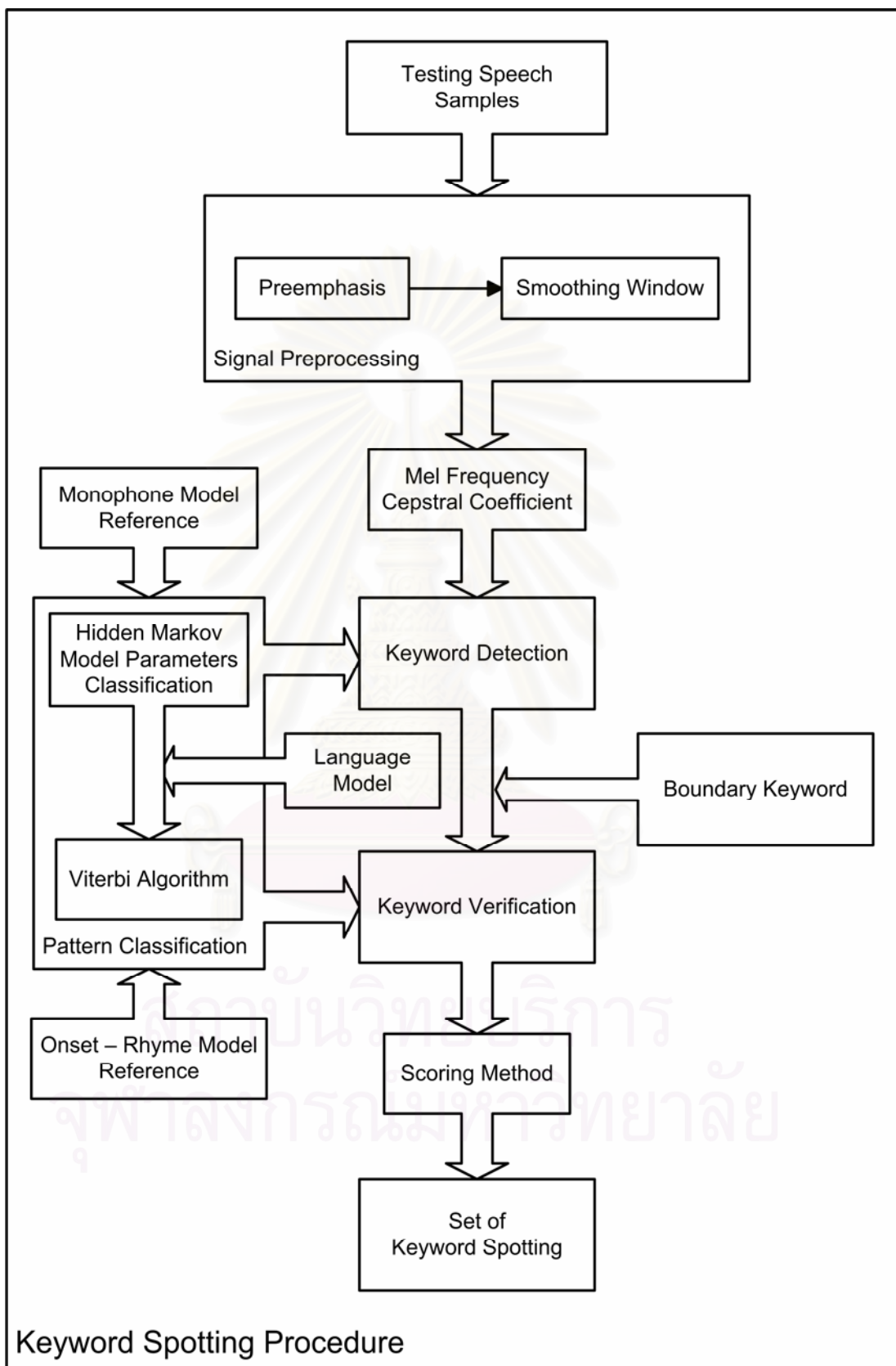
รูปที่ 3. 1 รายละเอียดของแบบจำลองระบบการรู้จำคำสำคัญภาษาไทย





รูปที่ 3.2 รายละเอียดขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.3 รายละเอียดขั้นตอนการทดสอบระบบการชี้เฉพาะคำสำคัญภาษาไทย

รายละเอียดขั้นตอนในระบบการชี้เฉพาะคำสำคัญภาษาไทย

ระบบการชี้เฉพาะคำสำคัญภาษาไทยในการวิจัยนี้มีรายละเอียดขั้นตอนต่าง ๆ ดังรูปที่ 3.1 ซึ่งสามารถแบ่งออกเป็น 2 ขั้นตอน ได้แก่ ขั้นตอนการฝึกฝนระบบการรู้จำคำสำคัญดังรูปที่ 3.2 และขั้นตอนการทดสอบระบบการชี้เฉพาะคำสำคัญภาษาไทยดังรูปที่ 3.3 โดยได้แสดงรายละเอียดของขั้นตอนต่าง ๆ ดังต่อไปนี้

1. ขั้นตอนการฝึกฝนระบบการรู้จำคำสำคัญ

ขั้นตอนการฝึกฝนระบบการรู้จำคำสำคัญนี้ จัดเป็นขั้นตอนในการสร้างชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ เพื่อใช้ในการรู้จำคำสำคัญ ซึ่งมีรายละเอียดดังนี้

1.1 ขั้นตอนการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

ขั้นตอนการประมวลผลสัญญาณเบื้องต้นเป็นกรรมวิธีในการจัดเตรียมข้อมูลจากข้อมูลเสียงที่ได้จากการบันทึกเสียงซึ่งเป็นข้อมูลดิบ นำมาผ่านกรรมวิธีประมวลผลสัญญาณเชิงเลขโดยแบ่งออกได้เป็น 2 กรรมวิธีย่อยได้แก่ กรรมวิธีเน้นล่วงหน้า (Preemphasis) กรรมวิธีวางกรอบสัญญาณ (Smoothing Window) ตามลำดับ

ในขั้นตอนการประมวลผลสัญญาณเบื้องต้นนั้น เนื่องจากสัญญาณเสียงพูดมีความแปรเปลี่ยนตามเวลา (Time-varying) และไม่เสถียร (Nonstationary) อีกทั้งยังเป็นสัญญาณสุ่มที่ไม่มีความเป็นเออร์годิก (Non-Ergodic) และไม่เป็นสัญญาณเฟ้นสุ่ม (Non-stochastic Signal) อีกด้วย ดังนั้นในการประยุกต์ใช้งานขั้นตอนวิธีการต่าง ๆ กับสัญญาณเสียงพูดจึงต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อย (Furui, S., 1985) เรียกว่า “กรอบเสียงพูด” (Speech Frame) โดยแต่ละกรอบเสียงพูดจะมีความยาวประมาณ 10 – 40 มิลลิวินาที (ms) ขึ้นอยู่กับความถี่ในการสุ่มตัวอย่าง (Sampling Frequency)

1.1.1 กรรมวิธีเน้นล่วงหน้า (Preemphasis)

ขั้นตอนกรรมวิธีเน้นล่วงหน้านี้เป็นการบีบอัดช่วงพิสัยพลวัต (Dynamic Range) ของสัญญาณเสียงพูด โดยการทำให้ความลาดเอียงในเชิงความถี่แบบราบลงซึ่งจะส่งผลให้ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าสูงขึ้น ในทางปฏิบัติแล้วจะนำสัญญาณผ่านตัวกรองเชิงเลขลำดับหนึ่ง (First-Order Digital Filter) ที่มีฟังก์ชันถ่ายโอน $H(z)$ ดังแสดงในสมการที่ (3-1) และ (3-2) (Furui, S., 1985) เมื่อ a เป็นสัมประสิทธิ์ของตัวกรอง $\tilde{v}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรรมวิธีเน้นล่วงหน้าที่ n และ $s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n และ $s(n-1)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าค่าก่อนหน้าที่ $n-1$ ดังนี้

$$H(z) = 1 - az^{-1} \dots\dots\dots(3-1)$$

$$\tilde{s}(n) = s(n) - as(n-1) \dots\dots\dots(3-2)$$

โดยกำหนดให้ค่าสัมประสิทธิ์ของตัวกรอง a มีค่าเข้าใกล้ 1 ในงานวิจัยนี้ กำหนดให้มีค่า $a = 0.95$ (วิศรุต อาชูปุตระ, 2539; เสาวลักษณ์ อารีพงศ์, 2538; เอกฤทธิ์ มณีน้อย, 2541)

1.1.2 กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

ขั้นตอนกรรมวิธีวางกรอบขนาดสัญญาณจัดเป็นขั้นตอนในการเตรียมข้อมูลในแต่ละกรอบข้อมูลเสียงพูด สัญญาณที่ผ่านกรรมวิธีเน้นล่วงหน้าจะถูกแบ่งออกเป็นกรอบเสียงพูดโดยการคูณสัญญาณเสียงพูดกับฟังก์ชันกรอบ ผลของการวางกรอบสัญญาณคือการลดทอนแอมพลิจูดอย่างช้า ๆ ที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อหลีกเลี่ยงความไม่ต่อเนื่องที่จุดปลาย ฟังก์ชันกรอบที่ใช้สำหรับการวิเคราะห์เสียงพูดในงานวิจัยนี้คือฟังก์ชันกรอบชนิด Hamming โดยที่กำหนดให้ขนาดของกรอบเท่ากับ 25 มิลลิวินาที และการเลื่อนของฟังก์ชันกรอบ เท่ากับ 10 มิลลิวินาที (Tolba, H., O'Shaughnessy, D., 1998)

1.2 การสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model Parameters Training)

ในหัวข้อนี้ได้แสดงรายละเอียดขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟดังแสดงในรูปที่ 3.2 ประกอบไปด้วย 2 ขั้นตอนได้แก่ ขั้นตอนการสกัดค่าลักษณะสำคัญโดยใช้สัมประสิทธิ์เซปสตรอลบนความถี่เมล และขั้นตอนการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model Parameters Training) ภายหลังจากการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟจะได้ชุดต้นแบบอ้างอิงของเสียงส่วนย่อยของพยางค์ (Subsyllable Reference)

1.2.1 ขั้นตอนการหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล (Mel Frequency Cepstral Coefficient)

ขั้นตอนการหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลจัดเป็นขั้นตอนในการลดจำนวนข้อมูลโดยการแสดงลักษณะของรูปคลื่นด้วยค่าพารามิเตอร์เพียงไม่กี่ค่าได้อย่างมีประสิทธิภาพ โดยนำข้อมูลเสียงพูดมาทำการวิเคราะห์ดังแสดงในบทที่ 2 หัวข้อ ข. ในการหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลสำหรับแต่ละเสียงพูดนั้น จะหาค่าสัมประสิทธิ์เฉพาะแต่ละกรอบข้อมูลเสียงพูด โดยการแทนที่แต่ละกรอบข้อมูลเสียงพูดด้วยเวกเตอร์ของสัมประสิทธิ์

เซปสตรอลบนความถี่เมล สำหรับงานวิจัยนี้จะกำหนดให้ค่าลำดับของสัมประสิทธิ์ที่ 16 ลำดับ (พงศืไท ทาสระคฺ, 2542)

1.2.2 ขั้นตอนวิธีการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ

ขั้นตอนวิธีการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟจัดเป็น กระบวนการสร้างชุดต้นแบบอ้างอิงของเสียงพูดแต่ละคำ โดยทำการปรับเปลี่ยนค่าพารามิเตอร์ ของแบบจำลองฮิดเดนมาร์คอฟ $\lambda = (A, B, \pi)$ ให้เป็นไปตามเสียงพูดแต่ละคำ ซึ่งอาศัยการแก้ไข ปัญหาพื้นฐานข้อที่ 1 และข้อที่ 3 ของแบบจำลองฮิดเดนมาร์คอฟด้วยกระบวนการไปหน้า – ย้อนกลับ (Forward – Backward Procedure) และการประมาณค่าซ้ำของ Baum – Welch (Baum – Welch Reestimation Procedure) ดังมีรายละเอียดแสดงในบทที่ 2 หัวข้อที่ 3.1 และ 3.3 ตามลำดับ

การสร้างชุดแบบจำลองฮิดเดนมาร์คอฟในการวิจัยนี้ กำหนดให้ แบบจำลองฮิดเดนมาร์คอฟเป็นชุดของ phoneme ต่าง ๆ ของคำสำคัญ ซึ่งสร้างขึ้นเป็น 3 ประเภท คือ

- 1) Monophone
- 2) Triphone
- 3) Onset – Rhyme

แบบจำลองทั้ง 3 ประเภทนี้ได้กล่าวถึงในบทที่ 2 หัวข้อที่ ๕. และเสียงพูดที่จะนำมาผ่านกรรมวิธี การฝึกฝนชุดพารามิเตอร์นั้น เป็นเสียงพูดของคำสำคัญที่ใช้ในระบบซึ่งเฉพาะคำสำคัญ ก่อนที่จะ นำมาฝึกฝนชุดพารามิเตอร์นั้นจะต้องผ่านกรรมวิธีแบ่งขอบเขตของคำในระดับ phoneme ในคำ สำคัญนั้น ๆ ดังที่กล่าวถึงในบทที่ 2 หัวข้อที่ ๕. เมื่อแบ่งขอบเขตของคำในระดับ phoneme ในคำ สำคัญได้แล้วจึงผ่านกรรมวิธีการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ แบบจำลองประเภทต่าง ๆ ที่ได้จะมีรายละเอียดแสดงในภาคผนวก ก

2. การทดสอบระบบการชี้เฉพาะคำสำคัญภาษาไทย

ขั้นตอนในการทดสอบระบบการชี้เฉพาะคำสำคัญภาษาไทยโดยใช้แบบจำลองฮิด เดน มาร์คอฟประกอบด้วย 5 ขั้นตอน ได้แก่ขั้นตอนการประมวลผลสัญญาณเบื้องต้น ขั้นตอนการ สกัดค่าลักษณะสำคัญ (Feature Extraction) ขั้นตอนการตรวจพบคำสำคัญ (Keyword

Detection) ขั้นตอนการตรวจสอบคำสำคัญ (Keyword Verification) และขั้นตอนการตัดสินใจ ในงานวิจัยนี้ได้แบ่งจำนวนคำสำคัญในการทดสอบระบบการชี้เฉพาะคำสำคัญนี้เป็นจำนวน 20 คำ 40 คำ และ 70 คำ เพื่อนำมาเปรียบเทียบผลลัพธ์ที่ได้ ซึ่งคำสำคัญในแต่ละชุดนี้ได้แสดงใน ภาคผนวก ก เนื่องจากในขั้นตอนการประมวลผลสัญญาณเบื้องต้นและขั้นตอนการสกัดค่า ลักษณะสำคัญมีรายละเอียดอยู่ในขั้นตอนการฝึกฝนชุดคำสำคัญ ในหัวข้อที่ 1 สำหรับในหัวข้อนี้ จึงแสดงเพียงรายละเอียดของขั้นตอนที่เหลือดังนี้

2.1 ขั้นตอนการตรวจพบคำสำคัญ (Keyword Detection)

เมื่อได้ค่าลักษณะสำคัญในขั้นตอนการสกัดค่าลักษณะสำคัญแล้วจึงนำมาผ่าน ขั้นตอนการตรวจพบคำสำคัญ ซึ่งใช้การแก้ปัญหาพื้นฐานข้อที่ 2 ของแบบจำลองฮิดเดน มาร์คอฟ โดยการใช้วิธีการ Viterbi (Viterbi Algorithm) ในบทที่ 2 หัวข้อที่ 3.2 โดยใช้แบบจำลองประเภท Monophone แต่ต้องกำหนดแบบจำลองภาษาขึ้นมาก่อน เนื่องจากว่าระบบการชี้เฉพาะคำสำคัญ ต้องการผลลัพธ์ที่ได้เมื่อผ่านระบบเป็นความถูกต้องของคำสำคัญเท่านั้น คำอื่น ๆ ที่ไม่ใช่คำสำคัญ ไม่มีความจำเป็นในการรู้จำของระบบ เพราะฉะนั้นการสร้างแบบจำลองภาษาจึงมีเฉพาะคำสำคัญ เท่านั้น ดังแสดงในรูปที่ 2.14

เมื่อนำค่าลักษณะสำคัญผ่านขั้นตอนนี้ผลลัพธ์ที่ได้จะประกอบด้วย คำสำคัญ ในเสียงพูด (ไม่มีคำอื่น ๆ เพราะว่าเป็นแบบจำลองภาษามีเฉพาะคำสำคัญเท่านั้น) ช่วงเวลาเริ่มต้น และสิ้นสุดของคำนั้น ๆ และค่าคะแนนของคำที่เกิดขึ้น ซึ่งเป็นค่าคะแนนของคำนั้น ๆ ในแบบจำลองภาษาที่ให้ค่าที่ดีที่สุดเมื่อเทียบกับคำอื่นในแบบจำลองภาษา โดยค่าคะแนนที่ได้ นั้นเป็นค่า logarithm ของค่าความน่าจะเป็นของคำ $P(W_n)$ ตัวอย่างดังแสดงในตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างของผลที่ได้จากการทดสอบระบบ

	Word	Interval (s)	Score
Input	“ห้องเที่ยวทะเลไทย”	0.00 – 1.43	-
Output	“หาด”	0.08 – 0.31	-1731.796021
	“เสมีด”	0.31 – 0.62	-2245.540283
	“หาด”	0.62 – 1.43	-6063.924316

2.2 ขั้นตอนการตรวจสอบคำสำคัญ (Keyword Verification)

ในขั้นตอนการตรวจสอบคำสำคัญจะพิจารณาจากผลลัพธ์ที่ได้จากขั้นตอนการ ตรวจพบคำสำคัญ ทำการวิเคราะห์ในแต่ละคำที่ได้นำมาดึงข้อมูลของค่าลักษณะสำคัญของ

เสียงพูดในช่วงเวลาที่ได้จากผลลัพธ์ของคำนั้นในขั้นตอนการตรวจพบคำสำคัญ เมื่อได้ค่าลักษณะสำคัญของเสียงพูดตามช่วงเวลาต่าง ๆ แล้วจึงนำค่าลักษณะสำคัญของเสียงพูดตามช่วงเวลามาผ่านกระบวนการขั้นตอนการตรวจสอบคำสำคัญ ซึ่งใช้การแก้ปัญหาพื้นฐานข้อที่ 2 ของแบบจำลองฮิดเดนมาร์คอฟ โดยการใช้วิธีการ Viterbi (Viterbi Algorithm) ในบทที่ 2 หัวข้อที่ 3.2 และจำแนกคำตามแบบจำลองภาษา ซึ่งงานวิจัยนี้ได้เสนอการใช้แบบจำลองประเภท Onset – Rhyme เพื่อใช้ในขั้นตอนการตรวจสอบคำสำคัญ และเพื่อเป็นการเปรียบเทียบประสิทธิภาพของระบบในขั้นตอนการตรวจสอบคำสำคัญจึงใช้แบบจำลองประเภท Triphone ซึ่งเป็น Context Dependent Phoneme ตามงานวิจัยของ Takeshi (1987) นำมาตรวจสอบคำสำคัญ

2.3 ขั้นตอนการตัดสินใจ

เนื่องจากผลที่ได้ทั้งหมดเป็นคำสำคัญของระบบ ในขั้นตอนนี้จึงเป็นการจำแนกผลลัพธ์ว่าเสียงพูดในช่วงเวลาใดที่เป็นคำสำคัญและเสียงพูดในช่วงเวลาใดไม่ได้เป็นคำสำคัญสามารถจำแนกได้โดยดูค่าคะแนนของแต่ละคำสำคัญของผลที่ได้จากขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone และ Onset – Rhyme แต่อย่างไรก็ตามการวิเคราะห์โดยการดูคะแนนเพียงอย่างเดียวไม่เพียงพอ เพราะว่าค่าคะแนนของคำสำคัญที่มีช่วงเวลามากจะมีค่าน้อยกว่าเมื่อเปรียบเทียบกับค่าคะแนนของคำสำคัญที่มีช่วงเวลาน้อย เช่น คำว่า “พัทยา” เมื่อเปรียบเทียบกับคำว่า “หาด” ค่าคะแนนของคำว่า “พัทยา” จะมีค่าน้อยกว่าค่าคะแนนของคำว่า “หาด” จึงจำเป็นต้องปรับค่าคะแนนใหม่ให้เป็นค่าคะแนนดังสมการที่ 3-3 (Mitchel, W., 1995)

$$\overline{Score}_n = \frac{Score_n}{EndTime_n - StartTime_n} \dots\dots\dots(3-3)$$

วิธีการจำแนกคำสำคัญในงานวิจัยนี้ใช้วิธี Log-Likelihood (Mitchel, W., 1995) เป็นการกำหนดค่าจุดเริ่มเปลี่ยนเพื่อจำแนกค่าคะแนนของคำที่ได้จากสมการที่ 3-3 ว่าเป็นคำสำคัญหรือไม่ โดยถ้าค่าคะแนนของคำนั้นมีค่าต่ำกว่าค่าจุดเริ่มเปลี่ยนจะถูกจำแนกว่าไม่จัดอยู่ในกลุ่มคำสำคัญและค่าคะแนนของคำสำคัญที่มีค่ามากกว่าหรือเท่ากับค่าจุดเริ่มเปลี่ยนจะถูกจำแนกว่าจัดอยู่ในกลุ่มคำสำคัญ

บทที่ 4

ผลการวิจัย

ในบทนี้จะกล่าวถึงรายละเอียดของผลการวิจัยเฉพาะคำสำคัญภาษาไทย ซึ่งการวิจัยนี้ได้ทำการเปรียบเทียบประสิทธิภาพของการใช้แบบจำลองประเภท Monophone เพื่อตรวจพบคำสำคัญ การใช้แบบจำลองประเภท Triphone และแบบจำลองประเภท Onset – Rhyme เพื่อตรวจสอบคำสำคัญ ในการทดสอบที่มีจำนวนคำสำคัญแตกต่างกันคือ 20 คำ 40 คำและ 70 คำ โดยใช้เสียงทดสอบจากคน 40 คนแบ่งออกเป็น 20 คนคือเสียงของผู้พูดที่อยู่ในชุดฝึกฝน และอีก 20 คนคือเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน

การวัดประสิทธิภาพของระบบที่เฉพาะคำสำคัญสามารถวิเคราะห์ได้จากความแม่นยำของระบบ (Accuracy) ค่า FOM (Figure of Merit) และ ROC Curve (Receiver Operating Characteristics) (Young, S., Kershaw, D., Odell, J. et al., 2000) ซึ่งการวัดประสิทธิภาพของระบบที่เฉพาะในแต่ละแบบมีรายละเอียดดังนี้

ความแม่นยำของระบบที่เฉพาะคำสำคัญ

ความแม่นยำของระบบที่เฉพาะคำสำคัญสามารถคำนวณได้จากสมการดังนี้

$$Accuracy = 100\% - \% False Accept - \% False Reject \dots\dots\dots(4-1)$$

อัตราความผิดพลาดจากการยอมรับ (False Accept) นั้นสามารถเกิดขึ้นได้จากการตัดสินใจที่ผิดพลาดเนื่องจากค่าคะแนนของคำพูดนั้นมีค่าสูงกว่าค่าผ่านระดับ จึงตัดสินใจให้เป็นคำสำคัญ มีโอกาสเกิดขึ้นได้ 2 กรณีคือ คำพูดนั้นไม่ได้เป็นคำสำคัญแต่ถูกตัดสินใจให้เป็นคำสำคัญ และคำพูดนั้นเป็นคำสำคัญแต่ถูกตัดสินใจให้เป็นคำสำคัญค่าอื่น สำหรับอัตราความผิดพลาดจากการไม่ยอมรับ (False Reject) นั้นสามารถเกิดขึ้นได้จากการตัดสินใจที่ผิดพลาดเนื่องจากค่าคะแนนของคำพูดนั้นมีค่าต่ำกว่าค่าผ่านระดับ จึงตัดสินใจว่าคำพูดนั้นไม่เป็นคำสำคัญ มีโอกาสเกิดขึ้นได้กรณีเดียวคือ คำพูดนั้นเป็นคำสำคัญแต่ถูกตัดสินใจไม่ให้เป็นคำสำคัญ ซึ่งการคำนวณหาอัตราความผิดพลาดทั้งสองชนิดนี้พิจารณาจากค่าทั้งหมดที่ปรากฏขึ้นจากขั้นตอนการตรวจสอบคำสำคัญ

ค่า FOM ของระบบที่เฉพาะคำสำคัญ

การวิเคราะห์ประสิทธิภาพการที่เฉพาะคำสำคัญสามารถวัดได้จากการคำนวณค่า Figure of Merit (FOM) ซึ่งค่านี้ถูกกำหนดโดย US National Institute of Standards and Technology

(NIST) (Young, S., Kershaw, D., Odell, J. et al., 2000) โดยการเปรียบเทียบผลของคำสำคัญที่ได้จากการทดลองกับชุดคำสำคัญจากเสียงพูดต้นแบบที่มีช่วงเวลาที่ถูกต้องของคำสำคัญ คือนำช่วงเวลาของคำสำคัญที่ได้จากการทดลองมาตรวจสอบกับช่วงเวลาของชุดต้นแบบของคำสำคัญ ถ้าช่วงเวลาของคำสำคัญที่ได้จากการทดลองตรงกับช่วงเวลาของชุดต้นแบบของคำสำคัญเรียกว่า Hit นอกจากนั้นจะเรียกว่า False – alarm (FA) ซึ่งค่า Figure of Merit เป็นการประมาณค่าขอบเขตบนของความแม่นยำเฉลี่ยของระบบการชี้เฉพาะในช่วง 1 – 10 False – alarm ต่อชั่วโมง ค่า FOM ถูกคำนวณโดยการกำหนดให้ช่วงเวลาทั้งหมดของเสียงพูดที่นำมาทดสอบคือ T ชั่วโมง ดังแสดงในสมการต่อไปนี้

$$FOM = \frac{1}{10T}(p_1 + p_2 + \dots + p_N + ap_{N+1}) \dots \dots \dots (4-2)$$

เมื่อ N คือจำนวนเต็มค่าแรกที่ $N \geq 10T - 0.5$

p_i คือเปอร์เซ็นต์ความถูกต้องก่อนพบ False alarm ลำดับที่ i

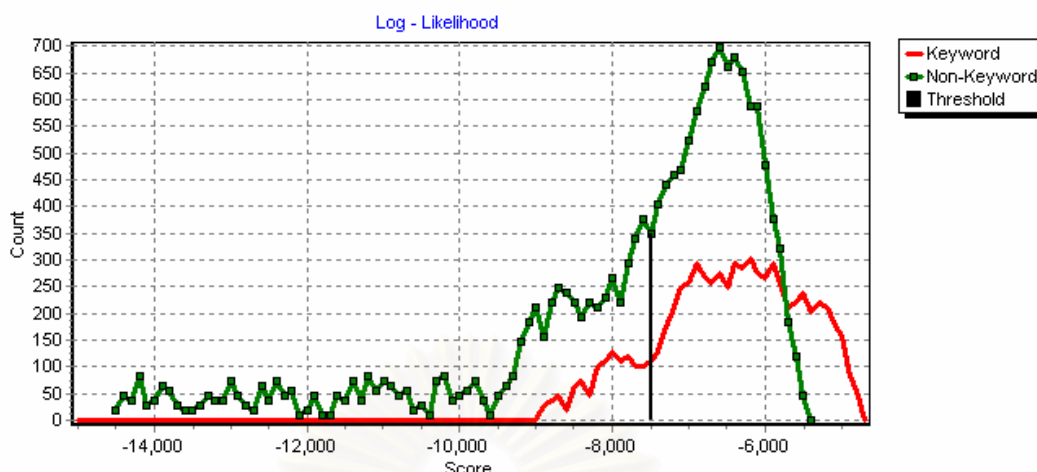
a คือเป็นตัวคูณมีค่าเป็น $a = 10T - N$ ซึ่งมีผลทำให้เป็น 10 False alarm ต่อชั่วโมง

ROC Curve

ROC Curve คือกราฟแสดงถึงค่าระหว่างอัตราการตรวจพบคำสำคัญ (เปอร์เซ็นต์) และจำนวน False alarm ตั้งแต่ 0 – 10 (Young, S., Kershaw, D., Odell, J. et al., 2000) เป็นกราฟซึ่งถูกกำหนดโดย US National Institute of Standards and Technology (NIST) เช่นกัน

ผลการชี้เฉพาะคำสำคัญภาษาไทย

ในขั้นตอนการวิเคราะห์ผลการวิจัยระบบการชี้เฉพาะคำสำคัญภาษาไทย เมื่อพิจารณาจากรูปที่ 4.1 ซึ่งเป็นกราฟ Log – Likelihood ของระบบการชี้เฉพาะคำสำคัญภาษาไทยในการวิจัยนี้ พบว่า ถ้าทำการจำแนกคำสำคัญโดยการคำนึงถึงความผิดพลาดน้อยที่สุด จะได้ระบบที่มีอัตราการตรวจพบคำสำคัญน้อยมาก และถ้าทำการจำแนกคำสำคัญโดยให้ตรวจพบคำสำคัญได้ทั้งหมด จะได้ระบบที่มีความผิดพลาดมาก เพื่อเป็นการศึกษาและวิเคราะห์ระบบการชี้เฉพาะคำสำคัญภาษาไทยในงานวิจัยนี้จึงกำหนดให้ค่าจุดเริ่มเปลี่ยนมีค่าเท่ากับ -7494 ซึ่งเมื่อพิจารณาจากกราฟที่เป็นคำสำคัญจะเป็นช่วงแบ่งกลุ่มคำสำคัญที่กราฟเริ่มมีการเปลี่ยนแปลงอย่างรวดเร็ว และพิจารณาจากกราฟที่ไม่เป็นคำสำคัญจะเป็นช่วงที่มีความหนาแน่นสูงกว่าช่วงอื่น ๆ เพื่อสามารถวิเคราะห์ถึงความผิดพลาดที่เกิดขึ้นจากระบบการชี้เฉพาะคำสำคัญภาษาไทยในงานวิจัยนี้



รูปที่ 4. 1 กราฟ Log - likelihood ของระบบการชี้เฉพาะคำสำคัญภาษาไทย

ผลการชี้เฉพาะคำสำคัญภาษาไทยได้แบ่งการทดสอบออกเป็น 3 ส่วนตามจำนวนของคำสำคัญคือ 20 คำ 40 คำ และ 70 คำ ซึ่งในบทนี้ได้แสดงถึงผลที่ได้จากการทดสอบในขั้นตอนการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone และผลที่ได้จากการทดสอบในขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone และ Onset – Rhyme ซึ่งได้แบ่งชุดเสียงสำหรับนำมาทดสอบกับระบบเป็น 2 ชุดคือ ชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน และชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน ซึ่งในแต่ละชุดมีจำนวนผู้พูด 40 คน ตารางที่ 4.1 และ 4.2 ได้แสดงอัตราความแม่นยำเฉลี่ยของระบบการชี้เฉพาะคำสำคัญภาษาไทยโดยตารางที่ 4.1 เป็นอัตราความแม่นยำเฉลี่ยที่ได้มาจากการใช้ชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน และตารางที่ 4.2 เป็นอัตราความแม่นยำเฉลี่ยที่ได้มาจากการใช้ชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน

ตารางที่ 4. 1 ผลอัตราความแม่นยำเฉลี่ยของชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน

ชุดคำสำคัญ	Monophone	Triphone	Onset – Rhyme
20 คำ	83.49 %	90.36 %	91.65 %
40 คำ	68.60 %	79.78 %	79.29 %
70 คำ	66.05 %	76.08 %	77.53 %

ตารางที่ 4. 2 ผลอัตราความแม่นยำเฉลี่ยของชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน

ชุดคำสำคัญ	Monophone	Triphone	Onset – Rhyme
20 คำ	81.63 %	89.41 %	90.21 %
40 คำ	58.60 %	78.49 %	79.13 %
70 คำ	56.98 %	75.76 %	76.40 %

สำหรับตารางที่ 4.3 – 4.5 ได้แสดงถึงจำนวน Hit จำนวน False alarm จำนวนคำสำคัญตามความเป็นจริง (Actual) และค่า FOM โดยค่าต่าง ๆ ในตารางที่ 4.3 เกิดขึ้นจากการขั้นตอนการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone สำหรับค่าต่าง ๆ ในตารางที่ 4.4 เกิดขึ้นจากขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone และค่าต่าง ๆ ในตารางที่ 4.5 เกิดจากขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Onset – Rhyme

ตารางที่ 4.3 รายละเอียดผลการทดสอบค่า FOM ของแบบจำลองเสียงระดับย่อยของพยางค์ประเภท Monophone

ชุดเสียงพูด	จำนวนคำสำคัญ	จำนวน Hit	จำนวน FA	จำนวนคำสำคัญตามความเป็นจริง	ค่า FOM
อยู่ในชุดฝึกฝน	20 คำ	168	1411	200	58.98
	40 คำ	1279	2443	1760	43.86
	70 คำ	1180	2593	1760	38.01
ไม่อยู่ในชุดฝึกฝน	20 คำ	163	1580	200	54.53
	40 คำ	889	2855	1760	34.46
	70 คำ	871	2992	1760	30.97

ตารางที่ 4.4 รายละเอียดผลการทดสอบค่า FOM ของแบบจำลองเสียงระดับย่อยของพยางค์ประเภท Triphone

ชุดเสียงพูด	จำนวนคำสำคัญ	จำนวน Hit	จำนวน FA	จำนวนคำสำคัญตามความเป็นจริง	ค่า FOM
อยู่ในชุดฝึกฝน	20 คำ	182	1174	200	66.13
	40 คำ	1524	2528	1760	52.43
	70 คำ	1467	2712	1760	49.34
ไม่อยู่ในชุดฝึกฝน	20 คำ	171	1309	200	61.30
	40 คำ	1151	2273	1760	49.45
	70 คำ	1069	2535	1760	46.31

ตารางที่ 4.5 รายละเอียดผลการทดสอบค่า FOM ของแบบจำลองเสียงระดับย่อยของพยางค์ประเภท Onset - Rhyme

ชุดเสียงพูด	จำนวน คำสำคัญ	จำนวน Hit	จำนวน FA	จำนวนคำสำคัญ ตามความเป็นจริง	ค่า FOM
อยู่ในชุดฝึกฝน	20 คำ	189	1034	200	66.62
	40 คำ	1602	2458	1760	53.18
	70 คำ	1517	2606	1760	49.90
ไม่อยู่ในชุดฝึกฝน	20 คำ	181	1208	200	62.55
	40 คำ	1423	2398	1760	49.04
	70 คำ	1341	2663	1760	45.40

ตารางที่ 4.6 และ 4.7 แสดงการเปรียบเทียบกันของค่า FOM โดยตารางที่ 4.6 เป็นการเปรียบเทียบค่า FOM ที่ได้มาจากการใช้ชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน และตารางที่ 4.7 เป็นการเปรียบเทียบค่า FOM ที่ได้มาจากการใช้ชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน

ตารางที่ 4.6 การเปรียบเทียบค่า FOM ของชุดเสียงของผู้พูดที่อยู่ในชุดฝึกฝน

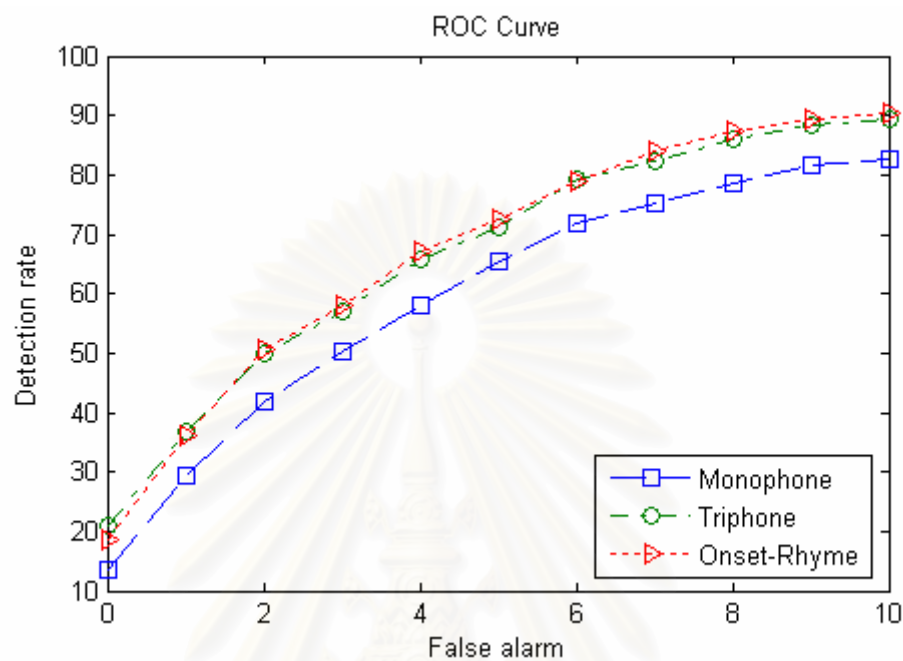
ชุดคำสำคัญ	Monophone	Triphone	Onset – Rhyme
20 คำ	58.98	66.13	66.62
40 คำ	43.86	52.43	53.18
70 คำ	38.01	49.34	49.90

ตารางที่ 4.7 การเปรียบเทียบค่า FOM ของชุดเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน

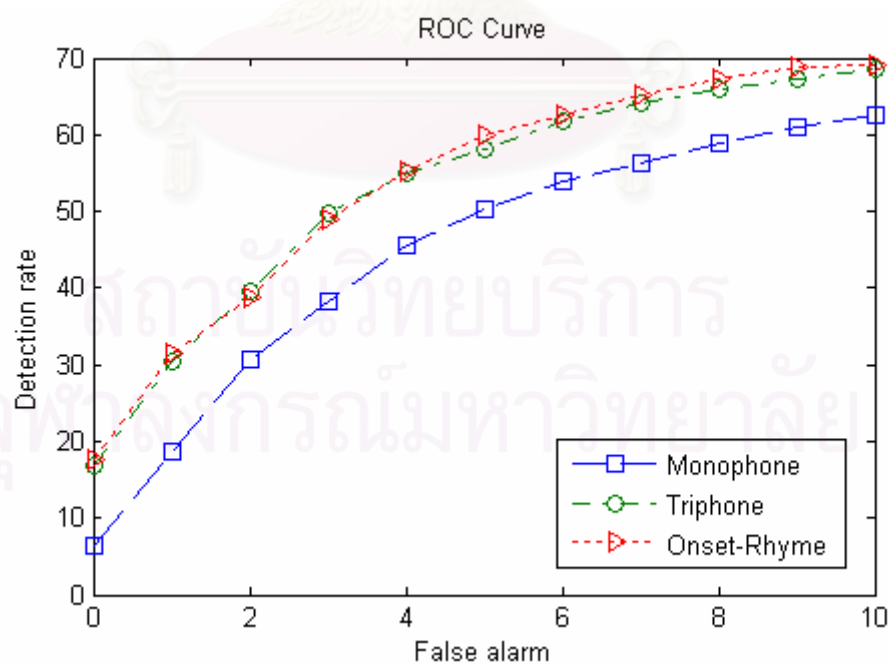
ชุดคำสำคัญ	Monophone	Triphone	Onset – Rhyme
20 คำ	54.53	61.30	62.55
40 คำ	34.46	49.45	49.04
70 คำ	30.97	46.31	45.40

การเปรียบเทียบด้วยกราฟ ROC Curve ดังรูปที่ 4.2 – 4.4 ได้แสดงถึงค่าอัตราการตรวจพบคำสำคัญต่อจำนวน False alarm ของขั้นตอนการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone และขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone และ Onset – Rhyme ในจำนวนของชุดคำสำคัญที่ต่างกันออกไปและใช้เสียงพูดในการทดสอบเป็นเสียงของผู้พูดที่อยู่ในชุดฝึกฝน โดยรูปที่ 4.2 เป็นกราฟที่ได้จากการใช้ชุดคำสำคัญจำนวน 20

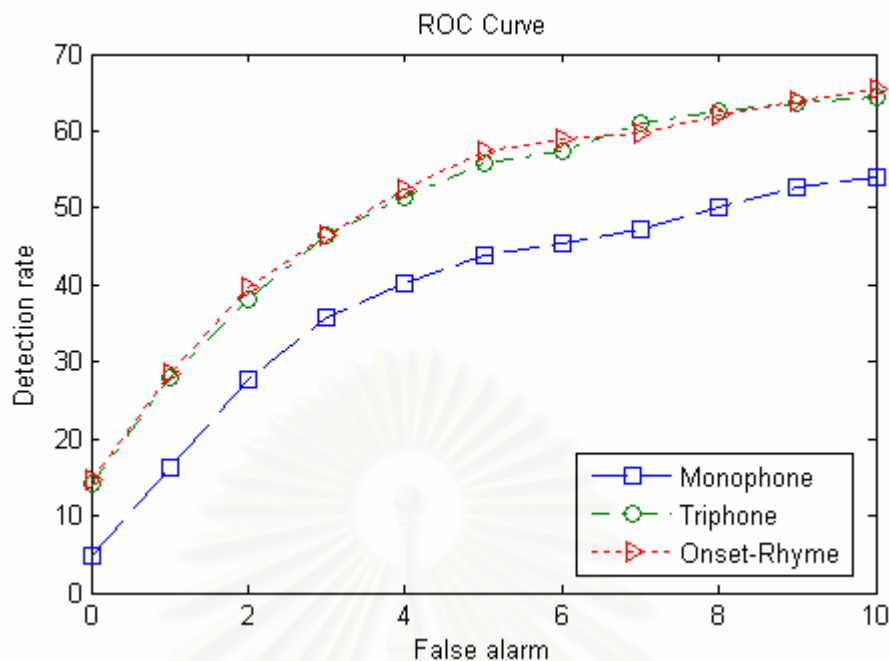
คำ สำหรับรูปที่ 4.3 เป็นกราฟที่ได้จากการใช้ชุดคำสำคัญจำนวน 40 คำ และรูปที่ 4.4 เป็นรูปที่ได้จากการใช้ชุดคำสำคัญจำนวน 70 คำ



รูปที่ 4.2 ROC Curve ของชุดคำสำคัญจำนวน 20 คำของเสียงของผู้พูดที่อยู่ในชุดฝึกฝน

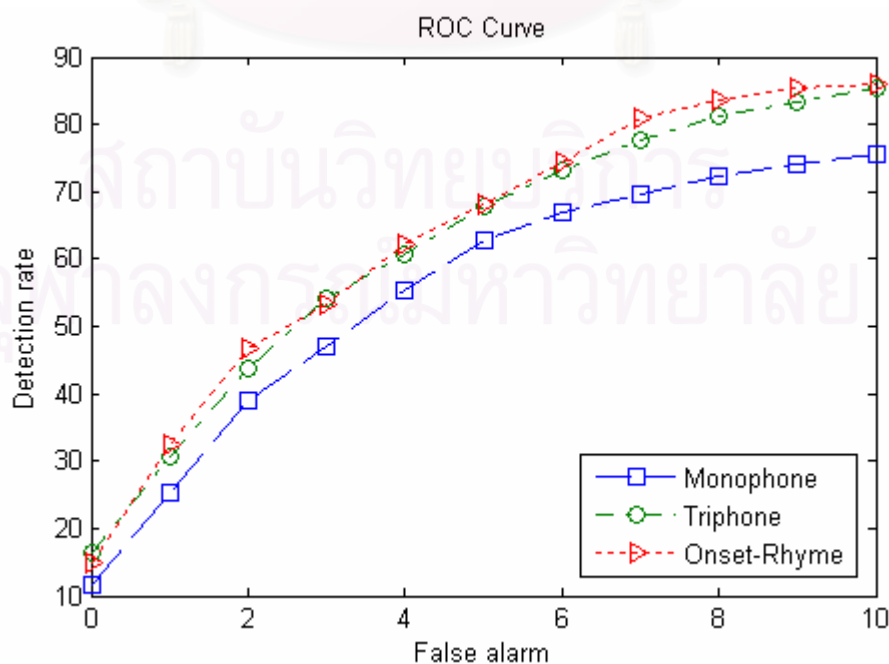


รูปที่ 4.3 ROC Curve ของชุดคำสำคัญจำนวน 40 คำของเสียงของผู้พูดที่อยู่ในชุดฝึกฝน

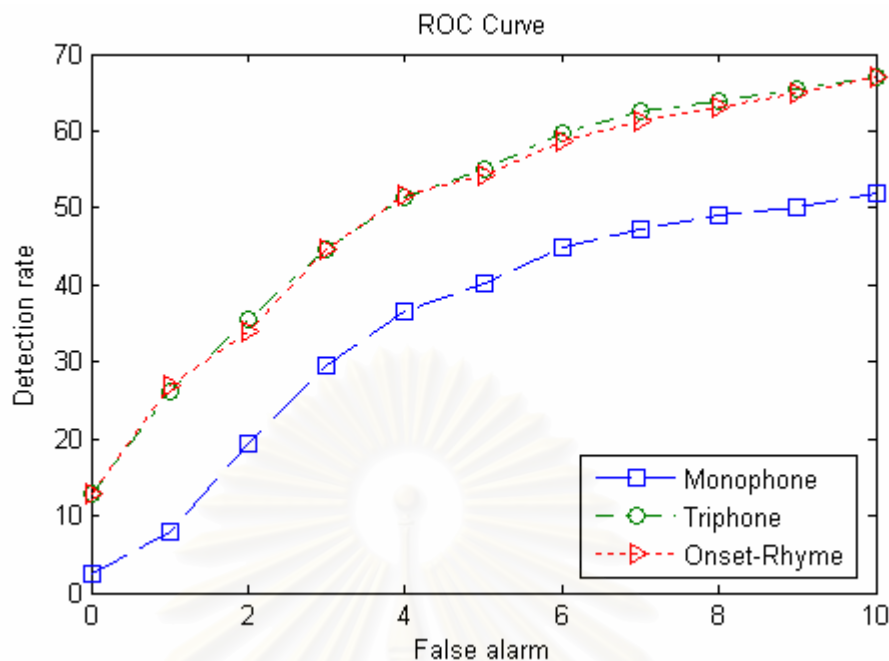


รูปที่ 4. 4 ROC Curve ของชุดคำสำคัญจำนวน 70 คำของเสียงของผู้พูดที่อยู่ในชุดฝึกฝน

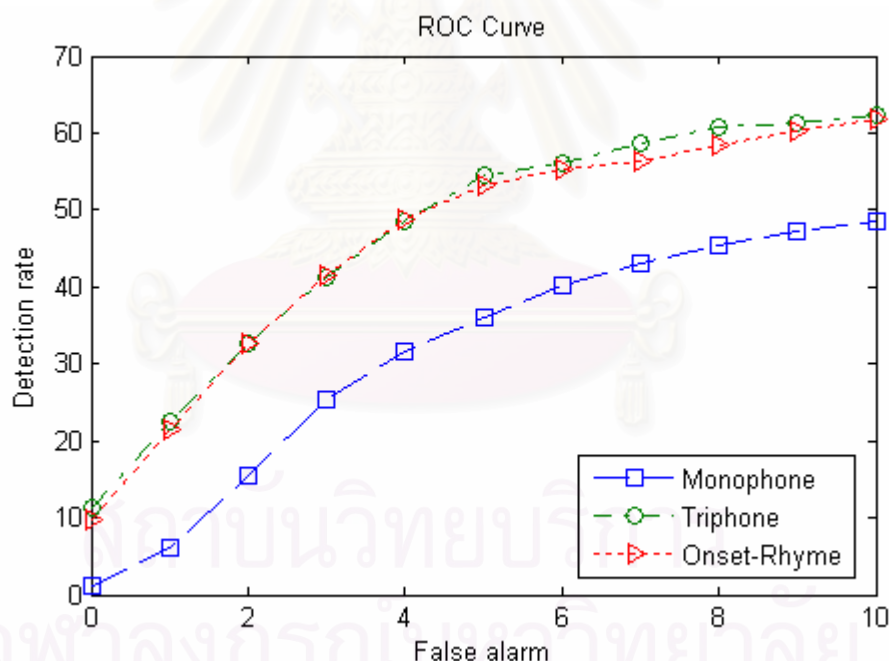
และสำหรับรูปที่ 4.5 – 4.7 ได้แสดงถึงค่าอัตราการตรวจพบคำสำคัญต่อจำนวน False alarm ของขั้นตอนการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone และขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone และ Onset – Rhyme ในจำนวนของชุดคำสำคัญที่ต่างกันออกไปและใช้เสียงพูดในการทดสอบเป็นเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน โดยรูปที่ 4.5 เป็นกราฟที่ได้จากการใช้ชุดคำสำคัญจำนวน 20 คำ สำหรับรูปที่ 4.6 เป็นกราฟที่ได้จากการใช้ชุดคำสำคัญจำนวน 40 คำ และรูปที่ 4.7 เป็นรูปที่ได้จากการใช้ชุดคำสำคัญจำนวน 70 คำ



รูปที่ 4. 5 ROC Curve ของชุดคำสำคัญจำนวน 20 คำของเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน



รูปที่ 4.6 ROC Curve ของชุดคำสำคัญจำนวน 40 คำของเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน



รูปที่ 4.7 ROC Curve ของชุดคำสำคัญจำนวน 70 คำของเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน

เมื่อพิจารณาความผิดพลาดในการชี้เฉพาะคำสำคัญภาษาไทยสามารถแยกประเภทของความผิดพลาดได้เป็น 3 ประเภทคือ ความผิดพลาดที่เกิดจากคำพูดนั้นไม่ได้เป็นคำสำคัญแต่ถูกตัดสนใจให้เป็นคำสำคัญ (Insert) คำพูดนั้นเป็นคำสำคัญแต่ถูกตัดสนใจให้เป็นคำสำคัญคำอื่น (Substitute) และคำพูดนั้นเป็นคำสำคัญแต่ถูกตัดสนใจไม่ให้เป็นคำสำคัญ (Delete) ดังตารางที่ 4.8 – 4.10 ได้แสดงถึงจำนวนคำสำคัญตามความเป็นจริง (Actual) จำนวน Hit จำนวน Insert

จำนวน Substitute และจำนวน Delete โดยค่าต่าง ๆ ในตารางที่ 4.8 แสดงถึงรายละเอียดของความผิดพลาดที่เกิดขึ้นจากการขั้นตอนการรู้จำโดยใช้แบบจำลองประเภท Monophone สำหรับค่าต่าง ๆ ในตารางที่ 4.9 แสดงถึงรายละเอียดของความผิดพลาดที่เกิดขึ้นจากการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone และค่าต่าง ๆ ในตารางที่ 4.10 แสดงถึงรายละเอียดของความผิดพลาดที่เกิดขึ้นจากการขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Onset - Rhyme และเมื่อพิจารณาความผิดพลาดที่เกิดขึ้นโดยวิเคราะห์จากจำนวนพยางค์ของคำสำคัญดังตารางที่ 4.11 - 4.13 ได้แสดงความผิดพลาดที่ปรากฏเป็นจำนวนพยางค์โดยตารางที่ 4.11 แสดงรายละเอียดของความผิดพลาดที่ปรากฏของแบบจำลอง Monophone สำหรับตารางที่ 4.12 แสดงรายละเอียดของความผิดพลาดที่ปรากฏของแบบจำลอง Triphone และตารางที่ 4.13 แสดงรายละเอียดของความผิดพลาดที่ปรากฏของแบบจำลอง Onset - Rhyme

ตารางที่ 4.8 รายละเอียดของความผิดพลาดของแบบจำลอง Monophone

ชุดเสียงพูด	จำนวนคำสำคัญ	Actual	Hit	Insert	Substitute	Delete
อยู่ในชุดฝึกฝน	20 คำ	200	168	1402	9	23
	40 คำ	1760	1279	2224	219	262
	70 คำ	1760	1180	2357	236	344
ไม่อยู่ในชุดฝึกฝน	20 คำ	200	163	1569	11	26
	40 คำ	1760	889	2681	174	697
	70 คำ	1760	871	2828	164	725

ตารางที่ 4.9 รายละเอียดของความผิดพลาดของแบบจำลอง Triphone

ชุดเสียงพูด	จำนวนคำสำคัญ	Actual	Hit	Insert	Substitute	Delete
อยู่ในชุดฝึกฝน	20 คำ	200	182	1166	8	10
	40 คำ	1760	1524	2506	22	214
	70 คำ	1760	1467	2670	42	251
ไม่อยู่ในชุดฝึกฝน	20 คำ	200	171	1298	11	18
	40 คำ	1760	1151	2101	172	437
	70 คำ	1760	1069	2356	179	512

ตารางที่ 4. 10 รายละเอียดของความผิดพลาดของแบบจำลอง Onset - Rhyme

ชุดเสียงพูด	จำนวนคำสำคัญ	Actual	Hit	Insert	Substitute	Delete
อยู่ในชุดฝึกฝน	20 คำ	200	189	1026	8	3
	40 คำ	1760	1602	2413	45	113
	70 คำ	1760	1517	2559	47	196
ไม่อยู่ในชุดฝึกฝน	20 คำ	200	181	1194	14	5
	40 คำ	1760	1423	2264	134	203
	70 คำ	1760	1341	2529	134	285

ตารางที่ 4. 11 รายละเอียดความผิดพลาดที่ปรากฏของแบบจำลอง Monophone

ชุดเสียงพูด	จำนวนคำสำคัญ	จำนวนคำผิดพลาดที่ปรากฏ		
		พยางค์เดียว	สองพยางค์	สามพยางค์
อยู่ในชุดฝึกฝน	20 คำ	1088	308	38
	40 คำ	2175	481	49
	70 คำ	2388	498	51
ไม่อยู่ในชุดฝึกฝน	20 คำ	1253	312	41
	40 คำ	2850	647	55
	70 คำ	2969	689	59

ตารางที่ 4. 12 รายละเอียดความผิดพลาดที่ปรากฏของแบบจำลอง Triphone

ชุดเสียงพูด	จำนวนคำสำคัญ	จำนวนคำผิดพลาดที่ปรากฏ		
		พยางค์เดียว	สองพยางค์	สามพยางค์
อยู่ในชุดฝึกฝน	20 คำ	948	213	23
	40 คำ	2146	554	42
	70 คำ	2416	513	34
ไม่อยู่ในชุดฝึกฝน	20 คำ	1082	219	26
	40 คำ	2148	523	39
	70 คำ	2430	564	53

ตารางที่ 4. 13 รายละเอียดความผิดพลาดที่ปรากฏของแบบจำลอง Onset - Rhyme

ชุดเสียงพูด	จำนวนคำสำคัญ	จำนวนคำผิดพลาดที่ปรากฏ		
		พยางค์เดียว	สองพยางค์	สามพยางค์
อยู่ในชุดฝึกฝน	20 คำ	824	194	19
	40 คำ	2030	499	42
	70 คำ	2230	522	50
ไม่อยู่ในชุดฝึกฝน	20 คำ	971	217	25
	40 คำ	2085	473	43
	70 คำ	2379	512	57

วิเคราะห์ผลการชี้เฉพาะคำสำคัญภาษาไทย

จากรายละเอียดผลการทดสอบระบบการชี้เฉพาะคำสำคัญภาษาไทยซึ่งได้ใช้การตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone และการตรวจสอบคำสำคัญโดยใช้แบบจำลองเสียงระดับย่อยของพยางค์ประเภท Triphone และ Onset - Rhyme ซึ่งทดสอบตามจำนวนคำสำคัญและแบ่งชุดเสียงทดสอบเป็นเสียงของผู้พูดที่อยู่ในชุดฝึกฝนและเสียงของผู้พูดที่ไม่อยู่ในชุดฝึกฝน ในหัวข้อนี้ได้วิเคราะห์ผลการชี้เฉพาะคำสำคัญภาษาไทยของชุดเสียงพูดทั้งหมดโดยรวมดังนี้

1. จากรูปที่ 4.1 แสดงให้เห็นว่าระบบการชี้เฉพาะคำสำคัญภาษาไทยในงานวิจัยนี้มีความจำเป็นต้องวิเคราะห์ค่าคะแนนของคำที่เกิดขึ้น เนื่องจากเส้นกราฟที่ไม่เป็นคำสำคัญทับกับพื้นที่ของเส้นกราฟที่เป็นคำสำคัญเป็นจำนวนมาก

2. จากตารางที่ 4.1 ถึง 4.7 และรูปที่ 4.2 ถึง 4.7 แสดงให้เห็นว่าประสิทธิภาพของระบบชี้เฉพาะคำสำคัญภาษาไทยในส่วนของขั้นตอนการตรวจสอบโดยใช้แบบจำลองประเภท Onset - Rhyme มีค่ามากกว่าแบบจำลองประเภท Triphone เนื่องจากคำหนึ่งพยางค์เมื่อนำมาสร้างแบบจำลองประเภท Onset - Rhyme จะประกอบด้วย 2 แบบจำลอง คือ Onset และ Rhyme และคำหนึ่งพยางค์เมื่อนำมาสร้างแบบจำลองประเภท Triphone จะประกอบด้วย 2 - 3 แบบจำลอง คือแบบจำลองของพยัญชนะต้น สระ และบางคำมีแบบจำลองของตัวสะกด ซึ่งมีผลต่อค่าความน่าจะเป็นของการคำนวณในขั้นตอนหาค่าความน่าจะเป็นของแบบจำลองฮิดเดน มาร์คอฟ และแบบจำลองประเภท Triphone เป็นแบบจำลองประเภท Context - dependent ซึ่งมีการวิเคราะห์เสียงที่อยู่ก่อนหน้าแบบจำลองและเสียงที่อยู่ถัดจากแบบจำลอง เมื่อคำสำคัญนั้นมีเสียงที่อยู่ก่อนหน้าคำสำคัญหรือมีเสียงที่อยู่ต่อจากคำสำคัญไม่ใช่เสียงเงียบ จึงส่งผลต่อค่าความ

น่าจะเป็นของแบบจำลอง เมื่อพิจารณาชุดเสียงพูดที่นำมาใช้ในการทดสอบจะเห็นว่า เสียงทดสอบเป็นเสียงของผู้พูดที่อยู่ในชุดฝึกฝนจะให้ผลการทดลองดีกว่าเสียงทดสอบที่เป็นเสียงของผู้พูดที่ไม่ได้อยู่ในชุดฝึกฝน เพราะว่าในการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟได้ปรับค่าพารามิเตอร์ให้มีความคล้ายคลึงกับเสียงต้นแบบที่ใช้ในการฝึกฝน จึงให้ค่าความน่าจะเป็นมีค่าสูงเมื่อใช้เสียงทดสอบที่เป็นเสียงของผู้พูดที่อยู่ในชุดฝึกฝน และเมื่อพิจารณาในส่วนของจำนวนคำสำคัญจะเห็นได้ว่า จำนวนชุดคำสำคัญที่มีจำนวนน้อยจะให้อัตราความแม่นยำเฉลี่ยดีกว่า เนื่องจากในแบบจำลองภาษาที่มีจำนวนคำสำคัญน้อย ทำให้ความน่าจะเป็นในการเลือกคำสำคัญที่เหมาะสมในช่วงเสียงทดสอบที่ไม่ใช่คำสำคัญมีค่าน้อย จึงทำให้ประสิทธิภาพของระบบขึ้นเฉพาะมีค่าสูง

3. จากตารางที่ 4.8 ถึง 4.13 แสดงให้เห็นว่าความผิดพลาดที่เกิดขึ้นส่วนใหญ่เกิดจากคำพูดนั้นไม่ได้เป็นคำสำคัญแต่ถูกตัดสินใจให้เป็นคำสำคัญ (Insert) เนื่องจากค่าคะแนนที่ได้จากเสียงพูดช่วงนี้มีค่าสูงกว่าระดับผ่าน จึงกำหนดให้เป็นคำสำคัญ และความผิดพลาดส่วนใหญ่เป็นคำพยางค์เดียว เพราะคำพยางค์เดียวมีองค์ประกอบของคำน้อย ทำให้มีความเป็นเอกลักษณ์ในคำนั้นน้อยเช่นกัน จึงมีผลในการรู้จำเสียงทดสอบทั้งในช่วงที่เป็นคำสำคัญและในช่วงที่ไม่เป็นคำสำคัญ ซึ่งทำให้ประสิทธิภาพของระบบไม่ดีเท่าที่ควร แสดงให้เห็นว่ากระบวนการหาค่าคะแนนที่ใช้ตัดสินใจคำสำคัญยังไม่มีประสิทธิภาพเพียงพอ และสำหรับชุดคำสำคัญ 20 คำเป็นชุดในการทดสอบที่ให้ประสิทธิภาพของระบบดีมาก เนื่องจากคำสำคัญในชุดนี้ส่วนใหญ่เป็นคำสำคัญที่มีจำนวนพยางค์หลายพยางค์ ทำให้มีความผิดพลาดที่เกิดจากคำพยางค์เดียวน้อย จึงส่งผลให้ค่าความแม่นยำและค่า FOM ของระบบมีค่าสูง

สรุปผลการวิจัยและข้อเสนอแนะ

สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาเกี่ยวกับระบบการชี้เฉพาะคำสำคัญภาษาไทย บนพื้นฐานของการตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์ โดยใช้แบบจำลองฮิดเดนมาร์คอฟเป็นกรรมวิธีสร้างชุดต้นแบบของเสียงระดับย่อยของพยางค์ ระบบประกอบด้วยขั้นตอนการตรวจหาคำสำคัญและขั้นตอนการตรวจสอบคำสำคัญ ซึ่งทำการทดสอบหาประสิทธิภาพของระบบจากชุดคำสำคัญ 70 คำ แบ่งเป็น 3 ชุดตามจำนวนคือ 20 คำ 40 คำ และ 70 คำ โดยใช้บทพูดซึ่งมีลักษณะเป็นประโยคต่อเนื่อง ดังแสดงในภาคผนวก ก

ในขั้นตอนการสกัดลักษณะสำคัญของเสียงจะอาศัยสัมประสิทธิ์เซปสตรอลบนความถี่เมลมีวงจรรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลจำนวน 20 วงจรเพื่อสร้างสัมประสิทธิ์เซปสตรอลบนความถี่เมลที่ 16 ลำดับ

ในขั้นตอนการสร้างชุดต้นแบบของเสียงระดับย่อยของพยางค์ประเภทต่าง ๆ โดยใช้แบบจำลองฮิดเดนมาร์คอฟมีการกระจายแบบเกาส์และมีจำนวนมิกเจอร์ 2 มิกเจอร์ เป็นการจับคู่ชุดพารามิเตอร์ของแบบจำลองสำหรับแต่ละเสียงระดับย่อยของพยางค์ เพื่อให้ในการรู้จำด้วยขั้นตอนวิธีการ Viterbi ซึ่งเป็นแบบจำลองประเภท Monophone, Triphone, Onset และ Rhyme และอาศัยหลักการของแบบจำลองภาษาในการรู้จำเสียงระดับย่อยของพยางค์เชื่อมต่อกันเป็นคำสำคัญ ผลลัพธ์ที่ได้จะเป็นค่าความน่าจะเป็นของคำสำคัญจากแบบจำลองภาษาที่ให้ค่าความน่าจะเป็นสูงสุด

ระบบการชี้เฉพาะคำสำคัญภาษาไทยมีขั้นตอนจากการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone ผลที่ได้จากขั้นตอนนี้แสดงเป็นคำสำคัญ ช่วงเวลาและค่าคะแนน ซึ่งมีลักษณะเป็นคำที่เป็นคำสำคัญต่าง ๆ เรียงต่อ ๆ กันตามช่วงเวลา จากนั้นนำค่าแต่ละช่วงเวลาผ่านขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Onset – Rhyme ผลที่ได้จากขั้นตอนนี้แสดงเป็นคำสำคัญ ช่วงเวลาและค่าคะแนน นำค่าคะแนนที่ได้หาค่าคะแนนใหม่จากสมการที่ 3-3 เพื่อจำแนกว่าเป็นคำสำคัญหรือไม่เป็นคำสำคัญโดยวัดจากค่าจุดเริ่มเปลี่ยน และสำหรับงานวิจัยนี้ได้เพิ่มการทดสอบระบบในขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone เพื่อเปรียบเทียบผลการทดลองเมื่อใช้แบบจำลองประเภท Onset – Rhyme

ผลการทดสอบอัตราความแม่นยำเฉลี่ยของขั้นตอนการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone มีค่าร้อยละ 69.23 ผลการทดสอบอัตราความแม่นยำเฉลี่ยของ

ขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone มีค่าร้อยละ 81.65 และผลการทดสอบอัตราความแม่นยำเฉลี่ยของขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Onset – Rhyme มีค่าร้อยละ 82.37 สำหรับค่า FOM เฉลี่ยของขั้นตอนการตรวจพบคำสำคัญโดยใช้แบบจำลองประเภท Monophone มีค่าร้อยละ 43.47 ค่า FOM เฉลี่ยของขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Triphone มีค่าร้อยละ 54.16 และค่า FOM เฉลี่ยของขั้นตอนการตรวจสอบคำสำคัญโดยใช้แบบจำลองประเภท Onset – Rhyme มีค่าร้อยละ 54.45 ซึ่งจะเห็นได้ว่าเมื่อใช้แบบจำลองของเสียงระดับย่อยของพยางค์ประเภท Onset – Rhyme ตรวจสอบคำสำคัญใช้ผลการวิจัยที่ดีที่สุด

ความผิดพลาดส่วนใหญ่เป็นการรู้จำคำสำคัญให้ผลเป็นคำพยางค์เดียว เนื่องจากมีองค์ประกอบในคำน้อยและมีช่วงเวลาของเสียงสั้น ๆ จึงมีความใกล้เคียงกับคำทั่วไปมาก มีผลต่อค่าความน่าจะเป็นของระบบการรู้จำ และทำให้เกิดความผิดพลาดจากการวิเคราะห์คำคะแนน

ผลการเปรียบเทียบเมื่อเพิ่มจำนวนคำสำคัญคือ 20 คำ 40 คำ และ 70 คำ ส่งผลให้ประสิทธิภาพของระบบลดลงตามลำดับ สำหรับประสิทธิภาพของระบบที่จำนวนคำสำคัญ 20 คำ มีค่าสูงมากเนื่องจากคำสำคัญในชุดนี้ส่วนมากเป็นคำหลายพยางค์ ทำให้เกิดความผิดพลาดน้อย

ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

ข้อเสนอแนะสำหรับแก้ไขข้อผิดพลาดของการรู้จำเฉพาะคำสำคัญภาษาไทยสำหรับงานวิจัยในอนาคตมีดังนี้

1. เพื่อเพิ่มความแม่นยำในการรู้จำ ควรทำการวิเคราะห์ในส่วนของค่าความน่าจะเป็นของค่าสังเกตที่มีการกระจายแบบเกาส์มิกเจอร์ เพื่อปรับจำนวนมิกเจอร์ให้มีความเหมาะสมกับค่าสัมประสิทธิ์ที่ได้จากการสกัดลักษณะสำคัญของเสียง
2. แบบจำลองฮิดเดนมาร์คอฟจำเป็นต้องอาศัยข้อมูลเสียงพูดเพื่อการฝึกฝนเป็นจำนวนมาก เพื่อเพิ่มความแม่นยำในการรู้จำจึงควรเพิ่มจำนวนเสียงพูดให้มากยิ่งขึ้น
3. ขั้นตอนการวิเคราะห์เพื่อจำแนกคำสำคัญในขั้นตอนการหาค่าคะแนน ควรใช้กระบวนการที่มีการวิเคราะห์ซับซ้อนมากกว่านี้ เช่น แบ่งการวิเคราะห์สำหรับคำสำคัญพยางค์เดียว หรือใช้กรรมวิธี N – Best เข้ามาช่วยในการวิเคราะห์ เป็นต้น
4. สำหรับการนำไปประยุกต์ใช้จริงควรมีการออกแบบระบบให้มีคำสำคัญที่ทำให้ระบบมีประสิทธิภาพ

รายการอ้างอิง

ภาษาไทย

- เจษฎา กานต์ประชา. การรู้จำเสียงพูดภาษาไทยอย่างคงทนโดยใช้สัมประสิทธิ์เมลเฟรีเควินที่ เซปสตรอลของค่าอัตราสัมพันธ์ของเสียงพูดที่มีสัญญาณรบกวน. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2545.
- ณัฐกร ทับทอง. การรู้จำคำพูดภาษาไทย โดยใช้ลักษณะแบ่งความต่างของหน่วยเสียง. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.
- ทวิ ประทุมทาน. การตรวจรู้เสียงพูดภาษาไทย. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2530.
- พงศ์ไท ทาสระคู. ระบบบ่งชี้ผู้พูดแบบระบบเปิดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบหลายชุดรหัส. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2542
- ไพศาล ธรรมโพธิทอง. ระบบการรับรู้เสียงพูดแบบต่างบุคคล. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2533.
- ระพีพัฒน์ เพ็ญศิริ. การรู้จำเสียงพูดสระภาษาไทยโดด ๆ ไม่ขึ้นกับผู้พูดโดยการวัดสเปกตรัมดิสแตนท์และใช้ไดนามิกไทม์วาร์ปิง. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.
- วิศรุต อาชูปุตร. ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2539.
- วิศรุต อาชูปุตร. การศึกษาหน่วยเริ่มของพยางค์เชิงกลศาสตร์ : พื้นฐานสำหรับการรู้จำเสียงพูดต่อเนื่องภาษาไทย. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2545.
- ศวิต กาสุริยะ. ระบบการรู้จำผู้พูด : การบ่งชี้ผู้พูดแบบขึ้นกับบทพูด. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2542.
- เสาวลักษณ์ อารีพงศา. การรู้จำเสียงพูดตัวเลขเป็นภาษาไทย แบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดนมาร์คอฟ โมเดลและเวกเตอร์ควอนไทซ์เซชัน. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.

เอกฤทธิ มณีน้อย. การรู้จำหน่วยเสียงสระภาษาไทยโดยใช้โครงข่ายประสาทเทียม.
วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, สาขาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย,
2541.

ภาษาอังกฤษ

Ahkuputa, V., Jitapunkul, S., Maneenoi, E., Kasuriya, S. and Amornkul, P. Comparison of Different Techniques On Thai Speech Recognition, Proceedings of Circuit and System on IEEE Asia-Pacific Conference (November 1998): 177-180.

Bo – Ren Bai, Chiu – Yu Tseng and Lin – Shan lee. A multi – phase approach for fast spotting of large vocabulary Chinese keywords from mandarin speech using prosodic information, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2 (Apr 1997): 903-906.

Chung – Hsien Wu and Yeou – Jiunn Chen. Multi – keyword Spotting of Telephone Speech using a Fuzzy Search Algorithm and Keyword – driven Two – level CBSM, Speech Communication 33 (2001): 197 – 212.

Dharanipragada, S. and Roukos, S. A Fast Vocabulary Independent Algorithm for Spotting Words in Speech, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 1 (May 1998): 233 – 236.

Foote, J.T., Young, S.J., Jones, G.J.F. and Jones, K.S. Unconstrained Keyword Spotting using Phone Lattices with Application to Spoken Document Retrieval, Computer Speech and Language 11 (1997): 207 – 224.

Foote, J.T. An Overview of Audio Information Retrieval, ACM Multimedia Systems 7 (1999): 2 – 10.

Furui, S. Digital Speech Processing, Synthesis, and Recognition Tokyo: Tokai University Press 1985.

James, D.A. and Young, S.J. A Fast Lattice – Based Approach To Vocabulary Independent Wordspotting, ICASSP (1994): 377 – 380.

Kawabata, T. and Kohda, M. Word spotting method based on top – down phoneme verification, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 12 (Apr 1987): 1438 – 1441.

- Knill, K.M. and Young, S.J. Fast Implementation Methods for Viterbi – Based Word – Spotting, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 1 (May 1996): 522-525.
- Knill, K.M. and Young, S.J. Low – cost Implementation of Open Set Keyword Spotting, Computer Speech and Language 13 (1999): 243 – 266.
- Maneenoi, E. An Acoustic Study of Subsyllable Rhymes : A Basis for Thai Continuous Speech Recognition System. Dissertation Proposal Department of Electrical Engineering Chulalongkorn University, 2002.
- Medress, M.F., Diller, T.C., Kloker, D.R., Lutton, L.L., Oredson, H.N. and Skinner, T.E. An Automatic Word Spotting System for Conversational Speech, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 3 (Apr 1978): 712 – 717.
- Mitchel, W. LVCSR Log-Likelihood Ratio Scoring for Keyword Spotting, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 1 (1995): 297-300.
- Nitta, T., Tanaka, S., Masai, Y. and Matsu'ura, H. Word – Spotting Based on Inter – Word and Intra – Word Diphone Models, Fourth International Conference on Volume 2 (Oct 1996): 1093 -1096.
- Rabiner, L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE Vol. 77, 2 (February 1989): 257-286.
- Rohlicek, J.R., Jeanrenaud, P., Ng, K., Gish, H., Musicus, B. and Siu, M. Phonetic Training and Language Modeling for Word Spotting IEEE Int. Conf on Acoustics, Speech and Signal Processing, Vol 2 (Apr 1993): 459 – 462.
- Tolba, H. and O'Shaughnessy, D. Automatic speech recognition based on cepstral coefficient and mel-based discrete energy operator, IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2 (May 1998): 973-976.
- Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V. and Woodland, P. The HTK Book (for HTK Version 3.0), 2000.



ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

รายการชุดคำสำคัญภาษาไทย

ในภาคผนวกบทนี้จะกล่าวถึงรายละเอียดบทพูดเพื่อใช้ทดสอบระบบชี้เฉพาะคำสำคัญภาษาไทยในงานวิจัยนี้ ซึ่งเป็นบทความที่เกี่ยวข้องกับการท่องเที่ยวทางทะเลในประเทศไทย และรายละเอียดของชุดคำสำคัญภาษาไทยที่เกี่ยวข้องกับสถานที่ท่องเที่ยวในเขตพื้นที่ทางทะเล ชื่อจังหวัดในประเทศไทยและกิจกรรมทางทะเลเป็นจำนวน 70 คำ สำหรับการบันทึกเสียงของผู้พูดเพื่อนำไปใช้ในงานวิจัยนี้ เสียงที่ถูกบันทึกจะเป็นการพูดบทความการท่องเที่ยวทางทะเลในประเทศไทย และคำสำคัญ 70 คำ

บทพูดเพื่อใช้ทดสอบระบบชี้เฉพาะคำสำคัญภาษาไทย

แหล่งท่องเที่ยวของไทย มีมากมายกว้างใหญ่ไพศาล มีความงดงามมหัศจรรย์ทั้งธรรมชาติป่าเขาลำเนาไพร กับธรรมชาติของท้องทะเลไทยที่มีชายฝั่งทะเลยาวเหยียดจากภาคตะวันออกเฉียงเหนือสุดแดนใต้ที่นราธิวาสกับชายฝั่งทะเลระนองฝั่งตะวันตกด้านทะเลอันดามัน ลัดเลาะลงมาถึงสตูลสุดแดนไทย บรรยากาศของท้องทะเลไทยที่น้ำทะเลเขียวใสสะอาดหาดทรายขาวเป็นแนวยาวลัดเลาะชายฝั่งตามเกาะแก่ง และแหล่งดำน้ำใต้ท้องทะเลสี่ครามที่อุดมด้วยทรัพยากรทางทะเลที่งดงามติดอันดับโลก ทั้งแนวปะการัง สัตว์น้ำทะเล ปลาหลากสีสัน ความมหัศจรรย์งดงามตระการตาที่น่าชมน่าสัมผัส

สถานที่ท่องเที่ยวตามชายฝั่ง เกาะแก่งรวมทั้งแหล่งดำน้ำของทะเลไทย มีอยู่ตามจังหวัดที่ติดทะเล 23 จังหวัด 2 ฝั่งทะเล คือ ฝั่งอ่าวไทยตะวันออกเฉียงเหนือ ภาคใต้ฝั่งอ่าวไทยและภาคใต้ฝั่งอันดามัน สำหรับกรุงเทพฯ ฉะเชิงเทรา และสมุทรสาคร ไม่มีแหล่งท่องเที่ยวติดชายทะเลที่สำคัญ

ชายฝั่งทะเลไทย มีอุทยานแห่งชาติทางทะเล 18 แห่ง เกาะแก่งกว่า 503 เกาะ ชายหาดที่ขาวสะอาดงดงามนับไม่ถ้วนกับโลกใต้ผืนน้ำกว้างไกลของท้องทะเลไทย ที่มีความสวยงามอัศจรรย์ไม่แพ้ทะเลใดในโลก พร้อมที่พักที่มีสิ่งอำนวยความสะดวกพร้อมมูลทุกจังหวัดกับเส้นทางเดินทางสู่แหล่งท่องเที่ยวสะดวกสบายและมิตรไมตรีที่ได้รับจากผู้คนตามท้องถิ่น

ท่องเที่ยวทะเลไทย จุดหมายของผู้คนที่นิยมธรรมชาติของสายลม แสงแดด ทะเลใสหาดทรายขาว ยังคงหลังไหลไปตามหมู่เกาะกลางทะเลในทุกวันหยุด ทั้งหมู่เกาะฝั่งทะเลตราดที่มีน้ำทะเล ชายหาด สวยไม่แพ้ฝั่งอันดามันกับชายหาดที่นิยมกันมาทุกยุคสมัยทั้งหาดบางแสน

พัทธา หัวหิน หรือเกาะสมุย เกาะพะงัน ท้องทะเลใต้กับเกาะชื่อดังอีกหลายเกาะ และ
สำหรับผู้พิสมัยการดำน้ำ อ่าวไทยก็มีแหล่งดำน้ำไม่ยิ่งหย่อนกว่าใครบริเวณเกาะเต่า

ทางด้านฝั่งทะเลอันดามันสี่ส้นความงามของท้องทะเลที่ธรรมชาติบรรจงสร้างเหนือกว่า
 คำบรรยายใดใดกับโลกใต้ท้องทะเลมหัศจรรย์แห่งธรรมชาติ สวรรค์ของนักท่องเที่ยวที่จะได้สัมผัส
หาดทรายละเอียดขาวนวล เกาะแก่งมากมาย เริ่มจากกระนงลัดเลาะชายทะเลสู่พังงา ดินแดน
แห่งเกาะที่สวยงามติดอันดับโลกที่หมู่เกาะสิมิลัน เกาะสุรินทร์ แล้วเลยมาถึงภูเก็ตศูนย์กลาง
ท่องเที่ยวฝั่งอันดามัน กระบี่กับหมู่เกาะพีพีอันเลื่องชื่อตามด้วยตรัง สตูล สุดเขตแดนไทย

การท่องเที่ยวทางทะเลในปัจจุบันมีปริมาณนักท่องเที่ยวเพิ่มมากขึ้น สิ่งเหล่านี้ย่อมเป็น
 ผลดีต่อการท่องเที่ยวในประเทศ แต่ก็มีผลกระทบต่อธรรมชาติของท้องทะเลแน่นอน การคง
 ธรรมชาติไว้จึงอยู่ที่ความร่วมมือของทุกคน เพื่อดำรงธรรมชาติที่งดงามเหล่านั้นไว้ตราบนานแสน
 นาน ด้วยจิตสำนึกของความเป็นไทยให้ทะเลไทยคงอยู่

รายละเอียดชุดคำสำคัญภาษาไทย

ในการทดสอบการชี้เฉพาะคำสำคัญภาษาไทยของงานวิจัยนี้ ได้แบ่งคำสำคัญเป็น 3 ชุด
 ตามจำนวนคือ 20 คำ 40 คำ และ 70 คำตามลำดับ รวมทั้งรายละเอียดของเสียงระดับย่อยในคำ
 สำคัญประเภท Monophone, Triphone และ Onset – Rhyme ในคำสำคัญต่าง ๆ ดังแสดงใน
 ตารางที่ ก.1 ถึง ก.10

สถาบันวิทยบริการ
 จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก. 1 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Monophone ของชุดคำสำคัญลำดับที่ 1 - 28

ลำดับ	คำสำคัญ	Monophone
1	จังหวัด	/c/ /a/ /ng^/ /w/ /a/ /t^/
2	ตราด	/tr/ /aa/ /t^/
3	จันทบุรี	/c/ /a/ /n^/ /th/ /a/ /b/ /u/ /r/ /ii/
4	ระยอง	/r/ /a/ /j/ /@@/ /ng^/
5	ชลบุรี	/ch/ /o/ /n^/ /b/ /u/ /r/ /ii/
6	ฉะเชิงเทรา	/ch/ /a/ /ch/ /qq/ /ng^/ /s/ /a/ /w^/
7	สมุทรปราการ	/s/ /a/ /m/ /u/ /t^/ /pr/ /aa/ /k/ /aa/ /n^/
8	กรุงเทพ	/kr/ /u/ /ng^/ /th/ /ee/ /p^/
9	สมุทรสาคร	/s/ /a/ /m/ /u/ /t^/ /s/ /aa/ /kh/ /@@/ /n^/
10	สมุทรสงคราม	/s/ /a/ /m/ /u/ /t^/ /s/ /o/ /ng^/ /khr/ /aa/ /m^/
11	เพชรบุรี	/ph/ /e/ /t^/ /b/ /u/ /r/ /ii/
12	ประจวบคีรีขันธ์	/pr/ /a/ /c/ /ua/ /p^/ /kh/ /i/ /r/ /ii/ /kh/ /a/ /n^/
13	ชุมพร	/ch/ /u/ /m^/ /ph/ /@@/ /n^/
14	สุราษฎร์ธานี	/s/ /u/ /r/ /aa/ /t^/ /th/ /aa/ /n/ /ii/
15	นครศรีธรรมราช	/n/ /a/ /kh/ /@@/ /n^/ /s/ /ii/ /th/ /a/ /m^/ /m/ /a/ /r/ /aa/ /t^/
16	สงขลา	/s/ /o/ /ng^/ /kh/ /aa/
17	ปัตตานี	/p/ /a/ /t^/ /t/ /aa/ /n/ /ii/
18	นราธิวาส	/n/ /a/ /r/ /aa/ /th/ /i/ /w/ /aa/ /t^/
19	ระนอง	/r/ /a/ /n/ /@@/ /ng^/
20	พังงา	/ph/ /a/ /ng^/ /ng/ /aa/
21	ภูเก็ต	/ph/ /uu/ /k/ /e/ /t^/
22	กระบี่	/kr/ /a/ /b/ /ii/
23	ตรัง	/tr/ /a/ /ng^/
24	สตูล	/s/ /a/ /t/ /uu/ /n^/
25	ทะเล	/th/ /a/ /l/ /ee/
26	อ่าว	/z/ /aa/ /w^/
27	เกาะ	/k/ /@/
28	หาด	/h/ /aa/ /t^/

ตารางที่ ก. 2 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Monophone
ของชุดคำสำคัญลำดับที่ 29 - 55

ลำดับ	คำสำคัญ	Monophone
29	อุทยาน	/z/ /u/ /t^/ /th/ /a/ /j/ /aa/ /n^/
30	ค้ำน้ำ	/d/ /a/ /m^/ /n/ /aa/ /m^/
31	บางแสน	/b/ /aa/ /ng^/ /s/ /xx/ /n^/
32	พัทยา	/ph/ /a/ /t^/ /th/ /a/ /j/ /aa/
33	หัวหิน	/h/ /uua/ /h/ /i/ /n^/
34	สมุย	/s/ /a/ /m/ /u/ /j^/
35	พะงัน	/ph/ /a/ /ng/ /a/ /n^/
36	เต่า	/t/ /a/ /w^/
37	สิมิถัน	/s/ /i/ /m/ /i/ /l/ /a/ /n^/
38	สุรินทร์	/s/ /u/ /r/ /i/ /n^/
39	พีพี	/ph/ /ii/ /ph/ /ii/
40	ลันตา	/l/ /a/ /n^/ /t/ /aa/
41	ลำน	/l/ /aa/ /n^/
42	จอมเทียน	/c/ /@@/ /m^/ /th/ /iia/ /n^/
42	สี่ช้าง	/s/ /ii/ /ch/ /a/ /ng^/
44	เขาสามมุ๊ก	/kh/ /a/ /w^/ /s/ /aa/ /m^/ /m/ /u/ /k^/
45	ลิบง	/l/ /i/ /b/ /o/ /ng^/
46	ถ้ำมรกต	/th/ /a/ /m^/ /m/ /@@/ /r/ /a/ /k/ /o/ /t^/
47	กูด	/k/ /uu/ /t^/
48	ช้าง	/ch/ /aa/ /ng^/
49	มันนอก	/m/ /a/ /n^/ /n/ /@@/ /k^/
50	มันใน	/m/ /a/ /n^/ /n/ /a/ /j^/
51	แหลมตะลุมพุก	/l/ /xx/ /m^/ /t/ /a/ /l/ /u/ /m^/ /ph/ /u/ /k^/
52	พระราชวังไกลกังวล	/phr/ /a/ /r/ /aa/ /t^/ /ch/ /a/ /w/ /a/ /ng^/ /kl/ /a/ /j^/ /k/ /a/ /ng^/ /w/ /o/ /n^/
53	เขาสามร้อยยอด	/kh/ /a/ /w^/ /s/ /aa/ /m^/ /r/ /@@/ /j^/ /j/ /@@/ /t^/
54	มะนาว	/m/ /a/ /n/ /aa/ /w^/
55	เขาตะเกียบ	/kh/ /a/ /w^/ /t/ /a/ /k/ /iia/ /p^/

ตารางที่ ก. 3 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Monophone
ของชุดคำสำคัญลำดับที่ 56 - 70

ลำดับ	คำสำคัญ	Monophone
56	เขาไกรลาส	/kh/ /a/ /w^/ /kr/ /a/ /j^/ /l/ /aa/ /t^/
57	เจ้าสำราญ	/c/ /aa/ /w^/ /s/ /a/ /m^/ /r/ /aa/ /n^/
58	ชะอำ	/ch/ /a/ /z/ /a/ /m^/
59	ปีกเตียน	/p/ /v/ /k^/ /t/ /ia/ /n^/
60	แหลมพรหมเทพ	/l/ /xx/ /m^/ /phr/ /o/ /m^/ /th/ /ee/ /p^/
61	ป่าตอง	/p/ /aa/ /t/ /@@/ /ng^/
62	แม่รำพึง	/m/ /xx/ /r/ /a/ /m^/ /ph/ /v/ /ng^/
63	ยอด	/j/ /@@/
64	หนู	/n/ /uu/
65	แมว	/m/ /xx/ /w^/
66	สมิหลา	/s/ /a/ /m/ /i/ /l/ /aa/
67	ตึกสุลต่านนท์	/t/ /i/ /n^/ /s/ /uu/ /l/ /aa/ /n/ /o/ /n^/
68	ตะรุเตา	/t/ /a/ /r/ /u/ /t/ /a/ /w^/
69	ดอนหอยหลอด	/d/ /@@/ /n/ /h/ /@@/ /j^/ /l/ /@@/ /t^/
70	เขาวิ้ง	/kh/ /a/ /w^/ /w/ /a/ /ng^/

ตารางที่ ก. 4 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 1 - 17

ลำดับ	คำสำคัญ	Triphone
1	จังหวัด	/c+a/ /c-a+ng [^] / /a-ng [^] +w/ /ng [^] -w+a/ /w-a+t [^] / /a-t [^] /
2	ตราด	/tr+aa/ /tr-aa+t [^] / /aa-t [^] /
3	จันทบุรี	/c+a/ /c-a+n [^] / /a-n [^] +th/ /n [^] -th+a/ /th-a+b/ /a-b+u/ /b-u+r/ /u-r+ii/ /r-ii/
4	ระยอง	/r+a/ /r-a+j/ /a-j+@@/ /j-@@+ng [^] / /@@-ng [^] /
5	ชลบุรี	/ch+o/ /ch-o+n [^] / /o-n [^] +b/ /n [^] -b+u/ /b-u+r/ /u-r+ii/ /r-ii/
6	ฉะเชิงเทรา	/ch+a/ /ch-a+ch/ /a-ch+qq/ /ch-qq+ng [^] / /qq-ng [^] +s/ /ng [^] -s+a/ /s-a+w [^] / /a-w [^] /
7	สมุทรปราการ	/s+a/ /s-a+m/ /a-m+u/ /m-u+t [^] / /u-t [^] +pr/ /t [^] -pr+aa/ /pr-aa+k/ /aa-k+aa/ /k-aa+n [^] / /aa-n [^] /
8	กรุงเทพ	/kr+u/ /kr-u+ng [^] / /u-ng [^] +th/ /ng [^] -th+ee/ /th-ee+p [^] / /ee-p [^] /
9	สมุทรสาคร	/s+a/ /s-a+m/ /a-m+u/ /m-u+t [^] / /u-t [^] +s/ /t [^] -s+aa/ /s-aa+kh/ /aa-kh+@@/ /kh-@@+n [^] / /@@-n [^] /
10	สมุทรสงคราม	/s+a/ /s-a+m/ /a-m+u/ /m-u+t [^] / /u-t [^] +s/ /t [^] -s+o/ /s-o+ng [^] / /o-ng [^] +khr/ /ng [^] -khr+aa/ /khr-aa+m [^] / /aa-m [^] /
11	เพชรบุรี	/ph+e/ /ph-e+t [^] / /e-t [^] +b/ /t [^] -b+u/ /b-u+r/ /u-r+ii/ /r-ii/
12	ประจวบคีรีขันธ์	/pr+a/ /pr-a+c/ /a-c+ua/ /c-ua+p [^] / /ua-p [^] +kh/ /p [^] -kh+i/ /kh-i+r/ /i-r+ii/ /r-ii+kh/ /ii-kh+a/ /kh-a+n [^] / /a-n [^] /
13	ชุมพร	/ch+u/ /ch-u+m [^] / /u-m [^] +ph/ /m [^] -ph+@@/ /ph-@@+n [^] / /@@-n [^] /
14	สุราษฎร์ธานี	/s+u/ /s-u+r/ /u-r+aa/ /r-aa+t [^] / /aa-t [^] +th/ /t [^] -th+aa/ /th-aa+n/ /aa-n+ii/ /n-ii/
15	นครศรีธรรมราช	/n+a/ /n-a+kh/ /a-kh+@@/ /kh-@@+n [^] / /@@-n [^] +s/ /n [^] -s+ii/ /s-ii+th/ /ii-th+a/ /th-a+m [^] / /a-m [^] +m/ /m [^] -m+a/ /m-a+r/ /a-r+aa/ /r-aa+t [^] / /aa-t [^] /
16	สงขลา	/s+o/ /s-o+ng [^] / /o-ng [^] +khl/ /ng [^] -khl+aa/ /khl-aa/
17	ปัตตานี	/p+a/ /p-a+t [^] / /a-t [^] +t/ /t [^] -t+aa/ /t-aa+n/ /aa-n+ii/ /n-ii/

ตารางที่ ก. 5 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 18 - 41

ลำดับ	คำสำคัญ	Triphone
18	นราธิวาส	/n+a/ /n-a+r/ /a-r+aa/ /r-aa+th/ /aa-th+i/ /th-i+w/ /i-w+aa/ /w-aa+t^/ /aa-t^/
19	ระนอง	/r+a/ /r-a+n/ /a-n+@@/ /n-@@+ng^/ /@@-ng^/
20	พังงา	/ph+a/ /ph-a+ng^/ /a-ng^+ng/ /ng^-ng+aa/ /ng-aa/
21	ภูเก็ต	/ph+uu/ /ph-uu+k/ /uu-k+e/ /k-e+t^/ /e-t^/
22	กระบี่	/kr+a/ /kr-a+b/ /a-b+ii/ /b-ii/
23	ตรัง	/tr+a/ /tr-a+ng^/ /a-ng^/
24	สตูล	/s+a/ /s-a+t/ /a-t+uu/ /t-uu+n^/ /uu-n^/
25	ทะเล	/th+a/ /th-a+l/ /a-l+ee/ /l-ee/
26	ฉะว	/z+aa/ /z-aa+w^/ /aa-w^/
27	เกาะ	/k+@/ /k-@/
28	หาด	/h+aa/ /h-aa+t^/ /aa-t^/
29	คูทยาน	/z+u/ /z-u+t^/ /u-t^+th/ /t^-th+a/ /th-a+j/ /a-j+aa/ /j- aa+n^/ /aa-n^/
30	ดำน้า	/d+a/ /d-a+m^/ /a-m^+n/ /m^-n+aa/ /n-aa+m^/ /aa-m^/
31	บางแสน	/b+aa/ /b-aa+ng^/ /aa-ng^+s/ /ng^-s+xx/ /s-xx+n^/ /xx- n^/
32	พัททยา	/ph+a/ /ph-a+t^/ /a-t^+th/ /t^-th+a/ /th-a+j/ /a-j+aa/ /j-aa/
33	หัวหิน	/h+uu/ /h-uu+h/ /uu-h+i/ /h-i+n^/ /i-n^/
34	สมุทร	/s+a/ /s-a+m/ /a-m+u/ /m-u+j^/ /u-j^/
35	พะงัน	/ph+a/ /ph-a+ng/ /a-ng+a/ /ng-a+n^/ /a-n^/
36	เต่า	/t+a/ /t-a+w^/ /a-w^/
37	ลิมิลัน	/s+i/ /s-i+m/ /i-m+i/ /m-i+l/ /i-l+a/ /l-a+n^/ /a-n^/
38	สุรินทร์	/s+u/ /s-u+r/ /u-r+i/ /r-i+n^/ /i-n^/
39	พีพี	/ph+ii/ /ph-ii+ph/ /ii-ph+ii/ /ph-ii/
40	ลันตา	/l+a/ /l-a+n^/ /a-n^+t/ /n^-t+aa/ /t-aa/
41	ล้าน	/l+aa/ /l-aa+n^/ /aa-n^/

ตารางที่ ก. 6 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 42 - 57

ลำดับ	คำสำคัญ	Triphone
42	จอมเทียน	/c+@@/ /c-@@+m^/ /@@-m^+th/ /m^-th+ia/ /th-ia+n^/ /ia-n^/
42	สี่ช้าง	/s+ii/ /s-ii+ch/ /ii-ch+a/ /ch-a+ng^/ /a-ng^/
44	เขาสามมุก	/kh+a/ /kh-a+w^/ /a-w^+s/ /w^-s+aa/ /s-aa+m^/ /aa-m^+m/ /m^-m+u/ /m-u+k^/ /u-k^/
45	ลิบง	/l+i/ /l-i+b/ /i-b+o/ /b-o+ng^/ /o-ng^/
46	ถ้ำมรกต	/th+a/ /th-a+m^/ /a-m^+m/ /m^-m+@@/ /m-@@+r/ /@@-r+a/ /r-a+k/ /a-k+o/ /k-o+t^/ /o-t^/
47	กูด	/k+uu/ /k-uu+t^/ /uu-t^/
48	ช้าง	/ch+aa/ /ch-aa+ng^/ /aa-ng^/
49	มันนอก	/m+a/ /m-a+n^/ /a-n^+n/ /n^-n+@@/ /n-@@+k^/ /@@-k^/
50	มันใน	/m+a/ /m-a+n^/ /a-n^+n/ /n^-n+a/ /n-a+j^/ /a-j^/
51	แหลมตะลุมพุก	/l+xx/ /l-xx+m^/ /xx-m^+t/ /m^-t+a/ /t-a+l/ /a-l+u/ /l-u+m^/ /u-m^+ph/ /m^-ph+u/ /ph-u+k^/ /u-k^/
52	พระราชวังไกลกังวล	/phr+a/ /phr-a+r/ /a-r+aa/ /r-aa+t^/ /aa-t^+ch/ /t^-ch+a/ /ch-a+w/ /a-w+a/ /w-a+ng^/ /a-ng^+kl/ /ng^-kl+a/ /kl-a+j^/ /a-j^+k/ /j^-k+a/ /k-a+ng^/ /a-ng^+w/ /ng^-w+o/ /w-o+n^/ /o-n^/
53	เขาสามร้อยยอด	/kh+a/ /kh-a+w^/ /a-w^+s/ /w^-s+aa/ /s-aa+m^/ /aa-m^+r/ /m^-r+@@/ /r-@@+j^/ /@@-j^+j/ /j^-j+@@/ /j-@@+t^/ /@@-t^/
54	มะนาว	/m+a/ /m-a+n/ /a-n+aa/ /n-aa+w^/ /aa-w^/
55	เขาตะเกียบ	/kh+a/ /kh-a+w^/ /a-w^+t/ /w^-t+a/ /t-a+k/ /a-k+ia/ /k-ia+p^/ /ia-p^/
56	เขาไกรลาส	/kh+a/ /kh-a+w^/ /a-w^+kr/ /w^-kr+a/ /kr-a+j^/ /a-j^+l/ /j^-l+aa/ /l-aa+t^/ /aa-t^/
57	เจ้าสำราญ	/c+aa/ /c-aa+w^/ /aa-w^+s/ /w^-s+a/ /s-a+m^/ /a-m^+r/ /m^-r+aa/ /r-aa+n^/ /aa-n^/

ตารางที่ ก. 7 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Triphone ของชุดคำสำคัญลำดับที่ 58 - 70

ลำดับ	คำสำคัญ	Triphone
58	ชะอำ	/ch+a/ /ch-a+z/ /a-z+a/ /z-a+m^/ /a-m^/
59	ปีกเตียน	/p+v/ /p-v+k^/ /v-k^+t/ /k^-t+ia/ /t-ia+n^/ /ia-n^/
60	แหลมพรหมเทพ	/l+xx/ /l-xx+m^/ /xx-m^+phr/ /m^-phr+o/ /phr-o+m^/ /o-m^+th/ /m^-th+ee/ /th-ee+p^/ /ee-p^/
61	ป่าตอง	/p+aa/ /p-aa+t/ /aa-t+@@/ /t-@@+ng^/ /@@-ng^/
62	แม่รำพึง	/m+xx/ /m-xx+r/ /xx-r+a/ /r-a+m^/ /a-m^+ph/ /m^-ph+v/ /ph-v+ng^/ /v-ng^/
63	ยอด	/j+@@/ /j-@@/
64	หนู	/n+uu/ /n-uu/
65	แมว	/m+xx/ /m-xx+w^/ /xx-w^/
66	สมิหลา	/s+a/ /s-a+m/ /a-m+i/ /m-i+l/ /i-l+aa/ /l-aa/
67	ดิณสูลานนท์	/t+i/ /t-i+n^/ /i-n^+s/ /n^-s+uu/ /s-uu+l/ /uu-l+aa/ /l-aa+n/ /aa-n+o/ /n-o+n^/ /o-n^/
68	ตะรุเตา	/t+a/ /t-a+r/ /a-r+u/ /r-u+t/ /u-t+a/ /t-a+w^/ /a-w^/
69	ดอนหอยหลอด	/d+@@/ /d-@@+n/ /@@-n+h/ /n-h+@@/ /h-@@+j^/ /@@-j^+l/ /j^-l+@@/ /l-@@+t^/ /@@-t^/
70	เขารัง	/kh+a/ /kh-a+w^/ /a-w^+w/ /w^-w+a/ /w-a+ng^/ /a-ng^/

ตารางที่ ก. 8 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Onset - Rhyme

ของชุดคำสำคัญลำดับที่ 1 - 27

ลำดับ	คำสำคัญ	Onset – Rhyme
1	จังหวัด	/c_a/ /a_ng^/ /w_a/ /a_t^/
2	ตราด	/tr_a/ /aa_t^/
3	จันทบุรี	/c_a/ /a_n^/ /th_a/ /a/ /b_u/ /u/ /r_i/ /ii/
4	ระยอง	/r_a/ /a/ /j_@/ /@@_ng^/
5	ชลบุรี	/ch_o/ /o_n^/ /b_u/ /u/ /r_i/ /ii/
6	ฉะเชิงเทรา	/ch_a/ /a/ /ch_q/ /qq_ng^/ /s_a/ /a_w^/
7	สมุทรปราการ	/s_a/ /a/ /m_u/ /u_t^/ /pr_a/ /aa/ /k_a/ /aa_n^/
8	กรุงเทพ	/kr_u/ /u_ng^/ /th_e/ /ee_p^/
9	สมุทรสาคร	/s_a/ /a/ /m_u/ /u_t^/ /s_a/ /aa/ /kh_@/ /@@_n^/
10	สมุทรสงคราม	/s_a/ /a/ /m_u/ /u_t^/ /s_o/ /o_ng^/ /khr_a/ /aa_m^/
11	เพชรบุรี	/ph_e/ /e_t^/ /b_u/ /u/ /r_i/ /ii/
12	ประจวบคีรีขันธ์	/pr_a/ /a/ /c_u/ /ua_p^/ /kh_i/ /i/ /r_i/ /ii/ /kh_a/ /a_n^/
13	ชุมพร	/ch_u/ /u_m^/ /ph_@/ /@@_n^/
14	สุราษฎร์ธานี	/s_u/ /u/ /r_a/ /aa_t^/ /th_a/ /aa/ /n_i/ /ii/
15	นครศรีธรรมราช	/n_a/ /a/ /kh_@/ /@@_n^/ /s_i/ /ii/ /th_a/ /a_m^/ /m_a/ /a/ /r_a/ /aa_t^/
16	สงขลา	/s_o/ /o_ng^/ /khl_a/ /aa/
17	ปัตตานี	/p_a/ /a_t^/ /t_a/ /aa/ /n_i/ /ii/
18	นราธิวาส	/n_a/ /a/ /r_a/ /aa/ /th_i/ /i/ /w_a/ /aa_t^/
19	ระนอง	/r_a/ /a/ /n_@/ /@@_ng^/
20	พังงา	/ph_a/ /a_ng^/ /ng_a/ /aa/
21	ภูเก็ต	/ph_u/ /uu/ /k_e/ /e_t^/
22	กระบี่	/kr_a/ /a/ /b_i/ /ii/
23	ตรัง	/tr_a/ /a_ng^/
24	สตูล	/s_a/ /a/ /t_u/ /uu_n^/
25	ทะเล	/th_a/ /a/ /l_e/ /ee/
26	อ่าว	/z_a/ /aa_w^/
27	เกาะ	/k_@/ /@/

ตารางที่ ก. 9 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Onset - Rhyme

ของชุดคำสำคัญลำดับที่ 28-54

ลำดับ	คำสำคัญ	Onset – Rhyme
28	หาด	/h_a/ /aa_t^/
29	อุทยาน	/z_u/ /u_t^/ /th_a/ /a/ /j_a/ /aa_n^/
30	ดำน้า	/d_a/ /a_m^/ /n_a/ /aa_m^/
31	บางแสน	/b_a/ /aa_ng^/ /s_x/ /xx_n^/
32	พทยา	/ph_a/ /a_t^/ /th_a/ /a/ /j_a/ /aa/
33	หัวหิน	/h_u/ /uua/ /h_i/ /i_n^/
34	สมุย	/s_a/ /a/ /m_u/ /u_j^/
35	พะงัน	/ph_a/ /a/ /ng_a/ /a_n^/
36	เต่า	/t_a/ /a_w^/
37	สิมิลัน	/s_i/ /i/ /m_i/ /i/ /l_a/ /a_n^/
38	สุรินทร์	/s_u/ /u/ /r_i/ /i_n^/
39	พีพี	/ph_i/ /ii/ /ph_i/ /ii/
40	ลันตา	/l_a/ /a_n^/ /t_a/ /aa/
41	ล้าน	/l_a/ /aa_n^/
42	จอมเทียน	/c_@/ /@@_m^/ /th_i/ /iia_n^/
42	สี่ช้าง	/s_i/ /ii/ /ch_a/ /a_ng^/
44	เขาสามมุก	/kh_a/ /a_w^/ /s_a/ /aa_m^/ /m_u/ /u_k^/
45	ลิบง	/l_i/ /i/ /b_o/ /o_ng^/
46	ถ้ำมรกต	/th_a/ /a_m^/ /m_@/ /@@/ /r_a/ /a/ /k_o/ /o_t^/
47	ภูต	/k_u/ /uu_t^/
48	ช้าง	/ch_a/ /aa_ng^/
49	มันนอก	/m_a/ /a_n^/ /n_@/ /@@_k^/
50	มันไฉ	/m_a/ /a_n^/ /n_a/ /a_j^/
51	แหลมตะลุมพุก	/l_x/ /xx_m^/ /t_a/ /a/ /l_u/ /u_m^/ /ph_u/ /u_k^/
52	พระราชวังไกลกังวล	/phr_a/ /a/ /r_a/ /aa_t^/ /ch_a/ /a/ /w_a/ /a_ng^/ /kl_a/ /a_j^/ /k_a/ /a_ng^/ /w_o/ /o_n^/
53	เขาสามร้อยยอด	/kh_a/ /a_w^/ /s_a/ /aa_m^/ /r_@/ /@@_j^/ /j_@/ /@@_t^/
54	มะนาว	/m_a/ /a/ /n_a/ /aa_w^/

ตารางที่ ก. 10 รายละเอียดขององค์ประกอบเสียงระดับย่อยของพยางค์ประเภท Onset-Rhyme

ของชุดคำสำคัญลำดับที่ 55-70

ลำดับ	คำสำคัญ	Onset – Rhyme
55	เขาตะเกียบ	/kh_a/ /a_w^/ /t_a/ /a/ /k_i/ /iia_p^/
56	เขาไกรลาส	/kh_a/ /a_w^/ /kr_a/ /a_j^/ /l_a/ /aa_t^/
57	เจ้าสำราญ	/c_a/ /aa_w^/ /s_a/ /a_m^/ /r_a/ /aa_n^/
58	ชะอำ	/ch_a/ /a/ /z_a/ /a_m^/
59	ปีกเตียน	/p_v/ /v_k^/ /t_i/ /iia_n^/
60	แหลมพรหมเทพ	/l_x/ /xx_m^/ /phr_o/ /o_m^/ /th_e/ /ee_p^/
61	ป่าตอง	/p_a/ /aa/ /t_@/ /@@_ng^/
62	แม่รำพึง	/m_x/ /xx/ /r_a/ /a_m^/ /ph_v/ /v_ng^/
63	ยอด	/j_@/ /@@/
64	หนู	/n_u/ /uu/
65	แมว	/m_x/ /xx_w^/
66	สมิหลา	/s_a/ /a/ /m_i/ /i/ /l_a/ /aa/
67	ดิณัฐลานนท์	/t_i/ /i_n^/ /s_u/ /uu/ /l_a/ /aa/ /n_o/ /o_n^/
68	ตะรุเตา	/t_a/ /a/ /r_u/ /u/ /t_a/ /a_w^/
69	ดอนหอยหลอด	/d_@/ /@@/ /@@/ /h_@/ /@@_j^/ /l_@/ /@@_t^/
70	เขาวัง	/kh_a/ /a_w^/ /w_a/ /a_ng^/

ประวัติผู้เขียนวิทยานิพนธ์

นายเฉลิมวุฒิ ไวชนะ เกิดเมื่อวันที่ 14 มีนาคม พ.ศ. 2523 ที่อำเภอพระประแดง จังหวัดสมุทรปราการ สำเร็จการศึกษา หลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้า มหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2544 และเข้าต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต แขนงวิชาการวิธีสัญญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2544



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย