

การหาตำแหน่งขอบเขตเสียงในเสียงพูดต่อเนื่องภาษาไทย



นางสาว ศิรินุช บุญสุข

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

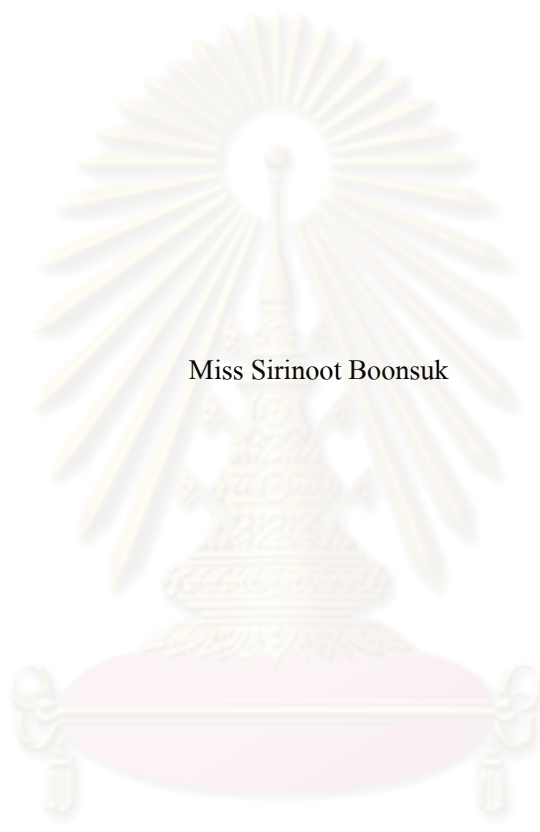
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

LOCATING ACOUSTIC BOUNDARIES IN THAI CONTINUOUS SPEECH



Miss Sirinoot Boonsuk

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Engineering Program in Computer Engineering  
Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2007

Copyright of Chulalongkorn University


หัวข้อวิทยานิพนธ์	การหาตำแหน่งขอบเขตเสียงในเสียงพูดต่อเนื่องภาษาไทย
โดย	นางสาวศิรินุช บุญสุข
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร. โปรคปราน บุญขุกกณะ
อาจารย์ที่ปรึกษาร่วม	อาจารย์ ดร.อดิวงค์ สุชาโต


คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย  
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต


  
..... คณะบดีคณะวิศวกรรมศาสตร์  
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

  
..... ประธานกรรมการ  
(รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

  
..... อาจารย์ที่ปรึกษา  
(อาจารย์ ดร. โปรคปราน บุญขุกกณะ)

  
..... อาจารย์ที่ปรึกษาร่วม  
(อาจารย์ ดร. อดิวงค์ สุชาโต)

  
..... กรรมการ  
(อาจารย์ ดร. พิษณุ คนองชัยยศ)

  
..... กรรมการ  
(ดร. ชัย วุฒิวิวัฒน์ชัย)

ศิรินุช บุญสุข : การหาค่าแหน่งขอบเขตเสียงในเสียงพูดต่อเนื่องภาษาไทย. (LOCATING ACOUSTIC BOUNDARIES IN THAI CONTINUOUS SPEECH) อ. ที่ปรึกษา : อ.ดร. ไพรพรรณ บุญขุกกณะ, อ.ที่ปรึกษาร่วม : อ.ดร.อดิวงค์ สุชาติ, 149 หน้า.

ความถูกต้องของการระบุตำแหน่งขอบเขตของเสียงมีบทบาทสำคัญที่สามารถเพิ่มประสิทธิภาพของการตัดแบ่งเป็นหน่วยเสียงในการรู้จำเสียง และเพิ่มคุณภาพของเสียงในการเลือกหน่วยเสียงสำหรับการสังเคราะห์เสียง การที่สามารถระบุขอบเขตเสียงได้โดยอัตโนมัติ จึงช่วยลดปัญหาของการสิ้นเปลืองแรงงานคนและลดเวลาในการพัฒนาการสร้างฐานข้อมูลเสียงที่ระบุตำแหน่งหน่วยเสียงด้วยคน

วิทยานิพนธ์นี้นำเสนอวิธีการในการหาค่าแหน่งขอบเขตเสียง โดยแบ่งเป็นสองขั้นตอนคือ 1) การระบุตำแหน่งขอบเขตเสียงแบบบังคับจากแบบจำลองฮิดเด้นมาคอฟเพื่อหาค่าแหน่งที่มีโอกาสเป็นขอบเขตเสียง 2) นำเสนอการปรับหาค่าแหน่งขอบเขตเสียงเพื่อปรับและหาค่าแหน่งขอบเขตเสียงที่ได้จากขั้นตอนแรกโดยละเอียด งานวิจัยนี้ได้ใช้ตัวจำแนกกลุ่มด้วยการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นที่ขึ้นกับบริบทเพื่อตรวจหาค่าแหน่งขอบเขตเสียง และได้ทำการจำแนกกลุ่มด้วย 21 ตัวจำแนกตามชนิดขอบเขตเสียง และในที่สุดได้เลือกค่าแหน่งที่มีค่าความน่าจะเป็นมากที่สุดมาเป็นผลลัพธ์ ซึ่งค่าความน่าจะเป็นคำนวณจากระยะทางของพื้นที่ที่แยกโดยฟังก์ชันดิสคริมิแนนต์

งานวิจัยนี้ใช้ฐานข้อมูลเสียงโตตัสในการประเมินประสิทธิภาพของวิธีที่เสนอ ซึ่งประกอบด้วยข้อมูลสัทลักษณะและข้อมูลแสดงเวลากำกับตำแหน่งขอบเขตเสียงสำหรับทุกหน่วยเสียงที่ระบุด้วยคน วิธีการหาค่าแหน่งขอบเขตเสียงที่เสนอได้รับความแม่นยำของการตรวจหาค่าแหน่งขอบเขตเสียงเท่ากับ 80.22% เมื่อใช้ระดับที่ยอมรับได้ 10 มิลลิวินาทีเพื่อนับเป็นตำแหน่งที่ถูกต้อง ความผิดพลาดการหาค่าแหน่งขอบเขตเสียงลดลง 43.42% เมื่อเทียบกับแบบอ้างอิงจากแบบจำลองฮิดเด้นมาคอฟ ค่าเฉลี่ยความคลาดเคลื่อนของตำแหน่งขอบเขตเสียงซึ่งเป็นจำนวนตำแหน่งเฟรมของตำแหน่งขอบเขตเสียงที่ตรวจหาได้คลาดเคลื่อนจากตำแหน่งขอบเขตเสียงที่ระบุด้วยคน มีค่าเฉลี่ยลดลงจาก 1.42 เฟรมเป็น 1.00 เฟรม เมื่อใช้เฟรมพิจารณาขนาด 10 มิลลิวินาที

จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา ..... วิศวกรรมคอมพิวเตอร์ .....

สาขาวิชา ..... วิศวกรรมคอมพิวเตอร์ .....

ปีการศึกษา ..... 2550 .....

ลายมือชื่อนิสิต .....  .....

ลายมือชื่ออาจารย์ที่ปรึกษา .....  .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....  .....

# # 487 07068 21 : MAJOR COMPUTER ENGINEERING

KEY WORD: PHONE BOUNDARY REFINEMENT/ SPEECH SEGMENTAION / SPEECH RECOGNITION

SIRINOOT BOONSUK : LOCATING ACOUSTIC BOUNDARIES IN THAI CONTINUOUS SPEECH. THESIS ADVISOR : PROADPRAN PUNYABUKKANA, Ph.D., THESIS COADVISOR : ATIWONG SUCHATO, Ph.D., 148 pp.

Precise phone boundary labeling plays an important role in improving segmentation performance in speech recognition, and increasing sound quality of unit selection in speech synthesis. Automatic phone alignment techniques are proposed to reduce the human efforts and time in the development of manually labeled speech corpus.

This thesis proposes an automatic method for locating acoustic boundaries. They can be divided into two steps: 1) HMM forced alignment is used to find the candidates phone boundaries, and 2) refinement of phone boundaries is proposed to adjust and fine-tune the boundaries obtained from the first step. The context-dependent Linear Discriminant Analysis (LDA) classifiers are used for phone boundary detection. The 21 specialized phone boundary classifiers are applied. The frame with maximum probability, calculated from distances in the space spanned by associated discriminant functions, is chosen as the output.

The LOTUS corpus (Large vOcabulary Thai continUous Speech recognition Corpus) is used to evaluate the proposed performance. It contains manual transcriptions with phone boundary information for every speech utterance. The proposed method yields the detection accuracy of 80.22% using 10 milliseconds tolerance level, considered as correct. The proposed refinement results in a 43.42% error reduction in locating phone boundaries compared to the baseline. The average deviation, the number of frame of the detected boundaries deviated from their corresponding manually labeled boundaries, is reduced from 1.42 to 1.0 frame when the frame size used is 10 milliseconds.

Department ..... Computer Engineering .....

Field of study .... Computer Engineering .....

Academic year ..... 2007.....

Student's signature..... 

Advisor's signature..... 

Co-advisor's signature..... 



## กิตติกรรมประกาศ

ในโอกาสนี้ข้าพเจ้าใคร่ขอขอบคุณ อาจารย์ ดร. โปรคปราน บุญยพุกกณะ อาจารย์ที่ปรึกษา วิทยานิพนธ์ และอาจารย์ ดร. อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งท่านทั้งสองได้ช่วยเหลือ ให้คำแนะนำ ข้อคิดที่เป็นประโยชน์ อันเป็นส่วนสำคัญที่ทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณ รศ. ดร. บุญเสริม กิจศิริกุล อ. ดร.พิชญ์ คนองชัยยศ และ ดร.ชัย วุฒิวิวัฒน์ชัย ที่ให้ข้อเสนอแนะดี ๆ และข้อคิดที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์นี้

นอกจากนี้ข้าพเจ้าขอขอบคุณ พี่ ๆ เพื่อน ๆ และน้อง ๆ ห้องปฏิบัติการ SLS ที่ได้ให้ความร่วมมือ สบับสนุน ช่วยเหลือ จนกระทั่งวิทยานิพนธ์นี้สำเร็จไปได้ด้วยดี

ท้ายที่สุดขอขอบคุณครอบครัวของข้าพเจ้า ที่ได้มอบกำลังใจ ความห่วงใย ช่วยเหลือแก่ข้าพเจ้าเสมอมา



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ .....	ช
สารบัญภาพ .....	ญ
สารบัญตาราง .....	ฎ
บทที่ 1 บทนำ .....	1
ความเป็นมาและความสำคัญของปัญหา .....	1
วัตถุประสงค์ของการวิจัย.....	2
ขอบเขตของการวิจัย .....	3
ประโยชน์ที่คาดว่าจะได้รับ.....	3
วิธีดำเนินการวิจัย.....	4
โครงสร้างวิทยานิพนธ์.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	6
แนวคิดและทฤษฎี.....	6
กลไกในการสร้างเสียงพูด.....	6
ทฤษฎีเกี่ยวกับความรู้เกี่ยวกับการตัดแบ่งหน่วยเสียง.....	7
ทฤษฎีเกี่ยวกับการวิเคราะห์สัญญาณเสียง.....	10
เสียงในภาษาไทย .....	18
ทฤษฎีทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล.....	29
งานวิจัยที่เกี่ยวข้อง.....	32
1. แนวทางแบบใช้สถิติ (Statistical Approach).....	32
2. แนวทางแบบใช้กฎ (Rule Based Approach).....	34
บทที่ 3 วิธีดำเนินการวิจัย.....	38
ขั้นตอนการดำเนินการวิจัย .....	38
เครื่องมือที่ใช้ในการวิจัย.....	38
ข้อมูลที่ใช้พัฒนาและทดสอบระบบ .....	38
1. ข้อมูลฝึกฝนและทดสอบ.....	40
2. การกำหนดขอบเขตหน่วยเสียงเพื่อสร้างฐานข้อมูล.....	41

การสร้างแบบอ้างอิง (Baseline).....	48
1. การฝึกฝนแบบจำลองฮิดเด้นมาคอฟ.....	49
2. การทดสอบประสิทธิภาพของแบบอ้างอิง.....	49
การสร้างระบบหาตำแหน่งขอบเขตเสียง.....	50
ประเมินผลระบบ (Evaluation).....	52
บทที่ 4 การปรับหาขอบเขตหน่วยเสียง.....	54
ลำดับขั้นตอนการวิจัย.....	54
1. การวิเคราะห์ชนิดขอบเขตเสียง (Boundary Analysis).....	54
2. การวิเคราะห์ค่าลักษณะสำคัญทางเสียง (Feature Analysis).....	55
3. การเลือกค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียง.....	58
3.1. การวิเคราะห์จากแผนภาพกล่อง.....	59
3.2. การวิเคราะห์จากผลทดสอบประสิทธิภาพของการจำแนกกลุ่ม.....	59
3.3. ความสัมพันธ์ของค่าลักษณะสำคัญทางเสียง (correlation).....	60
4. การวิเคราะห์ขนาดกรอบหน้าต่างการปรับ (Refining window).....	61
5. การปรับหาตำแหน่งขอบเขตหน่วยเสียง (Phone boundary refinement).....	63
5.1. การปรับหาตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ [43].....	63
5.1.1. ขั้นตอนการสร้างกฎตัดสินใจ.....	63
5.1.2. ขั้นตอนการปรับหาขอบเขตเสียงโดยใช้กฎตัดสินใจ.....	64
5.1.3. การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ.....	65
5.2. การปรับหาตำแหน่งขอบเขตเสียงโดยใช้สถิติ [45].....	66
5.2.1. ขั้นตอนแบบออฟไลน์ (off-line process).....	68
5.2.1.1. ขั้นตอนการสร้างต้นแบบการจำแนก.....	68
5.2.1.2. ขั้นตอนการวิเคราะห์พารามิเตอร์.....	69
5.2.1.3. ขั้นตอนทดสอบต้นแบบการจำแนก.....	76
5.2.2. ขั้นตอนแบบออนไลน์ (on-line process).....	77
5.2.2.1. ขั้นตอนการปรับหาขอบเขตเสียงโดยใช้สถิติ.....	77
5.2.2.2. การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้สถิติ.....	79
บทที่ 5 ผลการวิจัย.....	80
ผลการวิจัย.....	80
1. ผลการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิง.....	80
2. ผลการวิเคราะห์ชนิดขอบเขตเสียง.....	82



3. ผลการวิเคราะห์ค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียง.....	88
4. ผลการวิเคราะห์ขนาดกรอบหน้าต่างการปรับ (Refining window) .....	96
5. ผลจากการปรับหาขอบเขตหน่วยเสียง.....	96
5.1. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้กฎตัดสินใจ.....	96
5.1.1. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้กฎตัดสินใจ เมื่อขนาดกรอบ พิจารณาต่างกัน .....	101
5.2. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้สถิติ.....	105
5.2.1. ผลวิเคราะห์ร้อยละความถูกต้องของผลที่ได้จากวิธีที่เสนอเทียบกับคำตอบที่ แท้จริง.....	105
5.2.2. ผลวิเคราะห์ความคลาดเคลื่อนของผลที่ได้จากวิธีที่เสนอเทียบกับคำตอบที่ แท้จริง.....	109
5.2.3. ผลวิเคราะห์เวลาที่ใช้ในการจำแนก .....	110
5.2.4. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้สถิติเมื่อขนาดกรอบ พิจารณาต่างกัน .....	112
6. ผลการวิเคราะห์คะแนนในการพิจารณาหาตำแหน่งขอบเขตเสียง .....	118
บทที่ 6 บทสรุปผลการวิจัยและข้อเสนอแนะ .....	121
สรุปผลการวิจัย.....	121
ข้อเสนอแนะ.....	123
รายการอ้างอิง .....	125
ภาคผนวก ก.....	129
ภาคผนวก ข.....	131
ภาคผนวก ค.....	133
ภาคผนวก ง.....	134
ประวัติผู้เขียนวิทยานิพนธ์.....	148

## สารบัญภาพ

หน้า

รูปที่ 2.1	แบบจำลองการแหล่งกำเนิดและตัวกรอง [2] .....	7
รูปที่ 2.2	โครงสร้างของสเปกโตรแกรม [6] .....	11
รูปที่ 2.3	ตัวอย่างของสเปกโตรแกรมของประโยค “คร . ประสาร อ่านเปรื่อง” .....	13
รูปที่ 2.4	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ดา” .....	19
รูปที่ 2.5	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ทา” .....	20
รูปที่ 2.6	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “มา” .....	21
รูปที่ 2.7	ตัวอย่างสเปกตรัมจากการวิเคราะห์ค่าสัมประสิทธิ์การประมาณฟังก์ชันเชิงเส้นของหน่วย เสียงนาสิกกับหน่วยเสียงสระ .....	22
รูปที่ 2.8	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ลา” .....	22
รูปที่ 2.9	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ซา” .....	23
รูปที่ 2.10	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “รา” .....	23
รูปที่ 2.11	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ฟา” .....	24
รูปที่ 2.12	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “สา” .....	25
รูปที่ 2.13	ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “วา” .....	25
รูปที่ 2.14	แสดงลักษณะของแผนภาพกล่อง .....	31
รูปที่ 3.1	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /n <sup>^</sup> -/vb/ .....	42
รูปที่ 3.2	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/th/ และ /sil/-/kh/ ตามลำดับ .....	43
รูปที่ 3.3	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/t/, /sil/-/k/ และ /sil/-/p/ ตามลำดับ .....	43
รูปที่ 3.4	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/m/ และ /j <sup>^</sup> -/ng/ ตามลำดับ .....	44
รูปที่ 3.5	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /xx/-/l/ .....	44
รูปที่ 3.6	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/c/ และ /sil/-/ch/ ตามลำดับ .....	45
รูปที่ 3.7	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /ng <sup>^</sup> -/r/ .....	45
รูปที่ 3.8	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /ng <sup>^</sup> -/f/ และ /sil/-/s/ ตามลำดับ .....	46
รูปที่ 3.9	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/j/ และ /n <sup>^</sup> -/w/ ตามลำดับ .....	47
รูปที่ 3.10	ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /m/-/xx/ และ /s/-/a/-/sil/ ตามลำดับ ...	47
รูปที่ 3.11	ตัวอย่างสเปกโตรแกรมของเสียง /n-aa-j <sup>^</sup> -s-a-ng-aa/ ของตำแหน่งขอบเขตเสียงที่แท้จริง (ภาพบน) และการระบุด้วยแบบบังคับของแบบจำลองฮิดเดินมาคอฟ (ภาพล่าง) .....	49

รูปที่ 3.12	แผนภาพรวมการทำงานของระบบการหาตำแหน่งขอบเขตเสียงที่เสนอ .....	51
รูปที่ 4.1	การสร้างกรอบพิจารณาการหาตำแหน่งขอบเขตเสียง .....	62
รูปที่ 4.2	กระบวนการทำงานของการปรับหาตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ.....	65
รูปที่ 4.3	การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ .....	66
รูปที่ 4.4	แผนภาพการทำงานการปรับหาตำแหน่งขอบเขตเสียงโดยใช้สถิติ.....	67
รูปที่ 4.5	กระบวนการสร้างต้นแบบการจำแนก.....	68
รูปที่ 4.6	ขั้นตอนการวิเคราะห์พารามิเตอร์.....	69
รูปที่ 4.7	ตัวอย่างการสร้างต้นแบบการจำแนกกลุ่ม .....	70
รูปที่ 4.8	การคำนวณค่าเปลี่ยนแปลงเพื่อสร้างเวกเตอร์ลักษณะเฉพาะของข้อมูล .....	72
รูปที่ 4.9	ตัวอย่างเวกเตอร์ลักษณะเฉพาะของข้อมูลฝึกฝนชุดบวกต่อชุดลบแบบ 1:2.....	73
รูปที่ 4.10	ตัวอย่างเวกเตอร์ลักษณะเฉพาะของข้อมูลฝึกฝนชุดบวกต่อชุดลบแบบ 3:4.....	74
รูปที่ 4.11	การปรับหาขอบเขตเสียงโดยใช้สถิติ .....	78
รูปที่ 5.1	ฮิสโตแกรมจำนวนของขอบเขตเสียงชนิดต่าง ๆ.....	84
รูปที่ 5.2	ฮิสโตแกรมความถูกต้องการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง ....	94
รูปที่ 5.3	ฮิสโตแกรมความน่าจะเป็นความผิดพลาดระหว่างกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ ขอบเขตเสียง.....	95
รูปที่ 5.4	ฮิสโตแกรมความถูกต้องการหาตำแหน่งขอบเขตเสียงด้วยวิธีกฎตัดสินใจในชุดข้อมูลที่ 2 .....	98
รูปที่ 5.5	ฮิสโตแกรมความถูกต้องการหาตำแหน่งขอบเขตเสียงด้วยวิธีกฎตัดสินใจในชุดข้อมูลที่ 3 .....	100
รูปที่ 5.6	ฮิสโตแกรมร้อยละความถูกต้องของการหาขอบเขตเสียงโดยใช้กฎตัดสินใจ เมื่อขนาดกรอบ พิจารณาแตกต่างกัน.....	104
รูปที่ 5.2	ความผิดพลาดที่ลดลงเมื่อใช้ขนาดกรอบพิจารณาหาตำแหน่งขอบเขตที่แตกต่างกัน ....	114
รูปที่ 5.3	ฮิสโตแกรมร้อยละความถูกต้องของการหาขอบเขตเสียงเมื่อใช้ขนาดกรอบพิจารณาเพื่อ ปรับหาตำแหน่งขอบเขตที่แตกต่างกัน .....	117
รูปที่ 5.4	ฮิสโตแกรมจำนวนขอบเขตเสียงที่ระบุขอบเขตได้ถูกต้องในช่วงคะแนนต่าง ๆ .....	118
รูปที่ 5.5	แสดงแนวโน้มช่วงคะแนนต่าง ๆที่สามารถระบุตำแหน่งขอบเขตเสียงได้ถูกต้อง.....	119

## สารบัญตาราง

หน้า

ตารางที่ 2.2 ลักษณะของสัญญาณเสียงบนสเปกโตรแกรมของหน่วยเสียงแต่ละประเภท.....	12
ตารางที่ 2.3 ช่วงองค์ประกอบความถี่ สำหรับการสร้างพลังงานช่วงความถี่.....	16
ตารางที่ 2.4 สรุปรูปหน่วยเสียงและสัญลักษณ์แทนหน่วยเสียงที่เปรียบเทียบกับหน่วยเสียงกับสัทอักษร สากล (International Phonetic Alphabet, IPA) [7].....	26
ตารางที่ 3.1 สรุปรูปหน่วยเสียงที่ใช้ในวิทยานิพนธ์.....	39
ตารางที่ 4.1 ชุดการทดลองการปรับหาขอบเขตโดยใช้สถิติ.....	76
ตารางที่ 5.1 ความถูกต้องของการระบุขอบเขตเสียงจากแบบอ้างอิง.....	81
ตารางที่ 5.2 จำนวนขอบเขตเสียงสำหรับแต่ละชนิดขอบเขตเสียง.....	82
ตารางที่ 5.3 ร้อยละของจำนวนขอบเขตเสียงแต่ละชนิดขอบเขตเสียงต่อจำนวนขอบเขตเสียง ทั้งหมด.....	85
ตารางที่ 5.4 สรุปรูปชนิดขอบเขตเสียงที่ใช้ในวิทยานิพนธ์นี้.....	87
ตารางที่ 5.5 ค่าลักษณะสำคัญทางเสียงที่เลือกมาฝึกฝนแบบจำลองสำหรับแต่ละชนิดขอบเขตเสียง .....	89
ตารางที่ 5.6 ผลการปรับหาดำแหน่งขอบเขตเสียงโดยวิธีใช้กฎตัดสินใจเมื่อใช้ชุดข้อมูลที่ 2.....	97
ตารางที่ 5.7 ผลการปรับหาดำแหน่งขอบเขตเสียงโดยวิธีใช้กฎตัดสินใจเมื่อใช้ชุดข้อมูลที่ 3.....	99
ตารางที่ 5.8 ผลการทดสอบประสิทธิภาพของการระบุตำแหน่งขอบเขตหน่วยเสียงโดยวิธีทางสถิติ .....	105
ตารางที่ 5.9 การระบุตำแหน่งขอบเขตหน่วยเสียงโดยวิธีทางสถิติ.....	112
ตารางที่ 5.10 ผลการระบุตำแหน่งขอบเขตหน่วยเสียงวิธีที่เสนอ ซึ่งใช้นาครอปปิจารณาต่างกัน .....	115

# บทที่ 1

## บทนำ

### ความเป็นมาและความสำคัญของปัญหา

ระบบการรู้จำเสียงพูดได้แบ่งประเภทตามหน่วยการรู้จำ เช่น การรู้จำโดยใช้ลักษณะทางเสียงของคำซึ่งอาจใช้พยางค์ คำ หรือ วลีเป็นหน่วยในการเปรียบเทียบ และการรู้จำโดยใช้ลักษณะทางเสียงของหน่วยเสียงย่อย (Phoneme) ซึ่งมักจะใช้หน่วยเสียงที่ย่อยลงไป เช่น เสียงพยัญชนะและเสียงสระ ซึ่งเป็นวิธีที่เหมาะสมเมื่อต้องการเรียนรู้อำนาจจำนวนมาก ๆ ปัจจุบันการศึกษาทางด้านรู้จำเสียงพูดมีความก้าวหน้าไปอย่างมาก และการพัฒนาให้ระบบการรู้จำให้มีประสิทธิภาพมากขึ้น จึงมีงานวิจัยที่เข้ามาเพิ่มประสิทธิภาพของการรู้จำ เช่น การตรวจสอบจำนวนพยางค์หรืองานวิจัยที่พัฒนากรรมวิธีการพิจารณาลักษณะบ่งความต่างของเสียง การหาขอบเขตเริ่มต้นและสิ้นสุดของสัญญาณเสียงพูด เป็นต้น เนื่องจากการเรียนรู้อำนาจเสียงที่เป็นที่ยอมรับในปัจจุบันใช้แบบจำลองฮิดเด้นมาร์คอฟ (Hidden Markov Model, HMM) ที่พิจารณาข้อมูลทางสถิติเป็นการศึกษาข้อมูลซึ่งแบบจำลองฮิดเด้นมาร์คอฟยังมีข้อจำกัดบางประการ เนื่องจากเวกเตอร์ลักษณะเฉพาะ (feature vector) ที่ใช้เป็นตัวแทนเสียง (speech representation) ถูกสกัดมาจากกรอบเวลาที่มีความยาวคงที่ ทำให้มีความยากลำบากในการประยุกต์ความรู้จากสวนสัตศาสตร์ (Acoustic-phonetics) เข้ามาช่วยโดยตรงในการเรียนรู้อำนาจเสียง ดังนั้นทำให้เกิดการศึกษาเพื่อพัฒนาระบบการแบ่งส่วน (Segmentation) เพื่อระบุหาขอบเขตของหน่วยเสียงและบอกชนิดของหน่วยเสียง รวมทั้งได้ความถูกต้องของตำแหน่งรอยต่อหน่วยเสียง จากการแบ่งส่วนหน่วยเสียงที่สำคัญจึงได้เกิดวิทยานิพนธ์นี้ ที่นอกจากเสนอความสำคัญของการระบุหาขอบเขตของหน่วยเสียงแล้วยังเสนอวิธีการระบุหาขอบเขตของหน่วยเสียงอีกด้วย

วิธีการระบุหาขอบเขตของหน่วยเสียง นอกจากจะเป็นประโยชน์ต่อการเรียนรู้อำนาจเสียงแล้ว ยังเป็นที่ทราบกันดีว่าเป็นการบอกระยะเวลาของตำแหน่งหน่วยเสียงทางสัทลักษณะ (phonetic alignment) ซึ่งมีความสำคัญต่อการสังเคราะห์เสียงเพราะสามารถเพิ่มคุณภาพให้ดีขึ้นเนื่องจากคุณภาพของเสียงที่สังเคราะห์ได้ขึ้นอยู่กับคุณภาพหน่วยเสียงที่เลือกมาสังเคราะห์ ทำให้ต้องรู้ตำแหน่งขอบเขตที่แน่ชัดของการแบ่งส่วนหน่วยเสียงของสัญญาณเสียงสังเคราะห์ ดังนั้นงานหลักของการสังเคราะห์เสียงคือการเก็บฐานข้อมูลของแต่ละหน่วยเสียง โดยทำเครื่องหมายระบุขอบเขตของแต่ละหน่วยเสียงที่บันทึกนั้น ซึ่งขั้นตอนการระบุบอกระยะเวลาแต่ละหน่วยเสียงพิจารณาจากตัดสินใจของมนุษย์ ซึ่งตำแหน่งขอบเขตแต่ละหน่วยเสียงที่ได้มีความแม่นยำและน่าเชื่อถือมาก แต่ข้อเสียของการตัดสินใจจากคนในการระบุและหาขอบเขตหน่วยเสียงนั้นคือ การตัดสินใจนี้เป็นวิธีที่ไม่แน่นอนขึ้นอยู่กับความรู้และความเชี่ยวชาญเฉพาะบุคคล การเตรียมข้อมูลใช้



การตัดสินใจจากคนและใช้เวลาในการระบุตำแหน่งขอบเขตนาน เมื่อต้องการเตรียมข้อมูลที่มีขนาดข้อมูลใหญ่ จึงทำให้ต้องใช้จำนวนคนเพื่อมาตัดสินใจมากขึ้นและใช้เวลาในขั้นตอนนี้มาก ดังนั้นจึงเป็นแรงจูงใจของงานวิจัยเกี่ยวกับการระบุ การตัดแบ่งส่วนหน่วยเสียงอัตโนมัติและการระบุระยะเวลาเพื่อบอกการเริ่มต้นและสิ้นสุดของเสียงแบบอัตโนมัติมากขึ้น

แนวทางการพัฒนาโดยใช้ฮิดเดินมาคอฟซึ่งได้รับการยอมรับกันอย่างกว้างขวาง เพื่อนำมาหาหน่วยเสียงด้วยวิธีการระบุหน่วยเสียงแบบบังคับ (force alignment) ต้องฝึกฝนแบบจำลองหน่วยเสียงจากสัญญาณเสียงที่มีการแสดงสัทลักษณะระยะเวลาของหน่วยเสียงต่าง ๆ บนสัญญาณเสียง จากนั้นจะได้แบบจำลองหน่วยเสียงที่ได้นำมาใช้ในการระบุหน่วยเสียงบนสัญญาณเสียงที่เข้ามาได้ ดังนั้นประสิทธิภาพและความถูกต้องของการระบุระยะเวลาหน่วยเสียงของระบบขึ้นอยู่กับแบบจำลองนี้ แต่เนื่องจากข้อจำกัดการฝึกฝนแบบจำลองฮิดเดินมาคอฟมีขีดจำกัด ถึงแม้ว่าจะใช้จำนวนข้อมูลฝึกฝนมากขึ้นเกินขีดที่สามารถพัฒนาแบบจำลองให้ดีขึ้นได้ก็ไม่สามารถปรับปรุงให้แบบจำลองหน่วยเสียงที่ได้มีคุณภาพดีขึ้น และวิธีการระบุหน่วยเสียงแบบบังคับของฮิดเดินมาคอฟใช้ค่าสัมประสิทธิ์เมลเฟรเคว็นซีเซปทรัล (Mel frequency cepstral coefficients: MFCC) เพียงค่าเดียวเท่านั้น ดังนั้นจำเป็นต้องใช้ความรู้ข้อมูลอื่นมาประกอบเพื่อหาบริเวณตำแหน่งที่เหมาะสมที่สุดในการบอกขอบเขตหน่วยเสียง จากตำแหน่งขอบเขตสมมติฐานที่ได้มาจากแบบจำลองฮิดเดินมาคอฟ โดยทั่วไปมีการใช้เทคนิคต่าง ๆ มาช่วย เช่น ใช้แบบจำลองฮิดเดนมาร์คอฟ ใช้แบบจำลองเกาส์เซียน (Gaussian Mixture Models, GMM) ใช้ข่ายงานประสาทเทียม (Artificial Neural Networks, ANN) ใช้ข่ายงานประสาทเทียมชนิดเพอเซปตรอนหลายชั้น (Multi Layer Perceptrons, MLP) เป็นต้น ซึ่งต้องใช้ข้อมูลที่มีการระบุระยะเวลาหน่วยเสียงด้วยคนมาฝึกฝนและเรียนรู้ลักษณะชนิดและตำแหน่งขอบเขตนั้น ๆ

วิทยานิพนธ์นี้จึงเสนอการพัฒนาการหาตำแหน่งขอบเขตเสียงด้วยวิธีการใช้ค่าลักษณะสำคัญต่าง ๆ มาพิจารณา ที่ทำให้เพิ่มความถูกต้องของการระบุตำแหน่งได้ตรงกับคำตอบมากขึ้น โดยปรับตำแหน่งขอบเขตของหน่วยเสียงจากตำแหน่งสมมติฐานที่ได้จากแบบจำลองฮิดเดินมาคอฟที่สามารถระบุตำแหน่งขอบเขตได้แม่นยำได้ระดับหนึ่ง โดยในวิทยานิพนธ์นี้เลือกใช้การตัดสินใจหาตำแหน่งขอบเขตโดยวิธีทางสถิติ

### วัตถุประสงค์ของการวิจัย

เนื่องจากความถูกต้องของตำแหน่งขอบเขตหน่วยเสียงมีความสำคัญในการเรียนรู้จำและการสังเคราะห์เสียง จึงเกิดแรงจูงใจในวิทยานิพนธ์นี้ วิทยานิพนธ์นี้ศึกษาลักษณะสำคัญทางเสียงของแต่ละหน่วยเสียงต่าง ๆ ในภาษาไทยเพื่อหาตัวแทนเสียงที่เหมาะสมในการหาขอบเขตหน่วยเสียง เพื่อพัฒนาระบบการแบ่งสัญญาณเสียงเป็นหน่วยต่าง ๆ และสามารถหาตำแหน่ง

ขอบเขตของหน่วยเสียงในสัญญาณเสียงพูดภาษาไทยได้ ซึ่งเป็นประโยชน์ต่อการเพิ่มประสิทธิภาพ ความถูกต้องของระบบการรู้จำเสียงพูดอัตโนมัติและเพิ่มความเป็นธรรมชาติให้กับการสังเคราะห์เสียงมากขึ้น วิทยานิพนธ์นี้มุ่งเน้นการสร้างต้นแบบการหาขอบเขตหน่วยเสียงจากการศึกษา พฤติกรรมของหน่วยเสียงต่าง ๆ ที่บ่งบอกถึงตำแหน่งและชนิดของขอบเขตระหว่างหน่วยเสียงใน สัญญาณเสียงพูดภาษาไทย ในกระบวนการหาขอบเขตหน่วยเสียงในวิทยานิพนธ์นี้ได้ถูกออกแบบ มาจากฐานข้อมูลเสียงที่ไม่มีเสียงรบกวน เพื่อประสิทธิภาพของการทำงานของการทำงานของการนำไปใช้งานต่อ จึงควรนำข้อมูลที่อยู่ในสภาพแวดล้อมเช่นเดียวกัน

### ขอบเขตของการวิจัย

1. ศึกษาลักษณะสำคัญทางเสียงของแต่ละหน่วยเสียงต่าง ๆ ในภาษาไทยเพื่อหา ตัวแทนเสียงที่เหมาะสมในการหาขอบเขตของหน่วยเสียงหน่วยเสียง
2. สร้างระบบปรับตำแหน่งขอบเขตหน่วยเสียง ซึ่งปรับตามชนิดของขอบเขต ระหว่างเสียงส่วนต่าง ๆ ในสัญญาณเสียง
3. สามารถบอกโอกาสความน่าจะเป็นของหรือคะแนนความมั่นใจของตำแหน่ง ขอบเขตของหน่วยเสียงในสัญญาณเสียงภาษาไทย
4. มีความแม่นยำและความถูกต้องของตำแหน่งขอบเขตเสียงที่น่าเชื่อถือได้

### ประโยชน์ที่คาดว่าจะได้รับ

ทราบถึงลักษณะสำคัญทางเสียงต่าง ๆ และหาตัวแทนเสียงที่เหมาะสมในการหา ขอบเขตระหว่างหน่วยเสียงในเสียงพูดภาษาไทยได้ สามารถนำไปใช้ในการพัฒนาระบบรู้จำ เสียงพูดภาษาไทยให้มีประสิทธิภาพสูงขึ้น และเพิ่มความเป็นธรรมชาติของเสียงที่ได้จากการ สังเคราะห์ เมื่อทราบตำแหน่งที่แน่ชัดของขอบเขตระหว่างหน่วยเสียงของระบบสังเคราะห์เสียงได้

ประโยชน์ของการหาดำแหน่งขอบเขตเสียงตามการพิจารณาจากอะคูสติกนี้ เนื่องจากบางงานที่ต้องการการศึกษาข้อมูลหรือสมบัติของเซกเมนต์ (segment) โดยต้องการ พิจารณาการเปลี่ยนแปลงหรือจุดที่เกิดความบกพร่อง ผิดปกติ ดังนั้นต้นแบบของเซกเมนต์ที่มี ค่าอะคูสติกภายในเซกเมนต์เหมือนกันจึงมีความสำคัญในการนำไปพิจารณาต่อไป นอกจากนี้จะ เห็นว่าในงานวิจัยเกี่ยวกับการจำแนกหน่วยเสียงนั้น ต้องการการจำแนกที่มีประสิทธิภาพสูง จึงต้อง มีการรวมกลุ่มหน่วยเสียงที่มีลักษณะเหมือนกันไว้ด้วยกัน ดังนั้น การสร้างเซกเมนต์ที่มีลักษณะ เหมือนกันที่พิจารณาจากอะคูสติกจึงมีประโยชน์ต่อการใช้งานการจำแนกหน่วยเสียง เพราะ เซกเมนต์มีสมบัติทางเสียงเหมือนกันในเซกเมนต์เดียว ทำให้ระบบรู้จำสามารถเรียนรู้รูปแบบที่ เหมือนกัน ซึ่งนำไปสู่การประสิทธิภาพการจำแนกหน่วยเสียงที่สูงขึ้น

## วิธีดำเนินการวิจัย

1. ศึกษาข้อมูลทางศาสตร์ของหน่วยเสียงต่าง ๆ ในภาษาไทย รวมทั้งศึกษา ทฤษฎีทางด้านภาษาศาสตร์
2. ศึกษาข้อมูลที่ปรากฏบนสเปกโตรแกรมจากนั้นหาตัวแทนแสดงข้อมูล รวมทั้งวิเคราะห์ลักษณะเด่นต่าง ๆ ที่ได้จากการแยกสัญญาณเสียง
3. ศึกษาทฤษฎี รวมทั้งงานวิจัยที่เกี่ยวข้องที่มีอยู่ในปัจจุบันเพื่อใช้เป็นข้อมูล อ้างอิง เช่น งานเกี่ยวกับการปรับตำแหน่งขอบเขตหน่วยเสียงวิธีต่าง ๆ การหา ขอบเขตหน่วยเสียงชนิดต่าง ๆ และการแบ่งส่วนหน่วยเสียงเป็นต้น
4. ศึกษาข้อมูล ณ ตำแหน่งขอบเขตหน่วยเสียงที่ได้จากวิธีการระบุหน่วยเสียง แบบบังคับของแบบจำลองฮิดเดินมาคอฟ
5. หาความสัมพันธ์ของจุดสำคัญของการเกิดหน่วยเสียงกับหน่วยเสียงใน ภาษาไทย เพื่อพิจารณาจุดสำคัญนั้นไปสู่การบอกขอบเขตหน่วยเสียงต่าง ๆ
6. ทดสอบความถูกต้องของตำแหน่งขอบเขตหน่วยเสียงที่ได้จากวิทยานิพนธ์นี้ และเปรียบเทียบตำแหน่งขอบเขตกับชุดทดลองอ้างอิง
7. ปรับปรุงและพัฒนาระบบปรับตำแหน่งขอบเขตหน่วยเสียง เพื่อหาตำแหน่ง ขอบเขตหน่วยเสียงที่แท้จริง เพื่อให้ระบบมีความถูกต้องและแม่นยำที่ น่าเชื่อถือได้
8. สรุปผลการทดสอบ พร้อมข้อเสนอแนะ
9. จัดทำรูปเล่มวิทยานิพนธ์

## โครงสร้างวิทยานิพนธ์

ในวิทยานิพนธ์ฉบับนี้จะแบ่งเนื้อหาในการนำเสนอออกเป็น 6 ส่วน คือ ในบทที่ 2 จะกล่าวถึง ทฤษฎีและงานวิจัยที่เกี่ยวข้องของวิทยานิพนธ์นี้ รวมทั้งทฤษฎีที่ใช้ในการหาลักษณะ สำคัญของเสียงสำหรับการหาคำแหน่งขอบเขตเสียงในเสียงพูดต่อเนื่องภาษาไทย

ในบทที่ 3 จะได้กล่าวถึงวิธีการดำเนินการวิจัย ฐานข้อมูลที่ใช้ในพัฒนาระบบและ วัตถุประสงค์ของระบบ การสร้างแบบอ้างอิง และอธิบายการประเมินผลเพื่อเปรียบเทียบ ประสิทธิภาพของระบบการหาขอบเขตเสียงที่เสนอในวิทยานิพนธ์นี้ ในบทที่ 4 อธิบายการ วิเคราะห์ปัจจัยของระบบการหาคำแหน่งขอบเขตเสียงเช่น การวิเคราะห์ชนิดขอบเขตเสียง วิธีการ การหาลักษณะที่สำคัญของเสียง เป็นต้น รวมทั้งอธิบายขั้นตอนการทำงานโดยรวม และวิธีการ ปรับหาคำแหน่งขอบเขตเสียงโดยการ ใช้กฎตัดสินใจและการ ใช้สถิติ ซึ่งปรับหาคำแหน่งที่

เหมาะสมกับขอบเขตเสียงที่แท้จริงภายในบริเวณที่ใกล้เคียงกับตำแหน่งที่ได้จากการระบุจากแบบจำลองฮิดเด็นมากอฟ

ในบทที่ 5 และบทที่ 6 จะได้กล่าวถึงการวิเคราะห์ผลการทดสอบการหาตำแหน่งขอบเขตเสียงตามปัจจัยวิธีการหาตำแหน่งขอบเขต และสรุปผลที่ได้จากวิทยานิพนธ์นี้ รวมทั้งให้ข้อเสนอแนะในการพัฒนางานวิจัยการหาตำแหน่งขอบเขตเสียงที่เหมาะสมในเสียงพูดนี้ต่อไป



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### แนวคิดและทฤษฎี

#### กลไกในการสร้างเสียงพูด

##### แบบจำลองแหล่งกำเนิด-ตัวกรอง (Source-Filter Model)

Muller Johannes [1] ได้อธิบายกลไกการเกิดเสียงด้วยการใช้แบบจำลองการแหล่งกำเนิดและตัวกรองซึ่งเป็นที่นิยมใช้กันในปัจจุบัน โดยแบ่งการเกิดเสียงออกเป็นสามส่วน คือ ต้นกำเนิดเสียงเกิดขึ้นเมื่อมีการสั่นของเส้นเสียง ท่อที่เป็นทางผ่านของเสียงซึ่งหมายถึงช่องเสียงและการกระจายของลม เมื่อลมออกจากช่องปากที่มีทิศทางแตกต่างกันทำให้เสียงที่เปล่งออกมาต่างกันไปด้วย แบบจำลองนี้อธิบายลักษณะที่ปรากฏบนสเปกตรัมของสัญญาณเสียงได้

ในกลไกการเกิดเสียงสามารถเกิดจากต้นกำเนิดเสียงได้ต่าง ๆ มากมาย การสั่นของเส้นเสียงเป็นจุดสำคัญที่เกิดเสียงและเป็นตัวกำหนดการเกิดเสียงก้อง เมื่อเราเปล่งเสียงก้อง ได้แก่ กลุ่มเสียงสระ (vowel) เสียงนาสิก (nasal) เสียงร้าว (trill) เสียงพยัญชนะเหลว (liquid) เป็นต้น การสั่นของเส้นเสียงจะสร้างชุดพลังงานออกมา ในเวลาที่เปล่งเสียงออกมาสามารถอธิบายการเพิ่มกระแสลมได้เมื่อมีการเปิดช่องระหว่างเส้นเสียง เมื่อช่องระหว่างเส้นเสียงปิดกระแสลมจะลดลงทันที ต้นกำเนิดเสียงที่สำคัญรองลงมา คือ ต้นกำเนิดที่เกิดเสียงสอดแทรก สร้างจากการบังคับลมให้ผ่านช่องแคบเล็ก ๆ ในช่องปาก กลุ่มเสียงสอดแทรก เช่น เสียง ส /s/ เสียง ฟ /f/ เป็นตัวอย่างที่ได้จากต้นกำเนิดเสียงนี้ คลื่นเสียงมีลักษณะไม่เป็นระเบียบ ไม่เกิดรูปแบบที่ซ้ำเดิม และต้นกำเนิดที่สามคือ การเกิดเสียงระเบิด เกิดขึ้นเมื่อแรงดันของกระแสลมถูกปิดกั้นหรือถูกบีบบังคับจากอวัยวะในปาก เช่น ลิ้น หรือ ริมฝีปาก แล้วกระแสลมถูกปล่อยออกมาอย่างรวดเร็วส่งผลให้เกิดการระเบิดของกระแสลม (burst) ที่ปล่อยออกมาจากปาก ตัวอย่างของกลุ่มเสียงที่เกิดจากการระเบิด ได้แก่ กลุ่มเสียงกัก เสียงกึ่งเสียดแทรก ต้นกำเนิดการระเบิดนี้จะปรากฏชัดจากการเพิ่มแรงดันลมให้มากขึ้น จากนั้นแรงดันก็หายไปดังแสดงออกมาในรูปแบบความแบนราบของรูปร่างสเปกตรัม

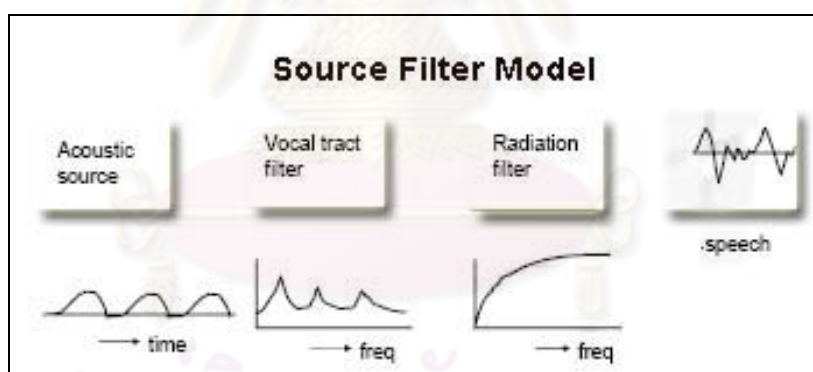
หน่วยเสียงต่าง ๆ เกิดเมื่อกระแสลมผ่านช่องเสียงหรือช่องจุมก แต่ละหน่วยเสียงเกิดขึ้นที่ตำแหน่งฐานกรณ์ (place of articulation) ลิ้น กรรไกร โดยมีตำแหน่งแน่นอน ซึ่งตำแหน่งเป็นตัวกำหนดรูปร่างของช่องเสียงที่ต่างกัน ส่งผลให้เกิดความถี่เรโซแนนซ์แตกต่างกันออกไปและเรียกความถี่เรโซแนนซ์นี้ว่า ความถี่ฟอร์มแมนต์ (formant) โดยค่าความถี่ฟอร์มแมนต์ขึ้นกับความยาวและรูปร่างของช่องเสียง หมายถึง เมื่อเสียงผ่านไปช่องเสียง จะถูกกรองด้วยตัวกรอง ทำให้เกิด



หน่วยเสียงที่มีองค์ประกอบความถี่ต่างกันไปตามรูปร่างของช่องทางเดินเสียงที่เปลี่ยนแปลงไปและแตกต่างกัน เช่น มีตำแหน่งของลิ้นแตกต่างกันทำให้ได้เสียงที่ต่างกัน เป็นต้น นอกจากนี้ค่าความถี่ฟอร์แมนที่เป็นตัวสำคัญในการระบุ หน่วยเสียง โดยเฉพาะเสียงสระต่าง ๆ ที่พิจารณาชนิดของเสียงสระแต่ละตัว จากความถี่ฟอร์แมนที่ต่ำสุด ได้แก่ ความถี่ฟอร์แมนที่อันดับที่หนึ่งและความถี่ฟอร์แมนที่อันดับที่สอง

ในขั้นตอนสุดท้ายของแบบจำลองการแหล่งกำเนิดและตัวกรอง ได้อธิบายทิศทางลมที่ออกมาจากช่องปาก เสียงที่เปล่งออกมาแตกต่างกันเนื่องจากทิศทางของลมต่างกัน เช่น กระแสลมออกทางจมูกหรือทางปาก ส่งผลให้ปรากฏเป็นลักษณะการกระจายตัวของเสียงในสเปกตรัม

จะเห็นได้ว่า แบบจำลองการแหล่งกำเนิดและตัวกรองเป็นรูปแบบสำคัญที่บอกลักษณะต่าง ๆ ที่เกิดขึ้นในสัญญาณเสียง เช่น รูปร่างสเปกตรัมของเสียง ตำแหน่งความถี่ฟอร์แมนที่ขึ้นกับรูปร่างของช่องเสียง เป็นต้น นอกจากนี้ยังอธิบายค่าความถี่ฟอร์แมนที่แตกต่างกันเมื่อผู้พูดต่างกัน หมายถึง ผู้หญิงมีค่าความถี่ฟอร์แมนที่แตกต่างกับผู้ชาย หรือค่าความถี่ฟอร์แมนของผู้หญิงสองคนแตกต่างกัน เพราะความยาวของช่องเสียงของแต่ละคนนั้นมีค่าต่างกัน จากรายละเอียดของแบบจำลองการแหล่งกำเนิดและตัวกรอง ได้สรุปออกเป็นแผนภาพดังรูปที่ 2.1



รูปที่ 2.1 แบบจำลองการแหล่งกำเนิดและตัวกรอง [2]

**ทฤษฎีเกี่ยวกับความรู้เกี่ยวกับการตัดแบ่งหน่วยเสียง**

การศึกษาทฤษฎีเกี่ยวกับความรู้เกี่ยวกับการตัดแบ่งหน่วยเสียงนี้เพื่อให้เป็นที่เข้าใจตรงกันในการนิยามความหมายของคำที่ใช้ในวิทยานิพนธ์นี้ รวมทั้งเพื่อศึกษากระบวนการและความสำคัญของการตัดแบ่งหน่วยเสียง (speech segmentation) และการบอกตำแหน่งของหน่วยเสียงอัตโนมัติ (automatic speech alignment) ซึ่งได้อธิบายดังนี้

## 1. ส่วนหน่วยเสียงหรือเซกเมนต์ (segment)

การนิยามหน่วยเสียงหรือเซกเมนต์ (segment) [3] ซึ่งให้นิยามขึ้นอยู่กับศาสตร์ที่ศึกษา ดังนี้

เมื่ออธิบายตามหลักของภาษาศาสตร์ [4] นิยามว่า หน่วยเสียงที่ใช้เพื่อเป็นตัวอ้างอิงหลักกับหน่วยที่ไม่ต่อเนื่องใด ๆ ที่เกิดขึ้นในสัญญาณเสียง โดยสามารถระบุชนิดหน่วยเสียงนั้นได้ ไม่ว่าจะผ่านทางกายภาพหรือทางด้านโสตประสาท

เมื่ออธิบายตามหลักของภาษาพูด หน่วยเสียงหมายถึง เสียงพยัญชนะ เสียงสระ เสียงวรรณยุกต์หรือการเน้นหนักเบา โดยที่ไม่มีความชัดเจนที่จะให้กำหนด สัญลักษณ์ของภาษามาเทียบกับเซกเมนต์ได้

เมื่ออธิบายตามหลักของภาษาการอ่าน หน่วยเสียงที่เป็นส่วนย่อยที่สุดที่สามารถบอกหรืออ่านได้คือ หน่วยเสียงต่าง ๆ (phone) ของภาษานั้น เช่น หน่วยเสียง ต /t/ มี 1 หน่วยเสียง เป็นต้น

จากการศึกษานิยามจากศาสตร์ต่าง ๆ ในวิทยานิพนธ์นี้พิจารณาจากอะคูสติก (acoustics) และการพิจารณาค่าลักษณะสำคัญทางเสียงเป็นหลักโดยนิยามว่า หน่วยเสียง หมายถึง ส่วนย่อยของเสียงที่มีขนาดเล็กที่สุด ต้องมีค่าลักษณะสำคัญทางเสียงของแต่ละหน่วยเสียงนั้นเหมือนกัน และสามารถระบุบอกชนิดหน่วยเสียงและอธิบายความหมายของหน่วยเสียงชนิดนั้นได้ สำหรับหน่วยเสียงที่ต่างชนิดกัน ต้องมีความแตกต่างของค่าลักษณะสำคัญทางเสียง หน่วยเสียงอาจจะเป็นหน่วยเสียงต่าง ๆ ของภาษานั้น ๆ หรือเล็กกว่าหน่วยเสียงก็ได้ เช่น เสียง น /n/ ในกลุ่มเสียงนาสิกของภาษาไทย เกิดเป็นหนึ่งหน่วยเสียง คือ ส่วนที่เป็นเสียง น /n/ ขณะที่เสียง ต /t/ ในภาษาไทย ซึ่งอยู่ในกลุ่มเสียงกัก มีลักษณะการเกิดเสียงเงียบ (เกิดจากการปิดกั้นลม) และมีส่วนเสียงระเบิด (เกิดเมื่อลมถูกปล่อยออกมา) ดังนั้น เมื่อใช้นิยามตามวิทยานิพนธ์นี้ จึงได้เซกเมนต์ในเสียง ต ประกอบเป็นส่วนเสียงเงียบ /sil/ และส่วนเสียงระเบิด /t/ เป็นต้น

เมื่อนำหน่วยเสียงย่อยต่าง ๆ มาประกอบกันเพื่อสร้างเป็นพยางค์หรือคำ จึงเกิดจากการประกอบกันของหน่วยเสียง คล้ายกับหลักการสังเคราะห์เสียง และเมื่อมีการประกอบกันของหน่วยเสียงย่อยต่าง ๆ จึงเกิดตำแหน่งที่เป็นรอยต่อระหว่างหน่วยเสียง ซึ่งเรียกว่า ตำแหน่งขอบเขตหน่วยเสียง

## 2. การแบ่งหน่วยเสียงออกเป็นส่วน ๆ (Segmentation)

การแบ่งหน่วยเสียงออกเป็นส่วน ๆ (Segmentation) คือ การระบุตำแหน่งของแต่ละส่วนย่อยของเสียง เพื่อบอกจุดเริ่มต้น สิ้นสุดหน่วยเสียง อาจจะเป็นการระบุขอบเขตระหว่างขอบเขตคำ พยางค์ และระหว่างหน่วยเสียงในสัญญาณเสียงพูดในธรรมชาติ การแบ่งหน่วยเสียงถึงแม้ว่าต้องใช้กระบวนการตัดสินใจของมนุษย์ หรือใช้ปัญญาประดิษฐ์ในการเข้าใจเสียงธรรมชาติ การระบุบอกตำแหน่งของหน่วยเสียงบนสัญญาณเสียงยังเป็นปัญหาที่ยาก ไม่มีคำตอบที่ถูกต้องแน่ชัด ในการบอกเส้นแบ่งระหว่างหน่วยเสียงรวมทั้งมีปัญหาเช่น มีการซ้อนทับของส่วนเสียงที่ตามมา หรือมีผลกระทบจากหน่วยเสียงก่อนหน้า เป็นต้น ถึงแม้ว่าการตัดแบ่งหน่วยเสียงไม่มีกฎเกณฑ์ แต่สามารถสร้างระบบที่ตัดสินใจและพิจารณาเส้นแบ่งจากข้อกำหนดได้อย่างอัตโนมัติ จากการศึกษาข้อมูลและเทียบประสิทธิภาพกับการตัดสินใจจากมนุษย์ที่มีความแม่นยำสูง

## 3. ทฤษฎีการทำงานแบบจำลองฮิดเดินมาคอฟ

แบบจำลองฮิดเดินมาคอฟเป็นที่นิยมใช้กันอย่างแพร่หลาย เนื่องจากแบบจำลองฮิดเดินมาคอฟ สามารถอธิบายได้โดยโครงสร้างทางคณิตศาสตร์ และสามารถประยุกต์ตามการใช้งานได้โดยเปลี่ยนแปลงทฤษฎีพื้นฐานให้เหมาะสม ดังนั้นวิทยานิพนธ์นี้ได้นำแบบจำลองฮิดเดินมาคอฟมาใช้ในขั้นตอนแรก เพื่อหาส่วนหน่วยเสียงในสัญญาณเสียงพูด จากการทำงานแบบจำลองฮิดเดินมาคอฟในวิธีการระบุหน่วยเสียงแบบบังคับ (force alignment) ซึ่งใช้ฟังก์ชันการทำงาน HVite เพื่อคำนวณและสร้างเครือข่ายของแต่ละหน่วยเสียงที่เข้ามา โดยใช้ข้อมูลที่ได้จากการแสดงสัทลักษณ์ (transcription) และพจนานุกรมจากนั้น แล้วเลือกจับคู่การออกเสียงพูดที่ดีที่สุด

วิธีการระบุหน่วยเสียงแบบบังคับ (force alignment) ใช้เพื่อหาตำแหน่งขอบเขตระหว่างหน่วยเสียงจากชุดข้อมูลที่รู้ลำดับแล้ว โดยมีขั้นตอนคือ นำแบบจำลองของแต่ละหน่วยเสียงที่รู้ลำดับแล้ว มาสร้างแบบจำลองโดยการประกอบกันของแต่ละแบบจำลองย่อยตามลำดับของข้อมูลที่ทราบจากการระบุชนิดหน่วยเสียงที่ได้ จากนั้นหาตำแหน่งที่ดีที่สุดของหน่วยเสียงในแบบจำลองจากชุดสถานะย่อย (state sequence) ที่นำมาประกอบ แล้วบันทึกเวลาขอบเขตระหว่างหน่วยเสียง โดยการระบุหน่วยเสียงแบบบังคับใช้อัลกอริทึมไวเทอบี (Viterbi) เพื่อหาชุดสถานะย่อยจากแบบจำลองที่ได้ฝึกฝนไปแล้ว อัลกอริทึมทำคล้ายกับสร้างเวกเตอร์ที่มีโครงสร้างเหมือนกับชุดเวกเตอร์ลักษณะเฉพาะที่ใช้ในขั้นตอนการสร้างแบบจำลอง แต่ละเวกเตอร์ลักษณะเฉพาะเวลากำกับ ทำให้สามารถระบุเวลาที่ตำแหน่งขอบเขตได้ เมื่อทราบเวลาที่แต่ละเวกเตอร์นั้น การหาตำแหน่งขอบเขตหน่วยเสียงเกิดขึ้นที่บริเวณรอยต่อระหว่างสถานะที่มีการประกอบกันของแบบจำลองสองหน่วยที่ต่างกัน การระบุเวลาในแต่ละเวกเตอร์อยู่ในรูป

เป็น  $n * s - c$  เมื่อ  $n$  คือตำแหน่งเวกเตอร์ที่อยู่ใน sequence (โดยเริ่มพิจารณาจาก 0)  
 $s$  คือขนาดกรอบสัญญาณเวลาที่เลื่อนไปตามสัญญาณเสียง  
 $c$  คือค่าคงที่ขึ้นอยู่กับจุดเริ่มต้นของตำแหน่งตรงกลางหน้าต่างของ  
 สัญญาณเสียง ที่พิจารณาโดยทั่วไปมีค่า  $c$  เท่ากับ  $s/2$  หรือ เท่ากับ 0

วิทยานิพนธ์นี้ใช้ฐานข้อมูลเสียงที่ชื่อ LOTUS (Large Vocabulary Thai Continuous Speech recognition Corpus) [5] ซึ่งเป็นข้อมูลที่สะอาดและไม่มีเสียงรบกวน ในขั้นตอนการฝึกฝนนั้นเพื่อสร้างแบบจำลองหน่วยเสียงที่สร้างภายใต้แบบจำลองฮิดเดินมาคอฟ โดยฐานข้อมูลนี้มีการระบุตำแหน่งและบอกระยะเวลาของหน่วยเสียงต่าง ๆ ด้วย

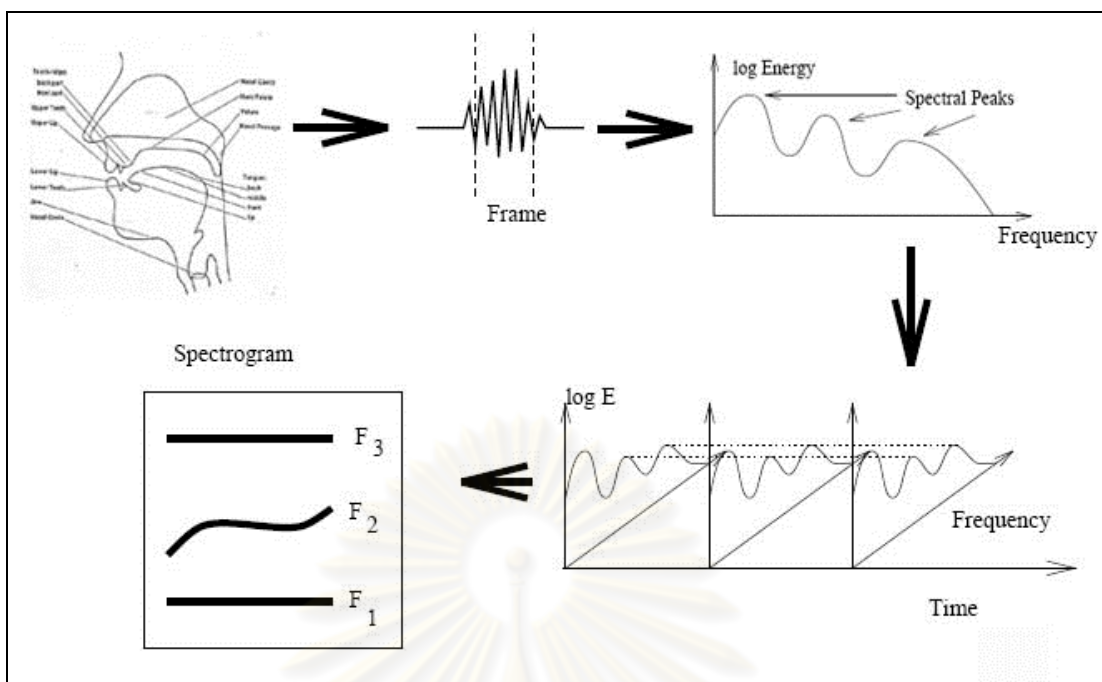
### ทฤษฎีเกี่ยวกับการวิเคราะห์สัญญาณเสียง

ได้แก่ ทฤษฎีที่ใช้ในการวิเคราะห์เสียงและหาค่าลักษณะสำคัญของเสียง และเครื่องมือที่ใช้ในการวิเคราะห์สัญญาณเสียง เช่น สเปกโตรแกรม เป็นต้น

#### 1. สเปกโตรแกรม (Spectrogram)

สเปกโตรแกรมเป็นตัวแทนของสัญญาณเสียงที่แสดงออกมาในรูปของกราฟิก ซึ่งเป็นเครื่องมือสำคัญอันหนึ่งที่บอกข้อมูลเกี่ยวกับหน่วยเสียงจากรูปแบบคลื่นในโดเมนความถี่ (Frequency domain) เนื่องจากเราไม่สามารถวิเคราะห์หน่วยเสียงต่าง ๆ ที่อยู่บนสัญญาณเสียงจากรูปแบบคลื่นในโดเมนทางเวลาได้ ดังนั้นวิทยานิพนธ์นี้จึงได้ศึกษาโครงสร้างของสเปกโตรแกรม และหลักในการอ่านสเปกโตรแกรมเพื่อคุณลักษณะของสัญญาณเสียงที่เกิดจากฐานกรณ์ที่แตกต่างกัน ก่อนที่จะอ่านลักษณะที่ปรากฏบนสเปกโตรแกรม ควรศึกษาโครงสร้างของสเปกโตรแกรม ดังแสดงในรูปที่ 2.2

ศูนย์วิทยทรัพยากร  
 จุฬาลงกรณ์มหาวิทยาลัย



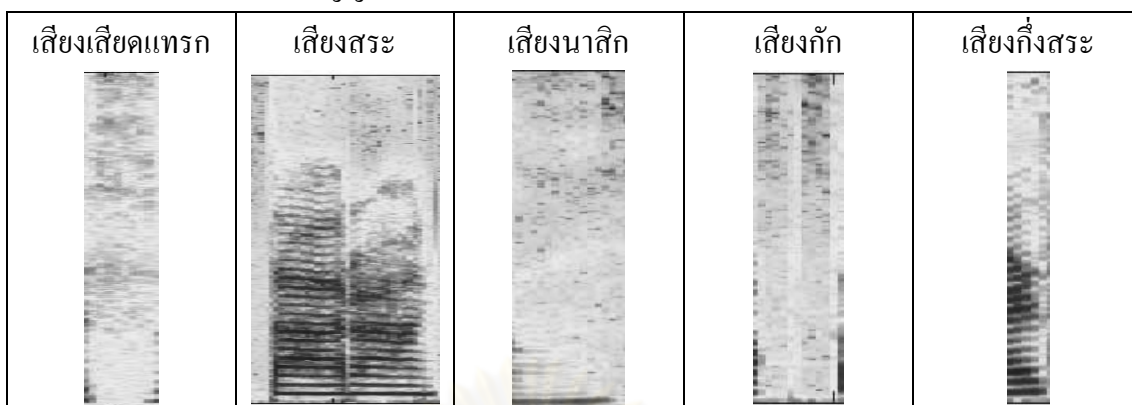
รูปที่ 2.2 โครงสร้างของสเปกโตรแกรม [6]

ข้อมูลที่ปรากฏบนสเปกโตรแกรม ประกอบด้วย

- 1) องค์ประกอบความถี่ของสัญญาณเสียง แสดงตามแกน  $y$
- 2) เวลาในการเคลื่อนที่ของเส้นที่สัมพันธ์กับเวลา แสดงตามแกน  $x$
- 3) ความเข้มสีแสดงแอมพลิจูด (amplitude) ของสัญญาณเสียง
- 4) ความเป็นเสียงก้อง (voicing) ของเสียงซึ่งนอกจากพิจารณาจากสัญญาณเสียง (waveform) แล้ว สามารถพิจารณาจากสเปกโตรแกรมได้ โดยดูจาก
  - a. ข้อมูลบนสเปกโตรแกรมแบบช่วงแคบ (narrowband spectrogram) จะปรากฏบริเวณที่เป็นเสียงก้อง ดูจากเส้นแถบสีตามแนวนอน
  - b. ข้อมูลบนสเปกโตรแกรมแบบช่วงกว้าง (Wideband spectrogram) จะปรากฏบริเวณที่เป็นเสียงก้อง ดูจากเส้นแถบสีตามแนวตั้ง
- 5) ตำแหน่งของลิ้นในปากและการปิดหรือเปิดของช่องปาก ส่งผลต่อค่าความถี่ฟอร์แมนท์ ซึ่งมีความสำคัญต่อเสียงสระ พบว่าความถี่ฟอร์แมนท์ที่หนึ่ง บอกลักษณะความสูงหรือต่ำของลิ้นในปาก ความถี่ฟอร์แมนท์ที่สองบอกตำแหน่งหน้าหรือหลังของลิ้นในช่องปาก เป็นต้น
- 6) สเปกโตรแกรมสามารถใช้ตรวจสอบความถูกต้องจากข้อมูลในไฟล์บอกเสียง แสดงสัญลักษณ์ ได้แก่ ระยะเวลา (duration) และความถี่ได้ แต่สเปกโตรแกรมไม่เหมาะสมในการแสดงข้อมูลบางอย่าง เช่น กรณีที่ต้องการดูความสัมพันธ์ของพลังงานของเสียงที่เกิดจากช่องปากเปรียบเทียบกับพลังงานของเสียงที่เกิดจากช่องจมูก

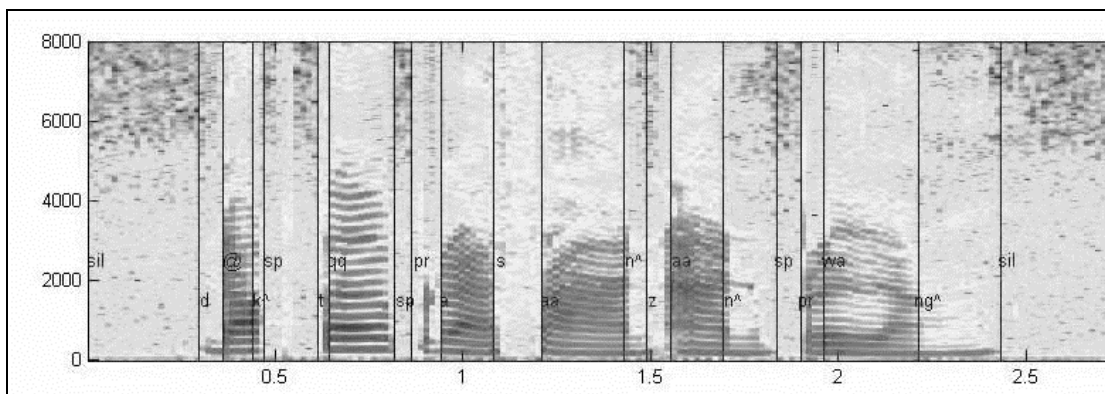


ตารางที่ 2.2 ลักษณะของสัญญาณเสียงบนสเปกโตรแกรมของหน่วยเสียงแต่ละประเภท



จากตารางที่ 2.2 แสดงลักษณะของเสียงบนสเปกโตรแกรมในกลุ่มเสียงที่ต่างกัน จะมีลักษณะต่างกันเนื่องจากมีตำแหน่งฐานกรณ์ต่างกัน รูปร่างของท่อเรโซแนนซ์ในช่องปากต่างกัน โดยเรียงจากกลุ่มเสียงเสียดแทรก เสียงสระ เสียงนาสิก เสียงกัก เสียงกึ่งสระ ตามลำดับ เมื่อพิจารณาความแตกต่างบนสเปกโตรแกรมในแต่ละช่วง พบว่าสามารถแบ่งขอบเขตของกลุ่มเสียงต่าง ๆ ได้ โดยจัดหมู่หน่วยเสียงที่มีลักษณะเหมือนกันไว้ในกลุ่มเสียงเดียวกันและแยกหน่วยเสียงที่มีลักษณะต่างกันในกลุ่มเสียงต่างกัน ซึ่งจะได้กลุ่มเสียงที่มีลักษณะร่วมคล้ายกันในบางกลุ่มเสียง

การอ่านสเปกโตรแกรมจะเริ่มจากแบ่งเสียงที่เปล่งออกมาเป็นส่วน ๆ จากการแบ่งช่วงเสียงจะได้ส่วนของหน่วยเสียงที่มีความเหมือนกันไว้ด้วยกันเป็นหน่วยเสียง สร้างขอบเขตของแต่ละส่วนโดยชี้ระบุด้วยเส้นในแกนตั้งดังตัวอย่างที่แสดงในรูปที่ 2.3 จากนั้นหาความสัมพันธ์ของแต่ละช่วงกับกลุ่มเสียงต่าง ๆ เช่น เสียงเสียดแทรก เสียงกัก เสียงเสียดแทรก เป็นต้น โดยอาศัยข้อมูลที่ปรากฏบนสเปกโตรแกรมควบคู่ไปกับการรู้ทางสัทศาสตร์เพื่ออธิบายลักษณะของหน่วยเสียงนั้น ๆ เช่น อาศัยความเป็นเสียงก้อง วิธีการออกเสียง และตำแหน่งของฐานกรณ์ ตัวอย่างเช่น เสียงพยัญชนะใช้การพิจารณาจากความเป็นเสียงก้อง ตำแหน่งของฐานกรณ์ อากาศของฐานกรณ์ ขณะที่เสียงสระแสดงถึงตำแหน่งความสูง ต่ำ หรือหน้า หลังของลิ้นและแสดงความกลมหรือการห่อริมฝีปาก เป็นต้น รูปที่ 2.3 แสดงตัวอย่างของหน่วยเสียงต่าง ๆ บนสเปกโตรแกรมของประโยค “ดร. ประสาร อ่านเปรื่อง” ซึ่งมีการแสดงสัญลักษณ์การอ่านตามการแสดงสัญลักษณ์การอ่านในฐานข้อมูล LOTUS คือ d-@-k^-3 t-q-q-z^-2|pr-a-z^-1 s-aa-n^-4|z-aa-n^-1 pr-vva-ng^-1| ซึ่งเปรียบเทียบหน่วยเสียงกับสัทอักษรสากล (International Phonetic Alphabet, IPA) [7] ในภาคผนวก ก



รูปที่ 2.3 ตัวอย่างของสเปกโตรแกรมของประโยค “คร. ประสาร อ่านเบรื่อง”

จากตัวอย่างที่แสดงดังรูปที่ 2.3 พบว่าจากการสังเกตลักษณะที่ปรากฏบนสเปกโตรแกรม สามารถบอกขอบเขตระหว่างหน่วยเสียงได้ในบางกลุ่มหน่วยเสียงแต่ยังไม่สามารถตัดสินใจตำแหน่งของขอบเขตได้แน่ชัดระหว่างขอบเขตบางหน่วยเสียง ซึ่งต้องใช้ลักษณะจากข้อมูลของสัญญาณเสียงมาช่วยพิจารณาเพิ่มเติม

## 2. ทฤษฎีที่ใช้ในการวิเคราะห์เสียงและหาค่าลักษณะสำคัญของเสียง

เป็นการศึกษาเกี่ยวกับการหาค่าลักษณะสำคัญของเสียง การวิเคราะห์ค่าลักษณะสำคัญของเสียงที่ได้เพื่อประโยชน์ในการนำไปใช้พัฒนาระบบ

### การหาค่าลักษณะสำคัญทางเสียง

การวัดหาค่าลักษณะสำคัญทางเสียง เป็นการหาตัวแทนของสัญญาณเสียงพูด โดยอาจจะแบ่งวิธีการพิจารณา แนวทางแรกคือแบ่งจากประโยชน์การนำไปใช้งาน เช่น การนำค่าลักษณะสำคัญทางเสียงเพื่อไประบุความเป็นเสียงก้อง แนวทางที่สองคือแบ่งค่าลักษณะสำคัญทางเสียงตามโดเมนที่กำลังพิจารณา เช่น สัญญาณเสียงที่พิจารณาแสดงอยู่ในโดเมนทางเวลา ซึ่งสามารถวัดค่าอัตสหสัมพันธ์ วัดค่าพลังงาน วัดค่าอัตราการตัดศูนย์ เป็นต้น ขณะที่สัญญาณเสียงที่พิจารณาแสดงอยู่ในรูปแบบทางความถี่ เช่น การค่าความถี่ฟอร์แมนท์ เป็นต้น ซึ่งในวิทยานิพนธ์นี้แบ่งตามแนวทางที่สอง

#### ก. วิธีการหาค่าลักษณะสำคัญทางเสียงตามโดเมนทางเวลา

วิธีการประมวลสัญญาณเสียงพูดโดยยึดการวิเคราะห์จากรูปคลื่นที่แปรตามแกนเวลามีหลายวิธี ซึ่งได้ ค่าลักษณะสำคัญทางเสียงที่พิจารณาในวิทยานิพนธ์นี้ ได้แก่ ค่าพลังงาน (Energy) ค่าพลังงานเมื่อแบ่งช่วงพลังงานต่าง ๆ ค่าอัตราการตัดศูนย์ (Zero-crossing rate) ค่าอัตสหสัมพันธ์ (Autocorrelation) เป็นต้น

### 1) การหาค่าพลังงาน (Energy)

ค่าลักษณะที่ใช้ในการวิเคราะห์เสียงพูดที่นิยมใช้กันแพร่หลาย คือ ค่าพลังงานของสัญญาณเสียงพูด เนื่องจากพลังงานของสัญญาณเสียงพูดเป็นค่าที่แสดงให้เห็นว่ามีสัญญาณเสียงเกิดขึ้น ณ เวลานั้นหรือไม่ อีกทั้งการคำนวณค่าพลังงานง่ายและรวดเร็ว การคำนวณค่าพลังงานจะทำที่ละกรอบเสียงพูด โดย  $E(m)$  คือค่าพลังงานของกรอบเสียงพูดที่จุด  $m$  และในแต่ละกรอบเสียงพูดจะทำที่ละกรอบเสียงพูดจำนวน  $n$  ตัวอย่าง ซึ่งใช้วิธีการคำนวณหาค่าพลังงานของสัญญาณ  $s[n]$  ใด ๆ ที่แปรตามเวลา ดังสมการ (2.1)

$$E = \sum_{n=-\infty}^{\infty} s^2[n] \quad (2.1)$$

เนื่องจากสัญญาณเสียงแปรเปลี่ยนอยู่ตลอดเวลา ไม่มีเสถียรภาพตามเวลา จึงพิจารณาสัญญาณเป็นช่วงเล็ก ๆ ตามแกนเวลาหรือแบ่งออกเป็นกรอบ ซึ่งในวิทยานิพนธ์นี้ใช้ขนาดกรอบประมาณ 20 มิลลิวินาที ดังนั้น การคำนวณพลังงานของเสียงในแต่ละกรอบ ตามสมการ (2.2)

$$E(m) = \sum_{n=0}^{N-1} [w(m)s(m-n)]^2 \quad (2.2)$$

โดยที่  $w(m)$  คือวินโดว์ฟังก์ชันที่ใช้กำหนดรูปร่างในการพิจารณาของสัญญาณเสียง  $s(n)$  ในหนึ่งกรอบ

$N$  คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ในกรอบหรือภายในกรอบของฟังก์ชันหน้าต่าง

### 2) อัตราการตัดผ่านศูนย์ (Zero-crossing rate)

อัตราการตัดผ่านศูนย์เป็นการวัดจำนวนครั้งของการเปลี่ยนเครื่องหมายทางคณิตศาสตร์ซึ่งเกิดจากการตัดผ่านแกนเวลาที่ระดับศูนย์ของสัญญาณเสียงพูด อัตราการตัดผ่านศูนย์เป็นเครื่องมืออย่างง่ายที่อธิบายความเป็นเสียงก้องหรือเสียงไม่ก้องเสียดแทรก (unvoiced fricative) ของสัญญาณเสียงพูด หากส่วนที่พิจารณาที่เป็นเสียงก้อง สัญญาณเสียงจะมีการเปลี่ยนแปลงเป็นรายแคบตัดผ่านแกนเวลา ทำให้อัตราการตัดศูนย์ต่ำขณะที่สัญญาณเสียงพูดแบบเสียงไม่ก้องเสียดแทรกมีอัตราการตัดผ่านศูนย์สูง การคำนวณอัตราการตัดผ่านศูนย์ดังสมการ

$$Z_n = \frac{1}{K} \sum_{i=1}^k \frac{|\text{sgn}\{S_n[i]\} - \text{sgn}\{S_n[i-1]\}|}{2} \quad (2.3)$$

เมื่อ  $\text{sgn}\{S_n[i]\} = \begin{cases} +1, & S_n[i] \geq 0 \\ -1, & S_n[i] < 0 \end{cases}$

โดยที่  $Z_n$  คืออัตราการตัดผ่านศูนย์ ที่กรอบเสียงพูด  $n$

$S_n[i]$  คือสัญญาณเสียงพูดที่  $i$  ในกรอบเสียงพูด  $n$

K คือความกว้างของกรอบเสียงพูด (จำนวนตัวอย่างเสียงพูดในแต่ละกรอบเสียงพูด)

3) ค่าวัดการเปลี่ยนแปลงรูปร่างของสเปกตรัม (Spectral transition measure: STM)

ค่าวัดการเปลี่ยนแปลงของสเปกตรัมนี้ได้ถูกเสนอเพื่อใช้ในบอกอัตราการเปลี่ยนแปลงของขนาดสเปกตรัม ดังนั้นตำแหน่งของขอบเขตหน่วยเสียงจะแสดงตำแหน่งที่จุดสูงสุดของค่าอัตราการเปลี่ยนแปลงของสเปกตรัม

$$STM(m) = (\sum_{i=1}^D a_i^2(m)) / D \quad (2.4)$$

โดยที่ D คือมิติของเวกเตอร์สเปกตรัม ในวิทยานิพนธ์นี้คือค่า 13

$a_i(m)$  คือค่าสัมประสิทธิ์ถดถอยหรือเป็นอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรอล ดังสมการที่ 2.4

m คือตำแหน่งเฟรมที่กำลังพิจารณา

$$a_i(m) = (\sum_{n=-I}^I MFCC(n+m) * n) / (\sum_{n=-I}^I n^2) \quad (2.5)$$

โดยที่ n คือตำแหน่งเฟรมซึ่งอ้างอิงเทียบกับกรอบที่กำลังพิจารณา

I คือจำนวนเฟรมที่ต้องการคำนวณค่าสัมประสิทธิ์ถดถอย เมื่ออ้างอิงกับตำแหน่งที่พิจารณาที่ตำแหน่งกึ่งกลางของกรอบพิจารณา ในวิทยานิพนธ์นี้ให้  $I=2$  ดังนั้นเมื่อใช้ขนาดเฟรม (frame step) 10 ms ทำให้มีช่วงระยะห่าง (interval) 40 ms

MFCC คือค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรัม

4) การคำนวณค่าพลังงานตามองค์ประกอบความถี่ต่าง ๆ

ค่าพลังงานตามองค์ประกอบความถี่ คำนวณได้จากผลรวมค่าความเข้มในช่วงองค์ประกอบความถี่ที่ปรากฏบนสเปกโตรแกรม ที่อยู่ในกรอบพิจารณาช่วงเล็กๆ ตามแกนเวลา ซึ่งพลังงานตามช่วงองค์ประกอบความถี่ต่าง ๆ แบ่งออกเป็น 7 ช่วงดังตารางที่ 2.3 โดยใช้สัญลักษณ์ F1 F2 และ F3 แทนความถี่ฟอร์แมนท์ที่หนึ่ง ความถี่ฟอร์แมนท์ที่สอง และความถี่ฟอร์แมนท์ที่สามตามลำดับ

ตารางที่ 2.3 ช่วงองค์ประกอบความถี่ สำหรับการสร้างพลังงานช่วงความถี่

ลำดับช่วง	ช่วงองค์ประกอบความถี่
1	[0, 400 เฮิรตซ์]
2	[F1, F2]
3	[F2, F3]
4	[F3, 5000 เฮิรตซ์]
5	[5000 เฮิรตซ์, 8000 เฮิรตซ์]
6	[0, 357 เฮิรตซ์]
7	[358 เฮิรตซ์, 5373เฮิรตซ์]

## 5) การคำนวณค่าอัตราส่วนพลังงาน

ค่าอัตราส่วนพลังงานที่เสนอในวิทยานิพนธ์นี้ คือ อัตราส่วนของพลังงานในองค์ประกอบความถี่ต่ำต่อพลังงานองค์ประกอบความถี่สูง เนื่องจากพลังงานในองค์ประกอบความถี่ต่ำของเสียงก้องจะมีค่ามากกว่าเสียงไม่ก้อง และพลังงานในองค์ประกอบความถี่สูงของเสียงก้องจะมีค่าน้อยกว่าเสียงไม่ก้อง ดังนั้นอัตราส่วนของพลังงานที่องค์ประกอบความถี่ต่ำต่อพลังงานที่องค์ประกอบความถี่สูงของเสียงก้องจึงมีค่ามากกว่าอัตราส่วนของพลังงานที่องค์ประกอบความถี่ต่ำต่อพลังงานที่องค์ประกอบความถี่สูงของเสียงไม่ก้อง

การคำนวณอัตราส่วนพลังงานสำหรับกลุ่มเสียงก้อง โซโนแรนท์ (sonorant) คำนวณจากอัตราส่วนพลังงานที่องค์ประกอบความถี่ช่วงที่ 6 และ 7 จากตารางที่ 2.3

## 6) การปรับค่าปกติลักษณะสำคัญ (Feature Normalization)

เนื่องจากต้องรวมหลายลักษณะสำคัญทางเสียงเข้าด้วยกัน ดังนั้นควรมีขั้นตอนการนอร์มอลไลซ์ (normalize) เพื่อให้ผลกระทบของน้ำหนักแต่ละลักษณะเท่าเทียมกัน โดยให้ค่าที่ทำนอร์มอลไลซ์ อยู่ในช่วง [0,1] ซึ่งในวิทยานิพนธ์นี้ใช้การคำนวณค่าปกติลักษณะสำคัญเทียบกับค่ามากที่สุดของค่าลักษณะสำคัญนั้นบนสัญญาณเสียงที่กำลังพิจารณา ดังสมการที่ 2.6

$$F_{normal_i} = F_i / F_{max} \quad (2.6)$$

โดยที่  $F_{normal}$  แทนค่าลักษณะสำคัญของตำแหน่งเฟรมที่กำลังพิจารณาที่ปรับให้เป็นค่าปกติลักษณะสำคัญนั้น

$i$  แทนตำแหน่งเฟรมที่กำลังพิจารณา

$F_{max}$  แทนค่าที่มากที่สุดของค่าลักษณะสำคัญบนสัญญาณเสียงที่กำลังพิจารณา



## ข. วิธีการหาค่าลักษณะสำคัญทางเสียงตามโดเมนทางความถี่ ได้แก่

### 1) การหาค่าความถี่ฟอร์แมนท์

ความถี่ฟอร์แมนท์สามารถคำนวณจากรากของพหุนามของสัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Predictive Coding, LPC) เนื่องจากแบบจำลองการประมาณพันธะเชิงเส้นเป็นแบบจำลองที่เสถียรภาพ ค่าโพลของฟังก์ชันถ่ายโอนจะอยู่ในวงกลมหนึ่งหน่วย เมื่อกำหนดให้  $z = \text{Re}(z) + j \text{Im}(z)$  เป็นรากของพหุนามสัมประสิทธิ์การประมาณพันธะเชิงเส้น[8] ความถี่ฟอร์แมนท์จะคำนวณได้จาก

$$F = \left( \frac{f_s}{2\pi} \right) \arctan \left( \frac{\text{Im}(z)}{\text{Re}(z)} \right) \quad (2.7)$$

โดยที่  $f_s$  เป็นค่าความถี่สุ่มตัวอย่าง (Sampling Frequency) ในการคำนวณความถี่ฟอร์แมนท์จะใช้รากของพหุนามสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่อยู่ในซีกบนของวงกลมหนึ่งหน่วยเท่านั้น

การวิเคราะห์จากรากค่าความถี่ฟอร์แมนท์ เป็นการพิจารณาจากการเคลื่อนที่ของความถี่ฟอร์แมนท์ (formant transition) เนื่องจากความถี่ฟอร์แมนท์ปรากฏเป็นเส้นต่อเนื่องในส่วนที่เป็นเสียงก้อง เช่น เส้นการเคลื่อนที่ของความถี่ฟอร์แมนท์ที่ปรากฏในเสียงสระจะมีการเคลื่อนที่ไม่เปลี่ยนแปลงความชันมากนัก หรือหากพิจารณาเสียงสระประกอบกับเสียงกึ่งสระ พบว่าเส้นการเคลื่อนที่ความถี่ฟอร์แมนท์ที่ต่อเนื่องจะค่อย ๆ มีการเปลี่ยนตำแหน่ง ไปจากตำแหน่งเริ่มต้นในสระเดิมไปสู่จุดสิ้นสุดของเสียงกึ่งสระ เป็นต้น ดังนั้นเส้นที่แสดงเส้นการเคลื่อนที่มีความชันของเส้นขึ้นกับชนิดหน่วยเสียงที่พิจารณา โดยเกิดจากจุดเริ่มต้นของตำแหน่งของแรกไปสู่ตำแหน่งสิ้นสุดหน่วยเสียงในเสียงถัดไป การวิเคราะห์ด้วยค่าความถี่ฟอร์แมนท์นี้นำไปใช้เป็นค่าลักษณะสำคัญของหาตำแหน่งเริ่มต้นของเสียงสระ ซึ่งเป็นประโยชน์ของการปรับตำแหน่งขอบเขตหน่วยที่มีสระตามมาได้

### 2) ค่าสัมประสิทธิ์เมลฟรีควีนซีเคปสตรัม

ค่าสัมประสิทธิ์เมลฟรีควีนซีเคปสตรัม (Mel frequency cepstral coefficients: MFCC) เป็นค่าลักษณะทางเสียงที่นิยมมากในการรู้จำเสียงพูดอัตโนมัติ ซึ่งมีวิธีการคำนวณเริ่มต้นจากการหาสเปกตรัมของสัญญาณเสียงโดยการแปลงฟูริเยร์แบบไม่ต่อเนื่อง (Discrete Fourier transformation) หรือการแปลงฟูริเยร์แบบเร็ว (Fast Fourier transformation) ซึ่งสเปกตรัมของสัญญาณเสียงเกิดจากส่วนประกอบ 2 ส่วนคือ เอนVELOPของสเปกตรัม (Spectral envelop) และโครงสร้างรายละเอียดของสเปกตรัม (Spectral fine structure) ซึ่งสัมประสิทธิ์เคปสตรัมเป็นการ

แทนสัญญาณในส่วนเอนเวโลปของสเปกตรัม โดยเคลปสตรัมหาได้จากการนำสเปกตรัมของสัญญาณเสียงผ่านตัวกรอง (Filter bank) ซึ่งกระจายตามสเกลเมล (Mel scale) ซึ่งเป็นการกระจายอยู่บนสเกลความถี่แบบไม่สม่ำเสมอซึ่งออกแบบให้เหมาะสมกับการรับฟังของหูของมนุษย์ การกระจายความถี่บนสเกลเมลคำนวณจากสมการที่ (2.8)

$$Mel = 1000 \log_2 \left( \frac{f}{1000} + 1 \right) \quad (2.8)$$

ดังนั้นจึงหาค่าพลังงานเสียง  $S_k$  สำหรับตัวกรองที่  $k$  จากนั้นหาค่าลอการิทึม (Logarithm) ของพลังงาน  $S_k$  แล้วนำของค่าลอการิทึมของพลังงานมาผ่านการแปลงโคไซน์แบบไม่ต่อเนื่อง (Discrete cosine transformation) จะได้ค่า MFCC ลำดับที่  $n$  ซึ่งแสดงในสมการที่ (2.9)

$$MFCC(n) = \sum_{k=1}^K \log(S_k) \cos\left(n(k - 0.5) \frac{\pi}{k}\right) \quad (2.9)$$

## เสียงในภาษาไทย

ทฤษฎีทางด้านภาษาศาสตร์ได้กำหนดหน่วยเสียงในภาษาไทย มีรายละเอียดดังนี้

เสียงภาษาไทยประกอบด้วย หน่วยเสียงสระมีทั้งหมด 21 หน่วยเสียง หน่วยเสียงพยัญชนะมีเพียง 21 หน่วยเสียงจากตัวอักษรภาษาไทยทั้งหมด 44 ตัว และเสียงวรรณยุกต์ 5 ระดับที่ทำให้เสียงในเสียงในภาษาไทยเป็นเสียงที่มีระดับเสียงสูงต่ำในคำ หากมีเสียงวรรณยุกต์แตกต่างกันจะส่งผลให้คำมีความหมายแตกต่างกัน สำหรับวิทยานิพนธ์นี้เน้นความแตกต่างของหน่วยเสียงจึงแบ่งออกเป็นกลุ่มเสียงเป็น 9 กลุ่ม พร้อมทั้งอธิบายการเกิดกลุ่มเสียง และลักษณะเด่นของแต่ละกลุ่มเสียง ดังนี้

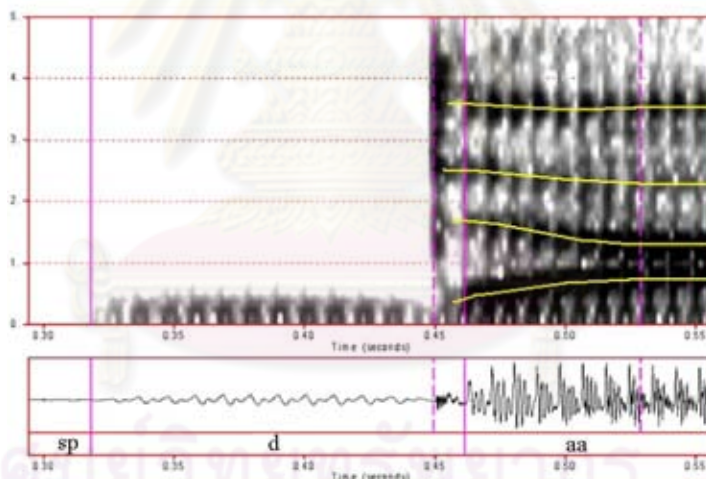
**1. เสียงกัก (Stops)** ฐานกรณ์ปิดสนิท มีการปิดกั้นลมไว้ในปากและจุดปิดกั้นอยู่ที่ริมฝีปาก หรือภายในปาก ส่งผลให้เกิดช่องว่างของเสียงเงียบ จากนั้นลมออกตรงกลางปากส่งผลให้เกิดเสียงระเบิดของความไม่เป็นคาบ เกิดเสียงรบกวนคล้ายกับกลุ่มเสียงเสียดแทรก

ข้อสังเกตลักษณะที่เด่นของกลุ่มนี้ คือ

- 1) การเคลื่อนที่ของความถี่ฟอร์แมนท์ (formant transition) โดยพิจารณาการเคลื่อนที่จากส่วนที่เป็นเสียงก้องของสระที่อยู่ตำแหน่งก่อนหน้ามายังเสียงกัก หรือเกิดจากเสียงกักแล้วเคลื่อนที่ไปยังเสียงสระที่ตามมา โดยพิจารณาความถี่ฟอร์แมนท์ที่ปรากฏว่ามีรูปแบบแน่ชัดหรือไม่
- 2) เนื่องจากแถบความเข้มของเสียงก้อง (Voice bar) ซึ่งให้เห็นการสั่นสะเทือนของเส้นเสียงเกิดขึ้นระหว่างการปิดฐานกรณ์หรือกักลม ซึ่งแถบเข้มที่แสดงความเป็นเสียงก้องปรากฏ

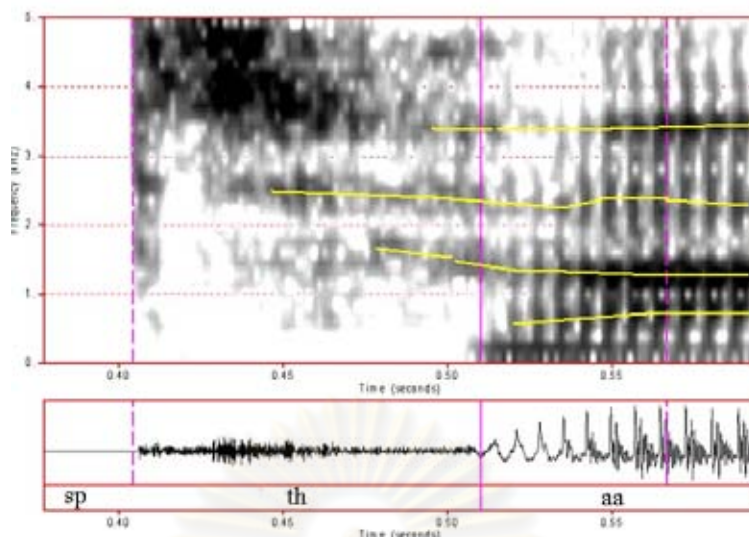
ในสเปกโตรแกรมที่บริเวณความถี่ต่ำเมื่อพิจารณาเสียงกักประเภทเสียงก้อง ขณะที่เสียงไม่ก้องจะไม่พบแถบนี้

- 3) เนื่องจากในเสียงกักมีการเกิดแรงดันลมที่ถูกกัก จึงมีช่องว่างของเสียงเงียบเกิดขึ้น
- 4) การปล่อยเสียงระเบิดมีลักษณะเป็นเส้นในแนวตั้งในช่วงความถี่สูง การเกิดเสียงระเบิดของช่วงเสียงไม่เป็นคาบหรือเสียงรบกวนจะตามมาจากช่องว่างเสียงเงียบ พบทั่วไปในตำแหน่งเริ่มต้นหรือกลางของเสียงกัก โดยเสียงกักประเภทเสียงไม่ก้องมากกว่าเสียงก้องเนื่องจากคุณสมบัติการออกเสียงชนิดของเสียงไม่ก้อง
- 5) เวลาเริ่มต้นช่วงเสียงก้อง (VOT) เสียงกักประเภทเสียงไม่ก้องมีช่วงเวลา 25 ถึง 100 มิลลิวินาที นานกว่าเสียงกักประเภทเสียงก้องประมาณ -20 ถึง 20 มิลลิวินาที
- 6) เสียงกักที่พ่นลม (Aspirated) สังเกตจากเสียงรบกวนในความถี่สูงที่ปรากฏกระจัดกระจายในองค์ประกอบความถี่สูงเกิดจากลมที่ถูกกักถูกปล่อยออกมา
- 7) เสียงกักไม่พ่นลม (unreleased) เกิดจากลมที่ถูกกัก ณ จุดใดจุดหนึ่งในช่องปากเท่านั้น แทนที่จะระเบิดออกมาก็กลับถูกกลืนลงไปเพราะจุดที่กักลมไว้ไม่เปิดออก จึงไม่มีเสียงออกมาเลย



รูปที่ 2.4 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ดา”

จุฬาลงกรณ์มหาวิทยาลัย



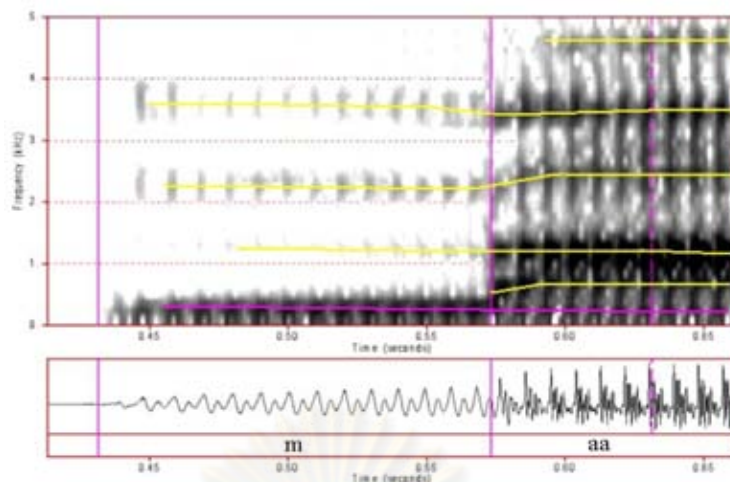
รูปที่ 2.5 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ทา”

จากรูปทั้งสองข้างต้นพบว่า มีช่วงเสียงเงียบปรากฏก่อนที่จะเกิดการปล่อยลมออกมา ซึ่งสังเกตได้จากช่วงที่ไม่มีพลังงานปรากฏออกมา เมื่อลมถูกปล่อยออกมา จะพบว่ารูปที่ 2.4 มีแถบความเข้มที่องค์ประกอบความถี่ต่ำปรากฏในสเปกโตรแกรมเมื่อออกเสียง ค /d/ ขณะที่ในรูปที่ 2.5 ที่แสดงสเปกโตรแกรมของเสียง ท /th/ ไม่มีแถบความเข้มของเสียงก้องปรากฏที่องค์ประกอบความถี่ต่ำเนื่องจากการพ่นลมออกมาส่งผลให้เกิดเส้นในแนวตั้งในช่วงองค์ประกอบความถี่สูง พบว่าช่วงเวลาที่เกิดในเสียง ท /th/ ซึ่งเป็นเสียงไม่ก้อง จะมีช่วงนานกว่าเสียง ค /d/ ซึ่งเป็นเสียงก้อง

การวิเคราะห์ในเสียงกักสามารถพิจารณาการค้นหาคำว่าเป็นเสียงก้องในกลุ่มเสียงกักมีตัวแปรสำคัญจากข้อมูลที่ได้จากการสังเกตจาก ช่วงเวลาก่อนเกิดเสียงก้อง (Prevoicing) หมายถึง ช่วงของเสียงก้องระหว่างส่วนที่มีการปิดกั้น (closure) ของเสียงกักก่อนที่จะเกิดการระเบิดของเสียงกักออกมา ซึ่งโดยทั่วไปจะพบในพลังงานที่องค์ประกอบความถี่ต่ำ ดังตัวอย่างในรูปที่ 2.4 ขณะที่รูปที่ 2.5 ไม่พบช่วงพลังงานดังกล่าว นอกจากนี้ ช่วงเวลาเริ่มต้นเสียงที่บอกระยะเวลาของช่วงที่เกิดการระเบิดเสียงกักเมื่อปล่อยลมออกมาถึงจุดเริ่มต้นของเสียงสระที่ตามมา เป็นต้น

2. เสียงนาสิก (Nasal) ฐานกรณปิดสนิท มีการปิดกั้นลมไว้ในปากหรือก้นที่ริมฝีปาก เพดานอ่อนห้อยลง วัลัมเปิด ลมจะถูกส่งออกผ่านโพรงจมูก เส้นเสียงสั้นเกิดเป็นเสียงก้อง ลักษณะเด่นของเสียงนาสิกคือความถี่ฟอร์แมนที่หนึ่งมีค่าต่ำและปรากฏไม่ชัดเจน มีพลังงานน้อยมาก หรือ ไม่มีพลังงานในช่วงของความถี่ฟอร์แมนที่สอง

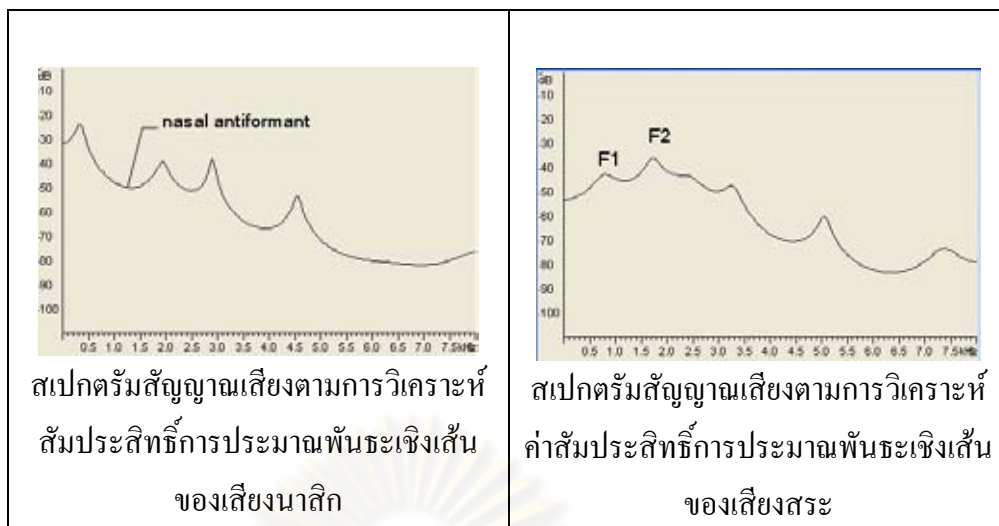




รูปที่ 2.6 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “มา”

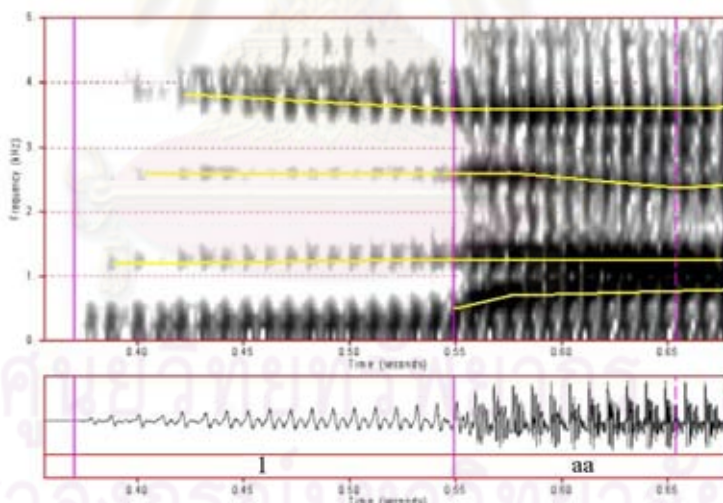
เสียงนาสิกมีความถี่เรโซแนนซ์ ในช่วง 200-400 เฮิรตซ์ นั้นจะพบว่ารูปแบบของความถี่ฟอร์แมนท์เกิดขึ้นคล้ายกับในเสียงสระ และเนื่องจากเกิดความถี่แอนไทฟอร์แมนท์ (antiformant) ในช่วงที่สูงกว่าความถี่ฟอร์แมนท์ที่สองจึงทำให้ไม่พบความถี่ฟอร์แมนท์ที่สองปรากฏในเสียงนาสิก ซึ่งทำให้เกิดบริเวณที่เป็นช่วงไม่ต่อเนื่อง (discontinuity) ระหว่างเสียงนาสิกและเสียงสระที่อยู่ติดกัน จากรูปที่ 2.6 ที่แสดงสเปกโตรแกรมของเสียง ม /m/ ที่ติดกับเสียงสระ ออ /aa/ สามารถหาตำแหน่งขอบเขตของหน่วยเสียงได้จากบริเวณที่ไม่ต่อเนื่องจากความถี่ฟอร์แมนท์ นอกจากนี้สามารถพิจารณาจากสเปกตรัมสัญญาณเสียงตามการวิเคราะห์จากค่าสัมประสิทธิ์การประมาณพหุเชิงเส้นที่แตกต่างกันของหน่วยเสียงนาสิกและเสียงสระด้วยเครื่องมือวิเคราะห์ wavesufer [9] พบว่า สเปกตรัมสัญญาณเสียงเมื่อวิเคราะห์ตามค่าสัมประสิทธิ์การประมาณพหุเชิงเส้นแตกต่างกันเนื่องจากความแตกต่างของความถี่ฟอร์แมนท์ที่หนึ่งและความถี่ฟอร์แมนท์ที่สองของเสียงนาสิกกับเสียงสระแตกต่างกัน ดังแสดงตามรูปที่ 2.7 ซึ่งแสดงตำแหน่งของความถี่ฟอร์แมนท์ที่หนึ่ง (F1) และความถี่ฟอร์แมนท์ที่สอง (F2) ปรากฏชัดเจนบนรูปร่างสัมประสิทธิ์การประมาณพหุเชิงเส้นของหน่วยเสียงสระ ขณะที่รูปร่างสัมประสิทธิ์การประมาณพหุเชิงเส้นของหน่วยเสียงนาสิกไม่มีความถี่ฟอร์แมนท์ที่องค์ประกอบความถี่สูง เกิดเป็นความถี่แอนไทฟอร์แมนท์





รูปที่ 2.7 ตัวอย่างสเปกตรัมจากการวิเคราะห์ค่าสัมประสิทธิ์การประมาณฟังก์ชันเชิงเส้นของหน่วยเสียงนาสิกกับหน่วยเสียงสระ

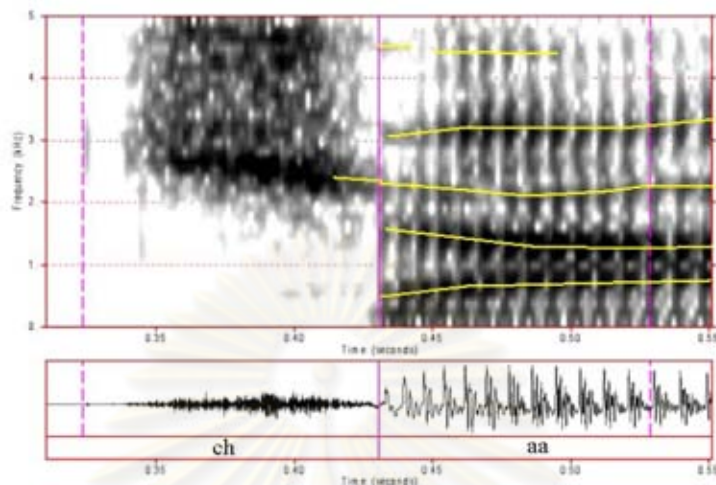
3. เสียงข้างลิ้น (**lateral**) ฐานกรณปิดสนิท มีการปิดกั้นลมไว้ในปาก และจุดปิดกั้นอยู่ที่ริมฝีปากหรือภายในปาก ลมออกข้างลิ้นอาจจะออกข้างเดียวหรือสองข้างก็ได้ เพดานอ่อนยกขึ้นปิดไว้ล้ม แตกต่างจากเสียงนาสิก เมื่อสังเกตการเคลื่อนที่ของความถี่ฟอร์แมนท์ ระหว่างความถี่ 1,200 เฮิรตซ์



รูปที่ 2.8 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ลา”

เนื่องจากเสียงข้างลิ้นนี้ เกิดจากการประกบกันของท่อเรโซแนนซ์ขนานกันสองท่อ โดยเกิดที่ข้างของลิ้นและอีกอันที่ได้ลิ้น ทำให้เกิดแอนติเรโซแนนซ์แสดงให้เห็นความถี่ฟอร์แมนท์ที่บาง ๆ ไม่ชัดเจน ดังปรากฏในรูปที่ 2.8 ในส่วนของเสียง ล /l/ ซึ่งจะเห็นว่ามีความถี่ฟอร์แมนท์ที่หนึ่งและความถี่ฟอร์แมนท์ที่สองที่ปรากฏชัดเจนขณะที่ความถี่ฟอร์แมนท์ที่สองและความถี่ฟอร์แมนท์ที่สามเบาบาง

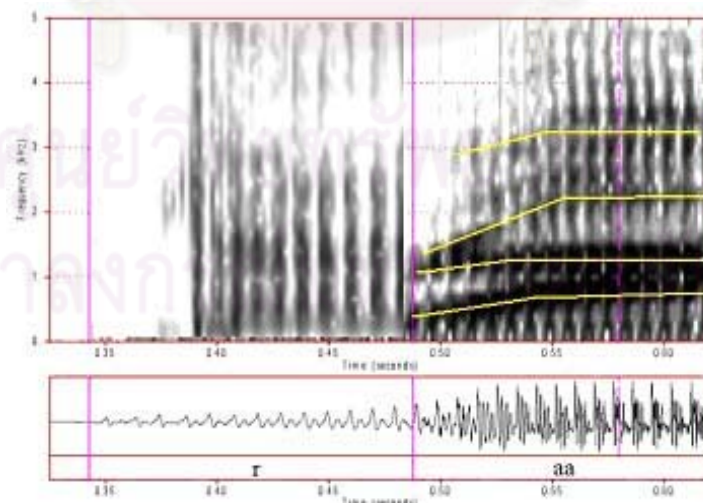
4. เสียงกึ่งเสียดแทรก (affricate) ฐานกรณ์ปิดสนิท เริ่มต้นด้วยการปิดกั้นลมไว้ในปากอย่างสนิท เช่นเดียวกับเสียงกัก แทนที่จะปล่อยให้ฟ่นออกมาทันทีแบบเสียงกัก อวัยวะในการออกเสียงจะเปิดออกช้า ๆ แล้วตามด้วยเสียงเสียดแทรก เพดานอ่อนยกขึ้นปิดช่องวิลิ้ม



รูปที่ 2.9 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ชา”

จากรูปที่ 2.9 แสดงเสียง ช /ch/ พบว่าลักษณะที่ปรากฏในสเปกโตรแกรมนั้นคล้ายกับเสียงเสียดแทรกที่แสดงลักษณะเป็นเสียงรบกวน โดยมีแถบเข้มปรากฏกระจายระจัดกระจายทั่วไป ไม่มีรูปแบบที่แน่ชัด แต่เนื่องจากการเกิดเสียงกึ่งเสียดแทรกคล้ายกับเสียงกัก จึงพบช่วงที่เป็นการระเบิดและหน่วยเสียงกึ่งเสียดแทรกมีการฟ่นลมออกมาเช่นกัน

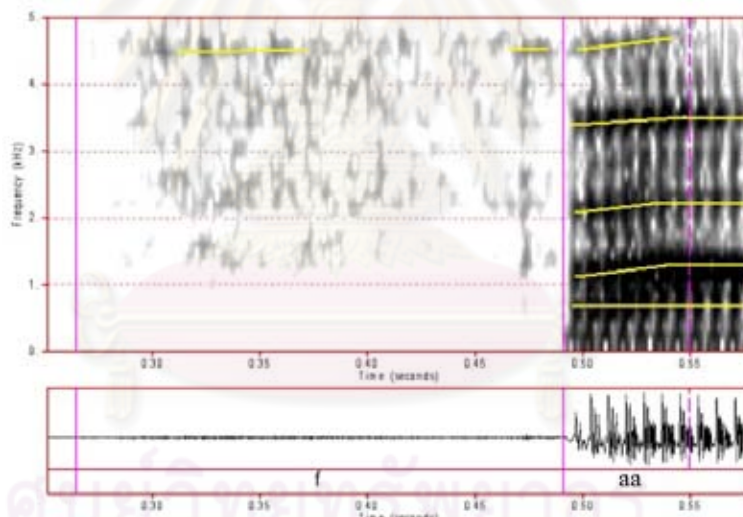
5. เสียงร้ว (trill) ฐานกรณ์เปิดปิดระยะสั้น เกิดจากปลายลิ้นกระดกขึ้นไปแตะปุ่มเหงือกอย่างรวดเร็ว และหลายครั้งจนได้ยินเสียงร้ว



รูปที่ 2.10 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “รา”

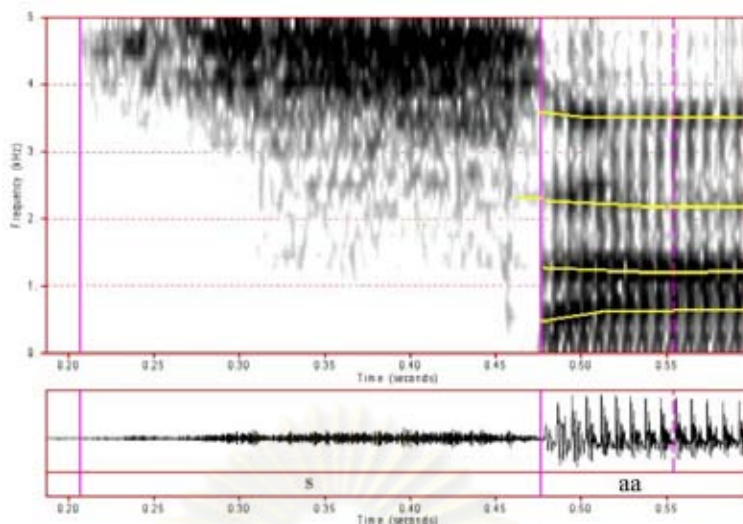
ตัวอย่างของเสียง ร /r/ ที่ปรากฏในรูปที่ 2.10 พบว่ามีการเปลี่ยนแปลงของแถบเส้นในแนวตั้งตลอดซึ่งเกิดจากการกระดกคลื่นไปมา ทำให้เกิดเป็นเสียงรัว หากพิจารณาลักษณะในช่วงของหน่วยเสียง ร นี้ พบว่ามีการเปลี่ยนแปลงอย่างอย่างรวดเร็ว ตลอดทั้งหน่วยเสียง และเมื่อสังเกตเส้นการเคลื่อนที่ความถี่ฟอร์แมนท์ที่สามเมื่อเปลี่ยนเข้าสู่เสียงสระที่ตามมานั้นมีความชันและความชันของเส้นการเคลื่อนที่ความถี่ฟอร์แมนท์ที่สี่มีความชันคล้ายกัน แต่จะชันน้อยกว่าความถี่ฟอร์แมนท์ที่สาม

6. เสียงเสียดแทรก (fricative) ฐานกรณ์เบียดชิดกัน เกิดจากการบีบตัวของช่องเสียงซึ่งแตกต่างกับเสียงกักที่เกิดจากการปิดกั้นไม่สมบูรณ์ในช่องเสียง ลมที่พ่นผ่านช่องเปิดแคบเล็ก ๆ ทำให้เกิดเป็นเสียงรบกวน ดังนั้นกลุ่มเสียงเสียดแทรกจึงมีลักษณะเด่นคือ คลื่นเสียงไม่มีรูปร่างเป็นระเบียบไม่มีการซ้ำเดิม ไม่เกิดขึ้นเป็นรายคาบ มีพลังงานเกิดขึ้นในช่วงองค์ประกอบความถี่แบบสุ่ม กระจายทั่วไป ไม่มีรูปแบบแน่ชัด กลุ่มเสียงเสียดแทรกเกิดได้จากสองวิธี ได้แก่ เกิดการเสียดแทรกกันระหว่างลมกับช่องทางผ่านแคบเล็ก ๆ คล้ายเสียง ฟ /f/ และเกิดได้จากลมกระจายผ่านช่องแคบและถูกบังคับให้ผ่านข้ามไปอีกที่อื่น เช่น กระแสลมที่ผ่านฟัน ซึ่งเกิดในเสียง ส /s/ เป็นต้น



รูปที่ 2.11 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “ฟา”

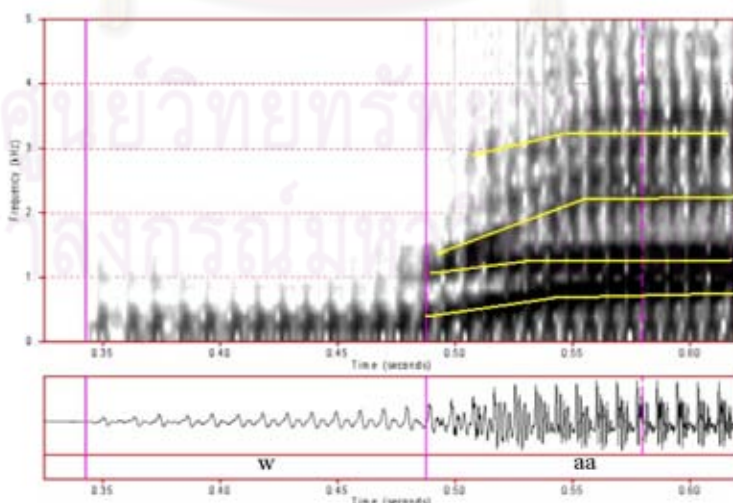
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.12 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “สา”

จากรูปที่ 2.11 และ 2.12 แสดงสเปกโตรแกรมของหน่วยเสียง ฟ /f/ กับ สระอา /aa/ และ แสดงสเปกโตรแกรมของหน่วยเสียง ส /s/ กับ สระอา /aa/ ตามลำดับ องค์ประกอบความถี่ที่ปรากฏบนสเปกโตรแกรมกระจายตัวทั่วไป ไม่มีรูปแบบแน่นอน ความเข้มเสียงแสดงแอมพลิจูดที่ปรากฏในองค์ประกอบความถี่สูงของเสียง ส /s/ มีความเข้มมากกว่าเสียง ฟ /f/ เพราะเกิดจากวิธีการเกิดเสียงเสียดแทรกที่ต่างกัน

7. เสียงกึ่งสระ (semivowel) ฐานกรณ์เปิด เสียงที่เกิดจากช่องว่างระหว่างฐานกรณ์ทั้งสองเปิดกว้างให้ลมผ่านสะดวกไม่มีการปิดกั้นลมหรือบังคับให้ลมแทรกออกมาตามช่องแคบ ๆ ในลักษณะเสียดสี ลักษณะเด่นคือ การเคลื่อนที่ของความถี่ฟอร์แมนที่น้อยกว่าเสียงสระเพราะมีการปิดบางส่วนในช่องเสียง



รูปที่ 2.13 ตัวอย่างของสเปกโตรแกรม (ภาพบน) และรูปคลื่น (ภาพล่าง) ของคำว่า “วา”



ตัวอย่างเสียง ว /w/ ที่แสดงในรูปที่ 2.13 เป็นเสียงกึ่งสระมีลักษณะคล้ายกับเสียงสระที่ลิ้นห่อ และตำแหน่งลิ้นอยู่ที่ตำแหน่งสูงก่อนไปทางด้านหลัง เมื่อสังเกตเส้นการเคลื่อนที่ของความถี่ฟอร์แมนทที่เริ่มจากการเคลื่อนที่จากส่วนเสียงกึ่งสระไปยังเสียงสระที่ตามมาจะพบว่า ค่อย ๆ มีการเปลี่ยนแปลงไป ไม่เกิดการเปลี่ยนแปลงทันที

8. **เสียงเงียบ (silence /short pause)** เกิดขึ้นได้สองกรณี คือ เกิดจากการหยุดออกเสียงระหว่างพยางค์ หรือเกิดจากการกักของลม ในเสียงกัก ทำให้ไม่ปรากฏความเข้มที่แสดงแอมพลิจูดบนสเปกโตรแกรม ไม่มีพลังงานปรากฏในช่วงหน่วยเสียงนี้หรือมีพลังงานน้อยมาก

9. **เสียงสระ (vowel)** จากรูปตัวอย่างของหน่วยเสียงต่าง ๆ พบว่าในสเปกโตรแกรมมีการเคลื่อนที่ของความถี่ฟอร์แมนทที่ปรากฏชัดเจน มีเส้นแนวนอนที่เข้มชัดซึ่งเป็นผลมาจากพลังงานที่ความถี่ฮาร์โมนิก และสามารถจำแนกชนิดเสียงสระเดี่ยวได้จากตำแหน่งความถี่ฟอร์แมนท ที่หนึ่ง ที่สอง และที่สาม สำหรับสระเสียงประสม มีการเคลื่อนที่ความถี่ฟอร์แมนทจากเสียงสระเสียงแรกไปยังเสียงกึ่งสระหรือเสียงสระอีกเสียงที่ตามมา มีการเปลี่ยนตำแหน่งของความถี่ฟอร์แมนท คล้ายกับการเคลื่อนที่ของลิ้นในปาก การเคลื่อนที่ความถี่ฟอร์แมนทไม่เป็นเส้นแนวนอนนิ่งตลอดทั้งเส้นเหมือนกับเสียงสระ

ตารางที่ 2.2 ที่ได้แสดงกลุ่มเสียง ชื่อของกลุ่มเสียงย่อ สัญลักษณ์แทนหน่วยเสียงที่ใช้ และตัวอย่างของแต่ละกลุ่มเสียง

ตารางที่ 2.4 สรุปหน่วยเสียงและสัญลักษณ์แทนหน่วยเสียงที่เปรียบเทียบกับหน่วยเสียงกับสัทอักษรสากล (International Phonetic Alphabet, IPA) [7]

กลุ่มเสียง	ชื่อ	หน่วยเสียง	สัทอักษรสากล	ตัวอย่างในภาษาไทย
เสียงกัก	เสียงกัก ฐานริมฝีปาก ชนิด อโฆมะ (voiceless aspirated bilabial stop)	/ph/	[ph]	พ ภ ผ
	เสียงกัก ฐานริมฝีปาก สติล อโฆมะ (voiceless unaspirated bilabial stop)	/p/	[p]	ป และ ตัวสะกดแม่ กบ
	เสียงกัก ฐานปุ่มเหงือก ชนิด อโฆมะ (voiceless aspirated apico alveolar stop)	/th/	[th]	ธ ท ผ ฑ ฐ ถ



กลุ่มเสียง	ชื่อ	หน่วยเสียง	สัทอักษรสากล	ตัวอย่างในภาษาไทย
	เสียงกัก ฐานปุ่มเหงือก สอดคล้อง อโฆษะ (voiceless unaspirated apico alveolar stop)	/t/	[t]	ต ฏ และ ตัวสะกดแม่ กด
	เสียงกัก ฐานเพดานอ่อน ชนิด อโฆษะ (voiceless aspirated dorso velar stop)	/kh/	[kh]	ข ค ฃ
	เสียงกัก ฐานเพดานอ่อน สอดคล้อง อโฆษะ (voiceless unaspirated dorso velar stop)	/k/	[k]	ก และ ตัวสะกดแม่ กด
	เสียงกัก ฐานริมฝีปาก สอดคล้อง โฆษะ (voiced bilabial stop)	/b/	[b]	บ
	เสียงกัก ฐานปุ่มเหงือก สอดคล้อง โฆษะ (voiced apico alveolar stop)	/d/	[d]	ด ฎ
	เสียงกัก ฐานเส้นเสียง (voiceless glottal stop)	/z/	[z]	อ
เสียงนาสิก	เสียงนาสิก ฐานริมฝีปาก (voiced bilabial nasal)	/m/	[m]	ม และ ตัวสะกดแม่ กม
	เสียงนาสิก ฐานปุ่มเหงือก (voiced apico alveolar nasal)	/n/	[n]	น และ ตัวสะกดแม่ กน
	เสียงนาสิก ฐานเพดานอ่อน (voiced dorso velar nasal)	/ng/	[ng]	ง และ ตัวสะกดแม่ กง
เสียงข้างลิ้น	เสียงเปิดข้างลิ้น ฐานปุ่มเหงือก (voiced apico alveolar lateral)	/l/	[l]	ล ฟ หล

กลุ่มเสียง	ชื่อ	หน่วยเสียง	สัทอักษรสากล	ตัวอย่างในภาษาไทย
เสียงกึ่งเสียดแทรก	เสียงกึ่งเสียดแทรก ฐานเพดานแข็ง ลิติด อโฆษะ (voiceless unaspirated laminal palatal affricate)	/c/	[c]	จ
	เสียงเสียดแทรก ฐานเพดานแข็ง ฐนิต อโฆษะ (voiceless aspirated laminal palatal affricate)	/ch/	[ch]	ช ฌ ฉ
เสียงร้ว	เสียงลิ้นร้ว ฐานปุ่มเหงือก (voiced apico alveolar trill)	/r/	[r]	ร
เสียงเสียดแทรก	เสียงเสียดแทรก ฐานปุ่มเหงือก (voiceless apico alveolar fricative)	/s/	[s]	ซ ส ศ ษ
	เสียงเสียดแทรก ฐานเส้นเสียง (voiceless glottal fricative)	/h/	[h]	ฮ ห
	เสียงเสียดแทรก ฐานริมฝีปากกับฟัน (voiceless labio dental fricative)	/f/	[f]	ฟ ฟ
เสียงกึ่งสระ	เสียงเปิดกึ่งสระ ฐานเพดานอ่อน (voiced fronto palatal approximant)	/j/	[j]	ย ญ และ ตัวสะกดแม่ เกย
	เสียงเปิดกึ่งสระ ฐานริมฝีปาก-เพดาน อ่อน (voiced labio velar approximant)	/w/	[w]	ว และ ตัวสะกดแม่ เกว

เสียงสระ (ซึ่งแสดงสัทอักษรสากล IPA ในภาคผนวก ก)

หน่วยเสียง	สระในภาษาไทย
/i/ /e/ /ɛ/ /ɜ/ /o/ /ɔ/ /a/ /u/ /o/ /@/	เสียงเดี่ยวสั้น (อิ, เอะ, แอะ, อี้, อะ, เออะ, อุ, โอะ, เออะ)
/ii/ /ee/ /xx/ /vv/ /qq/ /aa/ /uu/ /oo/ /@@/	เสียงเดี่ยวยาว (อี, เอ, แอ, อี้, เออ, อา, อุ, โอ, ออ)
/ia/ /iia/ /va/ /vva/ /ua/ /uua/	เสียงสระประสม (เอียะ, เอีย, เอือะ, เอือ, อัวะ, อิว)

## ทฤษฎีทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล

### 1. การวิเคราะห์ความแปรปรวน (Analysis of variance: ANOVA) [10]

การวิเคราะห์ความแปรปรวนเป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงกลุ่มเพื่อทดสอบความแตกต่างของข้อมูลที่ได้รับปัจจัยที่ต่างระดับกัน ซึ่งในวิทยานิพนธ์ หมายถึง การพิจารณาความแปรปรวนในค่าการเปลี่ยนแปลงลักษณะทางเสียง เพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยในค่าการเปลี่ยนแปลงลักษณะทางเสียงระหว่างสองกลุ่ม(นั่นคือ ระหว่าง กลุ่มตำแหน่งขอบเขตเสียงและกลุ่มที่ไม่ใช่ตำแหน่งขอบเขตเสียง) ซึ่งใช้การวิเคราะห์จากรูปภาพของแผนภาพกล่อง และตรวจสอบเงื่อนไขว่า ความแตกต่างระหว่างค่าเฉลี่ยระหว่างกลุ่ม ซึ่งประเมินความแตกต่างกันระหว่างกลุ่มว่าแตกต่างกันมากน้อยเท่าไร ซึ่งการตรวจสอบเงื่อนไขการพิจารณาสมมติฐาน โดยที่  $\mu_k$  คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่  $k$  (เมื่อ  $k = 1$  คือกลุ่มที่เป็นตำแหน่งขอบเขตเสียง และ  $k = 2$  คือกลุ่มที่ไม่ใช่ตำแหน่งขอบเขตเสียง)

สมมติฐานที่  $H_0 : \mu_1 = \mu_2$  (หมายถึง ค่าเฉลี่ยของกลุ่มตำแหน่งขอบเขตเสียงเท่ากับค่าเฉลี่ยของกลุ่มไม่ใช่ตำแหน่งขอบเขตเสียง)

สมมติฐานที่  $H_1 : \mu_1 \neq \mu_2$  (หมายถึง ค่าเฉลี่ยของกลุ่มตำแหน่งขอบเขตเสียงไม่เท่ากับค่าเฉลี่ยของกลุ่มไม่ใช่ตำแหน่งขอบเขตเสียง)

ถ้าผลการทดสอบสรุปว่า  $H_0$  จริง แสดงว่า ค่าแปรปรวนไม่ต่างกัน นั่นคือ (ค่าเฉลี่ยของข้อมูลทั้งสองกลุ่มไม่แตกต่างกัน) ในทางตรงข้ามถ้าผลการทดสอบสรุปว่า  $H_1$  จริง แสดงว่า ค่าแปรปรวนต่างกัน (นั่นคือ ค่าเฉลี่ยของข้อมูลทั้งสองกลุ่มแตกต่างกัน ข้อมูลนี้สามารถใช้อำนาจความแตกต่างระหว่างกลุ่มที่เป็นขอบเขตเสียงและไม่ใช่ขอบเขตเสียงได้) ซึ่งการพิจารณาในวิทยานิพนธ์นี้ต้องการหาข้อมูลที่มาเป็นค่าบังชี้หรืออำนาจกลุ่มตำแหน่งขอบเขตเสียงและกลุ่มที่ไม่ใช่ตำแหน่งขอบเขตเสียงออกจากกัน ดังนั้น ต้องปฏิเสธสมมติฐาน  $H_0$  และยอมรับสมมติฐานที่  $H_1$

การพิจารณาสมมติฐานใช้ค่าสถิติ  $F$  ดังสมการ (2.10) ซึ่งเป็นค่าตรวจสอบ

$$F = \frac{\left[ \frac{SS_{boundary}}{a - 1} \right]}{\left[ \frac{SS_{Error}}{N - a} \right]} \quad (2.10)$$

โดยที่  $a$  คือจำนวนกลุ่ม (ในวิทยานิพนธ์นี้  $a$  มีค่าเท่ากับ 2)

$N$  คือจำนวนข้อมูลที่ใช้วิเคราะห์ทั้งหมด

$$SS_{Error} = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{ij})^2 \text{ ซึ่ง } SS_{Error} \text{ คือความแปรผันภายในกลุ่ม}$$

$$SS_{boundary} = \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y}_{ij})^2 \text{ ซึ่ง } SS_{boundary} \text{ คือความแปรผันระหว่าง}$$

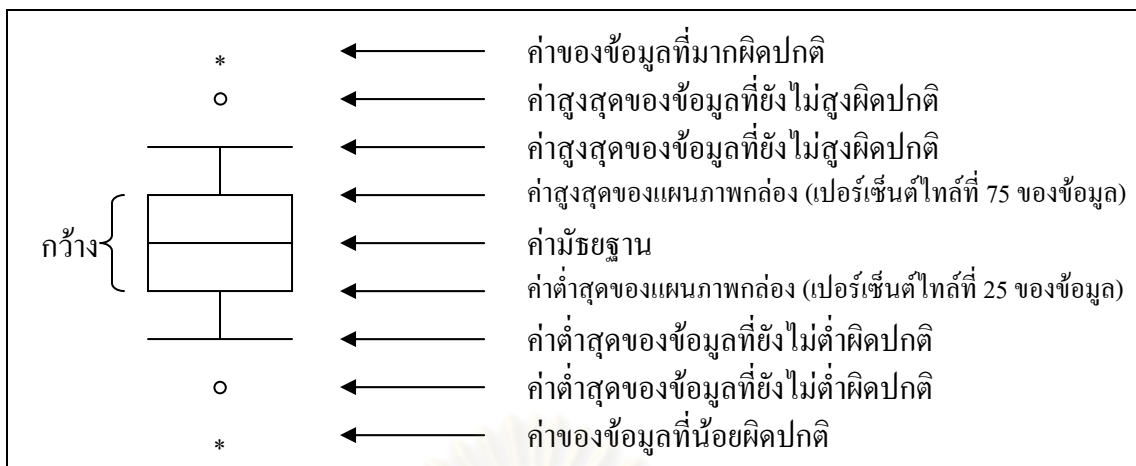
กลุ่ม เมื่อ  $\bar{y}_i$  คือค่าเฉลี่ยตัวอย่างชุดที่  $i$  และค่า  $\bar{y}_{ij}$  คือเฉลี่ยของตัวอย่างทั้งหมด ค่า  $y_{ij}$  คือค่าตัวอย่างตัวที่  $j$  ในกลุ่มที่  $i$  และ  $n_i$  คือจำนวนข้อมูลของกลุ่มที่  $i$

เมื่อได้ค่าสถิติ  $F$  จึงเปรียบเทียบกับ Sig (ค่าทางสถิติเมื่อระดับนัยสำคัญที่กำหนด  $\alpha$  คือ 0.05)

## 2. การพิจารณาลักษณะสำคัญทางสถิติด้วยแผนภาพกล่อง

แผนภาพกล่องเป็นเทคนิคที่ให้รายละเอียดเกี่ยวกับการแจกแจงของข้อมูลเป็นค่าสถิติเพื่อตรวจสอบการแจกแจงข้อมูล แสดงค่ากลางหรือค่ามัธยฐาน ค่าเปอร์เซ็นต์ไทล์ที่ 25 ( $Q_1$ ), 75 ( $Q_3$ ) และให้ค่าข้อมูลที่มีค่าผิดปกติ หมายถึง ค่าที่สูงมากหรือต่ำมาก (outlier) จากค่ากลาง ความกว้างของแผนภาพกล่อง แสดง 50% ของข้อมูลอยู่ในแผนภาพกล่อง ดังแสดงตามรูปที่ 2.14

- 1) ค่าสูงสุดของข้อมูลที่ยังไม่สูงผิดปกติคือค่าสูงของข้อมูลชุดนั้น ๆ ที่มีค่าไม่เกิน  $Q_3 + 1.5$  เท่าของความกว้างของแผนภาพกล่อง
- 2) ค่าต่ำสุดของข้อมูลที่ยังไม่ต่ำผิดปกติคือต่ำสุดของข้อมูลชุดนั้น ๆ ที่มีค่าไม่ต่ำกว่า  $Q_1 - 1.5$  เท่าของความกว้างของแผนภาพกล่อง
- 3) ค่าสูงสุดของข้อมูลที่มากกว่า  $Q_3 + 3$  เท่าของความกว้างของแผนภาพกล่อง จึงเรียกว่า Extreme
- 4) ค่าต่ำสุดของข้อมูลที่น้อยกว่า  $Q_1 - 3$  เท่าของความกว้างของแผนภาพกล่อง จึงเรียกว่า Extreme



รูปที่ 2.14 แสดงลักษณะของแผนภาพกล่อง

### 3. สัมประสิทธิ์สหสัมพันธ์

ค่าสัมประสิทธิ์สหสัมพันธ์ เป็นสถิติที่ใช้ทดสอบความสัมพันธ์ระหว่างตัวแปรเชิงกลุ่ม 2 ตัว สัญลักษณ์ที่ใช้คือ  $r$  สามารถระบุทิศทางความสัมพันธ์ว่ามีทิศทางเดียวกันหรือตรงข้ามกัน แสดงความสัมพันธ์อยู่ในรูปคะแนนเบี่ยงเบนซึ่งคำนวณตามสมการที่ 2.11

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.11)$$

โดยที่  $r$  แทนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร  $x$  และ  $y$  เมื่อรวบรวมข้อมูลของตัวแปร  $x$  และ  $y$  มาจำนวน  $n$  คู่ คือ  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

$\bar{x}$  แทนค่าเฉลี่ยของตัวแปร  $x$

$\bar{y}$  แทนค่าเฉลี่ยของตัวแปร  $y$

ค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าในช่วงระหว่าง -1 ถึง 1 ดังนั้นค่าที่ใช้ในการวิเคราะห์ความสัมพันธ์จึงใช้ค่ากำลังสองของค่าสัมประสิทธิ์สหสัมพันธ์แทนด้วย  $r^2$  ซึ่งค่า  $r^2$  นี้ นำมาเทียบกับร้อยละความแปรปรวนของตัวแปรหนึ่งกับอีกตัวหนึ่งได้ โดยพิจารณาค่ายกกำลังสองที่ได้คูณกับ 100 หากค่าผลลัพธ์มีค่าเกิน 80 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์กันมาก เช่น ค่า  $r = 0.5$  จะได้  $r^2 = 0.25$  ดังนั้นจึงประมาณค่าที่จะไปใช้เปรียบเทียบคือมีความสัมพันธ์กัน 25 % เป็นต้น



#### 4. วิธีการวิเคราะห์หัดิสคริมิแนนต์เชิงเส้น (Linear Discriminant Analysis, LDA)

การจำแนกกลุ่ม (Classify) ด้วยวิธีการวิเคราะห์หัดิสคริมิแนนต์เชิงเส้น (Linear Discriminant Analysis, LDA) แบ่งออกเป็นกลุ่มที่ต้องการและกลุ่มที่ไม่ต้องการ เช่น วิทยานิพนธ์นี้ ใช้จำแนกกลุ่มตำแหน่งที่เป็นขอบเขตเสียงและกลุ่มของตำแหน่งที่ไม่ใช่ขอบเขตเสียง โดยที่ เวกเตอร์ลักษณะสำคัญ (Feature Vector) ของชุดข้อมูลฝึก (Training Set) และชุดข้อมูลทดสอบ (Testing Set) แต่ละเวกเตอร์ลักษณะสำคัญที่ใช้นั้นได้มาจากลักษณะสำคัญทางเสียงที่ได้ในขั้นตอนการวิเคราะห์หลักเกณฑ์สำคัญตามหัวข้อ “การหาตำแหน่งขอบเขตเสียงโดยใช้สถิติ” ในบทที่ 3

#### 5. การคำนวณความน่าจะเป็นจากการจำแนก

$$P(g = c_i | x) = \frac{e^{-\frac{d_i^2}{2}}}{\sum_{k=1}^K e^{-\frac{d_k^2}{2}}} \quad (2.12)$$

โดยที่  $d_i$  คือระยะห่าง (distance in linear discriminant space) ระหว่างค่า  $x$  ที่สนใจกับจุดศูนย์กลางความหนาแน่นของประชากรในกลุ่มที่ต้องการจำแนก  $c_i$  โดยพิจารณาตามแต่ละชนิดตัวจำแนก

#### 6. การวัดระยะห่างยูคลิเดียน (Euclidean distance measure)

เมื่อวัดระยะห่างระหว่าง 2 เวกเตอร์ ได้แก่ เวกเตอร์  $P = [p_1 \ p_2 \ \dots \ p_n]$  และเวกเตอร์  $Q = [q_1 \ q_2 \ \dots \ q_n]$  ซึ่งคำนวณระยะห่างตามสมการ (2.13)

$$d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.13)$$

โดยที่  $d$  คือระยะห่างระหว่าง 2 เวกเตอร์ เมื่อ  $i$  แทนลำดับมิติของเวกเตอร์ และ  $n$  แทนข้อมูลลำดับมิติสุดท้ายของเวกเตอร์

#### งานวิจัยที่เกี่ยวข้อง

กระบวนการทางการรู้จำเสียงมีหลายวิธี เมื่อจำแนกส่วนการแบ่งหน่วยเสียงออกเป็น ส่วน ๆ นั้น ได้มีงานวิจัยที่สอดคล้องกับการแบ่งหน่วยเสียงแต่ละวิธี ดังนี้

##### 1. แนวทางแบบใช้สถิติ (Statistical Approach)

แบบจำลองฮิดเด้นมาคอฟมีประสิทธิภาพและศักยภาพสูงในการเรียนรู้จำด้วยเสียงอัตโนมัติ แต่ยังมีข้อจำกัดบางอย่างในงานสังเคราะห์เสียง เช่น ข้อมูลที่ใช้ในการฝึกฝน ต้องมีการระบุขอบเขตของหน่วยเสียงที่แสดงหลักเกณฑ์ด้วยมือ และการหาขอบเขตด้วยการใช้แบบจำลอง

ฮิดเด็นมาคอฟเป็นการพิจารณาเป็นกรอบคงที่ทางเวลาไม่สามารถนำข้อมูลอย่างอื่นมาร่วมพิจารณาด้วย แต่ยังเป็นที่ยอมรับกันอย่างแพร่หลายในการนำมาใช้เพื่อแบ่งส่วนหน่วยเสียง เช่น Yeon-Jun Kim [11] ได้เสนอการนำฮิดเด็นมาคอฟที่ใช้ฝึกฝนแบบวนซ้ำเพิ่มขึ้นเพื่อบอกขอบเขตหน่วยเสียงที่ได้จากการใช้แบบจำลองฮิดเด็นมาคอฟให้ผลแม่นยำกว่ามารวมการหาขอบเขตของสเปกตรัม เป็นต้นบางงานวิจัยใช้คุณสมบัติต่าง ๆ บนสัญญาณเสียงและพิจารณาลักษณะต่าง ๆ เพื่อบอกจุดสำคัญ (landmark) และใช้การตัดสินใจแบบแยกเป็นสองทาง (binary decision) เช่น Espy-Wilson [12] ได้เสนอวิธีการนำความรู้การสวนศาสตร์มารวมกับการเรียนรู้เชิงสถิติด้วยการทำงานของซัพพอร์ตเวกเตอร์แมชชีน (support vector machine, SVM) เพื่อแบ่งหน่วยเสียงของสัญญาณเป็น 5 หมวด ได้แก่ เสียงสระ เสียงเสียดแทรก เสียงกักและเสียงเงียบโดยใช้ตัวจำแนก 5 ตัว วางตามโครงสร้างแบบขนานรับข้อมูล 13 ค่าที่มีความสัมพันธ์กับลักษณะการออกเสียง (manner) พบว่าผลความถูกต้องดีขึ้นร้อยละ 12.8 เป็นร้อยละ 82.4 และผลความแม่นยำเพิ่มขึ้นร้อยละ 3.4 จากวิธีจากผลการทดลองกับฮิดเด็นมาคอฟที่ใช้ค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรอล 39 ค่าที่มีความแม่นยำร้อยละ 64.9 แต่พบปัญหาคือไม่สามารถการจำแนกสองหน่วยเสียงที่มี ลักษณะเดียวกันที่อยู่ติดกันได้ ดังนั้น Espy-Wilson [13] จึงใช้ตัวจำแนก 4 ตัวเพื่อทำการแยกกลุ่มเสียง 5 กลุ่ม และแบ่งการจำแนกเป็นระดับชั้นและคำนวณความน่าจะเป็นตามระดับชั้น จากโครงสร้างแบบแบ่งชั้นมีข้อดีคือไม่จำเป็นต้องใช้ทุกค่าลักษณะสำคัญทางเสียงเพื่อตัดสินใจแยกกลุ่มในทุกกลุ่มเสียง ตัวอย่างเช่น การสุ่มสัญญาณของเสียงสระที่ไม่ต้องฝึกฝนกับทุกกลุ่มเสียง ต้องการเพียงการฝึกฝนกับกลุ่มเสียงที่หนึ่งเท่านั้น จากนั้นจะได้ค่าความน่าจะเป็นของส่วนที่เสียงสระกับกลุ่มเสียงที่หนึ่งที่บอกความสัมพันธ์กัน เป็นต้น การแบ่งระดับชั้นนี้ยังง่ายต่อการชี้แจงหรือหาความผิดพลาดของแต่ละตัวจำแนก นำไปสู่การปรับความรู้การสวนศาสตร์ให้เหมาะสมกับแต่ละตัวจำแนกต่อไป แต่ผลเสียคือความผิดพลาดที่เกิดจากชั้นบนจะทำให้ส่งผลเกิดความไม่ถูกต้องในการจำแนกระดับชั้นล่าง งานนี้วัดความถูกต้องดีขึ้นร้อยละ 10.2 เป็นร้อยละ 79.8 และความแม่นยำเพิ่มขึ้นร้อยละ 3.2 เปรียบเทียบผลการทดลองกับฮิดเด็นมาคอฟที่ใช้ค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรอล 39 ค่าที่ได้ร้อยละ 64.9 จากทั้งสองงานวิจัยนี้ได้ศึกษาค่าพารามิเตอร์ต่างหรือลักษณะเด่นที่เหมาะสมของหน่วยเสียงแต่ละกลุ่มได้ ได้แก่ ความน่าจะเป็นของความเป็นเสียงก้อง ค่าอัตราการตัดศูนย์ พลังงานช่วงองค์ประกอบความถี่ต่าง ๆ ความถี่ฟอร์แมนท์ที่สาม ค่าอัตราส่วนพลังงาน เป็นต้น

การศึกษามีงานเกี่ยวกับการจำแนกกลุ่มเสียงต่าง ๆ ช่วยให้สามารถศึกษาธรรมชาติของแต่ละหน่วยเสียงและหาตัวแทนเสียงสำหรับกลุ่มเสียงต่าง ๆ ได้ ตัวอย่างเช่นการจำแนกกลุ่มเสียงนาสิก ที่บอกลักษณะที่เหมาะสมในของกลุ่มเสียงนาสิก โดย Tarun [14] ต้องการจะจำแนกกลุ่มเสียงนาสิกและเสียงกึ่งสระออกจากกัน เสนอการจับคุณสมบัติธรรมชาติของกลุ่มเสียงนาสิกที่มีลักษณะคล้ายเสียงเป็นเสียงบ่นพึมพำ และใช้ค่าอัตราส่วนพลังงานบางองค์ประกอบความถี่ ค่า

ความหนาแน่นของความถี่ฟอร์แมนท์และวัตถุประสงค์สูงสุดของสเปกตรัมเพื่อค้นหาเสียงนาสิก ซึ่งใช้ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนเป็นเครื่องมือในการจำแนกลักษณะ และ Atiwong [15] เสนอวิธีการจำแนกประเภทของเสียงกักออกเป็นชนิดหน่วยเสียงต่าง ๆ ซึ่งจากงานวิจัยนี้ทำให้ได้แรงจูงใจในการวิเคราะห์คุณสมบัติของลักษณะสำคัญทางเสียงของแต่ละกลุ่มเสียงก่อนที่จะเลือกลักษณะสำคัญที่สำคัญที่เหมาะสมสำหรับชนิดขอบเขตเสียงนั้น ๆ มาใช้ในการจำแนกกลุ่มต่อไป

นอกจากนี้ยังมีงานวิจัยที่เสนอวิธีการปรับตำแหน่งขอบเขตหน่วยเสียงให้มีความถูกต้องแม่นยำมากยิ่งขึ้น เช่น ใช้กฎในการตัดสินจากการพิจารณาค่าลักษณะสำคัญทางเสียง ได้แก่ Cheng-Yuan [16] ได้เสนอการใช้คุณลักษณะสำคัญต่าง ๆ เพื่อช่วยในการปรับตำแหน่งขอบเขตหน่วยเสียงที่ได้จากระบบรู้จำโดยอัตโนมัติมาคอฟใช้ค่าอัตราการตัดศูนย์ ค่าพลังงานลือก ค่าพิท ค่าองค์ประกอบความถี่ และค่าระดับของการระเบิดของเสียงเพื่อช่วยให้ตำแหน่งขอบเขตเสียงมีความแม่นยำมากยิ่งขึ้น ในงาน Lijuan Wang [17] เสนอการใช้ความรู้บริบทมาช่วยในการปรับตำแหน่งขอบเขตให้ดีขึ้นหลังจากที่ได้ตำแหน่งขอบเขตที่เป็นสมมติฐานจากการใช้การทำงานการระบุหน่วยเสียงของอัตโนมัติมาคอฟวิธีการระบุหน่วยเสียงแบบบังคับ ในงานนี้ใช้ตัวแทนการแสดงคล้ายกับไตรโฟน (triphone) โดยแบ่งเป็นสามส่วน ประกอบด้วย หน่วยเสียงที่อยู่ทางซ้ายของขอบเขต ส่วนขอบเขตระหว่างหน่วยเสียง และหน่วยเสียงทางขวาของขอบเขตแล้วฝึกฝนแบบจำลองด้วยเกาส์เซียน สำหรับแต่ละคู่ขอบเขตหน่วยเสียงที่ใช้ชุดของไตรโฟนแตกต่างกัน แนวคิดที่ต้องการเพิ่มประสิทธิภาพความถูกต้องของตำแหน่งขอบเขตจากพิจารณาตามคู่ ๆ ชนิดขอบเขตหน่วยเสียงเมื่อปรับตำแหน่งขอบเขตของ Abhinav Sethy [18] ที่ใช้วิธีไดนามิกไทม์วอร์ปิง (Dynamic time warping: DTW) และ Doroteo Toledano [19] ซึ่งตัดสินใจโดยใช้ข่างานประสาทเทียมจากคุณลักษณะสำคัญทางเสียง จึงเป็นแรงจูงใจในวิทยานิพนธ์นี้ให้ศึกษาวิธีการหาตำแหน่งขอบเขตเสียงตามคู่ ๆ ชนิดขอบเขตหน่วยเสียง โดยศึกษาลักษณะสำคัญทางเสียงที่ต่างกันในกลุ่มหน่วยเสียงนั้น

## 2. แนวทางแบบใช้กฎ (Rule Based Approach)

แนวทางนี้ศึกษาโดยอาศัยการใช้ข้อมูลและคุณสมบัติของสัญญาณเสียงต่าง ๆ มาเป็นลักษณะในการตัดสินใจ โดยไม่ใช้หลักทางสถิติมาช่วยในการสร้างตัวต้นแบบจากวิธีการสวนศาสตร์ (acoustic-phonetic) ในการเรียนรู้จำเสียงนั้นมีการใช้การตัดสินใจ

เนื่องจากพลังงานเป็นคุณสมบัติสำคัญอย่างหนึ่งของสัญญาณเสียงและการดูค่าความแตกต่าง การเปลี่ยนแปลงของพลังงานเป็นที่สนใจในงานวิจัยต่าง ๆ รวมทั้งเป็นค่าที่เสนอในวิทยานิพนธ์นี้เช่นกัน โดยเกิดได้แรงจูงใจจากงาน Espy-Wilson [12], [13], [20] ใช้ระดับความแตกต่างในความแตกต่างของพลังงานเพื่อวัดและแยกเสียงพยัญชนะเสียงที่มีเสียงสั้นหรือรัวน้อยกว่าสระ (sonorant) ขณะที่ Bitar [21] ที่ใช้ความแตกต่างของพลังงาน เพื่อหาจุดระเบิดของเสียงกัก

และวิธีการแบ่งหน่วยเสียงเป็นส่วน ๆ นี้มีลักษณะคล้ายกับงานของ Liu [22] ที่ศึกษาเป็นสองส่วน ดังนี้ ส่วนแรก หาจุดเปลี่ยนแปลงพลังงาน โดยคำนวณค่าพลังงานจากสเปกโตรแกรม โดยใช้ช่วงองค์ประกอบความถี่เป็นค่าที่กำหนดชัดเจน โดยแบ่งเป็นช่วงพลังงานต่าง ๆ ออกเป็น 6 ช่วงดังนี้ ช่วงที่หนึ่ง : 0.0-0.4 กิโลเฮิร์ตซ์ ช่วงที่สอง : 0.8-1.5 กิโลเฮิร์ตซ์ ช่วงที่สาม : 1.2-2.0 กิโลเฮิร์ตซ์ ช่วงที่สี่ : 2.0-3.5 กิโลเฮิร์ตซ์ ช่วงที่ห้า : 3.5-5.0 กิโลเฮิร์ตซ์ ช่วงที่หก : 5.0-8.0 กิโลเฮิร์ตซ์ ส่วนที่สอง ใช้การหาจุดสูงสุดของพลังงานเพื่อหาชนิดของจุดสำคัญบนสัญญาณเสียงจากแต่ละช่วงพลังงานเช่น จุดการระเบิดของหน่วยเสียง จุดเริ่มต้นของเสียงที่สั้น เป็นต้น โดยดูลักษณะเด่นพิเศษที่มีการเปลี่ยนแปลงทันที เช่น ส่วนกักและปล่อยเสียงกักหรือตำแหน่งการสิ้นสุดการสั้นของเส้นเสียง เป็นต้น เพื่อเป็นปัจจัยในการหาจุดสำคัญ งานนี้เป็นพื้นฐานแนวคิดของการสร้างต้นแบบการศึกษาข้อมูลจากสเปกโตรแกรมและโดยได้แบ่งพลังงานออกเป็นช่วงต่าง ๆ แต่ความแตกต่างกับที่เสนอในวิทยานิพนธ์นี้คือเสนอวิธีแบ่งช่วงพลังงานแบบไม่กำหนดค่าคงที่และเสนอชนิดของจุดสำคัญที่ต้องการค้นต่างกัน นอกจาก Liu ที่ได้นำข้อมูลบนสเปกโตรแกรมแล้ว Pat Lu [23] ได้อธิบายเกี่ยวกับลักษณะเด่นที่พบบนสเปกโตรแกรมของแต่ละกลุ่มเสียง เช่น หาจุดสูงสุดในการระเบิดของเสียงหยุดและสรุปลักษณะสำคัญทางเสียงเด่นสิบสามข้อที่มีประสิทธิภาพในการค้นหาหน่วยเสียงในกลุ่มเสียงหยุดจากสเปกโตรแกรม ซึ่งเสียงที่ใช้ในการวิเคราะห์ข้อมูลไม่ได้จากเสียงธรรมชาติแต่ใช้เสียงสังเคราะห์มาทดสอบแทน จากวิทยานิพนธ์นี้สนับสนุนให้ใช้สเปกโตรแกรมมาเป็นเครื่องมือสำคัญของการวิเคราะห์หน่วยเสียงและคุณภาพการเปลี่ยนแปลงที่รอยต่อของหน่วยเสียงได้

งานที่ศึกษาเกี่ยวกับลักษณะของแต่ละกลุ่มเสียงช่วยให้สามารถนำลักษณะเด่นของแต่ละกลุ่มเสียงอันมีประโยชน์ในการหาขอบเขตของกลุ่มเสียงบางกลุ่มได้ นอกจากงานจำแนกกลุ่มเสียงนาสิก ของ Marilyn Y Chen [24] ที่พัฒนาส่วนค้นหาเสียงนาสิกของระบบ LAFF (Lexical Access From Features) [25] ที่มีความถูกต้องของการจำแนกร้อยละ 83.6 สำหรับการจำแนกเสียงสระและเสียงนาสิก ในบางงานวิจัยได้เสนอการรวมหรือผสมลักษณะต่าง ๆ มาใช้ในการระบุขอบเขตหน่วยเสียงอัตโนมัติ นั่นคือ เสนอการระบุตำแหน่งขอบเขตหน่วยเสียงอัตโนมัติโดย Paul Micallef [26] เปรียบเทียบการใช้ลักษณะต่าง ๆ ร่วมกันกับการใช้แคปสโตรลเพียงอย่างเดียว และเสนอการวัดประสิทธิภาพการพบตำแหน่งขอบเขตตามชนิดต่าง ๆ จากผลวิจัยที่สนับสนุนว่าการผสมลักษณะต่าง ๆ มีประสิทธิภาพดีกว่าการใช้ลักษณะสำคัญทางเสียงอย่างเดียว

นอกจากนี้ความเป็นเสียงก้องหรือเสียงไม่ก้องนี้เป็นคุณสมบัติที่สำคัญเพื่อใช้หาขอบเขตของหน่วยเสียงในวิทยานิพนธ์นี้ ซึ่งงานวิจัยมากมายได้เสนอวิธีการหาส่วนที่เป็นเสียงก้อง เสียงไม่ก้องและเสียงเงียบ เช่น Zhou Zhijie [27] ใช้ค่าที่ช่วงพลังงานต่ำ ค่าอัตราการตัดศูนย์ ค่าอัตราส่วนพลังงาน เป็นคุณสมบัติเพื่อพิจารณาความเป็นเสียงก้องหรือเสียงไม่ก้อง ขณะที่ Atal และ



Rabiner [28] เสนอค่าที่เหมาะสมในการแยกกลุ่มเสียงก้อง เช่น ค่าพลังงานเสียง ค่าพลังงานจากความผิดพลาดการประมาณค่า ค่าอัตราการตัดศูนย์ ค่าความสัมพันธ์ระหว่างการแซมเปิลเสียงที่อยู่ติดกัน ค่าสัมประสิทธิ์การประมาณพันธุเชิงเส้น เป็นต้น ซึ่งในงานของ V.V.S.Sarma [29] ได้ใช้บางลักษณะข้างต้น เพื่อใช้ในการจำรูปแบบ แยกกลุ่มเสียงก้อง เสียงไม่ก้องและเสียงเงียบเช่นกัน งานนี้ใช้ค่าลือกของพลังงานและค่าอัตราสัมพันธ์แทนการไม่ใช้ค่าจากสัมประสิทธิ์การประมาณพันธุเชิงเส้นเพื่อลดการคำนวณโดยไม่ทำให้ความผิดพลาดเพิ่มขึ้น ต่อมางาน Siegel [30] ได้เสนอค่าพลังงานเฉลี่ย (Root Mean Square, RMS) ค่าอัตราส่วนพลังงานเสียงที่ช่วงองค์ประกอบความถี่สูงกับองค์ประกอบความถี่ต่ำ ค่าแอมพลิจูดที่จุดสูงสุดใช้เป็นลักษณะในวิธีการรู้จำรูปแบบ ขณะที่ Kannan [31] ที่เสนอค่าพิท (pitch) เป็นลักษณะหนึ่งของการช่วงเสียงที่เป็นเสียงก้อง จากงานข้างต้นนี้เป็นแนวทางการศึกษาลักษณะที่แยกความเป็นเสียงก้องและไม่ก้องนำไปสู่การหาความแตกต่างจากหน่วยเสียงที่อยู่ติดกันได้ที่มีลักษณะความเป็นเสียงก้องต่างกัน ได้

Jan P.H.van Santen [32] อธิบายความสำคัญของการค้นหาขอบเขตของหน่วยเสียงว่า เหตุผลหลักที่ระบบรู้จำเสียงอัตโนมัติไม่ได้ทำให้การตัดแบ่งหน่วยเสียงมีความแม่นยำสูงนั้น เพราะการแบ่งหน่วยเสียงถูกสร้างขึ้นมาเพื่อที่จะระบุหน่วยเสียงไม่ใช่ค้นหาขอบเขตระหว่างหน่วยเสียง ดังนั้นจึงเสนอการค้นหาขอบเขตโดยใช้หลักการคล้ายกับเทคนิคการหาขอบในกระบวนการทางรูปภาพ อีกทั้งเสนอการรวมกลุ่มของหน่วยเสียงที่มีคุณสมบัติร่วมกันไว้ในกลุ่มเดียวกัน จากหลักการหาความสัมพันธ์ของเสียงและรวมเสียงที่มีลักษณะเหมือนกันไว้ด้วยกัน ได้เสนอในอีกแนวคิดโดย Glass [33] สร้างอัลกอริทึม เดนโดแกรม (dendrogram) เพื่อแบ่งสัญญาณเสียงให้อยู่ในรูปของหน่วยเสียงย่อยโดยหาความสัมพันธ์ระหว่างแต่ละกรอบสัญญาณที่อยู่ติดกัน เกิดขอบเขตระหว่างกรอบสัญญาณเมื่อมีการเปลี่ยนแปลงความสัมพันธ์ระหว่างกรอบสัญญาณที่อยู่ก่อนหน้ากับกรอบสัญญาณถัดไป โดยอัลกอริทึมนี้แบ่งความสัมพันธ์ออกเป็นหลายระดับชั้น เปรียบเทียบส่วนของหน่วยเสียงย่อยภายในระดับเดียวกันหากมีความสัมพันธ์กันจะรวมกลุ่มให้ของส่วนหน่วยเสียงย่อยเดียวกัน และแสดงความสัมพันธ์ออกไปในระดับชั้นที่อยู่สูงถัดขึ้นไป เป็นต้น

การศึกษางานวิจัยที่เกี่ยวข้องซึ่งเสนอวิธีการหาดำแหน่งขอบเขตเสียงเมื่อมีลำดับของหน่วยเสียงสัญญาณเสียงเป็นข้อมูลเริ่มต้น โดยแบ่งขั้นตอนการทำงานเป็น 2 ขั้นตอน ได้แก่ การทำงานส่วนแรก คือ ขั้นตอนการหาดำแหน่งขอบเขตเสียงเริ่มต้นโดยใช้แบบจำลองฮิดเดินมาคอฟ และการทำงานส่วนที่สองคือ ขั้นตอนการปรับหาดำแหน่งเฟรมที่เหมาะสมกับตำแหน่งขอบเขตเสียงมากที่สุดซึ่งใกล้เคียงกับตำแหน่งขอบเขตเสียงเริ่มต้นที่ได้จากการทำงานส่วนแรก ซึ่งจากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า งานวิจัยที่เกี่ยวข้องนั้นได้เสนอวิธีการปรับหาดำแหน่งขอบเขตในขั้นตอนการทำงานที่สองที่แตกต่างกัน เช่น Toledano [19] เสนอการใช้ข่ายงานประสาท



เทียม โดยมีการใช้ข่างงานประสาทเทียมชนิดเพอเซปตรอนหลายชั้นโดย kim [34] และ Eun-Young [35] ขณะที่ Lijuan [17] เสนอใช้แบบจำลองเกาส์เซียน นอกจากนี้ Hung [36] มีเสนอการใช้ซัพพอร์ตเวกเตอร์แมชชีนและ Sérgio กับ Luis [37] เสนอการใช้ไดนามิกโทมวอร์ปปีง เป็นต้น วิทยานิพนธ์นี้จึงเสนอวิธีการปรับหาตำแหน่งขอบเขตหน่วยเสียงโดยใช้กฎการตัดสินใจที่ตั้งขึ้นเอง และเสนอการใช้วิธีการวิเคราะห์ดีสคริมีแนนต์เชิงเส้นมาช่วยในการค้นหาตำแหน่งขอบเขตเสียงซึ่งอธิบายต่อไป



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 3

### วิธีดำเนินการวิจัย

#### ขั้นตอนการดำเนินการวิจัย

ขั้นตอนการดำเนินงานวิจัยในบทนี้จะกล่าวถึง เครื่องมือที่ใช้ในการวิจัย ข้อมูลที่ใช้พัฒนาและทดสอบระบบในวิทยานิพนธ์ ชุดข้อมูลฝึกฝนและทดสอบ การกำหนดขอบเขตหน่วยเสียงเพื่อสร้างฐานข้อมูล การสร้างแบบอ้างอิงซึ่งประกอบด้วย การฝึกฝนแบบจำลองฮิดเด้นมาร์คอฟและการทดสอบประสิทธิภาพของแบบอ้างอิง อธิบายการสร้างระบบหาตำแหน่งขอบเขตหน่วยเสียง และเสนอวัดประสิทธิภาพและประเมินผลของระบบ เป็นต้น

#### เครื่องมือที่ใช้ในการวิจัย

- 1) MATLAB 7.0 [38] ใช้เป็นเครื่องมือใช้เพื่อพัฒนาและทดสอบระบบ
- 2) HTK toolkit [39] เป็นเครื่องมือที่พัฒนาโดยมหาวิทยาลัยเคมบริดจ์ (Cambridge University) ในวิทยานิพนธ์นี้ใช้เครื่องมือนี้เพื่อสร้างเป็นแบบทดลองอ้างอิง เพื่อระบุหาตำแหน่งหน่วยเสียงต่าง ๆ บนสัญญาณเสียงรับ ข้อมูลของลำดับหน่วยเสียงบนสัญญาณเสียงนั้น
- 3) Praat [40] เป็นเครื่องมือที่พัฒนาโดย Paul Boersma และ David Weenink ของสถาบัน the Institute of Phonetic Sciences มหาวิทยาลัยอัมสเตอร์ดัม เครื่องมือนี้ใช้เพื่อวิเคราะห์และศึกษาลักษณะทางโดเมนทางเวลาและโดเมนทางความถี่ของสัญญาณเสียง
- 4) Wavesurfer [8] เป็นเครื่องมือที่พัฒนาโดยศูนย์กลางเทคโนโลยีเสียงพูด (Centre for Speech Technology:CTT) ที่มหาวิทยาลัยเคทีเอช (Kungliga Tekniska Högskolan:KTH) วิทยานิพนธ์นี้ใช้เครื่องมือนี้สำหรับวิเคราะห์และศึกษาสเปกโตรแกรมของสัญญาณเสียงต่าง ๆ เพื่อระบุตำแหน่งขอบเขตหน่วยเสียงบนสัญญาณเสียงนั้น ซึ่งใช้เป็นเครื่องมือช่วยในการสร้างฐานข้อมูลสัญลักษณ์ (transcription) ที่มีการระบุข้อมูลทางเวลา

#### ข้อมูลที่ใช้พัฒนาและทดสอบระบบ

ฐานข้อมูลที่ใช้ในวิทยานิพนธ์นี้ ใช้ฐานข้อมูล LOTUS [5] และ [41] ชุดสมมูล (phonetically distributed: PD) ที่บันทึกเสียงจากสภาพแวดล้อมแบบห้องเงียบ วิทยานิพนธ์นี้ใช้ 1400 ประโยค จากการศึกษาตามข้อมูลที่ได้จากการแสดงสัญลักษณ์เดิมของฐานข้อมูล LOTUS

พบว่ามีการเปลี่ยนแปลงของอะคูสติคภายในเซกเมนต์เดียวกัน เพื่อเป็นการสอดคล้องกับแนวคิดของวิทยานิพนธ์นี้ จึงสร้างฐานข้อมูลใหม่เพื่อให้สอดคล้องกับสมบัติทางเสียงของแต่ละหน่วยเสียง ซึ่งมีรายละเอียดของชุดหน่วยเสียงใหม่ ประกอบด้วย เสียงสระ 24 หน่วยเสียง เสียงนาสิก 3 หน่วยเสียง เสียงกึ่งสระ 2 หน่วยเสียง เสียงเสียดแทรก 3 หน่วยเสียง เสียงกึ่งเสียดแทรก 2 หน่วยเสียง เสียงข้างลิ้น 1 หน่วยเสียง และเสียงร้าว 1 หน่วยเสียง ตามที่แสดงในตารางที่ 2.4 และกำหนดกลุ่มเสียงที่แก้ไขและเปลี่ยนแปลง ได้แก่ เสียงกักก้องที่แสดงการระเบิด 2 เสียง ได้แก่ /b/ /d/ เสียงกักไม่ก้อง 6 หน่วยเสียง และเสียง voiced closure 1 หน่วยเสียง ซึ่งแทนด้วยสัญลักษณ์ /vb/ ซึ่งมีรายละเอียดของการระบุแต่ละชนิดต่อไป และสรุปหน่วยเสียงที่ใช้ในวิทยานิพนธ์นี้ตารางที่ 3.1

เนื่องจากอยู่เชื่อมระหว่างบางหน่วยเสียงเช่น เสียงระเบิด (ทำหน้าที่เป็นตัวนำ) และเสียงสระ เสียงสระและเสียงกัก (ทำหน้าที่เป็นตัวสะกด) นั้น มีช่วงเวลาดังมาก และไม่สามารถสังเกตเห็นการเปลี่ยนแปลงทางสเปกโตรแกรมได้ และไม่สามารถชี้จุดเชื่อมต่อได้แน่ชัด ดังนั้นข้อกำหนดของการหาตำแหน่งขอบเขตเสียงในวิทยานิพนธ์นี้ จึงได้รวมบริเวณที่ไม่ชัดเจนดังกล่าวเข้าไปในหน่วยเสียงสระซึ่งมีช่วงเวลานานสามารถหาสมบัติทางเสียงเพื่อบอกชนิดหน่วยเสียงได้

ตารางที่ 3.1 สรุปหน่วยเสียงที่ใช้ในวิทยานิพนธ์

กลุ่มเสียง	หน่วยเสียง	สัทอักษรสากล	ตัวอย่างในภาษาไทย
เสียงกัก	/b/	[b]	บ
	/d/	[d]	ด ฎ
	/ph/	[ph]	พ ภ ผ
	/p/	[p]	ป และตัวสะกดแม่กบ
	/t/	[t]	ต ฏ และตัวสะกดแม่กด
	/th/	[t <sup>h</sup> ]	ธ ท ฒ ฑ ฐ ถ
	/k/	[k]	ก และตัวสะกดแม่กก
	/kh/	[k <sup>h</sup> ]	ข ค ฃ
เสียงvoiced closure	/vb/		
เสียงนาสิก	/m/	[m]	ม และตัวสะกดแม่กม
	/n/	[n]	น และตัวสะกดแม่กน
	/ng/	[ŋ]	ง และตัวสะกดแม่กง
เสียงข้างลิ้น	/l/	[l]	ล พ หล
เสียงกึ่งเสียดแทรก	/c/	[t]	จ

กลุ่มเสียง	หน่วยเสียง	สัญลักษณ์สากล	ตัวอย่างในภาษาไทย
	/ch/	[tʰ]	ช ฉ ฌ
เสียงร้ว	/r/	[r]	ร
เสียงเสียดแทรก	/s/	[s]	ซ ส ศ ษ
	/h/	[h]	ฮ ห
	/f/	[f]	ฟ ฟ
เสียงกึ่งสระ	/j/	[j]	ย ญ และตัวสะกดแม่เกย
	/w/	[w]	ว และตัวสะกดแม่เกอว
เสียงสระ	/i/ /e/ /x/ /v/ /q/ /a/ /u/ /o/ /@/	ตามภาคผนวก ก	เสียงเดี่ยวสั้น (อิ,เอะ,แอะ, อี, อะ,เออะ อุ, โอะ, เออะ)
	/ii/ /ee/ /xx/ /vv/ /qq/ /aa/ /uu/ /oo/ /@@/	ตามภาคผนวก ก	เสียงเดี่ยวยาว (อี, เอ, แอ, อี้อ, เออ, ออ, อุ, โอ, ออ)
	/ia/ /iia/ /va/ /vva/ /ua/ /uua/	ตามภาคผนวก ก	เสียงสระประสม (เอียะ, เอีย, เอือะ, เอือ, อัวะ, อัว)
เสียงเงียบ	/sil/		

### 1. ข้อมูลฝึกฝนและทดสอบ

จากฐานข้อมูลชุด PD นี้ได้แบ่งข้อมูลออกเป็น 3 ชุด คือ

ชุดที่ 1 ชุดข้อมูลที่ใช้ในการฝึกฝนแบบอ้างอิงและฝึกฝนแบบจำลอง ประกอบด้วย 700 ประโยค

ซึ่งเป็นเสียงของผู้หญิง 350 ประโยค

ซึ่งเป็นเสียงของผู้ชาย 350 ประโยค

ชุดที่ 2 ชุดข้อมูลที่ใช้ในการทดสอบแบบจำลองและพัฒนาระบบ ประกอบด้วย 350 ประโยค

ซึ่งเป็นเสียงของผู้หญิง 175 ประโยค

ซึ่งเป็นเสียงของผู้ชาย 175 ประโยค

ชุดที่ 3 ชุดข้อมูลที่ใช้ในการทดสอบระบบ ประกอบด้วย 350 ประโยค

ซึ่งเป็นเสียงของผู้หญิง 175 ประโยค

ซึ่งเป็นเสียงของผู้ชาย 175 ประโยค

ฐานข้อมูลที่ใช้ในวิทยานิพนธ์นี้แบ่งข้อมูลเป็นชุด ๆ เพื่อแยกข้อมูลที่ใช้ในการฝึกฝน ชุดข้อมูลที่ทดสอบและพัฒนาระบบ และชุดข้อมูลทดสอบระบบการหาขอบเขตเสียงที่เสนอออกจากกัน เพื่อป้องกันไม่ให้เกิดการใช้ข้อมูลในการสร้างต้นแบบการหาตำแหน่งขอบเขตเสียงที่ลำเอียงขึ้นกับข้อมูลที่ใช้ฝึกฝนนี้เท่านั้น

## 2. การกำหนดขอบเขตหน่วยเสียงเพื่อสร้างฐานข้อมูล

การกำหนดขอบเขตหน่วยเสียงเพื่อสร้างฐานข้อมูลในวิทยานิพนธ์นี้ สร้างเป็นกฎการตัดสินใจการหาตำแหน่งขอบเขตเสียง โดยพิจารณาจากลักษณะพิเศษของแต่ละเซกเมนต์ ซึ่งภายในเซกเมนต์จะแสดงอะคูสติคเหมือนกันและมีสมบัติทางเสียงเหมือนกัน ไม่มีการเปลี่ยนแปลงอย่างทันทีปรากฏภายในเซกเมนต์เดียวกัน ดังนั้นจุดเริ่มต้นของเซกเมนต์ คือ ตำแหน่งขอบเขตของหน่วยเสียงทางซ้ายที่ติดกับหน่วยเสียงก่อนหน้า และจุดสิ้นสุดของเซกเมนต์ คือ ตำแหน่งขอบเขตของหน่วยเสียงทางขวาที่ติดกับหน่วยเสียงที่ตามมา

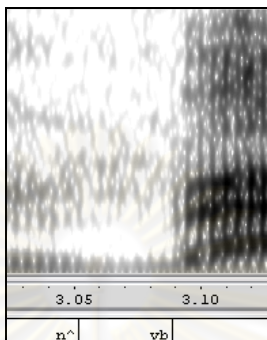
จากการพิจารณาคำแห่งของกลุ่มเสียงในขั้นตอนการกำหนดขอบเขตเสียงนี้ ทำให้สามารถชี้ความแตกต่างของสมบัติทางเสียงแต่ละหน่วยเสียงได้ ดังนั้นจึงศึกษาค่าลักษณะทางเสียงและเสนอการเลือกค่าลักษณะสำคัญทางเสียงที่เหมาะสมในการบอกความแตกต่างระหว่างกลุ่มเสียงที่พิจารณากับกลุ่มเสียงอื่นควบคู่ไปด้วย เช่น เลือกใช้ค่าสัมประสิทธิ์เมลเฟรีควีนซีเคปสตรัมซึ่งสกัดเป็นตัวแทนเสียงสำหรับการพิจารณาในทุกขอบเขตเสียง เสนอให้เลือกใช้ค่าพลังงานเพื่อบอกความแตกต่างของแต่ละกลุ่มเสียง เป็นต้น จากนั้นรวบรวมและสรุปค่าลักษณะทางเสียงที่ศึกษาและเสนอเลือกในขั้นตอนนี้ เพื่อวิเคราะห์หาค่าที่เหมาะสมโดยขั้นตอนการเลือกค่าลักษณะทางเสียงเพื่อใช้ในการสร้างระบบ ซึ่งอธิบายในบทที่ 4 ต่อไป

**2.1 เสียงกัก** ซึ่งนิยามใช้ในวิทยานิพนธ์นี้หมายถึงหน่วยเสียง /b/ /d/ /ph/ /th/ /kh/ /p/ /t/ และ /k/ ตามที่กำหนดในขั้นตอนนี้เท่านั้น โดยแยกจากหน่วยเสียง /vb/ อย่างชัดเจน ซึ่งการแบ่งวิธีสังเกตการระบุขอบเขตหน่วยเสียง เป็นกลุ่มเสียงย่อย ดังนี้

**2.1.1 เสียงกักแบบโฆษะ** ได้แก่ หน่วยเสียง /b/ และ /d/ (ตามตารางที่ 3.1) พบว่าหน่วยเสียงกลุ่มนี้มีพลังงานที่ถูกส่งผ่านออกมาในอากาศ ผ่านออกมาจากการแผ่รังสีจากกระพุ้งแก้ม มิได้เกิดจากลมจากช่องปากโดยตรง จึงพบแถบความเข้มของเสียงก้อง (Voiced bar) เกิดเป็นพลังงานที่องค์ประกอบความถี่ต่ำจากการสั่นที่ของเส้นเสียงในขณะที่เกิดช่องปิด ก่อนที่จะเกิดการพ่นลม ดังนั้นจึงแบ่งหน่วยเสียงตาม



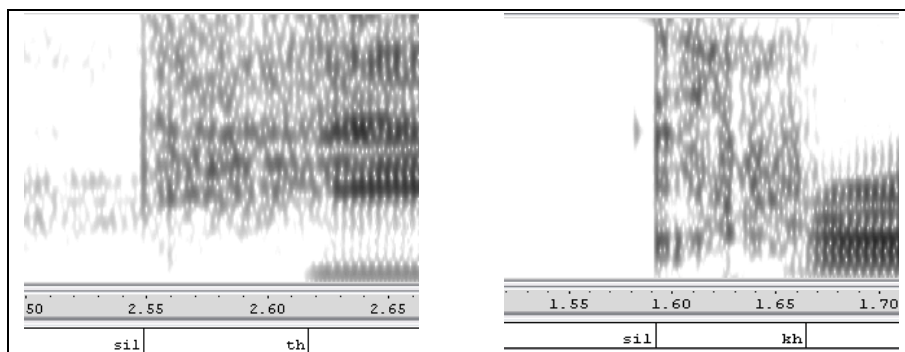
ข้อมูลแสดงสัทลักษณะเดิมตามอะคูสติกที่ปรากฏ ออกเป็น 2 ส่วนตามลำดับ คือ /vb/ แทนแถบความเข้มที่องค์ประกอบความถี่ต่ำปรากฏในสเปกโตรแกรม และ /b/ คือการระเบิดของเสียงกักออกมา ซึ่งระบุเมื่อมีเส้นในแนวตั้งปรากฏตาม องค์ประกอบความถี่สูงและสิ้นสุดเมื่อเริ่มเข้าสู่เสียงสระที่ตามมา โดยพิจารณาจาก แถบความถี่ฟอร์แมนท์ของเสียงสระ



รูปที่ 3.1 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /n^/-/vb/

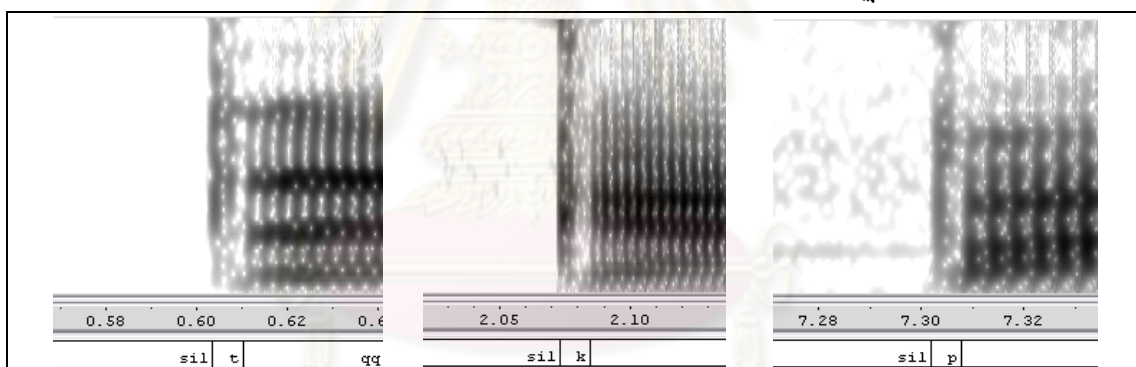
จากการกำหนดตำแหน่งขอบเขตเสียงของ vb พบว่าเสียง vb นี้อยู่ตำแหน่งเริ่มต้นของพยางค์ดังนั้นขอบเขตระหว่างขอบเขตเสียง vb จะปรากฏระหว่างเสียงสระและปรากฏระหว่างเสียงเงียบ ดังนั้นจึงเสนอค่าลักษณะทางเสียงที่เหมาะสมในการเลือกมาพิจารณาความแตกต่างระหว่างกลุ่มเสียง ได้แก่ ค่าความสัมพันธ์ระหว่างเฟรมซึ่งใช้บ่งชี้ความแตกต่างของสัญญาณเสียงเพื่อบอกเฟรมที่มีโอกาสเป็นตำแหน่งขอบเขตเสียง นอกจากนี้มีค่าลักษณะทางเสียงค่าอื่นๆ ได้แก่ ค่าพลังงาน และพลังงานตามช่วงองค์ประกอบความถี่ เนื่องจากเสียง vb มีค่าพลังงานต่ำ และพลังงานในช่วงองค์ประกอบความถี่ต่างๆ แตกต่างจากกลุ่มเสียงที่ติดกัน

- 2.1.2 เสียงกักแบบอโฆษะ ฟันลม เช่น /ph/ /th/ /kh/ พบว่ามีเสียงรบกวนปรากฏกระจายในองค์ประกอบความถี่สูงเนื่องจากลมที่ถูกกักถูกปล่อยออกมา โดยในกลุ่มนี้ไม่มีแถบความเข้มของเสียงก้องปรากฏที่องค์ประกอบความถี่ต่ำเหมือนกับกลุ่มเสียงกักแบบโฆษะ ดังนั้นระบุจุดเริ่มต้น จากเส้นในแนวตั้งในช่วงความถี่สูง ลักษณะเสียงรบกวนที่ปรากฏกระจายและระบุจุดสิ้นสุดหน่วยเสียงจากตำแหน่งที่ไม่พบความเป็นเสียงรบกวนหรือเริ่มเข้าสู่เสียงสระที่ตามมา



รูปที่ 3.2 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/th/ และ /sil/-/kh/ ตามลำดับ

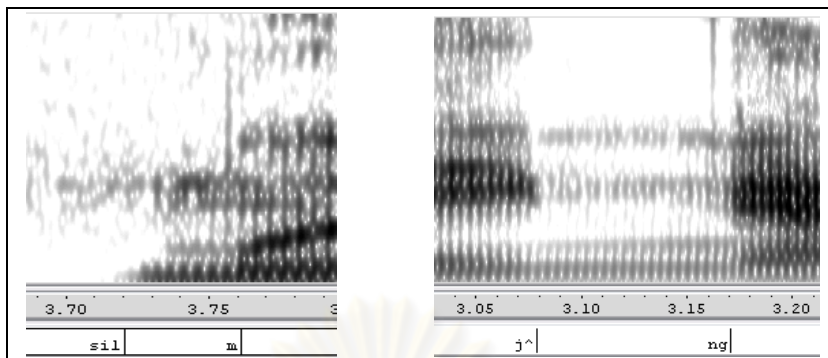
- 2.1.3 เสียงกักแบบอโฆยะ ไม่พ่นลม เช่น /p/ /t/ /k/ เนื่องจากจุดที่กักลมไว้ไม่เปิด ลมที่ถูกกักที่ไม่ได้ระเบิดออกมากลับถูกกลืนลงไป ดังนั้นจึงไม่มีเสียงออกมา จึงพบเพียงเสียงเสียบเชิงกลปรากฏก่อนที่จะเกิดหน่วยเสียงนี้ ในขณะที่ช่องปิดถูกปล่อยออกมาอย่างรวดเร็ว อาจเกิดเป็นลักษณะเสียงรบกวนปรากฏสั้น ๆ ซึ่งสังเกตจากพลังงานที่มีรูปร่างยาวออกไปทางแนวตั้งในสเปกโตรแกรม หลังจากช่วงที่เป็นช่องปิด ดังนั้น จึงเกิดเป็น /sil/ ที่เป็นเสียงเสียบ และตามด้วย /t/ ที่เกิดเป็นช่วงสั้น ๆ อาจจะมีเส้นตามแนวตั้งที่เกิดจากการปล่อยลมปรากฏ



รูปที่ 3.3 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/t/, /sil/-/k/ และ /sil/-/p/ ตามลำดับ

- 2.1.4 เสียงกัก ฐานเส้นเสียง ได้แก่ ตัว อ /z/ ในวิธานิพนธ์นี้ได้รวมตัว อ /z/ (ซึ่งอยู่ตำแหน่งก่อนเสียงสระ) เข้าไว้กับเสียงสระที่ตามมา
- 2.2 เสียงนาสิก ได้แก่ /m/ /n/ /ng/ สามารถระบุจุดเริ่มต้นได้จากการปรากฏของความถี่ฟอร์แมนท์ที่ความถี่ต่ำ จากการสูญเสียพลังงานในโพรงจมูกทำให้พลังงานในองค์ประกอบความถี่กลางถึงสูงจะลดต่ำลงจากพลังงานของสระที่อยู่ใกล้เคียง ดังนั้นให้พิจารณาจากการเคลื่อนที่ของความถี่ฟอร์แมนต์ในสระที่ติดกันเป็นหลัก โดยพิจารณาเริ่มต้นเสียงนาสิกที่ตามหลังสระ จากช่วงความกว้างของระยะห่างของความถี่ฟอร์แมนท์ที่หนึ่ง จะกว้างกว่าหรือความถี่ฟอร์แมนท์ที่องค์ประกอบความถี่

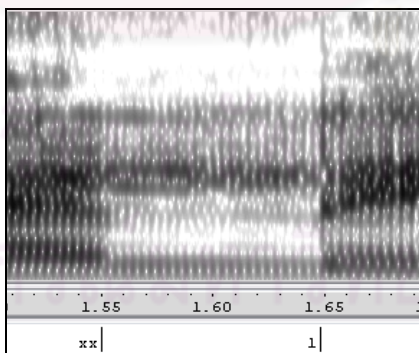
สูงไม่ชัด และพิจารณาจุดสิ้นสุดของเสียงนาสิกเมื่อมีเสียงสระตามมา พบว่า ความถี่ฟอร์แมนท์ที่องค์ประกอบความถี่สูงจะปรากฏในส่วนสระที่ตามมา



รูปที่ 3.4 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil-/m/ และ /j^/-ng/ ตามลำดับ

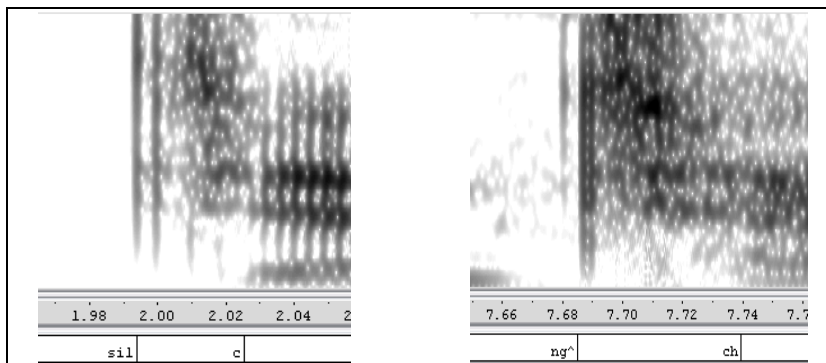
จากการกำหนดตำแหน่งขอบเขตเสียงนาสิก พบว่ากลุ่มเสียงนาสิกอยู่ตำแหน่งเริ่มต้นและอยู่ตำแหน่งท้ายพยางค์ ดังนั้นขอบเขตเสียงนาสิก จะปรากฏระหว่างเสียงสระ เสียงเสียดแทรกและระหว่างเสียงเงียบ ดังนั้นค่าลักษณะทางเสียงที่เสนอ เพื่อบอกความแตกต่างระหว่างกลุ่มเสียง ได้แก่ ค่าความถี่ฟอร์แมนท์ ค่าพลังงาน พลังงานตามช่วงองค์ประกอบความถี่ และอัตราส่วนระหว่างพลังงานบางช่วง องค์ประกอบความถี่ เนื่องจากเสียงนาสิกมีค่าพลังงานที่องค์ประกอบความถี่ต่ำ ซึ่งแตกต่างจากเสียงในกลุ่มอื่น

2.3 เสียงข้างลิ้น /l/ ระบุจุดเริ่มต้นจากการสังเกตความถี่ฟอร์แมนท์ที่หนึ่งต่ำ ขณะที่ความถี่ฟอร์แมนท์ที่สองที่ค่าปานกลาง บอจุดสิ้นสุด เมื่อเปลี่ยนจาก /l/ ไปสู่เสียงสระจากการพิจารณาการเคลื่อนที่อย่างรวดเร็วของความถี่ฟอร์แมนท์ที่สอง



รูปที่ 3.5 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /xx/-l/

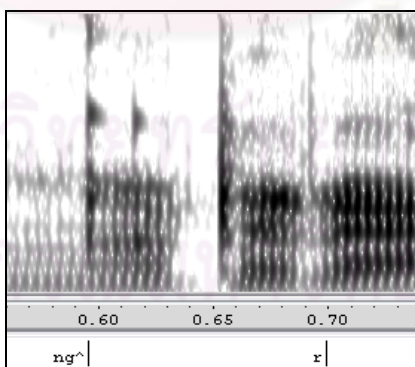
2.4 เสียงกึ่งเสียดแทรก ได้แก่ /c/ /ch/ สามารถระบุจุดเริ่มต้นจากเส้นตามแนวตั้ง ปรากฏบนสเปกโตรแกรมเป็นเช่นเดียวกับลักษณะการที่เกิดจากการสร้างช่องปิด ในเสียงกัก แล้วตามด้วยลักษณะเสียงรบกวน ปรากฏกระจายระจายทั่วไป (ซึ่งเหมือนเสียงเสียดแทรก) หลังจากทีช่องปิดถูกปล่อยออกมา



รูปที่ 3.6 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil/-/c/ และ /sil/-/ch/ ตามลำดับ

จากการกำหนดตำแหน่งขอบเขตเสียงกึ่งเสียดแทรก พบว่ากลุ่มเสียงกึ่งเสียดแทรก อยู่ตำแหน่งเริ่มต้นพยางค์ ดังนั้นจะปรากฏขอบเขตเสียงกึ่งเสียดแทรกระหว่างเสียงสระและระหว่างเสียงเงียบ ดังนั้นค่าลักษณะทางเสียงที่เสนอเพื่อบอกความแตกต่างระหว่างกลุ่มเสียง ได้แก่ ค่าพลังงาน พลังงานตามช่วงองค์ประกอบความถี่ ซึ่งใช้บอกความแตกต่างจากเสียงนี้ออกจากกลุ่มอื่น ค่าอัตราการตัดศูนย์สามารถใช้บอกความเป็นเสียงก้องไม่ก้อง ซึ่งกลุ่มเสียงกึ่งเสียดแทรกมีค่าอัตราการตัดศูนย์สูงกว่ากลุ่มเสียงที่ติดกัน และค่าความถี่ฟอร์แมนท์ซึ่งใช้พิจารณาบอกความแตกต่างจากกลุ่มเสียงสระได้ดี

- 2.5 เสียง ร /r/ เนื่องจากการร่วล้นมีการเปิด-ปิดอย่างรวดเร็ว จึงระบุขอบเขตหน่วยเสียงนี้เมื่อเกิดปรากฏแถบเส้นในแนวตั้ง สลับกับการขาดช่วงพลังงานหรือเกิดเป็นเสียงเงียบเชิงกล หรือพิจารณาจากแถบพลังงานสลับกับส่วนที่ไม่มีพลังงาน เกิดเปลี่ยนแปลงสลับกันไปมาอย่างรวดเร็ว

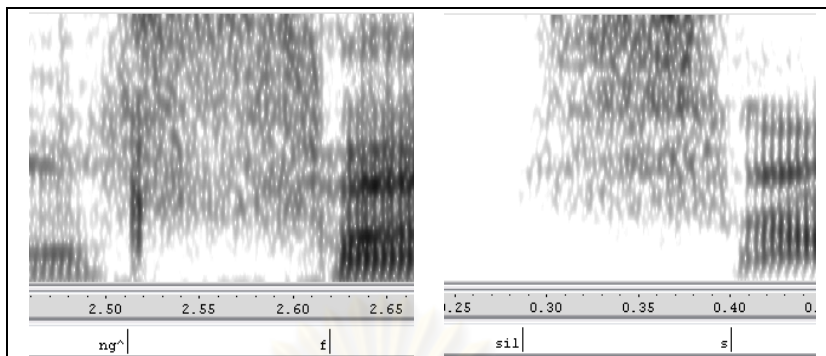


รูปที่ 3.7 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /ng^/-/r/

- 2.6 เสียงเสียดแทรก ได้แก่ /s/ /f/ /h/ ซึ่งระบุหาตำแหน่งเริ่มต้นของเสียดแทรกจากเสียงรบกวนที่ปรากฏกระจายทั่วไปบนสเปกโตรแกรม โดยรูปร่างของเสียงรบกวนที่ปรากฏแตกต่างกันบอกชนิดของหน่วยเสียงที่ต่างกัน นั่นคือ เสียง ฟ /f/ มีสเปกตรัมที่ค่อนข้างแบนในทุกแบนองค์ประกอบความถี่ และไม่มีจุดสูงสุดที่



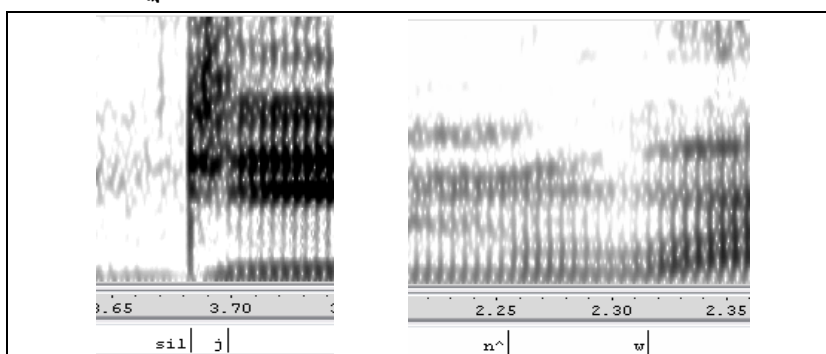
เด่นชัดในช่วงความถี่ใดความถี่หนึ่ง จึงมีเสียงรบกวนกระจายทั่วไป ขณะที่เสียง  $s/s/$  พบลักษณะเสียงรบกวนปรากฏที่องค์ประกอบความถี่สูงเป็นส่วนใหญ่



รูปที่ 3.8 ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง  $/ng^/-/f/$  และ  $/sil/-/s/$  ตามลำดับ

จากการกำหนดตำแหน่งขอบเขตเสียงเสียดแทรก พบว่ากลุ่มเสียงกึ่งเสียดแทรกอยู่ตำแหน่งเริ่มต้นพยางค์ ดังนั้นจะปรากฏขอบเขตเสียงเสียดแทรกระหว่างเสียงสระ เสียงกึ่งสระ เสียงนาสิกและระหว่างเสียงเจียบ ดังนั้นค่าลักษณะทางเสียงที่เสนอเพื่อบอกความแตกต่างระหว่างกลุ่มเสียง ได้แก่ ค่าพลังงาน พลังงานตามช่วงองค์ประกอบความถี่ โดยเฉพาะองค์ประกอบความถี่สูง เนื่องจากค่าพลังงานในองค์ประกอบความถี่สูงของเสียงเสียดแทรกมีค่ามาก ค่าอัตราการตัดศูนย์ซึ่งกลุ่มเสียงเสียดแทรกมีค่าอัตราการตัดศูนย์สูง นอกจากนี้เสนอค่าซึ่งใช้พิจารณาบอกความเป็นเสียงก้อง เช่น ความถี่มูลฐานและค่าความถี่ฟอร์แมนท์เพื่อบอกความแตกต่างจากกลุ่มเสียงสระ เสียงนาสิก และเสียงกึ่งสระ ได้ดี

- 2.7 เสียงกึ่งสระ ได้แก่  $/w/$  มีการเคลื่อนที่ความถี่ฟอร์แมนท์ที่รวดเร็ว มีการเปลี่ยนแปลงมากกว่ากับการเคลื่อนที่ของความถี่ฟอร์แมนท์ที่อยู่ในเสียงสระ การพิจารณาจุดเริ่มต้นและสิ้นสุดจึงพิจารณาการเปลี่ยนแปลงของความถี่ฟอร์แมนท์ที่หนึ่ง ความถี่ฟอร์แมนท์ที่สองและความถี่ฟอร์แมนท์ที่สาม และพิจารณาจากระยะความกว้างระหว่างของความถี่ฟอร์แมนท์ประกอบ ในกรณีนี้เสียงกึ่งสระอยู่ก่อนเสียงสระ สามารถระบุจุดสิ้นสุดจากความถี่ฟอร์แมนท์ของเสียงสระที่ตามมาที่ปรากฏชัดเจน

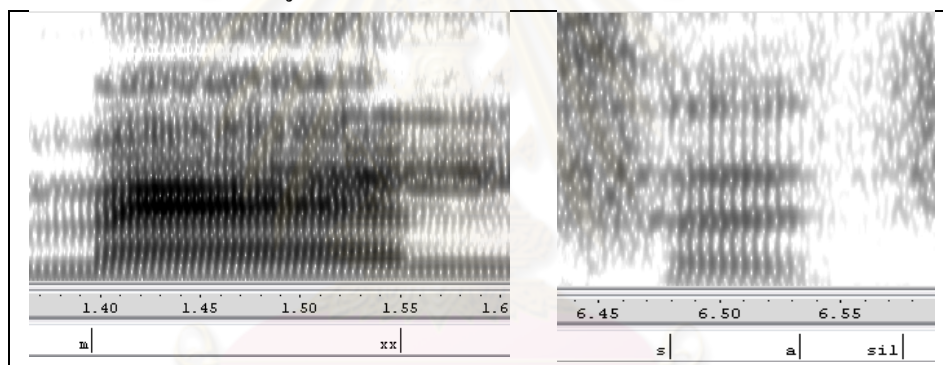




**รูปที่ 3.9** ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /sil-/j/ และ /m^/-/w/ ตามลำดับ

จากการกำหนดตำแหน่งขอบเขตเสียงกึ่งสระ พบว่ากลุ่มเสียงกึ่งสระอยู่ตำแหน่งเริ่มต้นและอยู่ตำแหน่งท้ายพยางค์ ดังนั้นจะปรากฏขอบเขตเสียงกึ่งสระระหว่างเสียงสระ เสียงเสียดแทรกและระหว่างเสียงเงียบ ดังนั้นค่าลักษณะทางเสียงที่เสนอเพื่อบอกความแตกต่างระหว่างกลุ่มเสียง ได้แก่ ค่าอัตราการตัดศูนย์ที่แยกเสียงก้องและเสียงไม่ก้อง ค่าพลังงาน พลังงานตามช่วงองค์ประกอบความถี่ต่าง ๆ นอกจากนี้เสียงกึ่งสระแสดงความถี่ฟอร์แมนท์ชัดเจน ดังนั้น ค่าความถี่ฟอร์แมนท์จึงสามารถบอกความแตกต่างเสียงกึ่งสระจากเสียงในกลุ่มอื่น

**2.8 เสียงสระ** พิจารณาจุดเริ่มต้นจากแถบความเข้มของพลังงานที่เด่นชัดและแถบความเข้มของพลังงานเด่นชัดที่ปรากฏอย่างชัดเจน ดังนั้นการระบุจุดเริ่มต้นและสิ้นสุดของเสียงสระจึงเป็นจุดที่ง่ายต่อการชี้ชัด และกลายเป็นกฎของการระบุหน่วยเสียงที่อยู่ก่อนหน้าเสียงสระหรือเสียงที่ตามมาจากเสียงสระได้



**รูปที่ 3.10** ตัวอย่างของสเปกโตรแกรมของจุดเชื่อมระหว่าง /m/-/xx/ และ /s/-/a/-/sil/ ตามลำดับ

จากการกำหนดตำแหน่งขอบเขตเสียงสระซึ่งเป็นส่วนสำคัญในพยางค์ ดังนั้นการพิจารณาค่าลักษณะทางเสียงเพื่อบอกความแตกต่างของกลุ่มเสียงนั้น จึงวิเคราะห์จากคุณสมบัติลักษณะของเสียงสระเปรียบเทียบกับอีกเสียงที่เป็นคู่พิจารณา เช่น วิเคราะห์จากความเป็นเสียงก้อง จึงเสนอใช้ค่าความถี่มูลฐาน หรือ วิเคราะห์จากลักษณะการออกเสียง จึงเสนอค่าความถี่ฟอร์แมนท์ ซึ่งในเสียงสระนี้ค่าความถี่ฟอร์แมนท์ชัดเจนเด่นชัดที่สุด นอกจากนี้เสนอค่าพลังงาน พลังงานตามช่วงองค์ประกอบความถี่ เป็นค่าลักษณะทางเสียงเพื่อบอกความแตกต่างระหว่างกลุ่มเสียง เป็นต้น

**2.9 เสียงเงียบ** ไม่มีแถบความเข้มใด ๆ ปรากฏอย่างชัดเจน ดังแสดงในรูปตัวอย่างของสเปกโตรแกรมของจุดเชื่อมของเสียงเงียบกับหน่วยเสียง

จากการกำหนดตำแหน่งขอบเขตเสียงเงียบ พบว่าลักษณะเด่นที่พบที่สามารถบอกความแตกต่างกับกลุ่มเสียงอื่น ได้แก่ ค่าพลังงาน

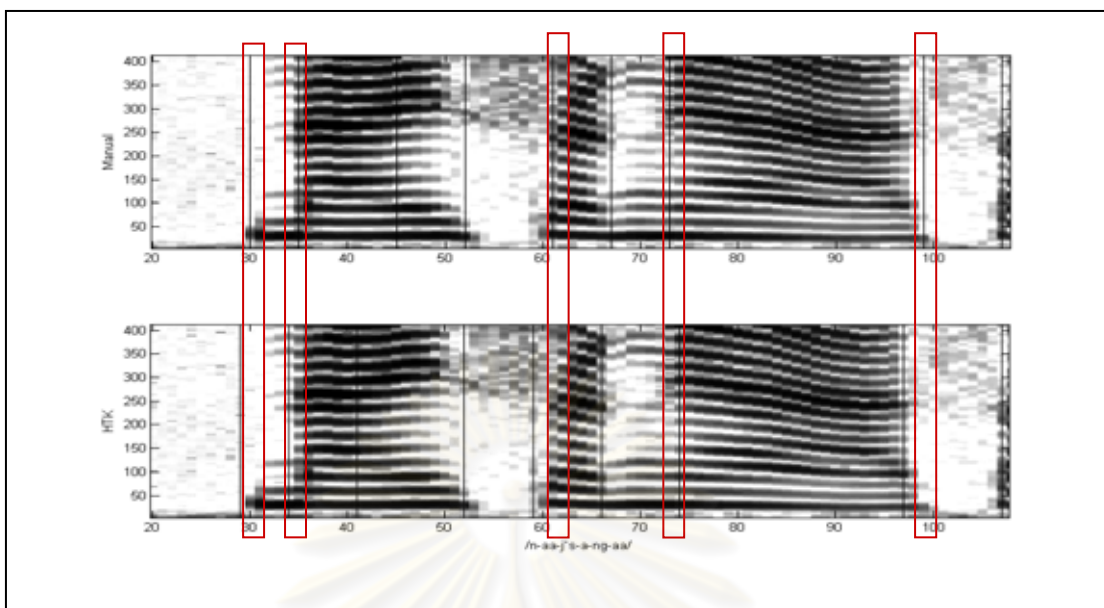
### การสร้างแบบอ้างอิง (Baseline)

วิทยานิพนธ์นี้สร้างแบบอ้างอิงเพื่อเปรียบเทียบประสิทธิภาพของการระบุตำแหน่งขอบเขตเสียงจากวิธีที่นำเสนอ และนำไปสู่การพัฒนากระบวนการหาตำแหน่งขอบเขตหน่วยเสียงที่มีประสิทธิภาพดีขึ้น เนื่องจากวิทยานิพนธ์นี้ใช้ผลจากการระบุหน่วยเสียงแบบบังคับของแบบจำลองฮิดเด้นมาร์คอฟจากเครื่องมือแบบจำลองฮิดเด้นมาร์คอฟ (Hidden Markov Model toolkit: HTK) [39] เพื่อใช้เป็นข้อมูลเบื้องต้นของการหาตำแหน่งขอบเขตหน่วยเสียงบนสัญญาณเสียง ดังนั้นผลการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิงของวิทยานิพนธ์นี้ หมายถึง ผลการระบุตำแหน่งขอบเขตเสียงจากการระบุหน่วยเสียงแบบบังคับของแบบจำลองฮิดเด้นมาร์คอฟ

การประเมินผลประสิทธิภาพของระบบนั้น พิจารณาเปรียบเทียบความแม่นยำของการระบุตำแหน่งขอบเขตเสียงที่ได้จากแบบอ้างอิง กับความแม่นยำของการระบุตำแหน่งขอบเขตเสียงที่ได้จากแบบวิธีที่เสนอ ซึ่งความแม่นยำถูกต้องของตำแหน่งขอบเขตหน่วยเสียงที่เป็นผลลัพธ์ของแบบอ้างอิงวัดเทียบกับตำแหน่งขอบเขตเสียงที่แท้จริง และความแม่นยำถูกต้องของตำแหน่งขอบเขตหน่วยเสียงที่ได้จากระบบที่เสนอเทียบกับตำแหน่งที่แท้จริงเช่นกัน

วิทยานิพนธ์ได้ศึกษาข้อบกพร่องที่เกิดขึ้นจากแบบอ้างอิงเพื่อพัฒนาระบบการหาขอบเขตเสียงให้แม่นยำและมีประสิทธิภาพดีขึ้น โดยวิเคราะห์ข้อบกพร่องที่ปรากฏบนสเปกโตรแกรมของผลการระบุหน่วยเสียงแบบบังคับของแบบจำลองฮิดเด้นมาร์คอฟ ดังแสดงตัวอย่างผลการระบุหน่วยเสียงตามรูปที่ 3.11 ซึ่งเปรียบเทียบกับการระบุตำแหน่งขอบเขตเสียงด้วยคน รายงานว่าตำแหน่งขอบเขตเสียงบางชนิดคลาดเคลื่อนจากตำแหน่งขอบเขตเสียงที่แท้จริง โดยตำแหน่งที่ได้จากการระบุด้วยแบบบังคับของแบบจำลองฮิดเด้นมาร์คอฟนั้น ไม่ได้ระบุตำแหน่งเฟรมที่เกิดการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงจริง

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.11 ตัวอย่างสเปกโตรแกรมของเสียง /n-aa-j<sup>h</sup>-s-a-ng-aa/ ของตำแหน่งขอบเขตเสียงที่แท้จริง (ภาพบน) และการระบุด้วยแบบบังคับของแบบจำลองฮิดเดินมาคอฟ (ภาพล่าง)

### 1. การฝึกฝนแบบจำลองฮิดเดินมาคอฟ

ชุดข้อมูลที่ใช้ในการฝึกฝนแบบจำลองของฮิดเดินมาคอฟนี้ ใช้ชุดข้อมูลที่ 1 ซึ่งเป็นชุดข้อมูลเพื่อฝึกฝนแบบจำลองต่าง ๆ ประกอบด้วยหน่วยเสียงทั้งหมด 55 หน่วยเสียง

การตั้งค่าพารามิเตอร์ในการฝึกฝนแบบจำลองฮิดเดินมาคอฟในขั้นตอนการฝึกฝนครั้งนี้ ใช้ค่าสัมประสิทธิ์เมลฟรีควีนซีเคปสตรอด 13 ค่าและใช้ค่าผลต่างอันดับที่หนึ่งและอันดับที่สองอีก 26 ค่า เป็นเวกเตอร์ลักษณะเฉพาะ 39 ค่า ซึ่งสกัดจากขนาดของหน้าต่างแบบแฮมมิง (Hamming) 20 มิลลิวินาที ด้วยอัตราทุก ๆ 10 มิลลิวินาที การฝึกฝนใช้ฮิดเดินมาคอฟแบบซ้ายไปขวา (left-to-right)

### 2. การทดสอบประสิทธิภาพของแบบอ้างอิง

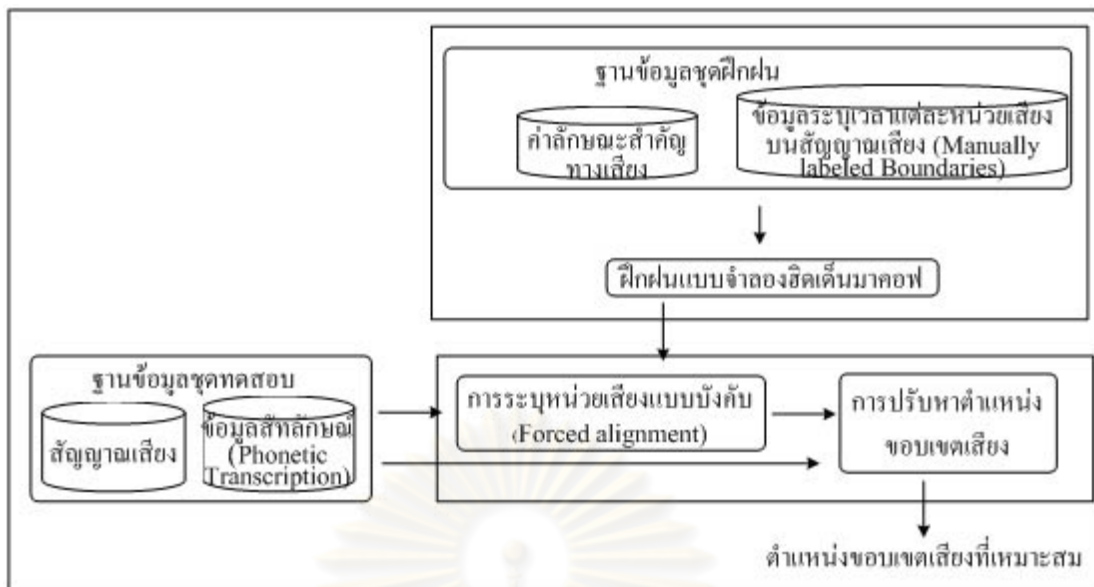
ขั้นตอนการทดสอบประสิทธิภาพของผลที่ได้จากแบบอ้างอิง ใช้ชุดข้อมูลที่ 3 เพื่อทดสอบ โดยข้อมูลชุดนี้แยกกันกับชุดข้อมูลที่ใช้ฝึกฝนแบบจำลองฮิดเดินมาคอฟ การวัดประสิทธิภาพของแบบอ้างอิงนี้เสนอตามวิธีประเมินประสิทธิภาพ โดยวัดความถูกต้องของการบอกตำแหน่งขอบเขตเสียงเมื่อเทียบกับคำตอบที่แท้จริง โดยรายงานตามระดับความคลาดเคลื่อนที่ยอมรับได้และวัดความคลาดเคลื่อนเฉลี่ยของการระบุตำแหน่งของ

แบบอ้างอิงเทียบกับคำตอบตำแหน่งขอบเขตเสียงที่แท้จริง ซึ่งจะอธิบายอย่างละเอียดในหัวข้อประเมินผลระบบ (Evaluation)

### การสร้างระบบหาตำแหน่งขอบเขตเสียง

กระบวนการทำงานในระบบการระบุหาตำแหน่งขอบเขตเสียงแบบอัตโนมัติ ประกอบด้วย การระบุตำแหน่งหน่วยเสียงบนสัญญาณเสียง การหาตำแหน่งขอบเขตเสียง และการปรับหาตำแหน่งขอบเขตเสียง ดังนั้น ระบบหาตำแหน่งขอบเขตของหน่วยเสียงในวิทยานิพนธ์นี้ แบ่งกระบวนการเป็น 2 ส่วนหลัก ดังแสดงตามรูปที่ 3.12 นั่นคือ ส่วนที่หนึ่งใช้การระบุหน่วยเสียงแบบบังคับจากแบบจำลองฮิดเดินมาคอฟโดยใช้ค่าสัมประสิทธิ์เมลเฟรีควีนซีเคปสตรอล เพื่อระบุหน่วยเสียงต่าง ๆ บนสัญญาณโดยอัตโนมัติเมื่อมีลำดับของหน่วยเสียงบนสัญญาณเสียงพูดนั้นเป็นข้อมูลพื้นฐาน โดยการทำงานของการระบุหน่วยเสียงแบบบังคับ (force alignment) ได้ผลลัพธ์ของกระบวนการเป็นการระบุช่วงเวลาแต่ละหน่วยเสียงแสดงบนสัญญาณเสียง จากช่วงเวลาของแต่ละหน่วยเสียงนี้ทำให้สามารถระบุตำแหน่งรอยต่อระหว่างหน่วยเสียง นำไปสู่การหาตำแหน่งขอบเขตหน่วยเสียงซึ่งเป็นตำแหน่งขอบเขตเสียงเริ่มต้น โดยใช้เป็นตำแหน่งสมมติฐานของตำแหน่งขอบเขตเสียงเท่านั้นซึ่งอาจจะเป็นตำแหน่งขอบเขตระหว่างหน่วยเสียงแท้จริงหรือไม่จริงก็ได้ ดังนั้นเมื่อต้องการระบุตำแหน่งขอบเขตหน่วยเสียงที่แท้จริง เนื่องจากการทำงานของการระบุหน่วยเสียงแบบบังคับจากแบบจำลองฮิดเดินมาคอฟเป็นเครื่องมือที่มีประสิทธิภาพในการรู้จำรูปแบบหน่วยเสียงและการแบ่งส่วนของหน่วยเสียง

วิทยานิพนธ์นี้เสนอขั้นตอนการปรับหาตำแหน่งขอบเขตเสียงในการทำงานส่วนที่สอง ซึ่งการทำงานส่วนที่สอง ทำหน้าที่ปรับหาตำแหน่งขอบเขตหน่วยเสียงให้ได้ตำแหน่งขอบเขตที่เหมาะสม โดยใช้ค่าลักษณะสำคัญทางเสียงต่าง ๆ เป็นพารามิเตอร์ในการพิจารณา และมีแนวคิดพื้นฐานการปรับตำแหน่งของเขตหน่วยเสียงทีละคู่ ๆ ตามชนิดขอบเขตหน่วยเสียงนั้นคล้ายกับ [16], [17] และ [18] ซึ่งข้อแตกต่างของวิทยานิพนธ์นี้กับงานวิจัยดังกล่าว คือ การเลือกใช้ค่าลักษณะสำคัญทางเสียงต่าง ๆ ที่แตกต่างกันสำหรับแต่ละชนิดขอบเขตเสียง และวิธีการตัดสินใจหารปรับหาตำแหน่งที่แตกต่างไปจากงานวิจัยอื่นที่ได้เสนอไว้ จากแนวคิดการปรับหาตำแหน่งขอบเขตเสียงตามชนิดของคู่ขอบเขตหน่วยเสียงนี้ ทำให้ต้องศึกษาและวิเคราะห์การเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงต่าง ๆ ตามแต่ละคู่ขอบเขตหน่วยเสียง ตามรายละเอียดการวิเคราะห์ค่าลักษณะสำคัญในบทที่ 4 ต่อไป



รูปที่ 3.12 แผนภาพรวมการทำงานของระบบการหาตำแหน่งขอบเขตเสียงที่เสนอ

ในการทำงานส่วนที่สองนี้ ประกอบด้วยขั้นตอนการทำงานดังนี้ การพิจารณาชนิดขอบเขตหน่วยเสียง การสกัดคำลักษณะสำคัญต่างๆ ที่กรอบพิจารณาปรับขอบเขตหน่วยเสียง และการตัดสินใจเพื่อเลือกตำแหน่งที่เหมาะสมกับตำแหน่งขอบเขตหน่วยเสียงที่แท้จริง โดยวิทยานิพนธ์นี้เสนอการวิธีการตัดสินใจเพื่อเลือกตำแหน่งที่เหมาะสมดังกล่าว ได้แก่

วิธีที่ 1 การหาตำแหน่งขอบเขตเสียงโดยใช้กฎการตัดสินใจที่ตั้งขึ้นเอง

วิธีหาตำแหน่งขอบเขตเสียงจากการพิจารณาดำเนินการที่มีการเปลี่ยนแปลงคำลักษณะสำคัญทางเสียงทันที ซึ่งตำแหน่งเฟรมดังกล่าวแสดงความแตกต่างของคำลักษณะสำคัญทางเสียงระหว่างสองเฟรมที่อยู่ติดกัน

วิธีที่ 2 การหาตำแหน่งขอบเขตเสียงโดยใช้สถิติ

วิธีหาตำแหน่งขอบเขตเสียงโดยสถิตินี้ พิจารณาจากผลที่ได้จากการจำแนกประเภทของกลุ่มที่เป็นตำแหน่งขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงที่แสดงในแต่ละเฟรม ดังนั้นประสิทธิภาพของการปรับหาตำแหน่งขอบเขตเสียงขึ้นอยู่กับผลลัพธ์การจำแนกจากตัวจำแนกซึ่งระบุประเภทกลุ่มขอบเขตเสียงและกลุ่มไม่ใช่ขอบเขตเสียงที่แสดงในแต่ละเฟรม ดังนั้นเพื่อสร้างตัวจำแนกประเภทให้มีประสิทธิภาพ จึงมีการวิเคราะห์ปัจจัยของการสร้างต้นแบบการจำแนกต่างๆ ได้แก่ การเลือกคำลักษณะสำคัญขั้นตอนการฝึกฝนแบบจำลอง การหาเวกเตอร์ลักษณะเฉพาะ เป็นต้น นอกจากนี้เสนอการทดสอบประสิทธิภาพของตัวจำแนก เพื่อวัดประสิทธิภาพของตัวจำแนก

ซึ่งมีรายละเอียดขั้นตอนทั้งหมดของการปรับหาขอบเขตหน่วยเสียงนี้ อธิบายในบทที่ 4 ต่อไป



## ประเมินผลระบบ (Evaluation)

การประเมินประสิทธิภาพของการระบุตำแหน่งขอบเขตเสียงของวิทยานิพนธ์เสนอวิธีการวัดทั้งการประเมินผลแบบพิจารณาผู้กระทำ (subjective evaluation) ที่วัดประสิทธิภาพของอัลกอริทึมที่เสนอ และการประเมินผลแบบพิจารณาผลของกระทำ (objective evaluation) ที่พิจารณาผลที่ได้จากการเสนอประสิทธิภาพของระบบได้ที่เสนอ เช่น วัดประสิทธิภาพระบบเมื่อมีการปรับขนาดของกรอบสัญญาณเสียงพิจารณา เป็นต้น

เมื่อได้ชุดแบบอ้างอิง แล้วออกแบบวิธีวัดประสิทธิภาพของอัลกอริทึมที่เสนอเปรียบเทียบกับชุดแบบอ้างอิง ซึ่งการคำนวณการวัดประสิทธิภาพของอัลกอริทึมที่เสนอจากการพิจารณาร้อยละความถูกต้องของตำแหน่งขอบเขตหน่วยเสียง ได้แก่

1) การวัดความแม่นยำของผลลัพธ์ที่ได้จากวิธีที่เสนอ ซึ่งคำนวณจากสมการที่

(3.1)

$$\% \text{ความแม่นยำของผลลัพธ์} = \frac{\text{จำนวนขอบเขตของผลลัพธ์จากวิธีที่เสนอที่ตรงกับคำตอบจริง}}{\text{จำนวนของขอบเขตของคำตอบทั้งหมด}} \times 100 \quad (3.1)$$

เนื่องจากการยากเมื่อต้องการชี้ว่าตำแหน่งขอบเขตระหว่างหน่วยเสียงที่ได้จากการระบุตำแหน่งนั้นเป็นตำแหน่งที่ดีหรือไม่ดี บางงานวิจัยเสนอว่ายอมให้มีการระบุผิดหน่วยเสียงผิดพลาดหรือคลาดเคลื่อนได้ 10 หรือ 15 มิลลิวินาที หรือในบางงานยอมให้ระยะที่ผิดพลาดมากกว่า 20 มิลลิวินาที ในวิทยานิพนธ์นี้เสนอการวัดประสิทธิภาพของการระบุตำแหน่งขอบเขตเสียง โดยยอมให้วัดผลคำตอบคลาดเคลื่อนจากคำตอบที่แท้จริง มีระดับความคลาดเคลื่อน 1 2 และ 3 เฟรม ซึ่งขนาดของเฟรมคือ 10 มิลลิวินาที โดยกำหนดขนาดของช่วงความคลาดเคลื่อนในการวัดประสิทธิภาพนั้นต้องไม่กว้างมากเกินไป เนื่องจาก ระยะหน่วยเสียงของบางหน่วยเสียงมีขนาดเล็กไม่เกิน 30 มิลลิวินาที หากขนาดของช่วงวัดผลกว้างเกินไปจะทำให้วัดผลการระบุผิดพลาด (วัดผลจากขอบเขตเสียงของหน่วยเสียงอื่น ซึ่งไม่ใช่หน่วยเสียงที่กำลังพิจารณา) ดังนั้น วิทยานิพนธ์นี้จึงเสนอช่วงระยะความคลาดเคลื่อนตามที่เสนอข้างต้น

เกณฑ์มาตรฐานของการวัดประสิทธิภาพของการทำงานแต่ละวิธีที่ใช้เปรียบเทียบผลการทดลองในวิทยานิพนธ์นี้ วัดผลการระบุตำแหน่งขอบเขตเสียงโดยยอมให้มีความคลาดเคลื่อนที่พิจารณาเทียบกับคำตอบที่แท้จริงมีขนาด 1 เฟรม หรือขนาด 10 มิลลิวินาที

การวัดการลดลงของค่าความผิดพลาดของระบบเปรียบเทียบกับค่าความผิดพลาดของชุดแบบอ้างอิง เป็นอีกเครื่องมือที่วัดประสิทธิภาพของอัลกอริทึมที่เสนอจากความถูกต้องของตำแหน่งขอบเขตหน่วยเสียง ตามสูตรดังสมการที่ (3.2)

$$\% \text{ความผิดพลาดที่ลดลง} = \frac{E(\text{Baseline}) - E(\text{Proposed})}{E(\text{Baseline})} \times 100 \quad (3.2)$$

โดยที่ %ความผิดพลาดที่ลดลง หมายถึง ร้อยละความผิดพลาดที่ลดลงเมื่อใช้วิธีที่เสนอ

E(Baseline) หมายถึงค่าความผิดพลาดของตำแหน่งขอบเขตหน่วยเสียงตามชุดแบบอ้างอิง ที่หาตำแหน่งขอบเขตหน่วยเสียงด้วยวิธีการระบุหน่วยเสียงแบบบังคับในแบบจำลองฮิดเดินมาคอฟ

E(Proposed) หมายถึงค่าความผิดพลาดของตำแหน่งขอบเขตหน่วยเมื่อใช้วิธีที่เสนอในการหาตำแหน่งหน่วยเสียง

## 2) การวัดค่าเฉลี่ยความคลาดเคลื่อนเมื่อเทียบกับคำตอบที่แท้จริง

กำหนดให้ “คำตอบที่แท้จริง” คือ การระบุตำแหน่งขอบเขตหน่วยเสียงด้วยคนเป็นตำแหน่งคำตอบที่ถูกต้อง ดังนั้นจึงวัดประสิทธิภาพของระบบจากค่าเฉลี่ยความคลาดเคลื่อนของคำตอบที่ได้จากระบบเทียบกับคำตอบแท้จริง แล้วเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนของระบบนี้กับ ค่าเฉลี่ยความคลาดเคลื่อนที่ได้จากแบบทดลองอ้างอิงซึ่งเทียบกับคำตอบแท้จริง ดังสมการที่ (3.3)

$$\text{ค่าเฉลี่ยความคลาดเคลื่อน} = \frac{\text{ผลรวมของ(ระยะที่คลาดเคลื่อน} \times \text{จำนวนขอบเขตเสียงชนิดนั้น)}}{\text{จำนวนขอบเขตเสียง}} \quad (3.3)$$

โดยที่ ระยะที่คลาดเคลื่อน หมายถึง ระยะห่างของตำแหน่งผลลัพธ์ที่ได้จากวิธีที่เสนอเทียบกับคำตอบที่แท้จริง

นอกจากการวัดประสิทธิภาพรวมทั้งหมดของระบบ ยังเสนอการวัดประสิทธิภาพของแต่ละส่วนประกอบย่อยของระบบ หรือตามสมมติฐานต่าง ๆ เพื่อหาข้อผิดพลาดที่เกิดขึ้นแต่ละส่วนได้ง่ายขึ้น หากกรณีที่มีความผิดพลาดจากส่วนประกอบย่อย ๆ ก็สามารถพัฒนาส่วนนั้นให้ดีขึ้นได้ก่อนที่จะประกอบเข้าเป็นต้นแบบที่สมบูรณ์ต่อไป เช่น การวัดผลที่ได้จากการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง เป็นต้น

วิทยานิพนธ์นี้วัดประสิทธิภาพ ตามระดับความคลาดเคลื่อนที่ยอมรับได้  $\pm 1$  เฟรม เป็นระดับมาตรฐานในการวัดผลการทดลองและเปรียบเทียบประสิทธิภาพวิธีที่เสนอ

## บทที่ 4

### การปรับหาขอบเขตหน่วยเสียง

#### ลำดับขั้นตอนการวิจัย

วิทยานิพนธ์นี้เสนอขั้นตอนการวิเคราะห์ 5 ขั้นตอน โดยวิเคราะห์และหาค่าพารามิเตอร์ที่เหมาะสมเพื่อสร้างต้นแบบของการปรับหาขอบเขตหน่วยเสียง จากนั้นนำต้นแบบที่ได้มาใช้ในขั้นตอนการปรับหาตำแหน่งที่เหมาะสมเพื่อเป็นตำแหน่งขอบเขตหน่วยเสียงที่แท้จริง โดยมีลำดับขั้นตอนทั้งหมดดังนี้

1. การวิเคราะห์ชนิดขอบเขตเสียง
2. การวิเคราะห์ค่าลักษณะสำคัญ
3. การเลือกค่าลักษณะสำคัญสำหรับแต่ละชนิดขอบเขตหน่วยเสียง
4. การวิเคราะห์ขนาดกรอบหน้าต่างการปรับขอบเขตหน่วยเสียง
5. การปรับหาตำแหน่งขอบเขตหน่วยเสียง

#### 1. การวิเคราะห์ชนิดขอบเขตเสียง (Boundary Analysis)

จากแนวคิดการปรับหาตำแหน่งขอบเขตเสียงของวิทยานิพนธ์นี้ ที่เสนอการพิจารณาค่าพารามิเตอร์ที่แตกต่างกันสำหรับแต่ละชนิดขอบเขตเสียง(ตามชนิดของกลุ่มเสียงที่ประกอบในส่วนทางซ้ายและส่วนขวาของขอบเขตเสียง) ดังนั้นการหาต้นแบบของการระบุตำแหน่งขอบเขตเสียงสำหรับทุกชนิดขอบเขตเสียงจึงใช้เวลาในการคำนวณประมวลผลนานและการวิเคราะห์ข้อผิดพลาดและศึกษาวิธีพัฒนาต้นแบบในแต่ละขั้นตอนต้องใช้เวลาาน และเมื่อศึกษาชนิดขอบเขตระหว่างหน่วยเสียงในชุดฐานข้อมูล PD ที่ใช้ในการพัฒนาระบบของวิทยานิพนธ์นี้พบว่า ฐานข้อมูลในวิทยานิพนธ์นี้ประกอบด้วยชนิดขอบเขตเสียงจำนวนหลายชนิด เมื่อพิจารณาจำนวนขอบเขตเสียงจากแต่ละชนิดขอบเขตเสียงนั้นพบว่า จำนวนขอบเขตเสียงของชนิดขอบเขตเสียงบางชนิดนั้นมีจำนวนขอบเขตไม่มากเมื่อเทียบกับจำนวนขอบเขตเสียงในฐานข้อมูลทั้งหมดที่ใช้ในวิทยานิพนธ์นี้ (หรือข้อมูลไม่เพียงพอ ทำให้ไม่สามารถนำมาสร้างกฎหรือต้นแบบการปรับหาตำแหน่งชนิดขอบเขตเสียงนั้นได้) ดังนั้นวิทยานิพนธ์จึงเสนอวิธีปรับหาตำแหน่งขอบเขตเสียงที่ได้จำนวนขอบเขตเสียงมากและมีประสิทธิภาพการระบุตำแหน่งให้ถูกต้องด้วย โดยพิจารณาเลือกปรับตำแหน่งขอบเขตเสียงเฉพาะบางชนิดขอบเขตเสียงที่มีจำนวนขอบเขตเสียงมาก ซึ่งปัจจัยการพิจารณาเลือกชนิดขอบเขตเสียง ได้แก่ จำนวนขอบเขตเสียง (เรียงลำดับจำนวนขอบเขตเสียงแต่ละชนิดจากมากไปหาน้อย) และ ลักษณะการเปลี่ยนแปลงของอะคูสติคที่เกิดขึ้นระหว่างชนิดขอบเขตเสียงนั้นเพื่อการวิเคราะห์หาคุณสมบัติการเปลี่ยนแปลงของค่าลักษณะ

สำคัญทางเสียงมาใช้ในการสร้างต้นแบบได้ต่อไป (ซึ่งลักษณะการเปลี่ยนแปลงที่เกิดขึ้นทันทีระหว่างคู่หน่วยเสียงที่ขอบเขตเสียงนั้น ต้องสามารถพิจารณาจากสเปกโตรแกรมที่มีการเปลี่ยนแปลงของค่าอะคูสติกและสามารถระบุตำแหน่งที่เกิดการเปลี่ยนแปลงได้อย่างชัดเจน)

ดังนั้นสามารถสรุปเงื่อนไขการพิจารณาเลือกชนิดขอบเขตเสียง เพื่อปรับหาค่าตำแหน่งขอบเขตเสียงที่แท้จริง ดังนี้

- 1) เรียงลำดับจำนวนขอบเขตเสียงจากมากไปหาน้อย เลือกชนิดขอบเขตเสียงที่อยู่ในอันดับของจำนวนขอบเขตเสียงที่มีจำนวนมาก เมื่อเปรียบเทียบกับจำนวนขอบเขตเสียงทั้งหมด โดยมีหลักการเลือกชนิดขอบเขตเสียงที่มีจำนวนของตัวอย่างข้อมูลของชุดฝึกฝนการจำแนกมากเพียงพอในการทดสอบจำแนกกลุ่มได้
- 2) เลือกชนิดขอบเขตเสียงที่สามารถระบุการเปลี่ยนแปลงของอะคูสติกอย่างชัดเจน โดยยกเว้นชนิดขอบเขตเสียงที่ประกอบด้วย กลุ่มเสียงรัว เสียงข้างลิ้น หรือชนิดขอบเขตเสียงที่เกิดจากการเชื่อมระหว่างกลุ่มเสียงสระและกลุ่มเสียงกึ่งสระ

## 2. การวิเคราะห์ค่าลักษณะสำคัญทางเสียง (Feature Analysis)

คุณสมบัติของค่าลักษณะสำคัญทางเสียงแสดงคุณสมบัติที่แตกต่างกัน ดังนั้นก่อนที่เลือกค่าลักษณะทางเสียงสำหรับแต่ละชนิดขอบเขตหน่วยเสียงที่แตกต่างกัน จึงวิเคราะห์แต่ละค่าลักษณะ ดังนี้

- 1) สัมประสิทธิ์เมลฟรีเควินซีเคปสตรัม

ค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรัมเป็นค่าลักษณะสำคัญทางเสียงที่นิยมนำมาในการเรียนรู้จำด้วยเสียง คำนวณมาจากการวิเคราะห์แบบเมลสเกล ซึ่งเป็นการวิเคราะห์สัญญาณเสียงตามความสามารถในการรับรู้ของมนุษย์ที่สามารถวิเคราะห์ข้อมูลออกมาได้ (ค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรัมแทนด้วยสัญลักษณ์ MFCC)

- 2) ค่าการเปลี่ยนแปลงรูปร่างของสเปกตรัม

เนื่องจากแนวคิดเสนอว่าตำแหน่งขอบเขตเสียงเกิดการเปลี่ยนแปลงเสียงจากเสียงหนึ่งเป็นอีกเสียงหนึ่ง ซึ่งจากทฤษฎีการเกิดเสียงนั้นอธิบายว่า การ

เปลี่ยนแปลงของรูปร่างของตัวกรองทำให้เกิดการเปลี่ยนแปลงรูปร่างของสเปกตรัมของสัญญาณเสียง ส่งผลให้เสียงที่เปล่งออกมาแตกต่างกัน ดังนั้นค่าการเปลี่ยนแปลงรูปร่างของสเปกตรัมนี้เป็นการวัดการเปลี่ยนแปลงของรูปร่างสเปกตรัมที่ตำแหน่งเฟรมพิจารณาเปรียบเทียบกับตำแหน่งเฟรมอื่นในกรอบพิจารณา และค่านี้สามารถระบุโอกาสของตำแหน่งขอบเขตเสียงได้ ในวิทยานิพนธ์นี้กำหนดให้ค่าการเปลี่ยนแปลงรูปร่างของสเปกตรัม แทนด้วยสัญลักษณ์ STM ซึ่งวิธีการคำนวณค่านี้อธิบายในบทที่ 2

### 3) ความถี่ฟอร์แมนต์

เนื่องจากการเปลี่ยนแปลงของชนิดและรูปร่างของช่องทางเดินเสียงทำให้เกิดการเปลี่ยนแปลงของรูปร่างทอเรียโซแนนท์ ดังนั้นเกิดความถี่ฟอร์แมนท์ที่แตกต่างกัน (อธิบายวิธีการหาค่าฟอร์แมนท์ในบทที่ 2) ซึ่งคุณสมบัติของความถี่ฟอร์แมนท์ที่สามารถใช้วิเคราะห์กลุ่มเสียงสระและกลุ่มเสียงก้องได้ดี ทำให้ใช้เป็นพารามิเตอร์นี้แยกตำแหน่งของกลุ่มเสียงก้องออกจากตำแหน่งกลุ่มเสียงอื่น ได้ ซึ่งวิทยานิพนธ์นี้เสนอการใช้ความถี่ฟอร์แมนท์ที่หนึ่ง สอง และสาม ซึ่งแทนด้วยสัญลักษณ์ F1 F2 และ F3 ตามลำดับ

นอกจากนี้ค่าความกว้างระหว่างความถี่ฟอร์แมนท์ที่หนึ่งและสอง ซึ่งแทนด้วยสัญลักษณ์ F2F1 เป็นค่าที่บ่งชี้การเปลี่ยนแปลงของความถี่ฟอร์แมนท์ที่หนึ่งและสอง บอกความสัมพันธ์ของเส้นทางการเคลื่อนที่ของความถี่ฟอร์แมนท์ที่หนึ่งและสองได้ ซึ่งเส้นทางการเคลื่อนที่นี้จะเป็นลักษณะที่ปรากฏเด่นชัดในหน่วยเสียงของกลุ่มเสียงก้อง เช่น เสียงสระ เสียงกึ่งสระ และเสียงนาสิก เป็นต้น ดังนั้นค่านี้สามารถเลือกใช้เป็นพารามิเตอร์แยกตำแหน่งที่เป็นกลุ่มเสียงก้องออกจากกลุ่มเสียงอื่นได้

### 4) ค่าพลังงาน

- a. ค่าพลังงานเสียง แทนด้วยสัญลักษณ์ TOTALENERGY ซึ่งค่าพลังงานเสียงสามารถบอกความเป็นลักษณะของกลุ่มเสียงต่างๆ โดยอธิบายวิธีการคำนวณค่านี้อธิบายในบทที่ 2
- b. ค่าพลังงานตามช่วงองค์ประกอบความถี่ต่างๆ แบ่งเป็นพลังงานในช่วงความถี่ที่หนึ่ง สอง สาม ลี ห้า แทนด้วยสัญลักษณ์ E1, E2, E3,



E4, E5 ตามลำดับ การเลือกใช้ค่าพลังงานในแต่ละช่วงเนื่องจากลักษณะการออกเสียงในแต่ละกลุ่มเสียงแตกต่างกัน เช่น เสียงในกลุ่มที่เป็นเสียงก้องจะมีพลังงานในองค์ประกอบความถี่ต่ำกว่าเสียงไม่ก้อง ทางตรงกันข้ามเสียงที่ไม่เป็นรายคาบ จะมีค่าพลังงานในองค์ประกอบความถี่สูงมากกว่า เป็นต้น

c. ค่าล็อกของพลังงาน ซึ่งคำนวณจากค่าล็อกการิทึมของพลังงาน จากการศึกษางานวิจัยที่เสนอการพิจารณาค่าลักษณะสำคัญทางเสียงที่เป็นพารามิเตอร์ช่วยในการจำแนกกลุ่มเสียงก้องและเสียงไม่ก้อง พบว่าค่าล็อกของพลังงานเป็นค่าชี้ความแตกต่างของสมบัติความเป็นเสียงก้องและเสียงไม่ก้องได้ แทนด้วยสัญลักษณ์ LOGEN

d. ค่าอัตราส่วนพลังงานในช่วงความถี่ [0, 357 เฮิรตซ์] และ [358, 5373 เฮิรตซ์] ซึ่งแทนด้วยสัญลักษณ์ ENSN ซึ่งการเลือกใช้อัตราส่วนพลังงานช่วงนี้เพื่อเป็นค่าบ่งชี้ความแตกต่างส่วนที่เป็นเสียงนาสิก เสียงกึ่งสระ และเสียงสระ เนื่องจากพลังงานในองค์ประกอบความถี่ต่ำของเสียงก้องจะมีค่ามากกว่าเสียงไม่ก้อง และพลังงานในองค์ประกอบความถี่สูงของเสียงก้องจะมีค่าน้อยกว่าเสียงไม่ก้อง ดังนั้นอัตราส่วนระหว่างพลังงานที่องค์ประกอบความถี่ต่ำกับพลังงานที่องค์ประกอบความถี่สูงของเสียงก้องจึงมีค่ามากกว่าอัตราส่วนระหว่างพลังงานที่องค์ประกอบความถี่ต่ำกับพลังงานที่องค์ประกอบความถี่สูงของเสียงไม่ก้อง

5) ความสัมพันธ์ระหว่างเฟรมที่สนใจและเฟรมที่อยู่ก่อนหน้า

การพิจารณาความสัมพันธ์ของสัญญาณเสียงระหว่างเฟรมที่ติดกัน แทนด้วยสัญลักษณ์ CORR ซึ่งค่านี้คำนวณจากความสัมพันธ์ระหว่างเฟรมที่ติดกัน มีค่าสูงแสดงว่า ตำแหน่งที่ติดกันมีโอกาสที่จะเป็นขอบเขตเสียงน้อย ขณะที่ค่าความสัมพันธ์ระหว่างเฟรมที่ติดกันมีค่าต่ำ แสดงว่ามีโอกาสที่เกิดเป็นขอบเขตเสียงมากกว่า

6) ความถี่มูลฐาน

ค่าความถี่มูลฐานสามารถบอกความเป็นคาบ ค่าความถี่มูลฐานในเสียง  
ก้องจะมีค่าสูงกว่าหน่วยเสียงไม่ใช่เสียงก้อง ในวิทยานิพนธ์นี้คำนวณ  
ค่าความถี่มูลฐานจากเครื่องมือ praat ซึ่งแทนด้วยสัญลักษณ์ F0

### 7) อัตราการตัดศูนย์

เนื่องจากสัญญาณเสียงมีการเปลี่ยนแปลงตัดผ่านแกนเวลา เสียงก้องจะ  
เกิดการเปลี่ยนแปลงเป็นรายคาบมีจำนวนครั้งตัดผ่านเวลาดำกว่าเสียงไม่ก้อง  
สรุปได้ว่า เสียงไม่ก้องเสียดแทรกมีค่าอัตราการตัดผ่านศูนย์สูง ขณะที่ค่าเสียง  
ก้องเสียดแทรกมีค่าอัตราการตัดผ่านศูนย์ต่ำ ดังนั้นค่าอัตราการตัดศูนย์เป็น  
พารามิเตอร์สำคัญเพื่อใช้บ่งชี้ความแตกต่างระหว่างกลุ่มเสียงก้องและเสียงไม่  
ก้องได้ ค่าอัตราการตัดศูนย์ซึ่งแทนด้วยสัญลักษณ์ ZCR (อธิบายวิธีการ  
คำนวณค่านี้อธิบายในบทที่ 2)

## 3. การเลือกค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียง

เนื่องจากทฤษฎีกำเนิดเสียงและตัวกรองของกลไกการเกิดเสียงของมนุษย์ เมื่อ  
เสียงเดินทางผ่านตามช่องทางเดินเสียง รูปร่างของสเปกตรัมถูกเปลี่ยนแปลงรูปร่างไปตามตัวกรอง  
และตัวกำเนิดเสียง ดังนั้นบริเวณรอยต่อระหว่างหน่วยเสียงจึงเกิดการเปลี่ยนแปลงของรูปร่าง  
สเปกตรัมทันทีส่งผลให้ค่าลักษณะสำคัญทางเสียงที่สกัดได้มีค่าเปลี่ยนแปลงตามเช่นกัน ดังนั้น  
วิทยานิพนธ์นี้ใช้ค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงมาเป็นข้อมูลเพื่อพิจารณาหา  
ตำแหน่งขอบเขตเสียง ซึ่งการพิจารณาค่าการเปลี่ยนแปลงลักษณะสำคัญทางเสียง จะวิเคราะห์เลือก  
ชุดค่าลักษณะสำคัญทางเสียงที่เหมาะสมสำหรับแต่ละชนิดขอบเขตเสียง และค่าลักษณะสำคัญทาง  
เสียงต่างๆที่นำมาวิเคราะห์ในวิทยานิพนธ์นี้ได้จากค่าที่เสนอในหัวข้อการวิเคราะห์ค่าลักษณะ  
สำคัญทางเสียง เช่น ค่าสัมประสิทธิ์เมลฟรีควีนซีเคปสตรอล ค่าพลังงาน ค่าพลังงานตามช่วง  
องค์ประกอบความถี่ ค่าอัตราการตัดศูนย์ ค่าความถี่มูลฐาน ค่าความสับสนพ้อง และค่าการ  
เปลี่ยนแปลงรูปร่างของสเปกตรัม [42] เป็นต้น

วัตถุประสงค์หลักของการเลือกชนิดค่าลักษณะสำคัญสำหรับแต่ละชนิดขอบเขตเสียง  
คือ เพื่อหาข้อมูลที่สามารถแบ่งแยกกลุ่มหน่วยเสียงระหว่างขอบเขตเสียงได้ (นั่นคือ สามารถชี้  
ความแตกต่างของกลุ่มเสียงทางซ้ายขอบเขตเสียงและกลุ่มเสียงทางขวาขอบเขตเสียงได้) การเลือก  
ค่าลักษณะทางเสียงที่เหมาะสมกับแต่ละชนิดขอบเขตเสียงนั้นใช้ชุดข้อมูลที่ 1 (ตามหัวข้อ “ข้อมูล  
ฝึกฝนและทดสอบ” ในบทที่ 3) ซึ่งใช้วิเคราะห์ความสามารถในการจำแนกกลุ่ม (การวิเคราะห์  
ประสิทธิภาพการจำแนกกลุ่มนั้นใช้ค่าการเปลี่ยนแปลงของลักษณะสำคัญค่าเป็นข้อมูลการแยก

กลุ่ม) วิเคราะห์แผนภาพกล่องของค่าการเปลี่ยนแปลงลักษณะสำคัญทางเสียงระหว่างกลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง หากค่าการวิเคราะห์ความแปรปรวน (ANOVA) เป็นต้น โดยขั้นตอนนี้มีเงื่อนไขพิจารณาแต่ละชนิดขอบเขตเสียง ซึ่งอธิบาย แต่ละวิธีการวิเคราะห์ ดังนี้

### 3.1. การวิเคราะห์จากแผนภาพกล่อง

เมื่อพิจารณาชนิดขอบเขตหน่วยเสียงที่แตกต่างกัน พบว่าค่าลักษณะสำคัญทางเสียงแสดงคุณสมบัติที่แตกต่างกัน ดังนั้นการพิจารณาตัวพารามิเตอร์เพื่อจำแนกกลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงสามารถวิเคราะห์ได้จากแผนภาพกล่องของค่าลักษณะสำคัญทางเสียง ซึ่งแสดงค่ากลางและการกระจายตัวของค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงสำหรับแต่ละกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงและใช้เปรียบเทียบการแจกแจงข้อมูล

แผนภาพกล่องอธิบายความสามารถการจำแนกกลุ่มข้อมูลชุดขอบเขตเสียงและชุดที่ไม่ใช่ขอบเขตเสียง เมื่อใช้ค่าพารามิเตอร์เป็นชุดข้อมูลการแยก โดยวิเคราะห์จากค่ากลางของแผนภาพกล่องและความกว้างของแผนภาพกล่อง เปรียบเทียบค่ากลางของแผนภาพกล่องระหว่างกลุ่มข้อมูลทั้งสองกลุ่ม พิจารณาค่ากลางของแผนภาพถ้าค่ามัธยฐานอยู่ที่ตำแหน่งกลางของกล่องแสดงว่า การแจกแจงของข้อมูลมีความสมมาตรไม่เบ้ไปทางข้อมูลด้านบนหรือด้านล่าง และพิจารณาความกว้างของแผนภาพกล่อง ถ้าความกว้างมากแสดงว่าข้อมูลมีการกระจายมาก ถ้าความกว้างน้อยแสดงว่าข้อมูลมีการกระจายน้อย (ซึ่งแสดงว่าข้อมูลในแต่ละกลุ่มกระจายน้อย ข้อมูลมีค่าใกล้เคียงกัน) ดังนั้นถ้าพิจารณาแล้วพบว่าค่ากลางของแต่ละแผนภาพกล่องแตกต่างกัน และช่วงกว้างของแผนภาพกล่องทั้งสองกลุ่มไม่ซ้อนทับกัน แสดงว่าโอกาสของการจำแนกกลุ่มระหว่างสองกลุ่มได้ดี ดังนั้นจึงเลือกค่าลักษณะทางเสียงนี้ไปพิจารณาในขั้นตอนถัดไป

### 3.2. การวิเคราะห์จากผลทดสอบประสิทธิภาพของการจำแนกกลุ่ม

การพิจารณาความสามารถของค่าลักษณะทางเสียงนั้น สามารถวิเคราะห์ได้จากการใช้ค่าลักษณะทางเสียงไปจำแนกกลุ่ม แล้ววัดผลความถูกต้องการจำแนกกลุ่ม จากการใช้ชุดข้อมูลที่ 2 เพื่อหาชุดค่าลักษณะสำคัญเป็นตัวแทนข้อมูลการจำแนก

- 1) ความแม่นยำการจำแนก

ค่าที่บ่งชี้ความสามารถการแบ่งแยกข้อมูลออกเป็นกลุ่ม สามารถพิจารณาได้จากประสิทธิภาพผลการจำแนกกลุ่มข้อมูล ซึ่งในวิทยานิพนธ์นี้ หมายถึง ร้อยละความถูกต้องของผลที่ได้จากการจำแนกกลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงเมื่อใช้แต่ละค่าลักษณะสำคัญทางเสียงเป็นข้อมูลชุดฝึกฝนและชุดทดสอบ ผลที่ได้จากการจำแนกนี้เปรียบเทียบกับคำตอบที่แท้จริงซึ่งคำตอบกลุ่มตำแหน่งขอบเขตเสียงได้จากตำแหน่งที่เป็นตำแหน่งขอบเขตเสียง และคำตอบกลุ่มที่ไม่ใช่ตำแหน่งขอบเขตเสียง ได้จากตำแหน่งที่ไม่ใช่ตำแหน่งขอบเขตเสียง

## 2) ความผิดพลาดของผลการจำแนก

การวิเคราะห์ความสามารถในการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง สามารถพิจารณาได้จากค่าความผิดพลาดของจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงเมื่อใช้ชุดค่าลักษณะทางเสียงมาเป็นตัวแทนข้อมูล ค่าความน่าจะเป็นความผิดพลาดการจำแนกนี้ บอกความสามารถของการแบ่งแยกกลุ่มที่พิจารณาออกเป็นสองส่วน เมื่อค่าความผิดพลาดการจำแนกมีค่ามากแสดงว่าความสามารถน้อยในการแยกข้อมูลสองกลุ่มออกจากกัน และเมื่อค่าความผิดพลาดการจำแนกน้อยหมายถึงระบบสามารถแยกหรือจำแนกข้อมูลสองกลุ่มออกจากกันได้ดี

### 3.3. ความสัมพันธ์ของค่าลักษณะสำคัญทางเสียง (correlation)

เมื่อเลือกได้ชุดค่าลักษณะสำคัญที่มีคุณสมบัติบ่งชี้ความแตกต่างของกลุ่มได้ดีแล้วจากการวิเคราะห์ประสิทธิภาพการจำแนกแล้ว นำมาวิเคราะห์ความสัมพันธ์ของค่าลักษณะสำคัญนี้เพื่อลดการซ้ำซ้อนของข้อมูล โดยในขั้นตอนนี้พิจารณาทีละคู่ค่าลักษณะสำคัญและวัดผลการหาความสัมพันธ์ของทั้งสองค่าลักษณะทางเสียงนั้น ซึ่งค่าความสัมพันธ์นี้สามารถคำนวณได้ตามสมการ (2.11) โดยมีเงื่อนไขการพิจารณาจากค่าความสัมพันธ์ คือ เมื่อค่าความสัมพันธ์ มีค่ามากกว่า 0.8 ถือว่า การพิจารณาค่าลักษณะสำคัญทางเสียงคู่นั้นมีความสัมพันธ์มาก

ขั้นตอนนี้พิจารณาค่าการกระจายตัวของแต่ละค่าลักษณะสำคัญและความสามารถการจำแนกกลุ่ม มีรายละเอียดดังนี้

- 1) พิจารณาชนิดขอบเขตเสียงชนิดที่หนึ่ง
- 2) เลือกค่าการเปลี่ยนแปลงลักษณะสำคัญแต่ละลักษณะมาพิจารณาค่าการกระจายตัวของแต่ละค่าลักษณะสำคัญจากแผนภาพกล่อง

3) พิจารณาค่าร้อยละความถูกต้องของผลการจำแนกและพิจารณาความน่าจะเป็นของความผิดพลาดที่เกิดขึ้นระหว่างสองกลุ่ม เมื่อใช้ค่าการเปลี่ยนแปลงลักษณะสำคัญแต่ละค่า มาเป็นข้อมูลจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง

4) วนซ้ำกลับไปทำข้อ 2) เพื่อเลือกค่าการเปลี่ยนแปลงลักษณะสำคัญ ชนิดถัดไป มาพิจารณาจนกระทั่งครบทุกชนิดของค่าลักษณะสำคัญที่เสนอ จากนั้นตัดสินใจเลือกชนิดค่าลักษณะสำคัญที่มีค่าเฉลี่ยของการกระจายตัวของค่าลักษณะสำคัญในชุดกลุ่มขอบเขตเสียงและกลุ่มไม่ใช่ขอบเขตเสียงที่มีค่าเฉลี่ยแตกต่างกัน มีค่าร้อยละความถูกต้องการจำแนกมากกว่าร้อยละ 60 และมีความน่าจะเป็นของความผิดพลาดของการจำแนกระหว่างสองกลุ่มน้อยกว่า 0.4

5) เมื่อเลือกได้ชุดค่าลักษณะสำคัญที่สามารถการจำแนกได้ดี จากนั้นวิเคราะห์ความสัมพันธ์ (correlation) เพื่อกำจัดค่าที่ซ้ำซ้อนออกไป โดยพิจารณาจากค่าความสัมพันธ์ที่ละคู่ ค่าลักษณะสำคัญทางเสียง หากค่าความสัมพันธ์เกิน 0.8 ถือว่าค่าลักษณะสำคัญทางเสียงในคู่ที่พิจารณา มีความสัมพันธ์มาก จึงคัดเลือกค่าลักษณะสำคัญที่ทำให้ผลความถูกต้องการจำแนกกลุ่มที่มีค่ามากกว่าเป็นค่าที่ลักษณะทางเสียงเป็นค่าที่ใช้ในชุดตัวแทนเสียง และกำจัดค่าลักษณะทางเสียงที่ให้ผลความถูกต้องน้อยกว่าออกไปจากชุดลักษณะทางเสียงของชนิดขอบเขตเสียงนั้น

6) ทำวนซ้ำในข้อ 1) เพื่อไปพิจารณาในชนิดขอบเขตเสียงอื่น ๆ จนครบทุกชนิดขอบเขตเสียง

ผลการวิเคราะห์จากแต่ละวิธีในขั้นตอนการเลือกค่าลักษณะสำคัญทางเสียงที่เหมาะสมสำหรับแต่ละชนิดขอบเขตเสียง แสดงผลอย่างละเอียดของแต่ละค่าลักษณะสำคัญสำหรับแต่ละชนิดขอบเขตเสียงในภาคผนวก ง

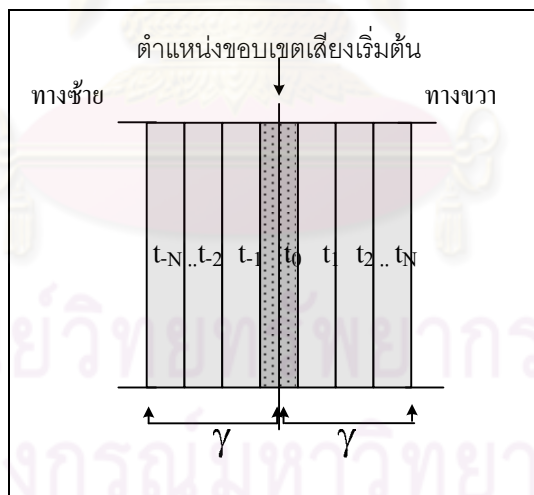
#### 4. การวิเคราะห์ขนาดกรอบหน้าต่างการปรับ (Refining window)

ขนาดกรอบหน้าต่างการปรับขอบเขตเสียงเป็นองค์ประกอบสำคัญต่อผลความแม่นยำของการหาตำแหน่งขอบเขตเสียง ขนาดที่เหมาะสมขึ้นอยู่กับลักษณะของตำแหน่งเริ่มต้นที่ได้จากแบบจำลองฮิดเดินมาคอฟ (ตำแหน่งของผลลัพธ์จากแบบจำลองฮิดเดินมาคอฟนำมาเป็นตำแหน่งขอบเขตเริ่มต้นของการปรับหาขอบเขตเสียงที่ถูกต้อง) ขนาดกรอบที่ใหญ่จะครอบคลุมตำแหน่งขอบเขตที่แท้จริงที่คลาดเคลื่อนห่างจากตำแหน่งเริ่มต้นที่ได้ โดยตำแหน่งขอบเขตเสียงที่แท้จริงในบางชนิดขอบเขตเสียงอยู่ห่างจากตำแหน่งเริ่มต้นมากกว่าขนาดกรอบหน้าต่างบริเวณที่พิจารณา แต่ข้อเสียของขนาดกรอบหน้าต่างการปรับที่ใหญ่คือ การตรวจพบตำแหน่งที่ผิดพลาด (detect the false alarm) ในทางตรงกันข้าม ขนาดกรอบหน้าต่างการปรับที่เล็กจะสามารถตรวจเลือกตำแหน่งที่มีโอกาสเป็นตำแหน่งขอบเขตเสียงที่แท้จริงที่อยู่ใกล้กับตำแหน่งขอบเขตเสียงเริ่มต้นที่ได้จากแบบจำลองฮิดเดินมาคอฟ แต่ข้อเสียของวิธีนี้คือ เป็นวิธีการหาคำตอบที่ไม่ยุติธรรมต่อการเลือกพิจารณาหาตำแหน่งที่น่าจะมีโอกาสเลือกเป็นตำแหน่งขอบเขตเสียง อาจจะทำให้ขนาด



กรอบหน้าต่างการปรับมีขนาดเล็กมากไม่ครอบคลุมตำแหน่งที่เป็นคำตอบที่แท้จริง ดังนั้นการพิจารณาขนาดกรอบหน้าต่างการปรับขอบเขตเสียง จึงต้องศึกษาและวิเคราะห์ลักษณะความผิดพลาดที่เกิดจากการระบุตำแหน่งขอบเขตเสียงที่ได้จากแบบจำลองฮิดเดินมาคอฟ ซึ่งพิจารณาค่าเฉลี่ยของความคลาดเคลื่อนได้จากการวัดระยะห่างจากการระบุขอบเขตเสียงคลาดเคลื่อนไปจากคำตอบที่แท้จริง และค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนดังกล่าว ดังนั้นเฟรมของสัญญาณเสียงที่อยู่ภายในกรอบหน้าต่างการปรับหาขอบเขตเสียงทุกเฟรม มีโอกาสเป็นตำแหน่งขอบเขตเสียงแท้จริง โดยมีหลักการพิจารณาเลือกเฟรมเพื่อเป็นตำแหน่งขอบเขตเสียงจากสมมติฐานของการพิจารณาสัญญาณเสียงในกรอบช่วงเวลาสั้นๆ แล้ว พบว่าเฟรมที่มีการเปลี่ยนแปลงของค่าอะคูสติคมากที่สุดเป็นตำแหน่งเฟรมที่เหมาะสมมากที่สุด ซึ่งจะอธิบายรายละเอียดการลักษณะการเปลี่ยนแปลงต่อไป

ตำแหน่งเริ่มต้นของขั้นตอนการปรับหาตำแหน่งขอบเขตเสียงจากผลลัพธ์การระบุตำแหน่งขอบเขตหน่วยเสียงจากฮิดเดินมาคอฟ ซึ่งแทนตำแหน่งเริ่มต้นด้วยสัญลักษณ์  $t_0$  จากนั้นสร้างกรอบพิจารณาการปรับหาตำแหน่งขอบเขตเสียงเป็นตำแหน่งอ้างอิงของกรอบหน้าต่างการปรับ โดยขยายออกไปทางซ้ายและขวาของตำแหน่งเริ่มต้นด้วยจำนวนกรอบ  $\gamma$  เฟรม ดังแสดงตามรูปที่ 4.1 ดังนั้นทุกตำแหน่งเฟรมภายในกรอบหน้าต่างการปรับเป็นตำแหน่งที่มีโอกาสเลือกตอบเป็นตำแหน่งขอบเขตเสียงที่แท้จริง



รูปที่ 4.1 การสร้างกรอบพิจารณาการหาตำแหน่งขอบเขตเสียง

กำหนดกรอบพิจารณาที่มีพื้นที่  $A(t_0)$  มีขนาด  $t$  เฟรม โดยคำนวณจากสมการที่ (4.1)

$$A(t_0) = \{t \mid t_0 - \gamma \leq t \leq t_0 + \gamma\} \quad (4.1)$$

โดยที่  $t_0$  แทนตำแหน่งขอบเขตเสียงสมมติฐานที่ได้จากการระบุแบบบังคับจากฮิดเดินมาคอฟ

$\gamma$  แทนจำนวนเฟรมที่ต้องการขยายในการปรับหาตำแหน่งขอบเขตเสียง

เพื่อวิเคราะห์หาค่าจำนวนกรอบที่เหมาะสม วิทยานิพนธ์นี้จึงวิเคราะห์ค่าจำนวนกรอบจากผลเฉลี่ยความคลาดเคลื่อนของผลลัพธ์แบบอ้างอิงเทียบกับคำตอบที่ระบุตำแหน่งขอบเขตด้วยคน ซึ่งได้รายงานค่าเฉลี่ยของความคลาดเคลื่อนผลลัพธ์แบบอ้างอิงเทียบกับคำตอบที่ตำแหน่งขอบเขตที่แท้จริงในบทต่อไป

## 5. การปรับหาค่าตำแหน่งขอบเขตหน่วยเสียง (Phone boundary refinement)

จากขั้นตอนการวิเคราะห์ชนิดขอบเขตเสียง การวิเคราะห์ค่าลักษณะสำคัญ การเลือกค่าลักษณะสำคัญสำหรับแต่ละชนิดขอบเขตเสียง และการวิเคราะห์ขนาดกรอบหน้าต่างการปรับขอบเขตหน่วยเสียง ขั้นตอนนี้จะนำค่าพารามิเตอร์ที่วิเคราะห์ได้มาสร้างระบบการปรับหาค่าตำแหน่งหน่วยเสียง เพื่อระบุหาค่าตำแหน่งที่เหมาะสมในกรอบพิจารณา โดยขั้นตอนนี้เสนอวิธีการปรับหาค่าตำแหน่งหน่วยเสียงด้วยการตัดสินใจด้วยกฎที่ตั้งขึ้นเอง และการใช้สถิติเพื่อระบุตำแหน่งขอบเขตเสียง ซึ่งมีรายละเอียดดังนี้

### 5.1. การปรับหาค่าตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ [43]

การปรับหาค่าตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจนี้ พิจารณาค่าตำแหน่งเฟรมที่มีการเปลี่ยนแปลงค่าลักษณะสำคัญทางเสียงทันที ดูจากความแตกต่างของค่าลักษณะสำคัญทางเสียงระหว่างสองกรอบสัญญาณเสียงที่อยู่ติดกัน (ซึ่งหมายถึงผลต่างของค่าลักษณะสำคัญทางเสียงตำแหน่งที่สนใจเทียบกับตำแหน่งก่อนหน้า) ในการหาค่าตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ ใช้ชุดข้อมูลที่ 2 ในขั้นตอนการสร้างกฎและพัฒนาระบบ จากนั้นเมื่อได้กฎสำหรับแต่ละชนิดขอบเขตหน่วยเสียง จึงนำไปทดสอบการระบุตำแหน่งขอบเขตเสียงเมื่อตัดสินใจจากกฎที่ตั้งขึ้นในชุดข้อมูลที่ 3

#### 5.1.1. ขั้นตอนการสร้างกฎตัดสินใจ

เมื่อสกัดค่าลักษณะสำคัญทางเสียงแล้ว จากนั้นศึกษาและวิเคราะห์คุณสมบัติของค่าเหล่านั้น ๆ เพื่อสร้างกฎการตัดสินใจพื้นฐานก่อนนำค่าลักษณะสำคัญนั้นไปใช้เป็นตัวแปรในการตัดสินใจ โดยในขั้นตอนการสร้างกฎตัดสินใจนี้ใช้ตำแหน่งคำตอบที่แท้จริงมาเป็นตำแหน่งที่พิจารณาคูสมบัตินี้ของค่าลักษณะสำคัญทางเสียงมาเป็นข้อมูลเบื้องต้น โดยอธิบายขั้นตอนการวิเคราะห์อย่างละเอียด คือ

- 1) พิจารณานิดขอบเขตหน่วยเสียง
- 2) ศึกษาพฤติกรรมของค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงทีละค่าตามชนิดขอบเขตเสียงนั้น ๆ หากมีการแสดงพฤติกรรมการเปลี่ยนแปลงค่า

นั่นทันที ที่ตำแหน่งของขอบเขตเสียงที่แท้จริงกับตำแหน่งเฟรมก่อนหน้า จึงเลือกค่าลักษณะสำคัญทางเสียงนั้นมาเพิ่มเข้าไปในการสร้างกฎการตัดสินใจ สำหรับแต่ละชนิดของขอบเขตเสียง (การศึกษาพฤติกรรมในขั้นตอนนี้พิจารณาจากสัญญาณเสียงต่าง ๆ บนฐานข้อมูลชุดที่ 2 ซึ่งสังเกตพฤติกรรมที่ปรากฏบนสัญญาณเสียงหลายไฟล์)

- 3) นอร์มอลไลซ์ค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียง
- 4) ทดสอบกฎเมื่อพิจารณาค่าลักษณะสำคัญตามที่เลือกในข้อ 2) เลือกตำแหน่งเฟรมที่มีการเปลี่ยนแปลงมากที่สุด ซึ่งอธิบายไว้ในหัวข้อ “การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ”
- 5) วัดผลการระบุหาตำแหน่งขอบเขตเสียง พิจารณาผลการระบุหาตำแหน่งขอบเขตเสียงหากมีประสิทธิภาพการระบุลดลง จึงสรุปว่าค่าลักษณะสำคัญทางเสียงค่านั้นไม่เหมาะสมสำหรับการปรับหาตำแหน่งขอบเขตเสียง และเลือกค่าลักษณะสำคัญค่าอื่นมาทดสอบแทน ทำซ้ำในข้อ 2) 3) 4) และ 5) ไปเรื่อย ๆ จนได้ผลการหาขอบเขตเสียงที่พอใจ ซึ่งหมายถึง เมื่อเพิ่มการพิจารณาค่าลักษณะสำคัญทางเสียงนั้นเข้าไปช่วยในกฎการปรับหาขอบเขตเสียงชนิดนั้นแล้ว ไม่สามารถเพิ่มประสิทธิภาพของการระบุหาตำแหน่งขอบเขตเสียงที่แท้จริงได้หรือไม่เกิดเปลี่ยนแปลงมากนัก
- 6) เมื่อได้หาชุดค่าลักษณะสำคัญทางเสียงตามชนิดของขอบเขตเสียงและกฎการระบุหาตำแหน่งขอบเขตเสียงจากการทดสอบบนชุดข้อมูลที่ 2 ซึ่งใช้สำหรับพัฒนาระบบแล้ว นำกฎการพิจารณาดังกล่าวไปทดสอบกับชุดข้อมูลที่ 3 เพื่อตรวจสอบความคงทนของกฎและยืนยันว่ากฎนั้นไม่ได้เฉพาะเจาะจง (over-fitting) กับชุดข้อมูลที่ 2 ที่ใช้พัฒนาเพียงชุดเดียวเท่านั้น

#### 5.1.2. ขั้นตอนการปรับหาขอบเขตเสียงโดยใช้กฎตัดสินใจ

เมื่อได้กฎการพิจารณาหาตำแหน่งที่เหมาะสมกับขอบเขตหน่วยเสียงจากขั้นตอนการสร้างกฎตัดสินใจแล้ว นำกฎที่ได้มาทดสอบกับชุดข้อมูลที่ 3 โดยมีเงื่อนไขว่า ในขั้นตอนนี้ต้องไม่มีการเปลี่ยนแปลงกฎที่ได้ ไม่ลดหรือเพิ่มค่าลักษณะสำคัญอื่นนอกเหนือจากเดิม

ข้อมูลพื้นฐานของขั้นตอนนี้ คือ ตำแหน่งของขอบเขตหน่วยเสียงบนสัญญาณเสียงซึ่งเป็นตำแหน่งเริ่มต้นที่ได้จากการระบุด้วยแบบจำลองฮิดเดินมาคอฟ นำมาพิจารณาชนิดของขอบเขตเสียง แล้วปรับหาตำแหน่งขอบเขต ดังแสดงในรูปที่ 4.2 ซึ่งมีรายละเอียด ดังนี้

- 1) สร้างกรอบการปรับหาขอบเขตหน่วยเสียง
- 2) พิจารณารนคขอบเขตเสียงที่ได้จากการระบุตำแหน่งหน่วยเสียงแบบบังคับ แล้วสกัดค่าลักษณะสำคัญทางเสียงแต่ละค่าด้วยวิธีเดียวกับใช้ในขั้นตอนการสร้างกฎ โดยสกัดทุกค่าลักษณะสำคัญทางเสียงที่เลือกตามชนิดขอบเขตเสียงนั้นในขั้นตอนการสร้างกฎ และหาค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงและนอร์มอลไลซ์ค่านั้น
- 3) พิจารณาเลือกตำแหน่งเฟรมที่เหมาะสมกับตำแหน่งขอบเขตเสียงที่แท้จริงจากการใช้กฎตัดสินใจ โดยเลือกตำแหน่งเฟรมที่แสดงการเปลี่ยนแปลงเด่นชัดที่สุด ซึ่งอธิบายไว้ในหัวข้อ “การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ”



รูปที่ 4.2 กระบวนการทำงานของการปรับหาตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ

### 5.1.3. การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ

การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจนี้ เมื่อได้ค่าการเปลี่ยนแปลงของแต่ละค่าลักษณะสำคัญทางเสียงสำหรับแต่ละตำแหน่งเฟรมในกรอบที่พิจารณา และทำนอร์มอลไลซ์ด้วยค่าการเปลี่ยนแปลงมากที่สุดของค่าลักษณะสำคัญในกรอบพิจารณานั้น ดังแสดงตัวอย่างการตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจตามรูปที่ 4.3 ที่กำหนดให้มีกรอบพิจารณา 4 เฟรมและแต่ละเฟรมมีค่าลักษณะสำคัญทางเสียงที่พิจารณา 4 ค่า ซึ่งเมื่อทำนอร์มอลไลซ์แต่ละค่าการเปลี่ยนแปลงด้วยค่าการเปลี่ยนแปลงที่มากที่สุดของค่าลักษณะในกรอบนั้น จะได้เวกเตอร์ลักษณะเฉพาะแทนข้อมูลที่แต่ละตำแหน่งเฟรมมีแต่ละค่าอยู่ในช่วง  $[0, 1]$  จากนั้นพิจารณาเลือกตำแหน่งเฟรมที่เหมาะสมกับตำแหน่งขอบเขตเสียงที่แท้จริงโดยวัดระยะห่างกับ “ตำแหน่งขอบเขตเสียงทางที่อ้างอิง”



รูปที่ 4.3 การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ

โดยกำหนดให้ ตำแหน่งขอบเขตอ้างอิง คือ ตำแหน่งที่มีการเปลี่ยนแปลงของทุกค่าลักษณะสำคัญ ซึ่งผ่านการออร์มอลโลซ์ค่าจะได้เวกเตอร์ลักษณะเฉพาะที่เป็นเวกเตอร์หนึ่ง  $[1 \ 1 \ 1 \ \dots \ 1]$  เนื่องจากแนวคิดที่เสนอตำแหน่งเฟรมที่เลือกเป็นตำแหน่งขอบเขตเสียงต้องเกิดการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงมากที่สุด

ดังนั้นการตัดสินใจเพื่อเลือกตำแหน่งเฟรมที่เหมาะสมเป็นคำตอบของตำแหน่งขอบเขตเสียง พิจารณาจากคะแนนที่ได้การเปรียบเทียบเวกเตอร์ลักษณะเฉพาะของตำแหน่งที่พิจารณาวัฏระยะห่างเทียบกับเวกเตอร์ลักษณะเฉพาะของตำแหน่งอ้างอิง ซึ่งการคำนวณระยะห่างอธิบายตามสมการ (2.13) ตำแหน่งที่พิจารณาเหมาะสมที่สุดในกรอบพิจารณามีค่าคะแนนระยะห่างน้อยที่สุด (หรือไม่แตกต่างจากตำแหน่งอ้างอิง มีการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงที่ตำแหน่งเฟรมที่พิจารณานี้มากที่สุด)

## 5.2. การปรับหาค่าตำแหน่งขอบเขตเสียงโดยใช้สถิติ [45]

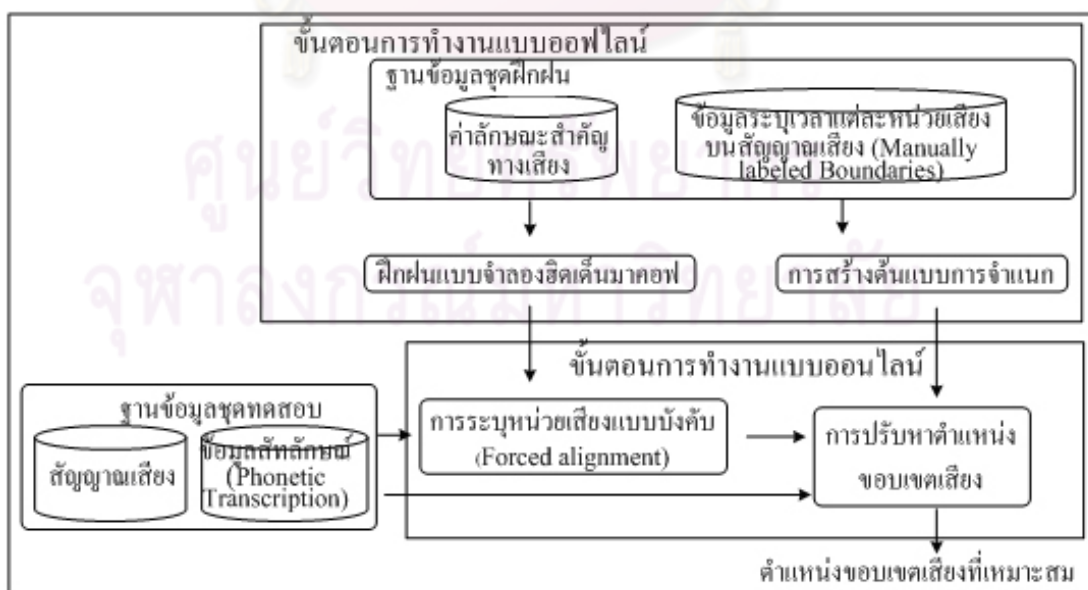
ขั้นตอนการปรับหาค่าตำแหน่งขอบเขตเสียงนี้เป็นขั้นตอนที่เสนอขึ้นมาเพื่อปรับหาค่าตำแหน่งเฟรมที่เหมาะสมกับตำแหน่งขอบเขตเสียงมากที่สุด ซึ่งวิเคราะห์ตำแหน่งจากการพิจารณาความน่าจะเป็นของตำแหน่งเฟรมที่พิจารณาคุณสมบัติความเป็นตำแหน่งขอบเขตเสียงและที่ไม่ใช่ขอบเขตเสียง งานวิจัยต่างๆ ได้เสนอเทคนิคการตัดสินใจต่างๆ เช่น ใช้ข่ายงานประสาทเทียมชนิดเพอเซปตรอนหลายชั้น [34] และใช้ข่ายงานประสาทเทียม [16] ที่ให้ประสิทธิภาพสูงในการจำแนกประเภทแต่อาจจะเจอเฉพาะขึ้นอยู่กับข้อมูลที่ใช้ฝึกฝนมากเกินไป (over-fitting) ประสิทธิภาพการจำแนกต่ำลง เมื่อใช้ข้อมูลที่ไม่มีความสัมพันธ์กับชุดที่นำมาฝึกฝน และในงานของ Hung-Yi Lo และ Hsin-min Wang [35] ใช้ซัพพอร์ตเวกเตอร์แมชชีนเพื่อจำแนกประเภทของกลุ่มที่เป็นขอบเขตเสียงออกจากกลุ่มที่ไม่ใช่ขอบเขตเสียงซึ่งมีแนวคิดคล้ายกับการวิทยานิพนธ์นี้



วิทยานิพนธ์นี้ใช้วิธีการวิเคราะห์ดิสক্রิมิแนนต์เชิงเส้นเป็นตัวจำแนก เนื่องจากในขั้นตอนการตัดสินใจเพื่อจำแนกกลุ่มข้อมูล ไม่มีความจำเป็นที่ต้องใช้เครื่องมือที่ตัดสินใจที่ซับซ้อน และข้อมูลที่ใช้จำแนกเป็นเวกเตอร์ลักษณะเฉพาะที่แสดงคุณสมบัติของเสียงที่ใช้ในการฝึกฝนหรือจำแนกประกอบด้วยมิติของเวกเตอร์ไม่มากนักเหมาะสมกับการนำมาใช้วิเคราะห์ด้วยดิสক্রิมิแนนต์เชิงเส้นใช้ ดังนั้นวิทยานิพนธ์นี้จึงเลือกใช้ตัวจำแนกวิธีการวิเคราะห์ดิสক্রิมิแนนต์เชิงเส้นเพื่อนำมาเป็นตัวระบุกลุ่มขอบเขตเสียงในการปรับหาขอบเขตเสียง

การใช้วิธีทางสถิติประกอบด้วยขั้นตอนการฝึกฝน การทดสอบตัวจำแนก การนำผลที่ได้จากการจำแนกประเภทจากวิธีการวิเคราะห์ดิสক্রิมิแนนต์เชิงเส้นมาตรวจหาตำแหน่งขอบเขตเสียง นอกจากนี้เพื่อสร้างการจำแนกให้มีประสิทธิภาพมากขึ้น วิทยานิพนธ์นี้เสนอขั้นตอนการเลือกค่าลักษณะสำคัญที่เหมาะสมกับชนิดขอบเขตเสียงเพื่อใช้ฝึกฝนแบบจำลองและจำแนกกลุ่มขอบเขตเสียงและเสนอขั้นตอนการหาชุดจำนวนข้อมูลฝึกฝนที่เหมาะสม และการใช้มิติของเวกเตอร์เพื่อเหมาะสมในการเป็นตัวแทนเสียงในการจำแนกกลุ่ม

การระบุหาตำแหน่งขอบเขตเสียงโดยใช้สถิติที่เสนอในวิทยานิพนธ์นี้ แบ่งเป็นขั้นตอนการทำงานแบบออฟไลน์ ซึ่งประกอบด้วยการฝึกฝนแบบจำลองฮิดเด้นมาร์คอฟและสร้างต้นแบบการจำแนก และขั้นตอนการทำงานแบบออนไลน์ ที่รวมกระบวนการทำงานเป็น 2 ส่วน (อธิบายในบทที่ 3) คือ ส่วนการทำงานของการระบุหน่วยเสียงแบบบังคับ (force alignment) และส่วนการปรับหาตำแหน่งขอบเขตหน่วยเสียง โดยมีรายละเอียดเพิ่มเติมของส่วนการทำงานปรับหาตำแหน่งขอบเขตหน่วยเสียงดังแสดงตามรูปที่ 4.4



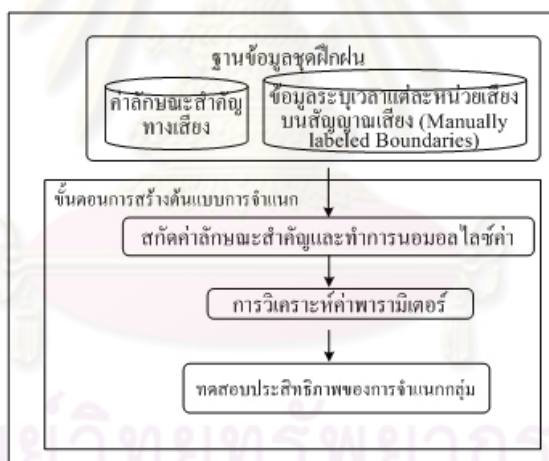
รูปที่ 4.4 แผนภาพการทำงานการปรับหาตำแหน่งขอบเขตเสียงโดยใช้สถิติ

### 5.2.1. ขั้นตอนแบบออฟไลน์ (off-line process)

เนื่องจากการหาตำแหน่งที่เหมาะสมเพื่อเป็นตำแหน่งขอบเขตเสียงนั้น พิจารณาจากตำแหน่งเฟรมที่อยู่ใกล้เคียงกับตำแหน่งสมมติฐานที่ได้จากการระบุแบบบังคับของอิดเดินมาคอฟโมเดล ดังนั้นในขั้นตอนของการทำงานแบบออฟไลน์นี้ ประกอบด้วย ขั้นตอนการสร้างแบบจำลองอิดเดินมาคอฟ (ซึ่งมีรายละเอียดการฝึกฝนแบบจำลองอิดเดินมาคอฟในบทที่ 3) ขั้นตอนการวิเคราะห์ค่าพารามิเตอร์ที่เหมาะสมในการสร้างต้นแบบการปรับหาดำแหน่งขอบเขตเสียง (ซึ่งหมายถึงการวิเคราะห์ 2 ขั้นตอน ที่ได้อธิบายไว้ในหัวข้อ การวิเคราะห์ค่าลักษณะสำคัญต่างๆ และการเลือกค่าลักษณะสำคัญสำหรับแต่ละชนิดขอบเขตหน่วยเสียง) และขั้นตอนการเตรียมตัวอย่างข้อมูลฝึกฝนและสร้างต้นแบบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงนำไปสู่การระบุบอกโอกาสความน่าจะเป็นในการพิจารณาเลือกตำแหน่งขอบเขตเสียงที่เหมาะสมต่อไป

#### 5.2.1.1. ขั้นตอนการสร้างต้นแบบการจำแนก

การเตรียมข้อมูลชุดฝึกฝนเพื่อใช้เป็นตัวอย่างการฝึกฝนตัวจำแนกกลุ่ม แสดงการทำงานตามรูปที่ 4.5 โดยอธิบายประกอบด้วยขั้นตอนดังนี้

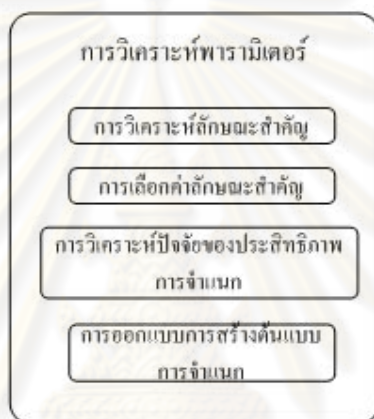


รูปที่ 4.5 กระบวนการสร้างต้นแบบการจำแนก

- 1) สกัดค่าลักษณะสำคัญทางเสียง (อธิบายรายละเอียดการสกัดค่าลักษณะสำคัญแต่ละค่าในบทที่ 2) จากนั้นคำนวณค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงนั้นและการนอร์มอลไลซ์ค่า
- 2) วิเคราะห์ค่าพารามิเตอร์ ดังอธิบายตามรูปที่ 4.6 ซึ่งประกอบด้วยการวิเคราะห์ค่าลักษณะสำคัญเพื่อเลือกค่าที่เหมาะสมกับตัวอย่าง การเลือกค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตหน่วยเสียง ซึ่งการเลือกค่าลักษณะสำคัญทางเสียงตามชนิดขอบเขตเสียง โดยพิจารณาตามผลที่ได้จากการ

วิเคราะห์ค่าลักษณะสำคัญทางเสียง (ซึ่งได้อธิบายตามหัวข้อที่ 2 และ 3) การวิเคราะห์ปัจจัยการจำแนก และการออกแบบการสร้างต้นแบบการจำแนก ซึ่งอธิบายรายละเอียดตามหัวข้อการวิเคราะห์พารามิเตอร์ต่อไป

- 3) ทดสอบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง จากเวกเตอร์ลักษณะเฉพาะของตัวแทนสัญญาณเสียงแต่ละเฟรม โดยใช้ข้อมูลตัวอย่างที่เป็นกลุ่มขอบเขตเสียงและตัวอย่างกลุ่มที่ไม่ใช่ขอบเขตเสียง มาสร้างเวกเตอร์ลักษณะเฉพาะนี้ซึ่งกำหนดค่าพารามิเตอร์แตกต่างกันตามหัวข้อปัจจัยประสิทธิภาพการจำแนก จากนั้นวัดประสิทธิภาพการจำแนกกลุ่มแต่ละชุดการทดลองตามชุดการทดลองที่ออกแบบขั้นตอนการวิเคราะห์พารามิเตอร์



รูปที่ 4.6 ขั้นตอนการวิเคราะห์พารามิเตอร์

ซึ่งรายละเอียดของแต่ละขั้นตอนการวิเคราะห์พารามิเตอร์ได้อธิบายดังนี้

#### 5.2.1.2. ขั้นตอนการวิเคราะห์พารามิเตอร์

ขั้นตอนการวิเคราะห์พารามิเตอร์นี้ ประกอบด้วย

- I. การวิเคราะห์ค่าลักษณะสำคัญทางเสียง (ซึ่งได้อธิบายตามหัวข้อที่ 2)
- II. การเลือกค่าลักษณะสำคัญทางเสียง (ซึ่งได้อธิบายตามหัวข้อที่ 3)
- III. การวิเคราะห์ปัจจัยของประสิทธิภาพการจำแนกกลุ่ม

ประสิทธิภาพของการจำแนกกลุ่มขึ้นอยู่กับต้นแบบการจำแนกกลุ่ม ดังนั้นเพื่อหาต้นแบบการจำแนกที่ดีจึงแยกตามปัจจัยการสร้างต้นแบบการจำแนก คือ

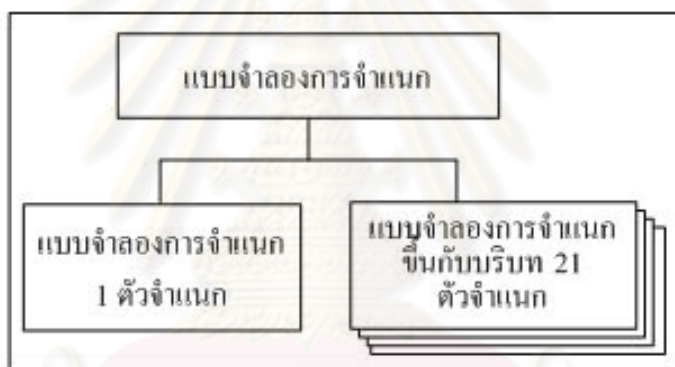
- A. ลักษณะต้นแบบการจำแนก
- B. ลักษณะของข้อมูลที่ใช้ฝึกฝน
- C. มิติของเวกเตอร์ลักษณะเฉพาะ

#### D. จำนวนของเวกเตอร์ลักษณะเฉพาะ

ซึ่งอธิบายรายละเอียดของแต่ละปัจจัยดังนี้

##### A. ลักษณะต้นแบบการจำแนก

วิทยานิพนธ์นี้เสนอการปรับหาขอบเขตหน่วยเสียงโดยพิจารณาตามชนิดขอบเขตหน่วยเสียง ดังนั้นการสร้างต้นแบบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง ต้องพิจารณาตามแต่ละชนิดขอบเขตหน่วยเสียง เพื่อทดสอบสมมติฐานว่าประสิทธิภาพการจำแนกกลุ่มขึ้นอยู่กับต้นแบบเฉพาะตามแต่ละชนิดขอบเขตเสียง จึงทดสอบผลการจำแนกเปรียบเทียบกับผลการจำแนกเมื่อใช้ตัวจำแนกกลุ่ม 1 ตัวที่สร้างขึ้นโดยที่ต้นแบบการจำแนกดังกล่าวไม่ได้พิจารณาตามแต่ละชนิดขอบเขตเสียง ดังแสดงการเปรียบเทียบของต้นแบบตามรูปที่ 4.7



รูปที่ 4.7 ตัวอย่างการสร้างต้นแบบการจำแนกกลุ่ม

##### B. ลักษณะของข้อมูลที่ใช้ฝึกฝน

ตามที่วิทยานิพนธ์นี้เสนอการปรับหาขอบเขตเสียงตามชนิดขอบเขตเสียง เนื่องจากลักษณะสำคัญทางเสียงต่างแต่ละค่านั้นแสดงลักษณะเฉพาะของกลุ่มเสียงที่แตกต่างกัน คุณสมบัติของแต่ละค่าลักษณะทางเสียงจึงเป็นองค์ประกอบที่ใช้เลือกใช้ค่าต่าง ๆ ตามแต่ละกลุ่มเสียง เมื่อพิจารณาชนิดขอบเขตเสียงแต่ละชนิด จึงพบพฤติกรรมที่แสดงแตกต่างกัน ยกตัวอย่างเช่น ชนิดขอบเขตเสียงที่มีการเปลี่ยนแปลงของค่าลักษณะทางเสียงเกี่ยวกับเสียงก้องพบความแตกต่างค่าที่บอกความเป็นเสียงก้องชัดเจน ในทางตรงข้ามเมื่อพิจารณาชนิดขอบเขตกลุ่มเสียงพูดต่าง ๆ กับเสียงเงียบพบความแตกต่างของการเปลี่ยนแปลงค่าพลังงานอย่างชัดเจน เป็นต้น ดังนั้นการพิจารณาเลือกค่าลักษณะทางเสียงสำหรับแต่ละชนิดขอบเขตเสียงจึงแตกต่างกัน ลักษณะของข้อมูลที่ใช้ในการฝึกฝนต้นแบบการจำแนกกลุ่ม

จึงเป็นปัจจัยสำคัญในการสร้างต้นแบบการจำแนกกลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง โดยการพิจารณาเลือกลักษณะสำคัญทางเสียงได้อธิบายในหัวข้อ “การวิเคราะห์ค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียง” แล้วนำค่าลักษณะสำคัญทางเสียงที่เลือกมาสร้างเป็นต้นแบบการจำแนกตามแต่ละชนิดขอบเขตเสียง ดังนั้น เวกเตอร์ลักษณะเฉพาะที่เป็นตัวแทนของข้อมูลที่ตำแหน่งเฟรมที่ใช้ฝึกฝน แทนด้วยสมการ (4.2)

$$\text{vector} = [\text{Feature}_1, \text{Feature}_2, \text{Feature}_3, \dots, \text{Feature}_n] \quad (4.2)$$

เมื่อ  $\text{vector}$  แทนเวกเตอร์ลักษณะเฉพาะที่ตำแหน่งเฟรมพิจารณา และ  $\text{Feature}_i$  แทนค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงต่าง ๆ แต่ละค่า โดยเรียงลำดับ  $i$  ลำดับ ซึ่งแทนลำดับที่ 1 ถึงลำดับที่  $n$  หมายถึง ค่าลักษณะสำคัญทางเสียงประกอบด้วยทั้งหมด  $n$  ค่าที่ได้ถูกพิจารณาเลือกตามชนิดขอบเขตเสียงนั้น

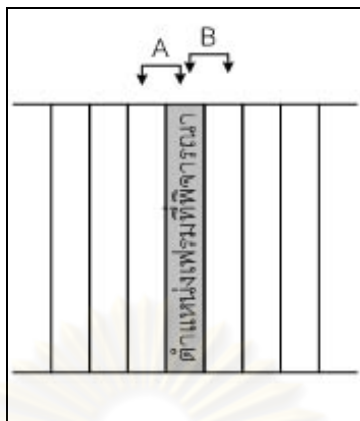
### C. มิติของเวกเตอร์ลักษณะเฉพาะ

เวกเตอร์ลักษณะเฉพาะของข้อมูลที่เสนอในวิทยานิพนธ์นี้ใช้ค่าเปลี่ยนแปลงของค่าลักษณะสำคัญต่าง ๆ ที่การคำนวณค่าการเปลี่ยนแปลงระหว่างจากค่าที่ตำแหน่งเฟรมที่พิจารณากับตำแหน่งเฟรมที่อยู่ติดกัน ซึ่งวิทยานิพนธ์นี้เสนอการคำนวณค่าเปลี่ยนแปลงเป็น 2 แบบ ซึ่งมีผลต่อจำนวนมิติของเวกเตอร์ลักษณะเฉพาะที่ใช้เป็นตัวแทนข้อมูล ได้แก่

- a) มิติเวกเตอร์ของข้อมูลที่แบบ 1 มิติ หมายถึง ค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญนั้นที่ได้จากผลต่างของค่าจากตำแหน่งเฟรมที่กำลังพิจารณากับตำแหน่งเฟรมก่อนหน้า เมื่อพิจารณาตามรูปที่ 4.8 พบว่าใช้เวกเตอร์ลักษณะเฉพาะจากค่า  $e$  (A) หรือแทนเวกเตอร์ลักษณะเฉพาะด้วย [A]
- b) มิติเวกเตอร์ของข้อมูลที่แบบ 2 มิติ หมายถึง ค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญนั้นที่ได้จากผลต่างของค่าจากตำแหน่งเฟรมกำลังพิจารณากับตำแหน่งเฟรมก่อนหน้า 1 ค่า และการเปลี่ยนแปลงค่าลักษณะสำคัญนั้นที่ได้จากผลต่างของค่าที่ตำแหน่งเฟรมกำลังพิจารณากับตำแหน่งเฟรมที่ตามมาอีก 1 ค่า รวมเป็น 2 ค่า เมื่อ



พิจารณาตามรูปที่ 4.8 พบว่าใช้เวกเตอร์ลักษณะเฉพาะจากค่าเอ (A) และค่าบี (B) หรือแทนเวกเตอร์ลักษณะเฉพาะด้วย [A B]



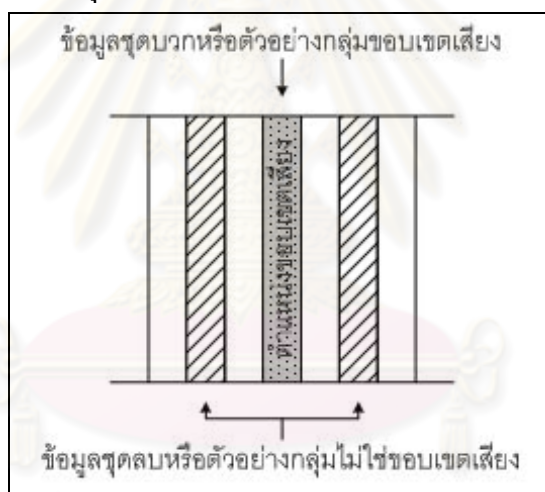
รูปที่ 4.8 การคำนวณค่าเปลี่ยนแปลงเพื่อสร้างเวกเตอร์ลักษณะเฉพาะของข้อมูล

เพื่อทดสอบสมมติฐานของปัจจัยดังกล่าว จึงเสนอการทดสอบเพื่อทดสอบประสิทธิภาพของระบบที่ได้จากการจำแนกกลุ่มเมื่อใช้จำนวนมิติของเวกเตอร์แตกต่างกัน ซึ่งเสนอในหัวข้อ “การออกแบบการทดลองปรับหาขอบเขตโดยใช้สถิติ” ต่อไป

#### D. จำนวนของเวกเตอร์ลักษณะเฉพาะ

เนื่องจากจำแนกกลุ่มในวิทยานิพนธ์นี้ใช้วิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น จำแนกกลุ่มของตำแหน่งขอบเขตเสียงและกลุ่มของตำแหน่งที่ไม่ใช่ขอบเขตเสียง โดยใช้ข้อมูลฝึกฝนประกอบด้วย ข้อมูลส่วนที่เป็นข้อมูลชุดบวกหรือตัวอย่างกลุ่มขอบเขตเสียง และข้อมูลส่วนที่แสดงชุดลบหรือตัวอย่างของกลุ่มที่ไม่ใช่ขอบเขตเสียง ดังนั้นจำนวนของข้อมูลที่เป็นตัวอย่างของชุดข้อมูลบวกและชุดข้อมูลลบจึงมีผลต่อประสิทธิภาพของการฝึกฝนต้นแบบการจำแนก โดยจำนวนเวกเตอร์ลักษณะเฉพาะของข้อมูลดังกล่าวพิจารณาจากจำนวนตัวอย่างของข้อมูลชุดลบระบุตัวอย่างของกลุ่มที่ไม่ใช่ขอบเขตเสียงและจำนวนตัวอย่างข้อมูลชุดบวกที่ระบุกลุ่มที่เป็นขอบเขตเสียงโดยอ้างอิงกับตำแหน่งขอบเขตเสียงที่แท้จริงของฐานข้อมูลชุดฝึกฝน ดังนั้นเสนอเป็นอัตราส่วนของจำนวนเฟรมข้อมูลชุดบวกกับจำนวนเฟรมข้อมูลชุดลบเพื่อมาใช้เป็นข้อมูลฝึกฝนแตกต่างกัน (หมายถึงจำนวนของข้อมูลที่ใช้ฝึกฝนแตกต่างกัน เมื่อใช้อัตราส่วนที่มีจำนวนข้อมูลที่มาก จะทำให้ได้ต้นแบบของการจำแนกที่มีต้นแบบข้อมูลเรียนรู้ที่มากขึ้น) ดังนั้นวิทยานิพนธ์นี้จึงเสนอจำนวนอัตราส่วนของตัวอย่างข้อมูลที่แตกต่างกัน ดังนี้

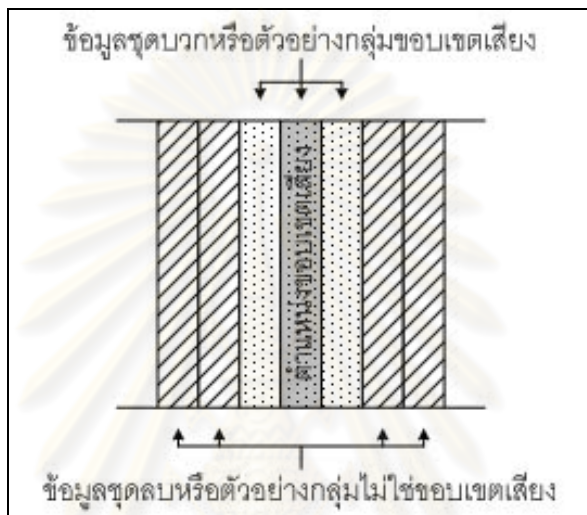
- a) อัตราส่วนจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบแบบ 1:2 หมายถึง การใช้ชุดข้อมูลตัวอย่างของกลุ่มที่เป็นขอบเขตเสียง 1 เฟรม ซึ่งได้จากคำตอบตำแหน่งขอบเขตเสียงที่แท้จริง (เป็นตำแหน่งอ้างอิง) และใช้ตัวอย่างของกลุ่มที่ไม่ใช่ขอบเขตเสียง 2 เฟรมซึ่งได้จากตำแหน่งเฟรมที่สองทางซ้ายที่อยู่ถัดออกไปจากตำแหน่งอ้างอิง 1 ค่าและได้จากตำแหน่งเฟรมที่สองทางขวาที่อยู่ถัดออกไปจากตำแหน่งอ้างอิงอีก 1 ค่า ดังแสดงตามรูปที่ 4.9 เมื่อได้ตำแหน่งขอบเขตเสียงแท้จริงอยู่ที่ตำแหน่ง  $t_0$  ดังนั้นเวกเตอร์ลักษณะเฉพาะได้จากค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงที่ตำแหน่ง  $t_0$  เพื่อเป็นตัวอย่างของกลุ่มบวกแสดงเป็นกลุ่มขอบเขตเสียง และได้เวกเตอร์ลักษณะเฉพาะจากค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงที่ตำแหน่ง  $t_2$  และตำแหน่ง  $t_2$  เป็นตัวอย่างของกลุ่มลบที่แสดงกลุ่มไม่ใช่ขอบเขตเสียง เป็นต้น



รูปที่ 4.9 ตัวอย่างเวกเตอร์ลักษณะเฉพาะของข้อมูลฝึกฝนชุดบวกต่อชุดลบแบบ 1:2

- b) อัตราส่วนจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบแบบ 3:4 หมายถึง การใช้ชุดข้อมูลตัวอย่างของกลุ่มที่เป็นขอบเขตเสียง 3 เฟรม ซึ่งได้จากคำตอบขอบตำแหน่งขอบเขตเสียงที่แท้จริง (เป็นตำแหน่งอ้างอิง) และอีก 2 เฟรมจากตำแหน่งที่อยู่ทางซ้ายและขวาของตำแหน่งขอบเขตที่แท้จริง จากนั้นหาตัวอย่างของกลุ่มที่ไม่ใช่ขอบเขตเสียง 4 เฟรม ซึ่งได้ 2 เฟรมจากสองเฟรมที่อยู่ตำแหน่งเฟรมที่สองและสามทางซ้ายที่อยู่ถัดออกไปจากตำแหน่งอ้างอิง 2 ค่าและได้จากตำแหน่งเฟรมที่สองและสามทางขวาที่อยู่ถัดออกไปจากตำแหน่งอ้างอิงอีก 2 ค่า ดังแสดงตามรูปที่ 4.10 เมื่อได้ตำแหน่ง

ขอบเขตเสียงแท้จริงอยู่ที่ตำแหน่ง  $t_0$  ดังนั้นเวกเตอร์ลักษณะเฉพาะได้จากค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงที่ตำแหน่ง  $t_0$  ตำแหน่ง  $t_1$  และ  $t_1$  รวม 3 เฟรมเพื่อเป็นตัวอย่างของกลุ่มบวกแสดงเป็นกลุ่มขอบเขตเสียง และได้เวกเตอร์ลักษณะเฉพาะจากค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงที่ตำแหน่ง  $t_2$   $t_3$   $t_2$  และ  $t_3$  รวมทั้งหมด 4 เฟรม และตำแหน่ง  $t_2$  เป็นตัวอย่างของกลุ่มลบที่แสดงกลุ่มไม่ใช่ขอบเขตเสียง เป็นต้น



รูปที่ 4.10 ตัวอย่างเวกเตอร์ลักษณะเฉพาะของข้อมูลฝึกฝนชุดบวกต่อชุดลบแบบ 3:4

IV. การออกแบบการสร้างต้นแบบการจำแนกกลุ่ม

เนื่องจากประสิทธิภาพของการจำแนกขึ้นอยู่กับการสร้างต้นแบบการจำแนกกลุ่ม เพื่อศึกษาและหาค่าพารามิเตอร์ที่เหมาะสมในการจำแนกกลุ่ม จึงออกแบบการทดลองเพื่อทดสอบประสิทธิภาพการจำแนก ซึ่งอยู่ในขั้นตอนการวิเคราะห์พารามิเตอร์ดังรูปที่ 4.6 เมื่อได้วิเคราะห์ปัจจัยของการสร้างต้นแบบการจำแนก

การทดสอบปัจจัยตามปัจจัยต่าง ๆ ซึ่งสรุปเป็นตารางการทดลองตามตารางที่ 4.1 ประกอบด้วย 8 ชุดการทดลอง ได้แก่

- a) การทดสอบตามปัจจัยชนิดต้นแบบการจำแนกกลุ่ม ซึ่งใช้ต้นแบบที่ได้จากการสร้างต้นแบบการจำแนกกลุ่มที่ไม่ขึ้นกับชนิดขอบเขตหน่วยเสียงทดสอบเปรียบเทียบกับต้นแบบที่ระบุตามชนิดขอบเขตหน่วยเสียง ทำให้ได้ต้นแบบการจำแนกที่เป็นแบบไม่ระบุชนิดขอบเขต 1 ตัวจำแนก และต้นแบบการจำแนกที่เป็นแบบระบุตามชนิดขอบเขตเสียงตามจำนวนชนิดขอบเขตเสียง ดังนั้น เมื่อ

- ค่าพารามิเตอร์อื่นคงที่ จะได้ชุดการทดลองที่ 1 และชุดการทดลองที่ 2 ทดลองเปรียบเทียบคู่กับชุดการทดลองที่ 3 และ 4 ตามลำดับ
- b) การทดสอบตามปัจจัยอัตราส่วนจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบ นั่นคือ มีเวกเตอร์ลักษณะเฉพาะจากข้อมูลที่ใช้ในการสร้างต้นแบบและการจำแนกที่เป็นตัวอย่างข้อมูลชุดบวกต่อข้อมูลชุดลบ ดังนั้นเมื่อกำหนดค่าพารามิเตอร์อื่น ๆ คงที่ ได้แก่ ชุดการทดลองที่ 1 ทดลองเปรียบเทียบกับชุดการทดลองที่ 2 ชุดการทดลองที่ 3 เปรียบเทียบกับชุดการทดลองที่ 4 ชุดการทดลองที่ 5 เปรียบเทียบกับชุดการทดลองที่ 6 และชุดการทดลองที่ 7 เปรียบเทียบกับชุดการทดลองที่ 8 ตามลำดับ
- c) การทดสอบตามปัจจัยลักษณะสำคัญทางเสียง ซึ่งใช้การเปรียบเทียบประสิทธิภาพของการจำแนกเมื่อใช้ต้นแบบการจำแนกที่สร้างจากข้อมูลฝึกฝนที่เลือกค่าลักษณะสำคัญทางเสียงต่าง ๆ ตามแต่ละชนิดขอบเขต คู่กับการจำแนกเมื่อสร้างต้นแบบจากทุกค่าลักษณะสำคัญทางเสียงทั้งหมด (หมายถึง ต้นแบบที่สร้างจากข้อมูลฝึกฝนที่เลือกค่าลักษณะสำคัญทางเสียงทุกค่า ไม่ได้เลือกค่าลักษณะสำคัญทางเสียงตามชนิดขอบเขตเสียงที่ปรับ ดังนั้น ต้นแบบการจำแนกแต่ละตัวตามชนิดขอบเขตเสียงนั้น ใช้ค่าลักษณะสำคัญทางเสียงทุกชนิดเหมือนกัน) ดังนั้นเมื่อกำหนดค่าพารามิเตอร์อื่น ๆ คงที่ ได้แก่ ชุดการทดลองที่ 3 ทดลองเปรียบเทียบกับชุดการทดลองที่ 5 และชุดการทดลองที่ 4 เปรียบเทียบกับชุดการทดลองที่ 6 ตามลำดับ พบว่าชุดการทดลองนี้ออกแบบเพื่อทดสอบสมมติฐานว่า การจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขต มีประสิทธิภาพขึ้นอยู่กับการเลือกใช้ค่าลักษณะสำคัญทางเสียงที่เพิ่มชนิดค่าลักษณะ (หมายถึง เปรียบเทียบผลการจำแนกเมื่อใช้ทุกชนิดค่าลักษณะสำคัญทางเสียง ที่มีชนิดค่าที่ใช้มากกว่าการเลือกเฉพาะบางค่าลักษณะสำคัญทางเสียง)
- d) การทดสอบตามปัจจัยมิติของเวกเตอร์ซึ่งใช้ต้นแบบที่ได้จากการสร้างต้นแบบการจำแนกกลุ่มที่ใช้ข้อมูลฝึกฝนเป็นค่าลักษณะความสำคัญทางเสียงแทนด้วยเวกเตอร์ลักษณะเฉพาะที่มีมิติเวกเตอร์ของข้อมูลที่แบบ 1 มิติ (หมายถึง ใช้ค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงที่ได้จากผลต่างของค่าตำแหน่งเฟรมที่กำลังพิจารณา กับเฟรมที่อยู่ก่อนหน้า 1 ค่า) เปรียบเทียบกับต้นแบบการจำแนกกลุ่ม

ที่ใช้ข้อมูลฝึกฝนเป็นเวกเตอร์ลักษณะเฉพาะที่มีมิติเวกเตอร์ลักษณะเฉพาะ 2 มิติ (หมายถึง ใช้ค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงที่ได้จากผลต่างของค่าตำแหน่งเฟรมที่กำลังพิจารณากับเฟรมที่อยู่ก่อนหน้า และผลต่างของค่าตำแหน่งเฟรมที่กำลังพิจารณากับเฟรมที่อยู่ถัดไป) ดังนั้นเมื่อกำหนดค่าพารามิเตอร์อื่น ๆ คงที่ได้แก่ ชุดการทดลองที่ 5 ทดลองเปรียบเทียบกับชุดการทดลองที่ 7 และชุดการทดลองที่ 6 เปรียบเทียบกับชุดการทดลองที่ 8 ตามลำดับ

ตารางที่ 4.1 ชุดการทดลองการปรับหาขอบเขตโดยใช้สถิติ

ชุดการทดลอง	1	2	3	4	5	6	7	8
ชนิดตัวจำแนก	แบบไม่ระบุชนิดขอบเขต		แบบระบุชนิดขอบเขต					
ต้นแบบตัวจำแนก (จำนวนตัวจำแนก)	1	1	21	21	21	21	21	21
อัตราส่วนจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบ	1 : 2	3 : 4	1 : 2	3 : 4	1 : 2	3 : 4	1 : 2	3 : 4
มิติเวกเตอร์ของข้อมูล (จำนวนข้อมูลต่อ 1 เฟรม)	2	2	2	2	2	2	1	1
ลักษณะสำคัญทางเสียง เฉพาะตามชนิดขอบเขต	ไม่	ไม่	ไม่	ไม่	เฉพาะ	เฉพาะ	เฉพาะ	เฉพาะ

หมายเหตุ : ต้นแบบตัวจำแนก 21 ตัวจำแนก แทนต้นแบบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มไม่ใช่ที่สร้างขึ้นตามชนิดขอบเขตเสียง 21 ชนิด (ซึ่งจะได้รายงานผลจำนวนชนิดขอบเขตหน่วยเสียงที่เลือกปรับหาค่าตำแหน่งขอบเขตเสียงในบทที่ 5)

### 5.2.1.3. ขั้นตอนทดสอบต้นแบบการจำแนก

ขั้นตอนการทดสอบต้นแบบการจำแนกนี้ เป็นการวัดประสิทธิภาพการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงเท่านั้น ไม่ใช่การวัดประสิทธิภาพที่ใช้ตัวจำแนกไประบุหาค่าตำแหน่งขอบเขตเสียง

เนื่องจากในขั้นตอนการเลือกชนิดค่าลักษณะสำคัญทางเสียงเลือกตามชนิดขอบเขตที่ละชนิด และการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขต



เสียงสำหรับแต่ละชนิดขอบเขตเสียงเลือกใช้คำลักษณะทางเสียงต่างกัน ดังนั้นการทดสอบในขั้นตอนนี้จึงพิจารณาทีละชนิดขอบเขตเสียง มีรายละเอียดการทดสอบดังนี้

- 1) เมื่อได้ชุดคำลักษณะสำคัญทางเสียงที่เหมาะสมสำหรับแต่ละชนิดขอบเขตเสียงแล้วสร้างชุดเวกเตอร์ลักษณะเฉพาะเพื่อเป็นตัวอย่างข้อมูลชุดฝึกฝน ซึ่งประกอบด้วย ข้อมูลตัวอย่างกลุ่มบวกและข้อมูลตัวอย่างกลุ่มลบ
- 2) สร้างชุดข้อมูลทดสอบตัวจำแนก โดยสกัดคำลักษณะสำคัญทางเสียงต่าง ๆ ตามชนิดเช่นเดียวกับชุดตัวอย่างข้อมูลฝึกฝน และสร้างเป็นชุดเวกเตอร์ลักษณะเฉพาะของตัวอย่างกลุ่มทดสอบ การสกัดคำข้อมูลทดสอบนี้ใช้ชุดข้อมูลที่ 2 ซึ่งเป็นข้อมูลชุดพัฒนาระบบและทดสอบ
- 3) กำหนดคำตอบของชุดข้อมูลฝึกฝน โดยกลุ่มขอบเขตเสียงได้มาจากตำแหน่งที่เป็นขอบเขตเสียงที่แท้จริง และกลุ่มที่ไม่ใช่ขอบเขตเสียงคือตำแหน่งที่ไม่ได้เป็นขอบเขตเสียงที่แท้จริง (ซึ่งอธิบายในหัวข้อปัจจัยประสิทธิภาพการจำแนก) โดยกำหนดเป็นคำตอบตัวอย่างกลุ่มบวกและคำตอบตัวอย่างกลุ่มลบ
- 4) นำชุดข้อมูลชุดฝึกฝน คำตอบของชุดการฝึกฝน และชุดข้อมูลทดสอบมาทดสอบการจำแนกกลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง
- 5) วัดผลการจำแนกกลุ่มขอบเขตเสียงและกลุ่มไม่ใช่ขอบเขตเสียง โดยเปรียบเทียบกับคำตอบที่แท้จริงเมื่อใช้ชุดข้อมูลทดสอบ
- 6) วัดร้อยละความถูกต้องการจำแนกจาก อัตราส่วนตำแหน่งเฟรมที่ตอบถูกต้องต่อตำแหน่งเฟรมทั้งหมด

### 5.2.2. ขั้นตอนแบบออนไลน์ (on-line process)

แบบจำลองฮิดเดินมาคอฟที่ฝึกฝนในขั้นตอนออฟไลน์นำมาใช้ระบุตำแหน่งสมมติฐานของขอบเขตเสียงเพื่อเป็นข้อมูลพื้นฐานจากการทำงานแบบบังคับของแบบจำลองฮิดเดินมาคอฟโมเดลในขั้นตอนของการทำงานแบบออนไลน์ และจากขั้นตอนการสร้างต้นแบบการจำแนกนี้ จะได้ต้นแบบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงนั้นเพื่อมาใช้ในขั้นตอนของการทำงานแบบออนไลน์ เพื่อระบุตำแหน่งขอบเขตเสียงที่เหมาะสมที่สุด (ซึ่งอธิบายในหัวข้อ “ขั้นตอนการปรับหาขอบเขตเสียงโดยใช้สถิติ”) จากนั้นทดสอบประสิทธิภาพการระบุตำแหน่งขอบเขตเสียงต่อไป

#### 5.2.2.1. ขั้นตอนการปรับหาขอบเขตเสียงโดยใช้สถิติ

เมื่อได้ต้นแบบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงสำหรับแต่ละชนิดขอบเขตเสียงแล้ว นำต้นแบบดังกล่าวมาสร้างระบบการระบุหาขอบเขตหน่วยเสียงของวิธีที่เสนอทดสอบกับชุดข้อมูลที่ 3 ซึ่งเป็นชุดข้อมูลเพื่อทดสอบประสิทธิภาพระบบจากการจำแนกประเภทระหว่างกลุ่มตำแหน่งขอบเขตเสียงและกลุ่มตำแหน่งที่ไม่ใช่ขอบเขตเสียง

ขั้นตอนการปรับหาขอบเขตหน่วยเสียงโดยใช้สถิตินี้มีการทำงานตามรูปที่ 4.11 โดยใช้ตำแหน่งของขอบเขตหน่วยเสียงบนสัญญาณเสียงซึ่งเป็นตำแหน่งเริ่มต้นที่ได้จากการระบุด้วยแบบจำลองฮิดเค็นมาคอฟ เป็นข้อมูลเบื้องต้นที่ในขั้นตอนนี้



รูปที่ 4.11 การปรับหาขอบเขตเสียงโดยใช้สถิติ

การปรับหาขอบเขตหน่วยเสียงโดยใช้สถิติ สามารถอธิบายการทำงานแบบละเอียดดังนี้

- 1) สร้างกรอบการปรับหาขอบเขตหน่วยเสียง
- 2) พิจารณาชนิดขอบเขตเสียงที่ได้จากการระบุตำแหน่งหน่วยเสียงแบบบังคับแล้วสกัดค่าลักษณะสำคัญทางเสียงแต่ละค่าตามค่าที่เลือกในการสร้างต้นแบบการจำแนกกลุ่มของชนิดขอบเขตเสียงนั้น ซึ่งได้จากขั้นตอนการสร้างต้นแบบการจำแนก
- 3) นำชุดค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงตามชนิดขอบเขตเสียงจากทุกตำแหน่งเฟรมภายในกรอบพิจารณาการปรับหาตำแหน่งขอบเขตเสียงไปจำแนกกลุ่ม
- 4) นำความน่าจะเป็นของการจำแนกกลุ่มที่ได้แต่ละตำแหน่งเฟรม ซึ่งคำนวณตามสมการที่ (2.12) ไปคำนวณเพื่อตัดสินใจเลือกตำแหน่งขอบเขตเสียง ซึ่งพิจารณาเลือกตำแหน่งขอบเขตเสียงที่เหมาะสมตามรายละเอียดในหัวข้อ “การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้สถิติ”

### 5.2.2.2. การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้สถิติ

การตัดสินใจเลือกตำแหน่งขอบเขตเสียงโดยใช้สถิติ หมายถึง การพิจารณาคำแหน่งที่เหมาะสมกับตำแหน่งขอบเขตเสียงพิจารณาจากผลลัพธ์จากการจำแนก (ซึ่งการทำงานขั้นตอนนี้คือ การตัดสินใจ (decision) ดังอธิบายไว้ในรูปที่ 4.11 ซึ่งแสดงการทำงานการปรับหาขอบเขตเสียงโดยใช้สถิติ) ผลลัพธ์ที่ได้จากการระบุกลุ่มขอบเขตเสียงนำมาคำนวณเป็นค่าความน่าจะเป็นของตำแหน่งที่พิจารณานั้นที่แสดงโอกาสของแต่ละตำแหน่งเฟรมนั้นจะถูกเลือกมาตอบเป็นตำแหน่งขอบเขตเสียงมากหรือน้อย ดังนั้นค่าความน่าจะเป็นที่ระบุว่าเป็นกลุ่มที่เป็นขอบเขตเสียง จึงใช้เป็นคะแนนการระบุความเหมาะสมกับตำแหน่งขอบเขตเสียงของแต่ละเฟรมนั้น คะแนนของแต่ละเฟรมภายในกรอบพิจารณาทำการนอร์มอลไลซ์ทำให้มีช่วงของคะแนนมีค่าตั้งแต่ 0-1 ถ้าคะแนนของตำแหน่งเฟรมที่พิจารณามีค่าสูงสุดในกรอบพิจารณา ตำแหน่งนั้นจะถูกเลือกมาเป็นคำตอบของตำแหน่งขอบเขตเสียง ในทางตรงข้ามถ้าคะแนนของเฟรมที่พิจารณามีค่าน้อยซึ่งหมายถึงตำแหน่งเฟรมนั้นมีโอกาสที่จะเป็นตำแหน่งขอบเขตเสียงน้อยเช่นกัน จะไม่ถูกเลือกพิจารณาเป็นตำแหน่งขอบเขตเสียง

## บทที่ 5

### ผลการวิจัย

#### ผลการวิจัย

วิทยานิพนธ์นี้ได้เสนอการระบุตำแหน่งขอบเขตเสียงเพื่อให้ได้ตำแหน่งที่เหมาะสมที่สุดจากวิธีที่เสนอเปรียบเทียบกับผลการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิง ซึ่งในการสร้างระบบระบุตำแหน่งขอบเขตเสียงเพื่อให้ได้ตำแหน่งที่เหมาะสมนี้ มีขั้นตอนการวิเคราะห์ชนิดขอบเขตเสียง การวิเคราะห์ค่าลักษณะสำคัญทางเสียง การวิเคราะห์เลือกค่าลักษณะสำคัญทางเสียง การวิเคราะห์ขนาดกรอบหน้าต่างการปรับหาตำแหน่งขอบเขตเสียง และการปรับหาตำแหน่งที่เหมาะสมกับตำแหน่งขอบเขตเสียง ดังนั้นวิทยานิพนธ์นี้จึงรายงานผลการวิเคราะห์ของแต่ละขั้นตอน ดังนี้

1. ผลการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิง
2. ผลการวิเคราะห์ชนิดขอบเขตเสียง
3. ผลการวิเคราะห์ค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียง
4. ผลการวิเคราะห์ขนาดกรอบหน้าต่างการปรับ (Refining window)
5. ผลจากการปรับหาขอบเขตหน่วยเสียง
6. ผลการวิเคราะห์คะแนนในการพิจารณาหาตำแหน่งขอบเขตเสียง

ซึ่งผลการวิเคราะห์รายงานอย่างละเอียดดังนี้

#### 1. ผลการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิง

แบบอ้างอิงที่ใช้ในเปรียบเทียบประสิทธิภาพของวิธีที่เสนอในวิทยานิพนธ์นี้ สร้างแบบอ้างอิงจากการระบุตำแหน่งแบบบังคับจากฮิดเด็นมาคอฟ วัดผลเทียบกับตำแหน่งขอบเขตของคำตอบจริง ตารางที่ 5.1 แสดงร้อยละความถูกต้องของการระบุตำแหน่งขอบเขตเสียงของแบบอ้างอิง โดยแบบอ้างอิงนี้สร้างจากค่าพารามิเตอร์ที่เหมาะสมที่สุดในขั้นตอนการฝึกฝนแบบจำลองฮิดเด็นมาคอฟ ซึ่งพิจารณาผลการระบุขอบเขตเสียงจากแบบอ้างอิงชุดนี้พบว่าร้อยละความถูกต้องของการระบุตำแหน่งขอบเขตเสียงมากที่สุด เมื่อเปรียบเทียบกับร้อยละความถูกต้องการระบุขอบเขตเสียงที่ได้จากชุดค่าพารามิเตอร์อื่นและใช้จำนวนรอบการฝึกฝนแบบจำลองฮิดเด็นมาคอฟที่แตกต่างกันไป (รายงานผลในภาคผนวก ก) และขนาดของเฟรมที่ใช้มีค่า 10 มิลลิวินาที

ตารางที่ 5.1 ร้อยละความถูกต้องของการระบุขอบเขตเสียงจากแบบอ้างอิง

ร้อยละความถูกต้องตามระดับความคลาดเคลื่อนที่ยอมรับ	±0 เฟรม	23.33
	±1 เฟรม	65.04
	±2 เฟรม	86.90
	±3 เฟรม	93.63
ความคลาดเคลื่อนเทียบกับคำตอบจริง (เฟรม)	ค่าเฉลี่ย	1.42
	ส่วนเบี่ยงเบนมาตรฐาน	1.55

จากตารางที่ 5.1 แสดงร้อยละความถูกต้องของผลที่ได้จากการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิงเทียบกับคำตอบที่แท้จริงมีค่า 23.33 65.04 86.90 และ 93.63 เมื่อรายงานตามระดับความคลาดเคลื่อนที่ยอมรับ 0 1 2 และ 3 เฟรมตามลำดับ ซึ่งการพิจารณาดำเน่องรอยต่อระหว่างหน่วยเสียงที่ได้จากผลจากการระบุตำแหน่งหน่วยเสียงแบบบังคับจากแบบจำลองฮิดเดินมาคอฟนั้น รายงานว่า บางชนิดขอบเขตเสียงไม่สามารถระบุได้ตรงกับตำแหน่งที่เกิดการเปลี่ยนแปลงของค่าอะคูสติกจริง จากการระบุหาตำแหน่งขอบเขตเสียงโดยไม่ทราบตำแหน่งของแต่ละหน่วยเสียงบนสัญญาณเสียงทำให้การระบุขอบเขตเสียงนั้นยาก ดังนั้นในงาน [17] [34] และ [36] จึงเสนอให้ใช้แบบจำลองฮิดเดินมาคอฟซึ่งเป็นเครื่องมือที่มีประสิทธิภาพในการระบุตำแหน่งของหน่วยเสียงบนสัญญาณเสียง จากนั้นพิจารณาหาตำแหน่งขอบเขตเสียงจากตำแหน่งที่เกิดการเปลี่ยนแปลงหน่วยเสียง ซึ่งเป็นวิธีที่วิทยานิพนธ์นี้เสนอ แต่เนื่องจากการทำงานของฮิดเดินมาคอฟทำหน้าที่สร้างแบบจำลองของแต่ละหน่วยเสียงเพื่อระบุตำแหน่งหน่วยเสียง จากนั้นนำผลลัพธ์มาวิเคราะห์ตำแหน่งขอบเขตเสียง ไม่ใช่การจำลองจากตำแหน่งขอบเขตเสียงจริงๆทำให้เกิดข้อผิดพลาดในการระบุตำแหน่งขอบเขตเสียง ดังนั้นวิทยานิพนธ์นี้จึงเสนอการพิจารณาปรับหาดำเน่องขอบเขตเสียงจากการวิเคราะห์คุณสมบัติตำแหน่งขอบเขตเสียง โดยใช้ผลลัพธ์การจำแนกกลุ่มของกลุ่มที่เป็นตำแหน่งขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง ซึ่งแบบจำลองกลุ่มขอบเขตเสียงที่เสนอในขั้นตอนที่สองของวิทยานิพนธ์นี้แตกต่างกับแบบจำลองหน่วยเสียงของฮิดเดินมาคอฟจึงทำให้เพิ่มประสิทธิภาพของการระบุหาตำแหน่งขอบเขตเสียงจากวิธีที่เสนอได้ นอกจากนี้แบบจำลองฮิดเดินมาคอฟเป็นแบบจำลองทางสถิติที่มีความซับซ้อน ไม่สามารถรับประกันได้ว่าตำแหน่งหน่วยเสียงที่ได้จากฮิดเดินมาคอฟแต่ละเฟรมนั้นถูกต้อง ส่งผลให้การพิจารณาดำเน่องขอบเขตเสียงผิดพลาดได้

นอกจากนี้ความคลาดเคลื่อนคำตอบที่ได้จากแบบอ้างอิงเทียบกับคำตอบจริงมีค่าเฉลี่ย 1.42 เฟรมและมีส่วนเบี่ยงเบน 1.55 เฟรม จากผลรายงานค่าเฉลี่ยความคลาดเคลื่อนนี้จึงเสนอขนาดกรอบหน้าต่างเพื่อปรับหาดำเน่องขอบเขตเสียงมีขนาด 5 7 และ 9 เฟรม จากการขยายจำนวนเฟรม 2 3 และ 4 เฟรมตามลำดับเมื่อกำหนดตำแหน่งขอบเขตเสียงเริ่มต้นอยู่จุดกึ่งกลางของ



กรอบหน้าต่างเนื่องจากเมื่อพิจารณาจากค่าเฉลี่ย 1.42 เฟรมนั้นสามารถสร้างกรอบที่ครอบคลุมตำแหน่งขอบเขตเสียงที่แท้จริงได้โดยห่างจากตำแหน่งขอบเขตเสียงเริ่มต้นมากกว่า 1 เฟรม และทดสอบปรับหาตำแหน่งขอบเขตเสียงด้วยขนาดกรอบหน้าต่างที่กว้างขึ้นซึ่งรายงานผลในหัวข้อถัดไป

## 2. ผลการวิเคราะห์ชนิดขอบเขตเสียง

ข้อมูลที่ใช้ในวิทยานิพนธ์นี้ประกอบด้วยจำนวนขอบเขตทั้งหมด 68,404 ขอบเขต ซึ่งแยกเป็นชนิดขอบเขตเสียงตามการแบ่งกลุ่มเสียงในบทที่ 3 ได้ 58 ชนิดขอบเขตเสียงดังตารางที่ 5.2 ซึ่งแสดงจำนวนขอบเขตเสียงแต่ละชนิดตามชนิดขอบเขตเสียงในรูปแบบ “เสียงเริ่มต้น-เสียงสิ้นสุด” และแสดงจำนวนขอบเขตเสียงแต่ละชนิดเรียงลำดับจากมากไปหาน้อยดังรูปที่ 5.1 (ตามรายละเอียดในภาคผนวก ข)

ตารางที่ 5.2 จำนวนขอบเขตเสียงสำหรับแต่ละชนิดขอบเขตเสียง

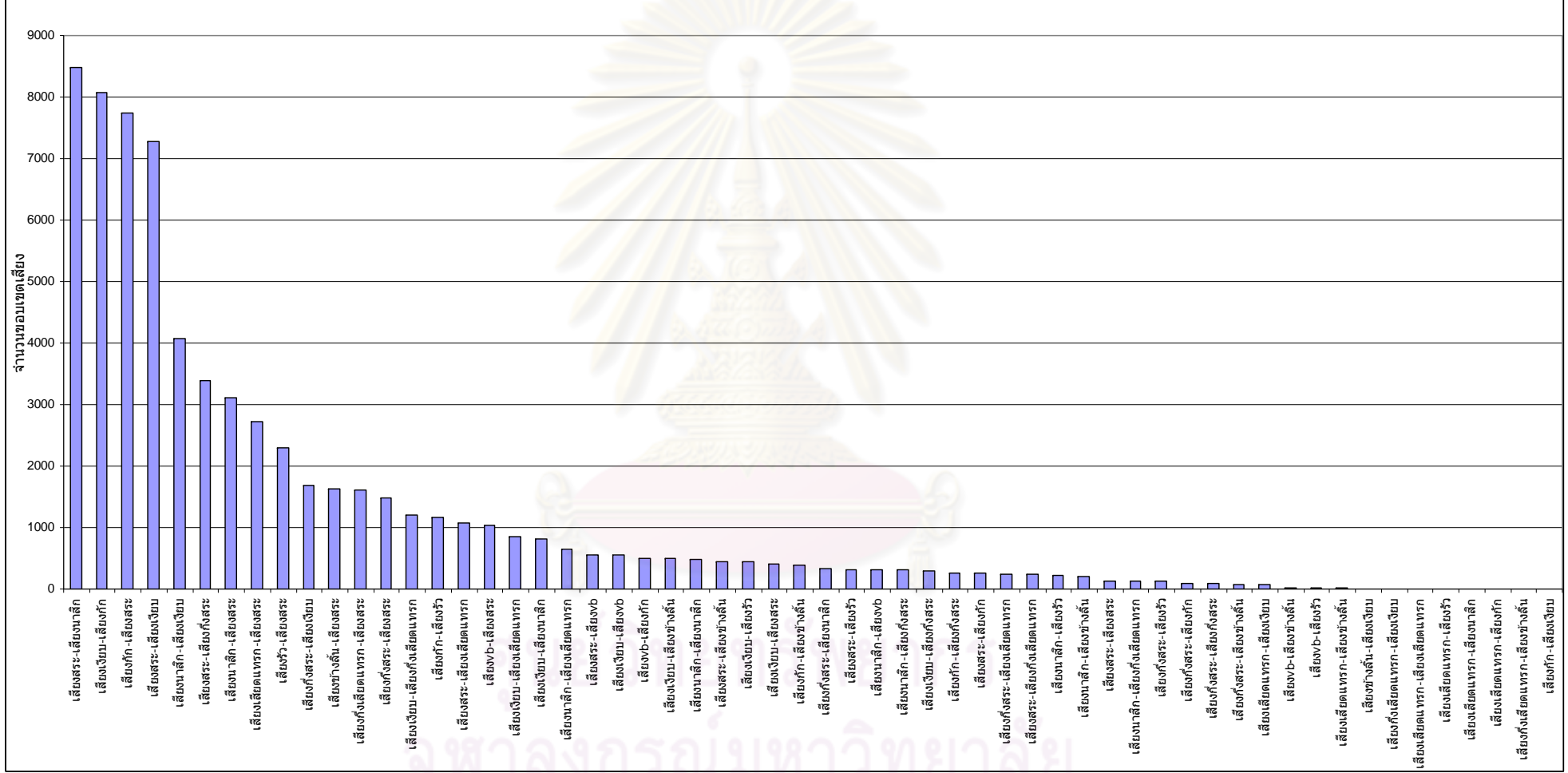
เสียงสิ้นสุด \ เสียงเริ่มต้น	เสียงเจียบ	เสียงสระ	เสียงนาสิก	เสียงกึ่งสระ	เสียงเสียดแทรก	เสียงกึ่งเสียดแทรก	เสียงกัก	เสียงvb	เสียงข้างลิ้น	เสียงร้ว
เสียงเจียบ		414	820	289	852	1,198	8,072	553	494	436
เสียงสระ	7,272	135	8,483	3,382	1,066	234	256	556	451	321
เสียงนาสิก	4,069	3,102	484	306	657	133		315	200	223
เสียงกึ่งสระ	1,693	1,474	325	90	244		97		78	128
เสียงเสียดแทรก	76	2,724	2		8		2		10	5
เสียงกึ่งเสียดแทรก	8	1,614							1	
เสียงกัก	1	7,737		260					381	1,174
เสียงvb		1,030					501		21	12
เสียงข้างลิ้น	9	1,627								
เสียงร้ว		2,299								

จากตารางที่ 5.2 พบว่า ชนิดขอบเขตเสียงระหว่างเสียงสระกับเสียงนาสิกมีจำนวนขอบเขตเสียง 8,483 ขอบเขตซึ่งมีจำนวนชนิดขอบเขตเสียงมากที่สุด หากสามารถระบุหาตำแหน่งขอบเขตเสียงได้ในชนิดขอบเขตเสียงที่มีจำนวนขอบเขตเสียงมากจะช่วยทำให้ประสิทธิภาพของการหาขอบเขตเสียงของระบบโดยรวมเพิ่มขึ้นอย่างเด่นชัด ขณะที่ชนิดขอบเขตเสียงที่น้อยที่สุด 1

ขอบเขต คือ ชนิดขอบเขตเสียงระหว่างเสียงกึ่งเสียดแทรกกับเสียงข้างสั้น และชนิดขอบเขตเสียงระหว่างเสียงกักกับเสียงเสียด ซึ่งเป็นชนิดขอบเขตเสียงที่มีจำนวนน้อยไม่เพียงพอการศึกษา จากจำนวนขอบเขตเสียงแต่ละชนิด สามารถพิจารณาเลือกชนิดในหัวข้อถัดไป จากการแสดงเรียงลำดับจากมากไปหาน้อยดังรูปที่ 5.1



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 5.1 ฮิตโตรแกรมจำนวนของขอเขตเสียงชนิดต่าง ๆ

วิทยานิพนธ์นี้เสนอการเลือกชนิดขอบเขตเสียงตามเงื่อนไขการเลือกชนิดขอบเขตเสียงที่อธิบายในหัวข้อการวิเคราะห์ชนิดขอบเขตเสียงในบทที่ 4 ซึ่งพิจารณาการเลือกชนิดขอบเขตเสียงที่มีจำนวนขอบเขตเสียงมาก โดยที่ยกเว้นขอบเขตเสียงที่ประกอบด้วยกลุ่มเสียงข้างลิ้นและเสียงร่ว และยกเว้นคู่กลุ่มเสียงที่ไม่สามารถระบุการเปลี่ยนแปลงของค่าอะคูสติกได้ (เนื่องจากการระบุตำแหน่งขอบเขตเสียงระหว่างกลุ่มเสียงดังกล่าว ไม่สามารถสังเกตเห็นตำแหน่งที่เกิดการเปลี่ยนแปลงทันที การเปลี่ยนแปลงของค่าลักษณะสำคัญไม่เกิดขึ้นอย่างชัดเจน) ดังนั้นผลการวิเคราะห์ชนิดขอบเขตเสียงได้ขอบเขตเสียงที่สามารถปรับหาดำแหน่งขอบเขตเสียง 21 ชนิดซึ่งมีจำนวนขอบเขตเสียงคิดเป็นร้อยละ 74.70 เมื่อเทียบกับจำนวนขอบเขตเสียงทั้งหมด ซึ่งแสดงเป็นสีแถบเข้มในตารางที่ 5.3

ตารางที่ 5.3 ร้อยละของจำนวนขอบเขตเสียงแต่ละชนิดขอบเขตเสียงต่อจำนวนขอบเขตเสียงทั้งหมด

เสียงเริ่มต้น \ เสียงสิ้นสุด	เสียงเงียบ	เสียงสระ	เสียงนาสิก	เสียงกึ่งสระ	เสียงเสียดแทรก	เสียงกึ่งเสียดแทรก	เสียงกัก	เสียงvb	เสียงข้างลิ้น	เสียงร่ว
เสียงเงียบ		0.61	1.20	0.42	1.25	1.75	11.80	0.81	0.72	0.64
เสียงสระ	10.63	0.20	12.40	4.94	1.56	0.34	0.37	0.81	0.66	0.47
เสียงนาสิก	5.95	4.53	0.71	0.45	0.96	0.19		0.46	0.29	0.33
เสียงกึ่งสระ	2.48	2.15	0.48	0.13	0.36		0.14		0.11	0.19
เสียงเสียดแทรก	0.11	3.98	0.00		0.01		0.00		0.01	0.01
เสียงกึ่งเสียดแทรก	0.01	2.36							0.00	
เสียงกัก	0.00	11.31		0.38					0.56	1.72
เสียงvb		1.51					0.73		0.03	0.02
เสียงข้างลิ้น	0.01	2.38								
เสียงร่ว		3.36								

ชนิดขอบเขตเสียงที่เลือกปรับหาดำแหน่งขอบเขตเสียงในวิทยานิพนธ์นี้ เป็นชนิดขอบเขตเสียงที่สามารถระบุตำแหน่งขอบเขตได้ชัดเจน ไม่มีความคลุมเครือในการตัดสินใจหาดำแหน่งขอบเขตเสียง ดังแสดงเป็นแถบสีเข้มในตารางที่ 5.3 แบ่งเป็น

- 1) ชนิดขอบเขตที่มีส่วนประกอบจากเสียงเงียบ ซึ่งสามารถแยกสมบัติของเสียงพูดและเสียงเงียบได้ออกจากกันอย่างชัดเจน พบว่าจำนวนขอบเขตเสียงที่เกิดจากเสียงเงียบตามชนิดขอบเขตเสียงที่เลือกนี้มีคิดเป็นร้อยละ 17.84 ของจำนวนขอบเขตเสียงทั้งหมด
- 2) ชนิดขอบเขตเสียงที่มีส่วนประกอบจากเสียงก้อง ซึ่งมีคุณสมบัติความเป็นเสียงที่มีการสั้นของเส้นเสียง ซึ่งประกอบด้วยเสียงนาสิก เสียงสระ และเสียงกึ่งสระ ดังนั้นจึงเป็นชนิดขอบเขตเสียงเหมาะสมในการพิจารณาเลือก เนื่องจากการเปลี่ยนแปลงคุณสมบัติที่เกิดขึ้นระหว่างหน่วยเสียงหนึ่งเป็นอีกหน่วยเสียงหนึ่งของชนิดขอบเขตเสียงนี้ เป็นการเปลี่ยนแปลงของหน่วยเสียงที่เป็นกลุ่มเสียงก้องและกลุ่มเสียงไม่ก้องสามารถพิจารณาการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงอย่างเด่นชัด หรือชนิดขอบเขตเสียงที่มีการเปลี่ยนแปลงระหว่างกลุ่มเสียงหนึ่งเป็นอีกกลุ่มหนึ่ง ซึ่งทั้งสองกลุ่มเป็นเสียงก้อง สามารถพิจารณาความแตกต่างระหว่างกลุ่มเสียงทั้งสองได้จากการพิจารณาค่าลักษณะของอาการออกเปล่งเสียง เช่น ค่าความถี่ฟอร์แมนท์ เป็นต้น นอกจากนี้จากการวิเคราะห์ทางสถิติ พบว่า ชนิดขอบเขตที่เป็นกลุ่มเสียงก้องตามชนิดขอบเขตเสียงที่เลือกนี้มีจำนวนขอบเขตเสียงคิดเป็นร้อยละ 39.09 ของจำนวนขอบเขตเสียงทั้งหมด (สำหรับการวิเคราะห์ชนิดขอบเขตเสียงที่เลือกใช้เพื่อการปรับตำแหน่งโดยใช้สถิติ) ซึ่งเป็นจำนวนขอบเขตเสียงที่เลือกนี้คิดเป็นจำนวนขอบเขตเสียงส่วนใหญ่ของจำนวนขอบเขตเสียงทั้งหมด
- 3) ชนิดขอบเขตที่มีส่วนประกอบจากเสียงเสียดแทรก เสียงกึ่งเสียดแทรกและเสียงกักซึ่งมีสมบัติการเกิดเสียดแทรก คล้ายลักษณะของเสียงรบกวนปรากฏกระจายทั่วหน่วยเสียง ซึ่งเป็นคุณสมบัติที่สามารถแยกหรือจำแนกออกจากเสียงที่ประกอบเป็นขอบเขตเสียง จากการวิเคราะห์ทางสถิติพบว่า จำนวนขอบเขตเสียงของชนิดขอบเขตเสียงตามทีเลือกนี้มีจำนวนมาก เป็นร้อยละ 19.16 เทียบกับจำนวนขอบเขตเสียงทั้งหมด (สำหรับการวิเคราะห์ชนิดขอบเขตเสียงที่เลือกใช้เพื่อการปรับตำแหน่งโดยใช้สถิติ)

จากเงื่อนไขการเลือกชนิดขอบเขตเสียง วิทยานิพนธ์นี้เลือก 19 ชนิดขอบเขตเสียงเพื่อหาขอบเขตหน่วยเสียงในโดยกฎตัดสิน และเลือก 21 ชนิดขอบเขตเสียงสำหรับการหาขอบเขตหน่วยเสียงโดยใช้สถิติ เนื่องจากการศึกษาความเปลี่ยนแปลงในชนิดขอบเขตเสียงในบางชนิดไม่สามารถบอกความแตกต่างของคุณสมบัติของกลุ่มเสียงได้อย่างชัดเจน เช่น ชนิดขอบเขตเสียงระหว่าง vb และ



เสียงสระ จึงทำให้ไม่สามารถหาค่าลักษณะสำคัญทางเสียงมาบ่งชี้เป็นพารามิเตอร์บอกความแตกต่างได้ ดังนั้นบางชนิดขอบเขตเสียงจึงยกเว้นการเลือกเพื่อปรับโดยกฎตัดสติใจ (ซึ่งจะอธิบายต่อไป) ดังตารางที่ 5.4

ตารางที่ 5.4 สรุปชนิดขอบเขตเสียงที่ใช้ในวิทยานิพนธ์นี้

ชนิดขอบเขตเสียง	การปรับหาค่าตำแหน่งขอบเขตเสียง	
	ตัดสติใจโดยใช้สถิติ	ตัดสติใจโดยใช้กฎ
เสียงเงิบ-เสียงเสียดแทรก	✓	✓
เสียงเงิบ-เสียงสระ	✓	✓
เสียงเงิบ-เสียงกัก	✓	✓
เสียงเงิบ-เสียงกึ่งสระ	✓	✓
เสียงเงิบ-เสียงกึ่งเสียดแทรก	✓	✓
เสียงเงิบ-เสียงนาสิก	✓	✓
เสียงเงิบ-vb	✓	✓
เสียงนาสิก-เสียงเสียดแทรก	✓	✓
เสียงนาสิก-เสียงสระ	✓	✓
เสียงนาสิก-เสียงเงิบ	✓	✓
เสียงสระ-เสียงเงิบ	✓	✓
เสียงสระ-เสียงเสียดแทรก	✓	✓
เสียงสระ-เสียงนาสิก	✓	✓
เสียงกึ่งสระ-เสียงเสียดแทรก	✓	✓
เสียงกึ่งสระ-เสียงเงิบ	✓	✓
เสียงเสียดแทรก-เสียงสระ	✓	✓
เสียงกัก-เสียงสระ	✓	✓
เสียงกึ่งเสียดแทรก-เสียงสระ	✓	✓
vb-เสียงสระ	✓	✗
เสียงสระ-vb	✓	✗
เสียงสระ-เสียงกัก	✓	✗
เสียงนาสิก-เสียงกึ่งเสียดแทรก	✗	✓

จากตารางที่ 5.3 พบว่า บางชนิดขอบเขตเสียงที่ไม่ได้เลือกปรับหาค่าตำแหน่งขอบเขตเสียงในการตัดสติใจจากกฎตัดสติใจ เนื่องจากในขั้นตอนการสร้างกฎตัดสติใจเพื่อปรับหาค่าตำแหน่งขอบเขต

เสียงต้องศึกษาค่าการเปลี่ยนแปลงลักษณะสำคัญที่ใช้ในการบ่งชี้ความแตกต่างของกลุ่มเสียง ซึ่งการวิเคราะห์ความแตกต่างของค่าลักษณะในขอบเขตเสียงระหว่างเสียง vb และเสียงสระ หรือในขอบเขตเสียงระหว่างเสียงสระและเสียง vb พบว่าไม่สามารถพิจารณาค่าลักษณะสำคัญที่ชี้ความแตกต่างได้เด่นชัด ถึงแม้ว่าค่าพลังงานที่สามารถบอกความแตกต่างของทั้งสองกลุ่มได้ เพราะกลุ่มเสียง vb มีพลังงานในช่วงความถี่ต่ำเท่านั้น ขณะที่กลุ่มเสียงสระมีค่าพลังงานในช่วงองค์ประกอบความถี่ต่างๆ แต่เมื่อนำมาสร้างกฎแล้วพบว่า ค่าพลังงานไม่สามารถนำมาสร้างกฎในการระบุหาตำแหน่งขอบเขตเสียงได้ดีนัก ขณะที่การวิเคราะห์แตกต่างของค่าลักษณะในขอบเขตเสียงระหว่างกลุ่มเสียงสระและเสียงกัก พบว่าขนาดของหน่วยเสียงกักนั้นมีขนาดเล็ก ในการศึกษาค่าลักษณะสำคัญที่เป็นตัวแทนเสียงไม่สามารถวิเคราะห์ค่าการเปลี่ยนแปลงลักษณะสำคัญที่ใช้บอกความแตกต่างของเสียงสระและกลุ่มเสียงกักได้ นอกจากนี้ตารางที่ 5.3 แสดงชนิดขอบเขตเสียงที่ไม่ได้เลือกปรับตำแหน่งขอบเขตเสียงในการตัดสินใจจากการใช้สถิติ เนื่องจากในขั้นตอนการสร้างต้นแบบการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงต้องใช้ข้อมูลที่เป็นตัวอย่างกลุ่มขอบเขตเสียงและตัวอย่างกลุ่มที่ไม่ใช่ขอบเขตเสียง จากการพิจารณาจำนวนตัวอย่างขอบเขตเสียงในขอบเขตเสียงระหว่างกลุ่มเสียงนาสิกและกลุ่มเสียงกึ่งเสียดแทรกพบว่า จำนวนขอบเขตเสียงไม่เพียงพอในการใช้เป็นตัวอย่างการสร้างต้นแบบการจำแนกได้ ดังนั้นชนิดขอบเขตเสียงดังกล่าวจึงไม่ได้รับการพิจารณา

### 3. ผลการวิเคราะห์ค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียง

ขั้นตอนการวิเคราะห์ค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียงนี้วิเคราะห์จากแผนภาพกล่องที่แสดงค่ากลางและการกระจายของค่าลักษณะทางเสียงของกลุ่มขอบเขตเสียงเปรียบเทียบกับกลุ่มไม่ใช่ขอบเขตเสียงแต่ละชนิดขอบเขต การวิเคราะห์ความแปรปรวน (ANOVA) และค่าประสิทธิภาพของการจำแนกเมื่อใช้ค่าลักษณะสำคัญ และวิเคราะห์ค่าลักษณะสำคัญทางเสียงเพื่อเลือกค่าที่เหมาะสมในการบ่งชี้ลักษณะเด่นของกลุ่มเสียงที่พิจารณาและบอกลักษณะความแตกต่างระหว่างสองกลุ่มเสียงที่คิดกันได้ จะได้ชุดค่าลักษณะสำคัญทางเสียงเพื่อสร้างชุดเวกเตอร์ลักษณะเฉพาะที่เป็นตัวแทนของกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงสำหรับแต่ละชนิดขอบเขตนั้น ซึ่งแสดงลักษณะทางเสียงที่เลือกสำหรับแต่ละชนิดขอบเขตเสียงดังตารางที่ 5.5 และนำชุดลักษณะทางเสียงที่เลือกมาทดสอบประสิทธิภาพของการจำแนก และวัดความน่าจะเป็นของความผิดพลาดที่เกิดขึ้นจากการจำแนก ตามแต่ละชนิดขอบเขตเสียงที่เลือกค่าลักษณะสำคัญทางเสียงแตกต่างกัน

ตารางที่ 5.5 ค่าลักษณะสำคัญทางเสียงที่เลือกมาฝึกฝนแบบจำลองสำหรับแต่ละชนิดขอบเขตเสียง

ชนิดขอบเขตเสียง	ค่าลักษณะสำคัญทางเสียงที่ถูกเลือก	ร้อยละความถูกต้อง การจำแนก	ความน่าจะเป็นของความ ผิดพลาดระหว่างสองกลุ่ม
เสียงเจียบ-เสียงเสียดแทรก	MFCC-STM-E1-E2-E3-E4-E5-ENSN- LOGEN-ZCR	71.37	0.27
เสียงเจียบ-เสียงสระ	MFCC-STM-F1 F2 F3-F2F1-E1-E2- E3-E4-LOGEN	73.02	0.24
เสียงกึ่งสระ-เสียงเจียบ	MFCC-STM-F1 F2 F3-CORR-E1- ZCR-ENSN-LOGEN-ZCR	73.18	0.26
เสียงนาสิก-เสียงเจียบ	MFCC-STM-F1 F2-F3-F2F1-E1- ENSN-LOGEN-ZCR	73.72	0.25
เสียงสระ-เสียงเจียบ	MFCC-STM-F1 -F2F1-E1-E2-E3- LOGEN	73.81	0.26
เสียงสระ-เสียงกัก	MFCC-F2F1-LOGEN-ENSN-E4-ZCR	73.85	0.29
เสียงสระ-เสียงเสียดแทรก	MFCC-F2-F3-F2F1- E1-E4-E5-ENSN- LOGEN-F0-ZCR	75.03	0.23
เสียงเจียบ-เสียงกัก	MFCC-STM-E1-E2-E3-E4-E5-LOGEN	75.05	0.25
เสียงเจียบ-เสียงกึ่งสระ	MFCC-STM-F1 F2 -F2F1-CORR-E1- E2-ENSN-LOGEN-F0-ZCR	76.65	0.18
เสียงกึ่งสระ-เสียงเสียดแทรก	MFCC-F1 F2 -F2F1- E1-E2-E3-E4-E5- ZCR-ENSN-LOGEN-F0-ZCR	76.75	0.19
เสียงสระ-vb	MFCC-STM-F1 -F2F1-CORR-E1-E2- E3-E4-ENSN-LOGEN-ZCR	76.92	0.23
เสียงสระ-เสียงนาสิก	MFCC-STM-F1-CORR-E1-E2-E3-E4- ENSN-LOGEN-ZCR	78.24	0.25
เสียงนาสิก-เสียงเสียดแทรก	MFCC-F1 F2 -F2F1-TOTALENERGY- E1-E3-E4-E5-ENSN-F0-ZCR	78.99	0.19

ชนิดขอบเขตเสียง	ค่าลักษณะสำคัญทางเสียงที่ถูกเลือก	ร้อยละความถูกต้อง การจำแนก	ความน่าจะเป็นของความ ผิดพลาดระหว่างสองกลุ่ม
เสียงนาสิก-เสียงสระ	MFCC-STM-F1 F2F1- TOTALENERGY -E1-E2-E3-E4- ENSN-LOGEN-ZCR	80.35	0.21
เสียงเจียบ-vb	MFCC-STM-F1 F2 F3 -F2F1-E1-E2- ENSN-LOGEN-ZCR	81.02	0.20
เสียงเสียดแทรก-เสียงสระ	MFCC-STM-F1 -CORR- E1-E2-E3- ENSN-LOGEN-F0-ZCR	82.12	0.18
เสียงกัก-เสียงสระ	MFCC-STM-F1 -F2-F3-F2F1-CORR- E1-E2-E3-LOGEN-F0-ZCR	82.47	0.21
เสียงเจียบ-เสียงกึ่งเสียดแทรก	MFCC-STM-CORR- E1-E3-E4-E5- ENSN-LOGEN-ZCR	83.66	0.18
Vb-เสียงสระ	MFCC-STM-F1 -F2F1-CORR-E1-E2- E3-E4-ENSN-LOGEN-ZCR	84.21	0.16
เสียงกึ่งเสียดแทรก-เสียงสระ	MFCC-STM-F1 F2 F3-F2F1-CORR- E1-E2-E3-LOGEN-F0-ZCR	84.44	0.17
เสียงเจียบ-เสียงนาสิก	MFCC-STM-F1 F2 -F2F1-CORR-E1- ENSN-LOGEN-F0-ZCR	84.54	0.17

จากตารางที่ 5.5 ใช้สัญลักษณ์แทนค่าลักษณะสำคัญทางเสียงต่าง ๆ ดังนี้ MFCC แทนค่าสัมประสิทธิ์เมลเฟรินซีเคปสตรัม STM แทนค่าการเปลี่ยนแปลงรูปร่างของสเปกตรัม F1 F2 F3 แทนความถี่ฟอร์แมนท์ที่หนึ่ง สอง สามตามลำดับ F2F1 แทนความกว้างระหว่างความถี่ฟอร์แมนท์ที่หนึ่งและสอง E1, E2, E3, E4, E5 แทนพลังงานในช่วงความถี่ที่หนึ่ง สอง สาม สี่ ห้า ตามลำดับ ENSN แทนอัตราส่วนพลังงานสำหรับช่วง LOGEN แทนลือกของพลังงาน TOTALENERGY แทนพลังงานเสียง CORR แทนความสัมพันธ์ระหว่างเฟรมที่สนใจและเฟรมที่อยู่ก่อนหน้า F0 แทนความถี่มูลฐาน ZCR แทนอัตราการตัดศูนย์ ค่าลักษณะสำคัญทางเสียงที่ใช้ในการวิเคราะห์นี้ได้อธิบายวิธีการคำนวณในบทที่ 2 และวิเคราะห์แต่ละค่าในหัวข้อการวิเคราะห์ลักษณะสำคัญ

### ผลการวิเคราะห์ค่าลักษณะสำคัญในการจำแนกกลุ่ม

ผลที่ได้จากการวิเคราะห์ค่าลักษณะสำคัญทางเสียงแต่ละค่าและใช้จำแนกกลุ่มซึ่งได้แสดงในตารางที่ 5.5 สรุปผลการวิเคราะห์ได้ ดังนี้

- 1) ค่าสัมประสิทธิ์เมลเฟร็ควีนซีเคปสตรัม (แทนด้วย MFCC ) เป็นค่าลักษณะสำคัญทางเสียงที่ถูกเลือกมาใช้ในทุกชนิดขอบเขตเสียง เนื่องจากค่า MFCC นิยมใช้เป็นตัวแทนเสียงในการรู้จำเสียงซึ่งถูกสกัดจากตัวกรองจำลองตามความสามารถได้ยินของมนุษย์ และเมื่อพิจารณาตามแต่ละชนิดขอบเขตเสียง ผลการวิเคราะห์จำแนกกลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงในทุกชนิดขอบเขตเสียงมีประสิทธิภาพการจำแนกกลุ่มสูง เมื่อใช้ค่า MFCC เป็นตัวแทนข้อมูล ดังนั้นค่า MFCC จึงได้เลือกมาใช้เป็นตัวแทน
- 2) จากข้อ 1) ซึ่งแสดงผลการวิเคราะห์ของค่า MFCC ส่งผลให้ค่า STM (ซึ่งได้จากการวัดค่าการเปลี่ยนแปลงของรูปร่างสเปกตรัมที่คำนวณจากค่า MFCC) สามารถบอกความแตกต่างและจำแนกกลุ่มขอบเขตเสียงออกจากกลุ่มที่ไม่ใช่ขอบเขตเสียงได้ดี ซึ่งได้ถูกเลือกในหลายชนิดขอบเขตเสียง เช่นกัน
- 3) จากตารางที่ 5.5 ค่าความถี่ฟอร์แมนต์ถูกเลือกมาใช้เป็นตัวแทนค่าลักษณะสำคัญทางเสียงในชนิดขอบเขตเสียงที่เป็นมีกลุ่มเสียงสระ เสียงนาสิก และเสียงกึ่งสระเป็นหน่วยเสียงที่ประกอบในขอบเขตเสียงซึ่งชนิดขอบเขตเสียงดังกล่าวมีคุณสมบัติของความเป็นเสียงก้อง เนื่องจากความถี่ฟอร์แมนต์จะปรากฏชัดในเสียงที่เกิดจากการสั่นเสี้ยนเสียงได้ดี และค่า F2F1 ซึ่งเป็นค่าความกว้างระหว่างความถี่ฟอร์แมนต์ที่หนึ่งและสองจึงถูกเลือกใช้ในชนิดขอบเขตเสียงที่มีกลุ่มเสียงสระและกลุ่มเสียงกึ่งสระเป็นส่วนประกอบด้วย เนื่องจากค่าความกว้างระหว่างความถี่ฟอร์แมนต์เป็นค่าที่ระบุความแตกต่างทิศทางของการเคลื่อนที่ของความถี่ฟอร์แมนต์ได้ทางอ้อม ดังนั้น ค่านี้อาจเลือกมาใช้เพื่อบอกความแตกต่างของทิศทางเคลื่อนที่ของความถี่ฟอร์แมนต์ ดังนั้นค่า F2F1 มีความสามารถในการระบุความแตกต่างของการเคลื่อนที่ของความถี่ฟอร์แมนต์(Formant contour) ที่ระบุจำแนกกลุ่มเสียงสระและกลุ่มเสียงกึ่งสระออกจากกัน
- 4) ค่าพลังงานเสียง ซึ่งแทนด้วย TOTALENERGY และค่าลือกของพลังงาน ซึ่งแทนด้วย LOGEN เป็นค่าที่ใช้เป็นตัวแทนของพลังงานของสัญญาณเสียงพูด ซึ่งได้รับการเลือกมาพิจารณาเป็นตัวแทนของค่าลักษณะสำคัญทางเสียงในทุกชนิดขอบเขต



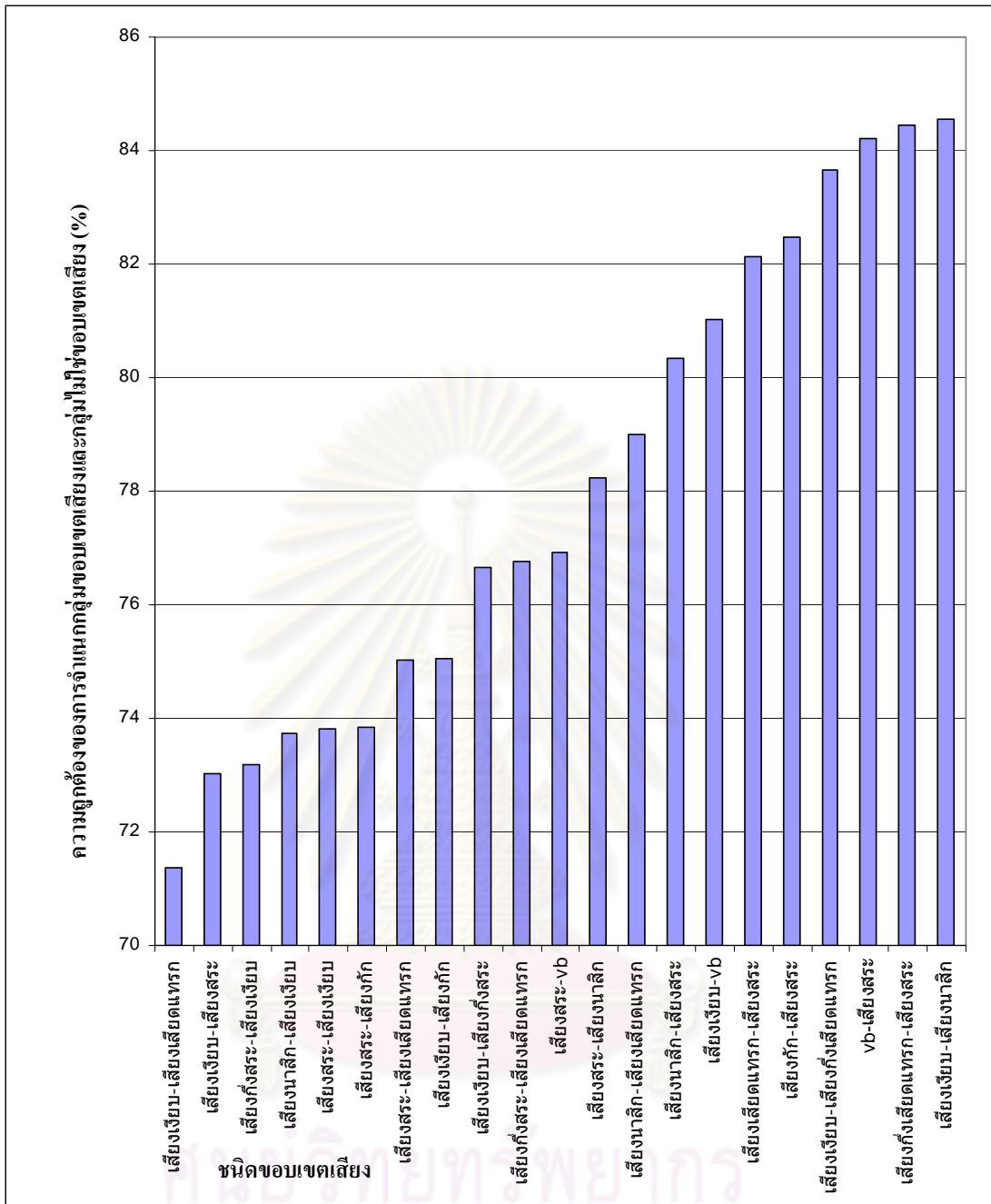
เสียง เนื่องจากหน่วยเสียงในกลุ่มเสียงที่แตกต่างกันมีค่าพลังงานต่างกันทำให้ที่ตำแหน่งขอบเขตเสียงเกิดการเปลี่ยนแปลงของค่าพลังงาน จึงนำค่าพลังงานเป็นค่าระบุการเปลี่ยนแปลงค่าลักษณะทางเสียงระหว่างเฟรมได้

- 5) จากการศึกษาพฤติกรรมค่าพลังงานเสียงตามช่วงความถี่ ถูกเลือกมาใช้ซึ่งการเปลี่ยนแปลงค่าพลังงานในช่วงความถี่ที่แตกต่างกัน ซึ่งแต่ละชนิดขอบเขตเสียงมีความแตกต่างของค่าพลังงานในแต่ละช่วง เช่น เสียงเสียดแทรกมีพลังงานในช่วงความถี่สูงเด่นชัดเมื่อเปรียบเทียบกับเสียงอื่น ดังนั้นค่าพลังงานในช่วง E4 E5 จึงถูกเลือกมาใช้เป็นค่าบ่งชี้ในชนิดขอบเขตเสียงที่ประกอบด้วย เสียงเสียดแทรกขณะที่ชนิดขอบเขตเสียงเงียบ ไม่มีค่าพลังงานหรือมีน้อยมาก ดังนั้นค่าพลังงานในช่วงองค์ประกอบความถี่ต่างๆและค่าพลังงานเสียงจึงถูกเลือกมาใช้เป็นตัวแทน และ ชนิดขอบเขตเสียงระหว่างเสียงสระกับเสียงนาสิกนั้น สามารถบ่งชี้ได้จากความแตกต่างกันของพลังงานในช่วงองค์ประกอบความถี่ต่ำ พลังงานในช่วงองค์ประกอบความถี่ตามฟอร์แมนท์ที่หนึ่ง และความถี่ตามฟอร์แมนท์ที่สอง เป็นต้น ซึ่งค่าพลังงานในแต่ละช่วงองค์ประกอบความถี่ถูกเลือกมาใช้แตกต่างกัน ตามสมบัติของกลุ่มเสียงที่พิจารณา ซึ่งใช้การวิเคราะห์ทางสถิติจากแผนภาพกล่องประกอบด้วย
- 6) จากตารางที่ 5.5 พบว่าในชนิดขอบเขตเสียงที่ประกอบด้วยกลุ่มเสียงนาสิก เสียงกึ่งสระ และเสียงสระ เช่น ระหว่างเสียงกึ่งสระ-เสียงเงียบ ระหว่างเสียงนาสิก-เสียงเงียบ เสียงสระ-เสียงกัก เสียงสระ-เสียงนาสิก ฯลฯ เลือกค่า ENSN มาใช้เป็นตัวแทนลักษณะทางเสียง ซึ่งค่า ENSN ได้จากการคำนวณอัตราส่วนของพลังงานองค์ประกอบความถี่ช่วงความถี่ 0-0.357 กิโลเฮิร์ตซ์ กับช่วงความถี่ 0.358-5.373 กิโลเฮิร์ตซ์ ค่า ENSN เหมาะสมในการระบุคุณสมบัติของกลุ่มเสียงนาสิก เสียงสระและเสียงกึ่งสระ ดังนั้นการใช้ค่าลักษณะสำคัญทางเสียงนี้จึงนิยมใช้ในขอบเขตเสียงที่ประกอบจากกลุ่มเสียงดังกล่าว
- 7) เนื่องจากค่าอัตราการตัดศูนย์มีค่าสูงในเสียงที่ไม่เกิดรายคาบ และเกิดการเสียดแทรกคล้ายเสียงรบกวน เช่น กลุ่มเสียงเสียดแทรก กลุ่มเสียงกึ่งเสียดแทรก และกลุ่มเสียงกัก ดังนั้นจากตารางที่ 5.5 จึงพบว่าชนิดขอบเขตเสียงประกอบด้วยกลุ่มเสียงดังกล่าว เลือกค่าอัตราการตัดศูนย์เป็นค่าลักษณะสำคัญทางเสียงในการบอกความแตกต่างระหว่าง เป็นส่วนมากของชนิดของเขตเสียงกลุ่มเสียงดังกล่าว

- 8) ค่า  $F_0$  เป็นค่าความถี่มูลฐานซึ่งแสดงความเป็นเสียงก้องได้ดี สามารถบอกความแตกต่างที่เกิดจากส่วนประกอบของกลุ่มเสียงที่เป็นเสียงก้องออกจากกลุ่มเสียงที่ไม่ใช่เสียงก้อง เช่น ชนิดขอบเขตระหว่างเสียงเสียดแทรก-เสียงสระ ดังนั้นตารางที่ 5.5 แสดงว่าค่า  $F_0$  เป็นค่าลักษณะสำคัญที่สามารถจำแนกกลุ่มได้ดี และค่า  $F_0$  ถูกเลือกเพื่อระบุตำแหน่งขอบเขตเสียงที่มีส่วนประกอบของเสียงเสียดแทรกกับเสียงก้อง
- 9) จากตารางที่ 5.5 ที่แสดงค่าความน่าจะเป็นของความผิดพลาดระหว่างการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงของมีค่าอยู่ระหว่าง 0.15 ถึง 0.30 และพบว่าชนิดขอบเขตเสียงที่มีประสิทธิภาพการจำแนกสูงสุดไม่ใช่ชนิดขอบเขตเสียงที่มีค่าความน่าจะเป็นความผิดพลาดต่ำที่สุด

จากตารางที่ 5.5 ซึ่งวัดประสิทธิภาพของค่าลักษณะสำคัญที่เลือกมาเป็นตัวแทนข้อมูลของการจำแนกตามหัวข้อ “ขั้นตอนการทดสอบต้นแบบการจำแนก” เมื่อพิจารณาความสามารถในการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง โดยเรียงลำดับค่าร้อยละความถูกต้องการจำแนกของแต่ละชนิดขอบเขตเสียง จากชนิดขอบเขตเสียงที่สามารถจำแนกที่มีค่าน้อย(หรือสามารถแยกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงออกจากกันได้น้อย)ไปหาชนิดขอบเขตเสียงที่ค่าความถูกต้องการจำแนกมาก (หรือสามารถจำแนกกลุ่มได้ดีมากที่สุด) ซึ่งเรียงลำดับจากซ้ายไปขวา แสดงในรูปที่ 5.2 พบว่าชนิดขอบเขตเสียงเงียบกับเสียงเสียดแทรกสามารถจำแนกกลุ่มได้ยาก และชนิดขอบเขตเสียงเงียบกับเสียงนาสิกสามารถจำแนกกลุ่มได้ดีที่สุด

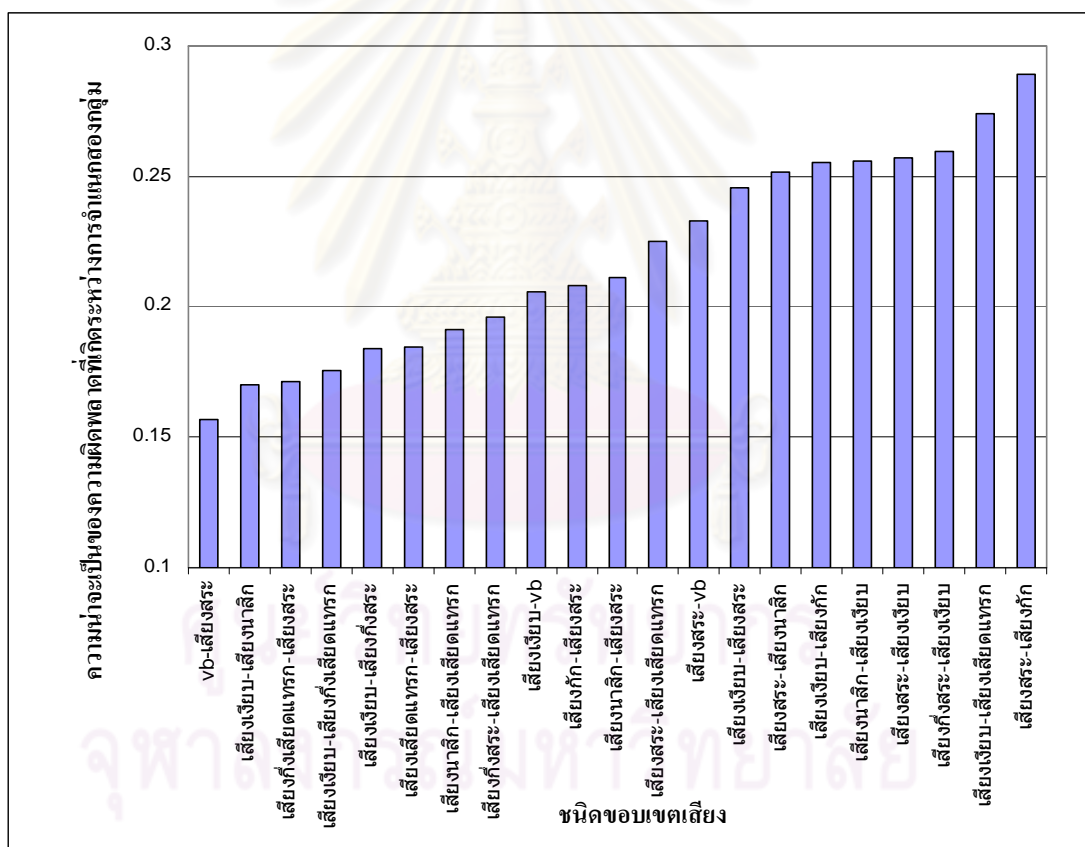
ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 5.2 อีซีโตรแกรมความถูกต้องของการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง

จากรูปที่ 5.2 ที่แสดงร้อยละความถูกต้องของการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง พบว่าชนิดขอบเขตเสียงระหว่างเสียงเงียบ-เสียงเสียดแทรกมีร้อยละความถูกต้อง 71.37 ซึ่งมีประสิทธิภาพน้อยที่สุดเมื่อเปรียบเทียบกับชนิดขอบเขตเสียงอื่นทั้งหมด 21 ชนิด ตามการวิเคราะห์เลือกขอบเขตเสียงในวิทยานิพนธ์ และชนิดขอบเขตเสียงระหว่างเสียงเงียบ-เสียงนาลิกมีร้อยละความถูกต้องของการจำแนก 84.54 ซึ่งสามารถจำแนกได้ดีที่สุด

นอกจากนี้การวิเคราะห์ความสามารถในการจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงสามารถพิจารณาได้จาก ค่าความผิดพลาดของจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงเมื่อใช้ชุดค่าลักษณะทางเสียงมาเป็นตัวแทนข้อมูล (ดังที่แสดงในตารางที่ 5.5) ค่าความผิดพลาดของจำแนกกลุ่มนี้ได้อธิบายในบทที่ 4 หัวข้อ “การวิเคราะห์จากผลทดสอบประสิทธิภาพของการจำแนกกลุ่ม” เมื่อพิจารณาเรียงลำดับผลการค่าความผิดพลาดน้อย (ซึ่งหมายถึงชนิดขอบเขตเสียงที่สามารถจำแนกกลุ่มได้ดี) ไปหาผลการค่าความผิดพลาดมาก (หรือสามารถแยกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงออกจากกันยาก) เรียงลำดับจากซ้ายไปขวา แสดงในรูปที่ 5.3 และจากการพิจารณาฮิสโตแกรมรูปที่ 5.2 และ 5.3 ซึ่งแสดงความสามารถการจำแนกของแต่ละชนิดขอบเขตเสียงนั้น การเรียงลำดับชนิดขอบเขตเสียงอยู่ในทิศทางที่สวนทางกัน นั่นคือ ชนิดขอบเขตเสียงที่มีความสามารถการจำแนกกลุ่มดีจะอยู่ในตำแหน่งฮิสโตแกรมแห่งทางขวาของรูปที่ 5.2 และอยู่ในตำแหน่งฮิสโตแกรมแห่งทางด้านซ้ายของรูปที่ 5.3



รูปที่ 5.3 ฮิสโตแกรมความน่าจะเป็นความผิดพลาดระหว่างกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง

จากรูปที่ 5.3 แสดงความน่าจะเป็นความผิดพลาดของจำแนกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง ซึ่งชนิดขอบเขตระหว่างเสียง vb-เสียงสระ มีค่าความผิดพลาดน้อยที่สุดและชนิดขอบเขตระหว่างเสียงสระ-เสียงกัก มีค่าความผิดพลาดมากที่สุด

เมื่อเลือกค่าลักษณะทางเสียงที่ใช้เป็นตัวแทนข้อมูลการจำแนกกลุ่มขอบเขตเสียง และกลุ่มที่ไม่ใช่ขอบเขตเสียง สามารถนำมาสร้างเวกเตอร์ลักษณะเฉพาะของข้อมูลชุดฝึกฝนและเตรียมข้อมูลสำหรับทดสอบของแต่ละชนิดขอบเขตเสียงได้ เพื่อเลือกเป็นตำแหน่งขอบเขตที่แท้จริงได้ต่อไป

#### 4. ผลการวิเคราะห์ขนาดกรอบหน้าต่างการปรับ (Refining window)

ผลการวิเคราะห์ขนาดกรอบหน้าต่างการปรับรายงานผลการวิเคราะห์ในหัวข้อการวัดประสิทธิภาพของการระบุตำแหน่งขอบเขตเสียงจากแบบอ้างอิงซึ่งแสดงความคลาดเคลื่อนคำตอบที่ได้จากแบบอ้างอิงเทียบกับคำตอบจริง และผลการวิเคราะห์ที่เกี่ยวเนื่องกับการใช้ขนาดหน้าต่างการปรับต่างกัน ซึ่งรายงานผลประสิทธิภาพระบบที่ใช้ขนาดกรอบหน้าต่างการปรับแตกต่างกันจะแสดงในตารางที่ 5.8 และตารางที่ 5.9 ต่อไป

#### 5. ผลจากการปรับหาขอบเขตหน่วยเสียง

##### 5.1. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้กฎตัดสินใจ

ชนิดขอบเขตเสียงที่เลือกปรับหาตำแหน่งขอบเขตเสียงด้วยวิธีนี้ เลือก 19 ชนิดขอบเขตเสียงดังตารางที่ 5.4 จากขั้นตอนการวิเคราะห์ชนิดขอบเขตเสียง และหาตำแหน่งขอบเขตเสียงการปรับหาตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจ ทดสอบประสิทธิภาพของวิธีที่เสนอการหาตำแหน่งขอบเขตเสียงจากการใช้กฎที่ตั้งขึ้นด้วยข้อมูลชุดข้อมูลที่ 2 (ชุดข้อมูลเพื่อพัฒนาระบบ) และเพื่อทดสอบความคงทนของกฎที่ตั้งขึ้นจึงทดสอบกับชุดข้อมูลที่ 3 (ชุดข้อมูลเพื่อใช้ทดสอบระบบ) ซึ่งแสดงร้อยละความถูกต้องของตำแหน่งขอบเขตเสียงในแต่ละชนิดขอบเขตเสียง โดยรายงานผลความถูกต้องที่ระดับความคลาดเคลื่อนที่ยอมรับได้ 1 เฟรมในตารางที่ 5.6 ที่ใช้กับชุดข้อมูลที่ 2 และ 5.8 ที่ใช้กับชุดข้อมูลที่ 3 และแสดงร้อยละตามค่าความผิดพลาดที่ลดลงจากการคำนวณความแตกต่างระหว่างร้อยละความถูกต้องของตำแหน่งขอบเขตเสียงที่ได้จากการหาตำแหน่งขอบเขตเสียงจากกฎที่ตั้งขึ้นกับร้อยละความถูกต้องของตำแหน่งขอบเขตเสียงที่ได้จากแบบอ้างอิง เปรียบเทียบกับร้อยละความถูกต้องของตำแหน่งขอบเขตเสียงที่ได้จากแบบอ้างอิงตามสมการที่ (3.2)

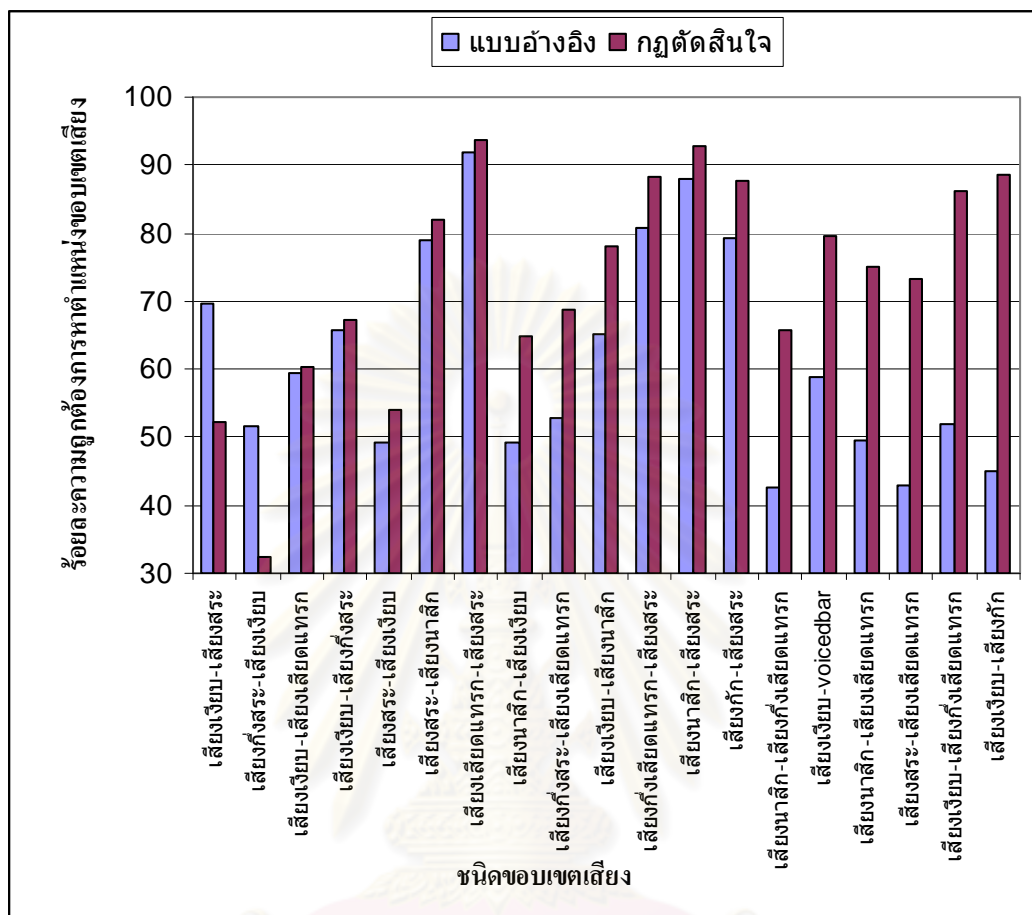


ตารางที่ 5.6 ผลการปรับหาค่าแหน่งขอบเขตเสียงโดยวิธีใช้กฎตัดสนใจเมื่อใช้ชุดข้อมูลที่ 2

ชนิดขอบเขตหน่วยเสียง	ร้อยละความถูกต้อง		ร้อยละของ ความผิดพลาดที่ลดลง
	แบบอ้างอิง	กฎตัดสนใจ	
เสียงเจียบ-เสียงสระ	69.70	52.17	-57.85
เสียงกึ่งสระ-เสียงเจียบ	51.75	32.50	-39.90
เสียงเจียบ-เสียงเสียดแทรก	59.45	60.23	1.92
เสียงเจียบ-เสียงกึ่งสระ	65.85	67.12	3.72
เสียงสระ-เสียงเจียบ	49.13	54.15	9.87
เสียงสระ-เสียงนาสิก	79.10	81.86	13.21
เสียงเสียดแทรก-เสียงสระ	92.01	93.77	22.03
เสียงนาสิก-เสียงเจียบ	49.15	64.82	30.82
เสียงกึ่งสระ-เสียงเสียดแทรก	52.94	68.63	33.34
เสียงเจียบ-เสียงนาสิก	65.30	78.11	36.92
เสียงกึ่งเสียดแทรก-เสียงสระ	80.92	88.30	38.68
เสียงนาสิก-เสียงสระ	88.13	92.76	39.01
เสียงกัก-เสียงสระ	79.30	87.57	39.95
เสียงนาสิก-เสียงกึ่งเสียดแทรก	42.50	65.85	40.60
เสียงเจียบ-voicedbar	58.96	79.65	50.41
เสียงนาสิก-เสียงเสียดแทรก	49.41	75.15	50.88
เสียงสระ-เสียงเสียดแทรก	42.86	73.41	53.47
เสียงเจียบ-เสียงกึ่งเสียดแทรก	52.05	86.10	71.01
เสียงเจียบ-เสียงกัก	45.07	88.46	78.99
รวมทั้งหมด	63.82	73.44	26.59

นอกจากนี้ตารางที่ 5.6 แสดงความผิดพลาดที่ลดลงของร้อยละความถูกต้องของตำแหน่งขอบเขตเสียง เรียบลำดับชนิดขอบเขตเสียงตามค่าความผิดพลาดที่ลดลงจากน้อยไปมาก ซึ่งชนิดขอบเขตเสียงระหว่างเสียงเจียบกับเสียงสระ สามารถปรับปรุงผลการหาขอบเขตเสียงเมื่อเทียบกับแบบอ้างอิงได้น้อยที่สุด (มีความผิดพลาดลดลงน้อยที่สุด) และสามารถหาขอบเขตเสียงได้มากที่สุดชนิดขอบเขตเสียงระหว่างเสียงเจียบกับเสียงกัก ที่มีค่าความผิดพลาดลดลงมากที่สุด

78.99 และในรูปที่ 5.4 แสดงค่าร้อยละความถูกต้องการหาตำแหน่งขอบเขตเสียงด้วยวิธีกฎตัดสติใจในชุดข้อมูลที่2 ที่เรียงลำดับชนิดขอบเขตเสียงตามความผิดพลาดที่ลดลงจากน้อยไปมาก



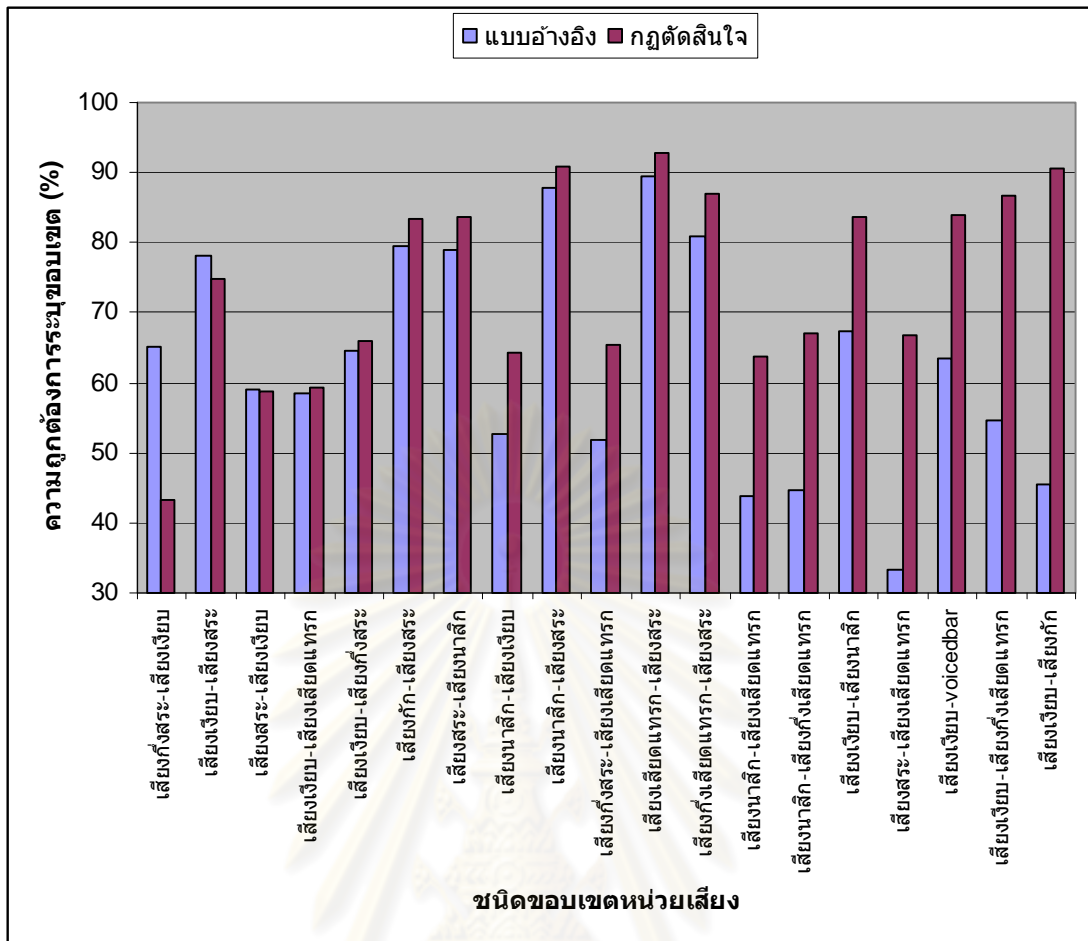
รูปที่ 5.4 ฮิสโตแกรมความถูกต้องการหาตำแหน่งขอบเขตเสียงด้วยวิธีกฎตัดสติใจในชุดข้อมูลที่2

เมื่อทดสอบกฎการพิจารณาหาตำแหน่งที่เหมาะสมกับขอบเขตหน่วยเสียงจากขั้นตอนการสร้างกฎตัดสติใจด้วยชุดข้อมูลที่2 แล้ว จึงนำกฎการตัดสติใจและวิธีการเตรียมชูปค่าลักษณะทางเสียงแต่ละชนิดขอบเขตที่ได้ มาทดสอบกับชุดข้อมูลที่ 3 โดยมีเงื่อนไขว่า ไม่เปลี่ยนแปลงกฎและชุดค่าลักษณะสำคัญที่ได้จากขั้นตอนการสร้างกฎ

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.7 ผลการปรับหาค่าแหน่งขอบเขตเสียงโดยวิธีใช้กฎตัดสินใจเมื่อใช้ชุดข้อมูลที่ 3 เมื่อใช้ขนาดกรอบพิจารณา 5 เฟรม

ชนิดขอบเขตหน่วยเสียง	ร้อยละความถูกต้อง		ร้อยละของความผิดพลาดที่ลดลง
	แบบอ้างอิง	กฎตัดสินใจ	
เสียงกึ่งสระ-เสียงเจียบ	65.23	43.18	-63.42
เสียงเจียบ-เสียงสระ	78.26	74.78	-16.01
เสียงสระ-เสียงเจียบ	59.12	58.68	-1.08
เสียงเจียบ-เสียงเสียดแทรก	58.45	59.29	2.02
เสียงเจียบ-เสียงกึ่งสระ	64.46	65.93	4.14
เสียงกัก-เสียงสระ	79.55	83.43	18.97
เสียงสระ-เสียงนาสิก	79.09	83.78	22.43
เสียงนาสิก-เสียงเจียบ	52.79	64.34	24.47
เสียงนาสิก-เสียงสระ	87.87	90.97	25.56
เสียงกึ่งสระ-เสียงเสียดแทรก	51.72	65.52	28.58
เสียงเสียดแทรก-เสียงสระ	89.62	92.78	30.44
เสียงกึ่งเสียดแทรก-เสียงสระ	80.98	87.07	32.02
เสียงนาสิก-เสียงเสียดแทรก	43.95	63.69	35.22
เสียงนาสิก-เสียงกึ่งเสียดแทรก	44.60	67.20	40.79
เสียงเจียบ-เสียงนาสิก	67.44	83.72	50.00
เสียงสระ-เสียงเสียดแทรก	33.20	66.80	50.30
เสียงเจียบ-voicedbar	63.50	83.94	56.00
เสียงเจียบ-เสียงกึ่งเสียดแทรก	54.66	86.82	70.93
เสียงเจียบ-เสียงกัก	45.44	90.68	82.92
รวมทั้งหมด	65.00	73.49	24.26



รูปที่ 5.5 อีตโรแกรมความถูกต้องการหาตำแหน่งขอบเขตเสียงด้วยวิธีกฎตัดสินใจในชุดข้อมูลที่ 3

ตารางที่ 5.7 แสดงผลการระบุหาตำแหน่งขอบเขตเสียงจากวิธีกฎตัดสินใจที่ทดสอบกับชุดข้อมูลที่ 3 ซึ่งใช้ขนาดกรอบพิจารณา 5 เฟรม โดยวัดความถูกต้องที่ระดับความคลาดเคลื่อนที่ยอมรับได้ 1 เฟรม และแสดงความผิดพลาดที่ลดลงของร้อยละความถูกต้องของตำแหน่งขอบเขตเสียง เรียงลำดับชนิดขอบเขตเสียงตามค่าความผิดพลาดที่ลดลงจากน้อยไปมาก ซึ่งชนิดขอบเขตเสียงระหว่างเสียงกึ่งสระกับเสียงเจี๊ยบ สามารถปรับปรุงผลการหาขอบเขตเสียงเมื่อเทียบกับแบบอ้างอิงได้น้อยที่สุด (มีความผิดพลาดลดลงน้อยที่สุด) และสามารถหาขอบเขตเสียงได้มากที่สุดชนิดขอบเขตเสียงระหว่างเสียงเจี๊ยบกับเสียงกัก ที่มีค่าความผิดพลาดลดลงมากที่สุด 82.92 และในรูปที่ 5.5 แสดงค่าร้อยละความถูกต้องการหาตำแหน่งขอบเขตเสียงด้วยวิธีกฎตัดสินใจในชุดข้อมูลที่ 3 เรียงลำดับชนิดขอบเขตเสียงตามความผิดพลาดที่ลดลงจากน้อยไปมาก เมื่อเปรียบเทียบผลการระบุตำแหน่งขอบเขตเสียงโดยใช้กฎตัดสินใจของตารางที่ 5.6 ซึ่งได้จากการทดสอบกับฐานข้อมูลที่ใช้พัฒนาระบบมีร้อยละของความผิดพลาดที่ลดลงเป็น 26.59 กับตารางที่ 5.7 ซึ่งได้จากการทดสอบกับฐานข้อมูลที่ใช้ทดสอบระบบ มีร้อยละของความผิดพลาดที่ลดลงเป็น 24.26 ซึ่งไม่แตกต่างกันนักเมื่อเทียบกับการทดลองที่ใช้ฐานข้อมูลชุดพัฒนาระบบ ดังนั้น การปรับตำแหน่งขอบเขตเสียงจากวิธีใช้กฎตัดสินใจเป็นวิธีที่วิธีการปรับหาตำแหน่งขอบเขตเสียงที่มีประสิทธิภาพ

และกฎที่ใช้ในการตัดสินใจมีความคงทน ไม่เฉพาะเจาะจงขึ้นกับข้อมูลที่พิจารณาและสร้างกฎเท่านั้น เนื่องจากเงื่อนไขของกฎที่สร้างขึ้นเกิดจากการศึกษาค่าการเปลี่ยนแปลงอะคูสติคและพิจารณาเลือกค่าลักษณะสำคัญทางเสียงสำหรับแต่ละชนิดขอบเขตเสียงจากชุดข้อมูลพัฒนาระบบ ซึ่งเป็นคนละชุดกับฐานข้อมูลทดสอบ ดังนั้นการวัดประสิทธิภาพของกฎตัดสินใจโดยทดสอบกฎกับข้อมูลที่ไม่เคยพบมาก่อนเช่นนี้และได้รับผลการระบุแม่นยำดี สะท้อนให้เห็นประสิทธิภาพของวิธีที่เสนอว่ากฎที่ได้ไม่ลำเอียง ไม่เฉพาะเจาะจงกับข้อมูล

### 5.1.1. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้กฎตัดสินใจ เมื่อขนาดกรอบพิจารณาต่างกัน

ตารางที่ 5.8 ผลการระบุตำแหน่งขอบเขตหน่วยเสียงโดยวิธีใช้กฎตัดสินใจ เมื่อใช้ขนาดกรอบพิจารณาต่างกัน

ขนาดกรอบพิจารณา		5 เฟรม	7 เฟรม	9 เฟรม
ร้อยละความถูกต้องตามระดับความคลาดเคลื่อนที่ยอมรับ	±0 เฟรม	36.58	36.98	36.30
	±1 เฟรม	73.43	73.39	72.05
	±2 เฟรม	86.93	86.50	85.73
	±3 เฟรม	93.08	92.82	92.01
ความคลาดเคลื่อนเทียบกับคำตอบจริง (เฟรม)	ค่าเฉลี่ย	1.21	1.22	1.26
	ส่วนเบี่ยงเบนมาตรฐาน	1.55	1.59	1.63

ตารางที่ 5.8 ซึ่งรายงานผลการระบุตำแหน่งขอบเขตหน่วยเสียงของแต่ละชนิดขอบเขตเสียงเมื่อใช้ขนาดกรอบหน้าต่างพิจารณาแตกต่างกัน พบว่าชนิดขอบเขตเสียงส่วนใหญ่เมื่อใช้ขนาดกรอบหน้าต่างพิจารณา 5 เฟรม ทำให้ผลการทดสอบมีผลการระบุขอบเขตเสียงถูกต้องมากกว่าการทดสอบที่ใช้ขนาดกรอบหน้าต่างพิจารณา 7 และ 9 เฟรม โดยวัดทุกระดับที่ยอมรับได้ของความคลาดเคลื่อนของคำตอบเทียบกับคำตอบจริง

สรุปได้ว่า จากสมมติฐานที่ต้องการพิสูจน์ขนาดกรอบพิจารณาว่ามีผลต่อความถูกต้องของการหาตำแหน่งขอบเขตเสียงที่แท้จริงนั้น พบว่า เป็นจริงตามสมมติฐาน เนื่องจากเมื่อใช้ขนาดกรอบที่มีขนาดกว้างมากเกินไปทำให้มีโอกาสค้นหาตำแหน่งที่ผิดพลาดได้ ซึ่งพิจารณาเลือกตำแหน่งขอบเขตเสียงอื่น มาเป็นตำแหน่งขอบเขตเสียงที่เหมาะสมทำให้ผลการหาตำแหน่งขอบเขตเสียงผิดไปจากความจริง ซึ่งจากผลการทดลองในวิทยานิพนธ์นี้รายงานขนาดกรอบพิจารณาที่มีค่า 5 เฟรมทำให้ความถูกต้องจากผลการระบุจากวิธีที่เสนอเทียบคำตอบจริงให้มากที่สุด



ตารางที่ 5.9 ผลการระบุตำแหน่งขอบเขตหน่วยเสียงวิธีที่ใช้กฎ เมื่อใช้ขนาดกรอบพิจารณาต่างกัน วัดความถูกต้องที่ระดับความคลาดเคลื่อนที่ยอมรับได้ 1 เฟรม

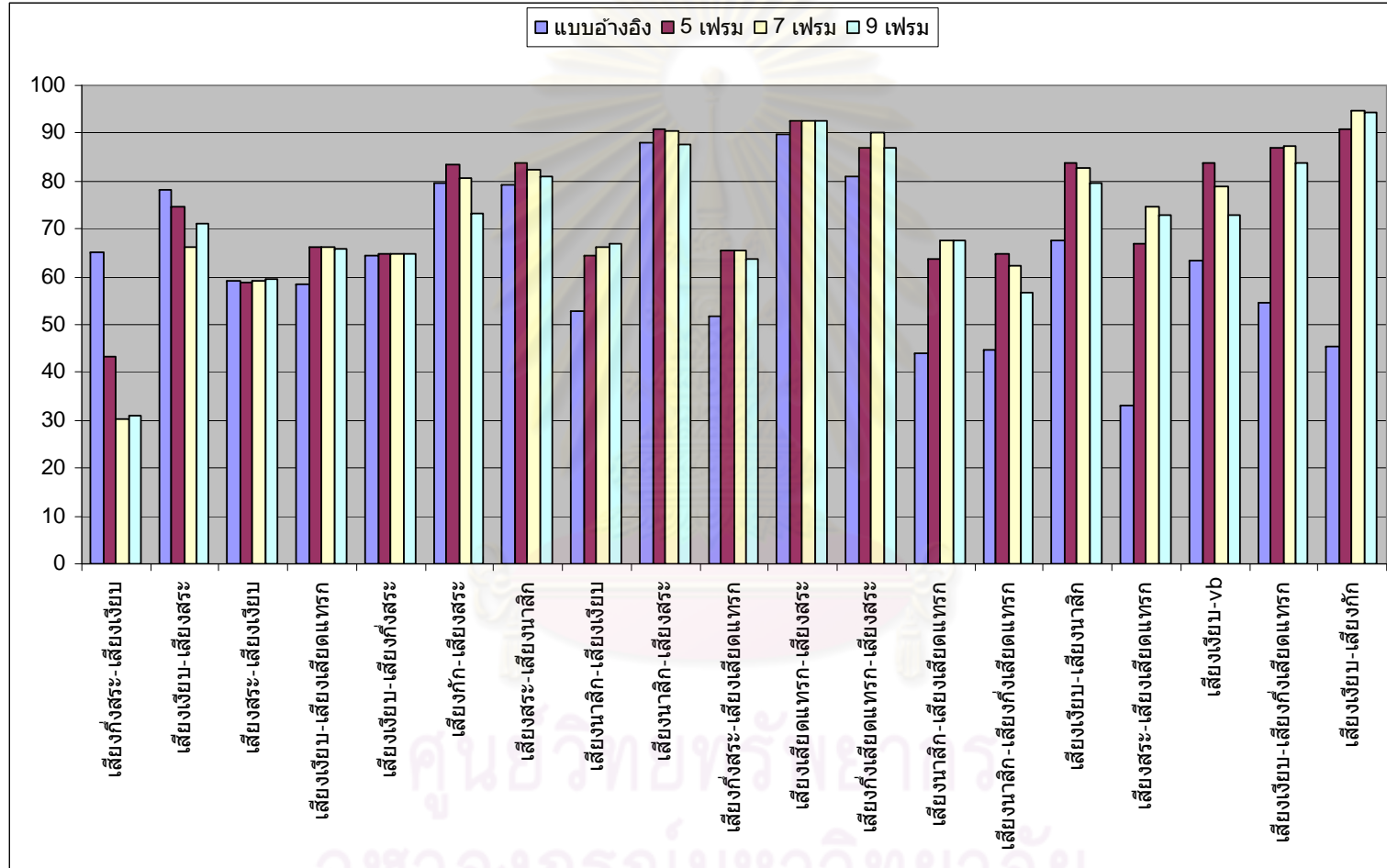
ชนิดขอบเขตหน่วยเสียง	ร้อยละความแม่นยำ				ความผิดพลาดลดลง		
	แบบ อ้างอิง	ขนาดกรอบพิจารณา			ขนาดกรอบพิจารณา		
		5 เฟรม	7 เฟรม	9 เฟรม	5 เฟรม	7 เฟรม	9 เฟรม
เสียงกึ่งสระ-เสียงเจียบ	65.23	43.18	30.23	30.91	-63.42	-100.66	-98.71
เสียงเจียบ-เสียงสระ	78.26	74.78	66.09	71.30	-16.01	-55.98	-32.01
เสียงสระ-เสียงเจียบ	59.12	58.68	59.28	59.34	-1.08	0.39	0.54
เสียงเจียบ-เสียงเสียดแทรก	58.45	66.36	66.36	65.91	2.02	19.04	17.95
เสียงเจียบ-เสียงกึ่งสระ	64.46	64.79	64.79	64.79	4.14	0.93	0.93
เสียงกัก-เสียงสระ	79.55	83.43	80.79	73.18	18.97	6.06	-31.15
เสียงสระ-เสียงนาสิก	79.09	83.78	82.42	81.12	22.43	15.93	9.71
เสียงนาสิก-เสียงเจียบ	52.79	64.34	66.16	66.77	24.47	28.32	29.61
เสียงนาสิก-เสียงสระ	87.87	90.97	90.32	87.74	25.56	20.20	-1.07
เสียงกึ่งสระ-เสียงเสียดแทรก	51.72	65.52	65.52	63.79	28.58	28.58	25.00
เสียงเสียดแทรก-เสียงสระ	89.62	92.78	92.63	92.78	30.44	29.00	30.44
เสียงกึ่งเสียดแทรก-เสียงสระ	80.98	87.07	90.00	87.07	32.02	47.42	32.02
เสียงนาสิก-เสียงเสียดแทรก	43.95	63.69	67.52	67.52	35.22	42.05	42.05
เสียงนาสิก-เสียงกึ่งเสียดแทรก	44.60	64.86	62.16	56.76	40.79	31.70	21.95
เสียงเจียบ-เสียงนาสิก	67.44	83.72	82.79	79.53	50.00	47.14	37.13
เสียงสระ-เสียงเสียดแทรก	33.20	66.80	74.70	72.73	50.30	62.13	59.18
เสียงเจียบ-vb	63.50	83.94	78.83	72.99	56.00	42.00	26.00
เสียงเจียบ-เสียงกึ่งเสียดแทรก	54.66	86.82	87.46	83.92	70.93	72.34	64.53
เสียงเจียบ-เสียงกัก	45.44	90.68	94.83	94.28	82.92	90.52	89.52
<b>รวมทั้งหมด</b>	<b>65.00</b>	<b>73.43</b>	<b>73.40</b>	<b>72.05</b>	<b>24.26</b>	<b>24.00</b>	<b>20.14</b>

จากตารางที่ 5.9 แสดงผลการระบุตำแหน่งขอบเขตเสียงโย้ใช้กฎตัดสินใจเมื่อใช้ขนาดกรอบพิจารณาแตกต่างกันในแต่ละชนิดขอบเขตเสียง วัดผลที่ระดับความคลาดเคลื่อนที่ 1 เฟรมเทียบกับคำตอบที่แท้จริง ซึ่งสามารถแสดงเป็นฮิสโตแกรมของร้อยละความถูกต้องของการหาตำแหน่งขอบเขตเสียงจากชนิดขอบเขตเสียงต่าง ๆ เมื่อใช้ขนาดกรอบพิจารณาแตกต่างกันโดยระดับความคลาดเคลื่อนที่ 1 เฟรมเทียบกับคำตอบที่แท้จริงดังรูปที่ 5.6 พบว่าทุกชนิดขอบเขตเสียงมีแผนภูมิแท่งของผลการระบุหาตำแหน่งขอบเขตเสียงถูกต้องเมื่อใช้ขนาดกรอบพิจารณามีค่า 5 เฟรม

สูงกว่าค่าแผนภูมิแท่ง ของผลการระบุหาตำแหน่งขอบเขตเสียงถูกต้องเมื่อใช้ขนาดรอบพิจารณา 7 และ 9 เฟรม ซึ่งผลการระบุหาตำแหน่งขอบเขตเสียงถูกต้องเมื่อใช้ขนาดรอบพิจารณา 9 เฟรมมีค่าต่ำที่สุดเนื่องจาก ในขั้นตอนการเลือกตำแหน่งเฟรมในรอบพิจารณา จากขนาดรอบที่กว้างจึงมีโอกาสที่จะเลือกตำแหน่งเฟรมนั้นอยู่ในตำแหน่งที่มีระยะห่างจากตำแหน่งเริ่มต้นที่ได้จากแบบจำลองฮิดเดินมากอฟหรือเลือกตำแหน่งเฟรมที่เป็นตำแหน่งขอบเขตเสียงอื่นที่ไม่ใช่ขอบเขตที่กำลังพิจารณา ส่งผลให้พิจารณาเลือกตำแหน่งเฟรมที่ไม่เหมาะสมเป็นคำตอบที่ผิดของวิธีที่เสนอ



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 5.6 ฮิสโตแกรมร้อยละความถูกต้องของการหาขอบเขตเสียงโดยใช้กฎตัดสิน เมื่อใช้ขนาดกรอบพิจารณาแตกต่างกัน

## 5.2. ผลจากการระบุตำแหน่งขอบเขตเสี่ยงจากวิธีใช้สถิติ

การออกแบบชุดการทดลองตามปัจจัยประสิทธิภาพของการจำแนก (ซึ่งอธิบายตามหัวข้อ 5.2.1.2. ขั้นตอนการวิเคราะห์พารามิเตอร์ ในบทที่ 4) สามารถเปรียบเทียบผลการระบุตำแหน่งขอบเขตเสี่ยงจากการทดสอบที่ใช้ค่าพารามิเตอร์ที่แตกต่างกันและเพื่อหาระบบการหาตำแหน่งขอบเขตเสี่ยงที่เหมาะสมกับคำตอบที่แท้จริง โดยในขั้นตอนการทดลองนี้ ได้ใช้ปัจจัยที่มีผลต่อประสิทธิภาพของการหาขอบเขตจากผลการจำแนก ทั้งหมด 8 การทดลอง ซึ่งแบ่งตาม ลักษณะต้นแบบการจำแนก ลักษณะของข้อมูลที่ใช้ฝึกฝน มิติของเวกเตอร์ลักษณะเฉพาะ จำนวนของเวกเตอร์ลักษณะเฉพาะ เป็นต้น ซึ่งผลจากการทดลองแสดงรายงานตามตารางที่ 5.10

ตารางที่ 5.10 ผลการทดสอบประสิทธิภาพของการระบุตำแหน่งขอบเขตหน่วยเสี่ยงโดยวิธีทางสถิติ

ชุดการทดลอง		1	2	3	4	5	6	7	8
ร้อยละความถูกต้องตามระดับความคลาดเคลื่อนที่ยอมรับ	±0 เฟรม	30.13	27.81	40.89	40.02	40.65	40.11	44.93	45.88
	±1 เฟรม	67.98	68.47	78.76	80.02	78.58	79.84	79.64	80.22
	±2 เฟรม	82.56	83.31	89.12	89.80	89.15	89.77	89.17	89.52
	±3 เฟรม	91.12	91.34	93.86	94.11	93.86	94.14	93.73	93.95
ความคลาดเคลื่อนเทียบกับคำตอบจริง (เฟรม)	ค่าเฉลี่ย	1.41	1.42	1.08	1.07	1.09	1.07	1.03	1.00
	ส่วนเบี่ยงเบนมาตรฐาน	1.70	1.70	1.59	1.56	1.59	1.56	1.60	1.58
เวลาที่ใช้จำแนก	7 ชม.	38 นาที	25 นาที	28 นาที	29 นาที	23 นาที	23 นาที	9 นาที	9 นาที
	44 วินาที	48 วินาที	17 วินาที	11 วินาที	34 วินาที	18 วินาที	19 วินาที	54 วินาที	

จากผลการวิเคราะห์ตามหัวข้อ“การทดสอบประสิทธิภาพของการจำแนก” สามารถอธิบายตามผลการวัดประสิทธิภาพได้ดังนี้

### 5.2.1. ผลวิเคราะห์ร้อยละความถูกต้องของผลที่ได้จากวิธีที่เสนอเทียบกับคำตอบที่แท้จริง

จากขั้นตอนการวิเคราะห์พารามิเตอร์ เพื่อวัดประสิทธิภาพของการจำแนกกลุ่มตามปัจจัยการสร้างต้นแบบการจำแนก ได้รายงานผลการวิเคราะห์ของแต่ละปัจจัย ดังนี้

#### A. ผลวิเคราะห์จากลักษณะต้นแบบการจำแนก

จากตารางที่ 5.10 พบว่า เมื่อเปรียบเทียบร้อยละความถูกต้องของการระบุตำแหน่งขอบเขตหน่วยเสียงทุกระดับความคลาดเคลื่อนที่ยอมรับได้ ของชุดการทดลองที่ 1 และ 2 ซึ่งเป็นผลจากการระบุตำแหน่งขอบเขตเสียงจากต้นแบบที่ไม่ระบุตามชนิดขอบเขตหน่วยเสียง เทียบกับ ชุดการทดลองที่ 3 4 5 6 7 และ 8 ซึ่งใช้ชนิดตัวจำแนกแบบระบุตามชนิดขอบเขตเสียง พบว่า ชุดการทดลองที่ 1 และ 2 จากชนิดตัวจำแนกแบบไม่ระบุชนิดขอบเขต (ใช้เพียง 1 ตัวจำแนก) มีผลร้อยละความถูกต้องต่ำกว่าชุดการทดลองที่ 3 4 5 6 7 และ 8 ซึ่งใช้ตัวจำแนกแบบระบุชนิดขอบเขต (แยกตามชนิดขอบเขตเสียงเป็น 21 ตัวจำแนก) เนื่องจากว่าการพิจารณาการเปลี่ยนแปลงของลักษณะสำคัญที่เกิดขึ้นในแต่ละชนิดขอบเขตเสียง พบว่าการเปลี่ยนแปลงแตกต่างกัน และเมื่อพิจารณาค่าการเปลี่ยนแปลงที่เป็นตัวแทนกลุ่มข้อมูล พบว่า การใช้ตัวจำแนกกลุ่มแบบตัวจำแนกเดียว มีข้อมูลที่ใช้เป็นตัวแทนกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง จากแต่ละชนิดขอบเขตเสียง (บางชนิดขอบเขตเสียงมีความการเปลี่ยนแปลงเด่นชัด ในทางตรงกันข้าม บางชนิดขอบเขตเสียงมีค่าการเปลี่ยนแปลงค่าลักษณะสำคัญเล็กน้อย) ทำให้ตัวแทนจากแต่ละชนิดขอบเขตเสียงแตกต่างกัน เป็นผลให้ค่ามีความหลากหลาย กระจายตัว ส่งผลให้ไม่สามารถชี้ความแตกต่าง ระหว่างกลุ่มขอบเขตเสียงและกลุ่มไม่ใช่ขอบเขตเสียงได้ ในทางกลับกัน ผลที่ได้จากการใช้ตัวจำแนกจากตัวจำแนกย่อย 21 แบบ (ซึ่งในชนิดขอบเขตเสียงมีข้อมูลที่ใช้เป็นตัวแทนกลุ่มไม่แตกต่างกันมากนักภายในกลุ่มเดียวกัน และบอกความแตกต่างระหว่างกลุ่มที่ต่างกันได้ดี)

สรุปได้ว่า การระบุตำแหน่งขอบเขตเสียงจากต้นแบบการจำแนกที่ระบุตามชนิดขอบเขตเสียง ดีกว่าการระบุจากต้นแบบการจำแนกที่ไม่ระบุตามชนิดขอบเขตเสียง

#### B. ผลวิเคราะห์จากลักษณะของข้อมูลที่ใช้ฝึกฝน



จากการพิจารณาผลวิเคราะห์ของร้อยละความถูกต้องการระบุหาตำแหน่งขอบเขตเสียงจากตัวจำแนกแบบระบุชนิดขอบเขตเสียงที่ใช้การเลือกค่าลักษณะเฉพาะตามชนิดของขอบเขตเสียง เปรียบเทียบกับตัวจำแนกแบบระบุแยกชนิดขอบเขตเสียงที่ไม่เลือกค่าลักษณะสำคัญทางเสียง(ใช้ทุกค่าลักษณะสำคัญทางเสียง) โดยเปรียบเทียบผลการระบุหาตำแหน่งขอบเขตเสียงเมื่อวัดผลตามระดับความคลาดเคลื่อนที่ยอมรับได้พบว่าที่ระดับความคลาดเคลื่อน 1 และ 2 เฟรม ชุดการทดลองที่ 3 เทียบกับชุดการทดลองที่ 5 และเปรียบเทียบชุดการทดลองที่ 4 กับชุดการทดลองที่ 6 ตามลำดับ พบว่า ต้นแบบจากการจำแนกซึ่งใช้ทุกค่าลักษณะสำคัญทางเสียงให้ค่าความถูกต้องการระบุตำแหน่งขอบเขตไม่แตกต่างจากการทดลองที่เลือกใช้ค่าลักษณะทางเสียง (ซึ่งเมื่อพิจารณาผลการระบุ พบว่าแตกต่างแบบไม่มีนัยสำคัญ)ทำให้ยังไม่สามารถสรุปเพื่อหาค่าพารามิเตอร์ที่เหมาะสมการสร้างต้นแบบการจำแนกได้ จึงเปรียบเทียบผลการระบุตำแหน่งขอบเขตเสียงจากต้นแบบการจำแนกเมื่อใช้การเลือกค่าลักษณะสำคัญตามชนิดขอบเขตเสียงทุกการทดลอง โดยรวมการทดลองที่ 7 และ 8 ประกอบ พบว่า ร้อยละความถูกต้องการผลการระบุตำแหน่งขอบเขตเสียง ของการทดลองที่ 7 และ 8 มีค่าสูงกว่าการทดลองที่ 3 และ 4

นอกจากนี้ยังมีข้อสนับสนุนจากผลการวิเคราะห์ในตารางที่ 5.5 ซึ่งสรุปลักษณะสำคัญสำหรับแต่ละชนิดขอบเขตเสียง ซึ่งแต่ละชนิดขอบเขตเสียงแตกต่างกัน และภาคผนวก ง ซึ่งแสดงผลการพิจารณาแผนภาพกล่อง และค่าความแปรปรวนของขั้นตอนการเลือกค่าลักษณะสำคัญ สรุปได้ว่า การเลือกค่าลักษณะสำคัญที่เหมาะสมสำหรับแต่ละชนิดขอบเขตเสียงแตกต่างกัน

สรุปได้ว่า การระบุหาตำแหน่งขอบเขตเสียงจากต้นแบบการจำแนกที่ระบุตามชนิดขอบเขตเสียงที่ใช้ข้อมูลจำแนกจากค่าลักษณะสำคัญแตกต่างกัน สำหรับแต่ละชนิดขอบเขตเสียง ดีกว่าการระบุจากต้นแบบการจำแนกที่ใช้ข้อมูลจำแนกจากทุกชนิดค่าลักษณะสำคัญ

### C. ผลวิเคราะห์จากมิติของเวกเตอร์ลักษณะเฉพาะ

จากสมมติฐาน การสร้างต้นแบบการจำแนกกลุ่มที่มีมิติของเวกเตอร์แตกต่างกันตัวแทนข้อมูลจำแนก จากค่าการเปลี่ยนแปลงของลักษณะสำคัญทางเสียง เป็นปัจจัยการสร้างต้นแบบที่แตกต่างกันนั้น มีผลต่อประสิทธิภาพการจำแนก ซึ่งจากการวิเคราะห์ผลการระบุตำแหน่งขอบเขตเสียง ดังนั้นจึงวัดผลการ

ระบุตำแหน่งขอบเขตเสี่ยงจากการชดเชยการทดลองที่ 7 เปรียบเทียบกับชุดการทดลองที่ 5 และชุดการทดลองที่ 8 เปรียบเทียบกับชุดการทดลองที่ 6 พบว่า เวกเตอร์ลักษณะเฉพาะที่มีมิติเวกเตอร์ของข้อมูลที่แบบ 1 มิติ มีค่าความแม่นยำมากกว่าตัวจำแนกที่ใช้มิติเวกเตอร์ของข้อมูล 2 มิติ จากผลการทดลอง วิเคราะห์ได้ว่า การใช้มิติเวกเตอร์ของข้อมูล 2 มิติ ซึ่งคำนวณค่าการเปลี่ยนแปลงที่ตำแหน่งพิจารณากับตำแหน่งก่อนหน้า และค่าการเปลี่ยนแปลงที่ตำแหน่งพิจารณากับตำแหน่งถัดไป พบว่าค่าการเปลี่ยนแปลงที่ตำแหน่งพิจารณากับตำแหน่งถัดไปค่านี้เป็นค่าที่เพิ่มเข้ามา ทำให้แตกต่างจากข้อมูลที่มีมิติเวกเตอร์แบบ 1 มิติ ซึ่งค่านี้ทำให้ผลการทดลองมีค่าความแม่นยำของลดลง เนื่องจากการคำนวณค่าการเปลี่ยนแปลง ที่ได้จากตำแหน่งพิจารณากับตำแหน่งถัดไป ไม่ใช่บริเวณที่เกิดการเปลี่ยนแปลงของอะคูสติค ทำให้ค่าที่คำนวณดังกล่าวเกิดเป็นข้อมูลที่ทำให้มีความแปรปรวนของข้อมูลมาก ตัวอย่างของข้อมูลแต่ละกลุ่มกระจาย และส่งผลให้ข้อมูลที่จำแนกที่มีมิติมากสามารถบอกความแตกต่างหรือแยกกลุ่มยาก ทำให้เกิดความผิดพลาดของการจำแนกต่อไป

สรุปได้ว่า เวกเตอร์ลักษณะเฉพาะของตัวแทนข้อมูล จากค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงแบบ 1 มิติ ได้จากการคำนวณค่าการเปลี่ยนแปลงที่ตำแหน่งพิจารณากับตำแหน่งก่อนหน้า เป็นข้อมูลเพียงพอเพื่อแสดงลักษณะการเปลี่ยนแปลงค่าอะคูสติคที่เกิดขึ้นที่ตำแหน่งขอบเขตเสี่ยงได้

#### D. ผลวิเคราะห์จากจำนวนของเวกเตอร์ลักษณะเฉพาะ

จากสมมติฐาน จำนวนของตัวอย่างข้อมูลฝึกฝนที่เป็นตัวแทนของชุดข้อมูลบวกและชุดข้อมูลลบมีผลต่อประสิทธิภาพของการจำแนกกลุ่ม จึงพิจารณาผลการระบุตำแหน่งขอบเขตเสี่ยงจากตัวจำแนกที่ใช้จำนวนของตัวอย่างข้อมูลฝึกฝนที่มีอัตราส่วนของจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบแบบ 1:2 เทียบกับ ตัวจำแนกที่ใช้จำนวนของการฝึกฝนที่มากที่มีอัตราส่วนของจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบแบบ 3:4 พบว่า ชุดการทดลองที่ 7 (ใช้จำนวนตัวอย่างน้อยกว่า) มีร้อยละความถูกต้องน้อยกว่าชุดการทดลองที่ 8 เมื่อวัดระดับความคลาดเคลื่อนที่ยอมรับได้มากกว่า 1 เฟรม

สรุปได้ว่า ต้นแบบที่มีข้อมูลตัวอย่างชุดฝึกฝนที่ประกอบด้วยข้อมูลจำนวนมากขึ้น ทำให้ประสิทธิภาพเพิ่มขึ้น (จากการใช้ตัวอย่างข้อมูลฝึกฝนที่มีอัตราส่วนของจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบแบบ 3:4 ทำให้ผลการ

จำแนกกลุ่มดีกว่าใช้จำนวนตัวอย่างข้อมูลที่น้อยจากตัวอย่างข้อมูลที่มีอัตราส่วนของจำนวนชุดข้อมูลบวกต่อจำนวนชุดข้อมูลลบแบบ 1:2) เนื่องจากการจำแนกกลุ่มด้วยวิเคราะห์ดิสคริมิแนนต์เชิงเส้นมี ข้อมูลตัวอย่างฝึกฝนที่มากขึ้นเพื่อใช้เรียนรู้ในการสร้างฟังก์ชันการจำแนกกลุ่ม ซึ่งจำนวนของตัวอย่างข้อมูลฝึกฝนที่มากจะทำให้การประมาณเงื่อนไขการจำแนกกลุ่มของข้อมูลสองกลุ่มที่แตกต่างกัน ได้ชัดเจนขึ้น โดยพิจารณาจากการกระจายตัวของข้อมูลของแต่ละกลุ่มและพิจารณาความแตกต่างระหว่างข้อมูลสองกลุ่มได้ และอีกข้อสนับสนุนคือ ข้อมูลที่เป็นตัวอย่างกลุ่มตำแหน่งขอบเขตเสี่ยงจาก 3 ตำแหน่งเฟรมนั้นครอบคลุมตัวอย่างที่เกิดการเปลี่ยนแปลงค่าอะคูสติคที่แท้จริง ขณะที่การใช้ข้อมูลตัวอย่างกลุ่มตำแหน่งขอบเขตเสี่ยงจาก 1 ตำแหน่งเฟรม อาจจะไม่ใช้ตำแหน่งที่เกิดการเปลี่ยนแปลงของอะคูสติค มีความคลาดเคลื่อนของข้อมูล ทำให้ตัวอย่างของข้อมูลกลุ่มตำแหน่งขอบเขตเสี่ยงไม่มีข้อมูลที่เป็นตำแหน่งที่เกิดการเปลี่ยนแปลงค่าอะคูสติค

#### สรุปผลการวิเคราะห์จากร้อยละความถูกต้องการระบุตำแหน่งขอบเขตเสี่ยง

ผลที่ได้จากการจำแนกที่ใช้ข้อมูลฝึกฝนที่มีมิติเวกเตอร์แตกต่างกัน ทำให้สามารถตรวจหาตำแหน่งขอบเขตเสี่ยงได้แตกต่างกัน สรุปได้ว่าการใช้ค่าการเปลี่ยนแปลงลักษณะสำคัญเป็นตัวแทนข้อมูลแบบ 1 ค่า สามารถแสดงการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงได้เพียงพอ โดยการสร้างตัวอย่างของข้อมูลที่มีอัตราส่วนของข้อมูลชุดบวก (กลุ่มตำแหน่งขอบเขตเสี่ยง) ต่อชุดข้อมูลลบ (กลุ่มที่ไม่ใช่ขอบเขตเสี่ยง) แบบ 3:4 ทำให้ได้การจำแนกเพื่อปรับหาตำแหน่งขอบเขตเสี่ยงมีร้อยละความถูกต้องดีกว่าแบบ 1:2

#### 5.2.2. ผลวิเคราะห์ความคลาดเคลื่อนของผลที่ได้จากวิธีที่เสนอเทียบกับคำตอบที่แท้จริง

- 1) พิจารณาค่าความคลาดเคลื่อนของตำแหน่งเฟรมที่เหมาะสมที่ได้จากผลลัพธ์การหาตำแหน่งจากวิธีที่เสนอเทียบกับตำแหน่งขอบเขตเสี่ยงที่แท้จริง พบว่า ผลลัพธ์ที่ได้จากการใช้การจำแนกแบบไม่ระบุชนิดขอบเขต 1 ตัวจำแนก(ในการทดลองที่ 1 และ 2) เพื่อหาขอบเขตเสี่ยงทุกชนิดขอบเขตเสี่ยงพบว่าค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนสูงมาก (มีค่าไม่แตกต่างจากชุดอ้างอิง) นั่นคือ ไม่สามารถปรับปรุงผลจากการระบุตำแหน่งขอบเขตเริ่มต้นจากแบบจำลองฮิดเด้นมาคอปได้เลย ในทางตรงข้าม เมื่อพิจารณาการทดลอง 3 4 5 6 7 และ 8 ที่สร้างจากตัวจำแนกที่พิจารณาตามชนิดขอบเขตเสี่ยง พบว่า มีค่าเฉลี่ยและส่วนเบี่ยงเบน

มาตรฐานของความคลาดเคลื่อนของขอบเขตเสียงเทียบกับคำตอบลดลงอย่างมาก ซึ่งเป็นผลการทดลองซึ่งสนับสนุนแนวคิดของการสร้างต้นแบบระบบหาขอบเขตเสียงที่พิจารณาตามชนิดขอบเขตเสียง เนื่องจากว่าการปรับหาดำแหน่งขอบเขตเสียงพิจารณาดำแหน่งเฟรมในกรอบที่มีขนาดกรอบที่ไม่กว้างมากนัก จึงไม่สามารถค้นหาตำแหน่งขอบเขตเสียงที่ผิดพลาดที่อยู่ห่างไปจากตำแหน่งของเขตเสียงเริ่มต้นได้ ดังนั้น ผลการระบุตำแหน่งขอบเขตเสียงที่ได้จากการทดลองต่างๆ จึงสามารถปรับปรุงผลที่ปรับหาดำแหน่งขอบเขตเสียงที่อยู่ในขนาดกรอบพิจารณาเท่านั้น จากผลการทดลองจากการใช้ตัวจำแนกตัวเดียว (การทดลองที่ 1 และ 2) ไม่สามารถหาดำแหน่งขอบเขตเสียงที่แท้จริงได้มากนัก เนื่องจากข้อมูลที่ใช้ฝึกฝนนั้น มีความแปรปรวน การกระจายตัวของข้อมูลของกลุ่มตำแหน่งที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงมีการกระจายมาก ข้อมูลซ้อนทับกัน ไม่สามารถแบ่งแยกกลุ่มได้จึงทำให้ผลการจำแนกไม่ดี

- 2) การทดลองที่ 8 ซึ่งเป็นชุดที่มีค่าร้อยละความถูกต้องการระบุตำแหน่งขอบเขตเสียงดีที่สุด มีค่าเฉลี่ยความคลาดเคลื่อน 1 เฟรมซึ่งมีขนาด 10 มิลลิวินาที หมายถึง วิธีที่เสนอนี้สามารถหาดำแหน่งขอบเขตเสียงได้ความคลาดเคลื่อนกับตำแหน่งขอบเขตเสียงที่แท้จริงมีโอกาสคลาดเคลื่อนเฉลี่ย 10 มิลลิวินาทีเท่านั้น ซึ่งสามารถลดลงจากชุดทดลองอ้างอิงจากแบบจำลองฮิดเด็นมาคอฟ สรุปว่า ระบบสามารถไปค้นหาตำแหน่งขอบเขตเสียงที่แท้จริงมาพิจารณาเป็นตำแหน่งขอบเขตเสียงที่เหมาะสมได้ เนื่องจาก การปรับหาดำแหน่งขอบเขตเสียงจากต้นแบบที่ใช้ข้อมูลการฝึกฝนขึ้นตามแต่ละชนิดขอบเขตเสียง ดังนั้นค่าการเปลี่ยนแปลงของลักษณะสำคัญภายในชนิดขอบเขตเสียงจึงไม่แตกต่างกันมากนัก การกระจายตัวของข้อมูลแต่ละกลุ่มจึงไม่มาก อีกทั้งข้อมูลที่ใช้ฝึกฝนมาก ทำให้ตัวจำแนกกลุ่มสามารถสร้างการประมาณแบ่งแยกความแตกต่างระหว่างสองกลุ่มได้ชัดเจน ส่งผลให้การจำแนกกลุ่มเพื่อหาดำแหน่งขอบเขตเสียงดีไปด้วย

### 5.2.3. ผลวิเคราะห์เวลาที่ใช้ในการจำแนก

- 1) จากผลการวิเคราะห์เวลานั้นพบว่า จำนวนตัวอย่างข้อมูลจำแนกมีผลต่อเวลาที่ใช้ในการประมวลผลจำแนก จากข้อสนับสนุนต่อไปนี้
  - a. การทดลองที่ 7 และ 8 ซึ่งใช้ข้อมูล 1 ค่าเป็นตัวแทนค่าการเปลี่ยนแปลงค่าลักษณะสำคัญตำแหน่งที่พิจารณาเทียบกับตำแหน่งก่อนหน้าจากแต่ละค่าลักษณะทางเสียงเพื่อสร้างเวกเตอร์ลักษณะเฉพาะ ใช้เวลาน้อยกว่าการใช้

ตัวแทน 2 ค่า จากการคำนวณค่าเปลี่ยนแปลงค่าลักษณะสำคัญตำแหน่งที่พิจารณาเทียบกับตำแหน่งก่อนหน้า และตำแหน่งที่พิจารณาเทียบกับตำแหน่งถัดไป จากการทดลองที่ 5 และ 6

- b. การทดลองที่ 5 และ 6 ซึ่งเลือกใช้ค่าลักษณะสำคัญทางเสียงบางชนิดค่าลักษณะสำคัญ เปรียบเทียบกับการทดลองที่ 3 และ 4 ซึ่งใช้ค่าลักษณะสำคัญทางเสียงทุกชนิดค่าลักษณะสำคัญ ทำให้มีจำนวนข้อมูลที่จำแนกมาก เป็นผลให้เวลาในการจำแนกน้อยกว่า
  - c. นอกจากนี้ จากการพิจารณาเวลาที่ใช้จำแนกจากต้นแบบการจำแนกตัวเดียวเป็นตัวแทนกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียงของทุกชนิดขอบเขตเสียง (การทดลองที่ 1 และ 2) ซึ่งมีข้อมูลที่ใช้เป็นตัวแทนมีขนาดใหญ่มาก เมื่อนำไปจำแนกจึงต้องใช้เวลาในการฝึกฝนนานและจำแนกนาน เมื่อเปรียบเทียบกับการใช้ตัวจำแนกตามชนิดขอบเขตเสียงซึ่งได้จากแต่ละตัวจำแนกย่อยตามชนิดขอบเขตเสียงในการทดลองที่ 3 4 5 6 7 และ 8 ซึ่งมีตัวอย่างข้อมูลฝึกฝนจากแต่ละชนิดขอบเขตเสียง (ซึ่งใช้เวลาในการจำแนกน้อยมากเมื่อเทียบกับการจำแนก 1 ตัวจำแนก)
- 2) จำนวนตัวอย่างข้อมูลฝึกฝน ไม่มีผลกระทบต่อเวลาที่ใช้จำแนก เช่น เปรียบเทียบเวลาที่ใช้จำแนกเมื่อใช้จำนวนข้อมูลที่เป็นอย่างกลุ่มบวกและกลุ่มลบในการฝึกฝนแบบ 1:2 เปรียบเทียบกับตัวอย่างกลุ่มบวกและกลุ่มลบในการฝึกฝนแบบ 3:4 ดัง การทดลองที่ 1 เปรียบเทียบกับการทดลองที่ 2 การทดลองที่ 3 เทียบกับการทดลองที่ 4 การทดลองที่ 5 เปรียบเทียบกับ 6 และการทดลองที่ 7 เปรียบเทียบกับ 8 ผลการพิจารณาเวลารายงานว่า เวลาที่ใช้ในการจำแนกไม่แตกต่างกันมากนัก ซึ่งสรุปได้ว่า จำนวนตัวอย่างของข้อมูลฝึกฝนไม่มีผลกระทบต่อเวลามากนัก เมื่อเทียบกับข้อมูลที่ใช้จำแนก
  - 3) ในการเลือกต้นแบบการจำแนกเพื่อสร้างระบบการระบุหาตำแหน่งขอบเขตเสียง จากการพิจารณาเวลาที่ใช้ในการจำแนกเป็นอีกปัจจัยในการสร้างต้นแบบ เมื่อเปรียบเทียบการใช้เวลาการจำแนกกลุ่ม จากการทดลองนี้พบว่า จากทดลองที่ 7 และ 8 ใช้เวลา 9 นาที 19 วินาที และ 9 นาที 54 วินาที ตามลำดับซึ่งใช้เวลาในการจำแนกกลุ่มน้อยที่สุดเมื่อเทียบกับชุดการทดลองต่างๆ ซึ่งเวลาที่ใช้ในการจำแนกของการทดลองที่ 7 ไม่แตกต่างจากการทดลองที่ 8 มากนัก



จากผลการวิเคราะห์ร้อยละความถูกต้องของผลที่ได้จากวิธีที่เสนอเทียบกับคำตอบที่แท้จริงซึ่ง ออกแบบตามชุดการทดลองต่างๆ พบว่าชุดการทดลองที่ 8 มีประสิทธิภาพในการปรับตำแหน่ง ที่เหมาะสมได้ดีที่สุด และพิจารณาเวลาที่ใช้ในการจำแนกประกอบพบว่า การทดลองที่ 7 ใช้เวลาใน การจำแนกน้อยที่สุดแต่ไม่แตกต่างกับการทดลองที่ 8 มากนัก ดังนั้น วิทยานิพนธ์นี้จึงเสนอ การสร้างต้นแบบการระบุตำแหน่งขอบเขตเสียงโยใช้ค่าพารามิเตอร์ตามการทดลองที่ 8 เพื่อปรับหา ตำแหน่งขอบเขตเสียงที่เหมาะสม

#### 5.2.4. ผลจากการระบุตำแหน่งขอบเขตเสียงจากวิธีใช้สถิติเมื่อขนาดกรอบพิจารณา ต่างกัน

การปรับหาขอบเขตเสียงให้ถูกต้องใกล้เคียงกับคำตอบจริงคือจุดมุ่งหมายของ วิทยานิพนธ์นี้ ดังนั้นตำแหน่งของเฟรมที่พิจารณาเลือกตำแหน่งที่เหมาะสมนี้ จึงขึ้นอยู่กับขนาด กรอบการปรับ ซึ่งใช้กำหนดโอกาสการเลือกของตำแหน่งเฟรมต่างๆ ให้ได้รับการเลือกเป็น ตำแหน่งที่เหมาะสมกับขอบเขตที่แท้จริง ซึ่งได้เปรียบเทียบผลการระบุขอบเขตเสียงที่ใช้ขนาด กรอบพิจารณาที่แตกต่างกัน คือ ใช้ขนาดกรอบ 5 เฟรม 7 เฟรมและ 9 เฟรมตามลำดับ ดังแสดงไป ในตารางที่ 5.8 ตามวิธีใช้กฎ และทดสอบผลการทดสอบการระบุหาตำแหน่งขอบเขตเสียงโดยใช้ สถิติ ดังตารางที่ 5.11

#### ตารางที่ 5.11 ผลการระบุตำแหน่งขอบเขตหน่วยเสียงโดยวิธีทางสถิติ เมื่อใช้ขนาดกรอบพิจารณา ต่างกัน

ขนาดกรอบพิจารณา		5 เฟรม	7 เฟรม	9 เฟรม
ร้อยละความถูกต้อง ตามระดับ ความคลาดเคลื่อนที่ ยอมรับ	±0 เฟรม	45.88	22.61	8.02
	±1 เฟรม	80.23	75.74	35.37
	±2 เฟรม	89.52	89.83	81.63
	±3 เฟรม	93.96	94.44	92.82
ความคลาดเคลื่อน เทียบกับคำตอบจริง (เฟรม)	ค่าเฉลี่ย	1.00	1.27	1.93
	ส่วนเบี่ยงเบน มาตรฐาน	1.58	1.46	1.40

จากตารางที่ 5.11 รายงานผลความถูกต้องของการระบุเมื่อใช้ขนาดกรอบพิจารณา 5 เฟรมมีผลร้อยละความถูกต้องมากที่สุด เมื่อเปรียบเทียบกับ การทดลองเมื่อใช้ขนาดกรอบพิจารณา 7 เฟรมและ 9 เฟรม โดยวัดรายงานผลตามระดับความคลาดเคลื่อนที่ยอมรับแตกต่างกันทุกระดับ และเมื่อพิจารณาผลเปรียบค่าเฉลี่ยความคลาดเคลื่อนเทียบกับคำตอบจริงพบว่า การระบุเมื่อใช้ขนาด

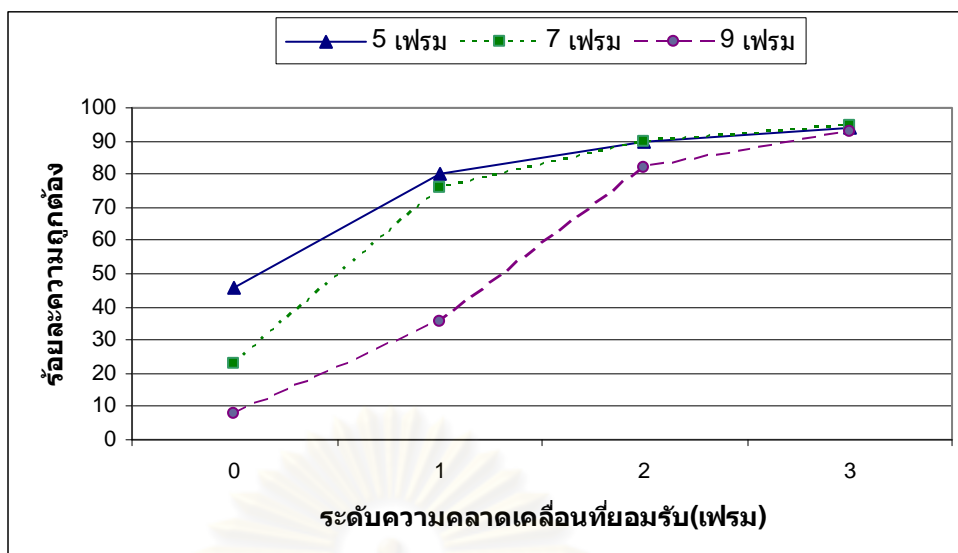
กรอบพิจารณา 5 เฟรมให้ผลค่าเฉลี่ยความคลาดเคลื่อน 1 เฟรม (เมื่อเฟรมมีขนาด 10 มิลลิวินาที) เท่านั้นซึ่งเป็นค่าที่ต่ำที่สุดเมื่อเปรียบเทียบกับขนาดกรอบ 7 และ 9 เฟรม ดังนั้นขนาดกรอบพิจารณา 5 เฟรมเหมาะสมในการค้นหาตำแหน่งขอบเขตเสียง ตามระดับความคลาดเคลื่อนที่ยอมรับได้ 1 เฟรม ซึ่งเหมาะสมกับวิธีการวัดที่ใช้ในวิทยานิพนธ์นี้

### สรุปผลวิเคราะห์การทดลองขนาดกรอบพิจารณาที่ต่างกัน

จากขนาดกรอบพิจารณาที่มีผลต่อความถูกต้องของการหาตำแหน่งขอบเขตเสียงที่แท้จริง เนื่องจากขนาดกรอบที่เหมาะสมจะมีโอกาสค้นหาตำแหน่งที่เหมาะสมภายในกรอบพิจารณา ถึงแม้ขนาดกรอบที่กว้างจะสามารถมีโอกาสเลือกพิจารณาตำแหน่งได้มากขึ้นแต่ในทางกลับกัน มีข้อเสียคือ ทำให้เกิดผลการพิจารณาตำแหน่งขอบเขตเสียงผิดพลาดไปจากความจริง ดังเช่น ผลจากการระบุตำแหน่งขอบเขตเสียงเมื่อใช้ขนาดกรอบพิจารณา 7 เฟรม และ 9 เฟรม

จากผลการทดลองจากตารางที่ 5.11 ซึ่งมีความสัมพันธ์ทิศทางเดียวกันกับตารางที่ 5.8 จึงสามารถสรุปผลการวิเคราะห์ว่าขนาดกรอบการพิจารณา 5 เฟรมเหมาะสมในการสร้างต้นแบบของกรอบพิจารณาหาตำแหน่งขอบเขตหน่วยเสียงที่เหมาะสม เนื่องจาก เมื่อใช้ขนาดกรอบพิจารณา 5 เฟรมซึ่งหมายถึงขนาดกรอบการพิจารณาที่เหมาะสมในการปรับหาตำแหน่งขอบเขตเสียงที่แท้จริงใช้จำนวนเฟรมที่ต้องการขยายไป 2 ตำแหน่งเฟรมซ้ายและขวาเมื่อเทียบกับตำแหน่งขอบเขตเสียงอ้างอิงเป็นขนาดกรอบที่ครอบคลุมความผิดพลาด ที่เกิดจากการวิเคราะห์ค่าเฉลี่ยคลาดเคลื่อนที่ของตำแหน่งขอบเขตเสียงที่ได้จากการใช้แบบจำลองฮิดเด้นมาคอฟเทียบกับตำแหน่งขอบเขตเสียงที่แท้จริงซึ่งมีค่า 1.42 เฟรม

นอกจากนี้ จากการวิเคราะห์ผลความถูกต้องของการระบุตำแหน่งขอบเขตเสียงจากการใช้ขนาดกรอบหน้าต่างการปรับหาตำแหน่งที่แตกต่างกัน ทำให้อนุมานสมมติฐาน การเสนอขนาดกรอบพิจารณาที่มีความกว้างมากขึ้น โดยขยายกรอบหน้าต่างพิจารณายืดหยุ่นซึ่งขยายจำนวนเฟรมจากตำแหน่งขอบเขตเสียงเริ่มต้นของขอบเขตเสียงที่พิจารณาก่อนหน้าถึงตำแหน่งขอบเขตเสียงเริ่มต้นของขอบเขตเสียงที่พิจารณาขอบเขตเสียงถัดไปเป็นขนาดกรอบพิจารณา ไม่สามารถเพิ่มผลความถูกต้องของการระบุตำแหน่งขอบเขตเสียง เนื่องจาก เมื่อใช้ขนาดกรอบที่กว้างกว่าขนาดเหมาะสมที่เสนอในวิทยานิพนธ์นี้ ทำให้การตรวจหาตำแหน่งขอบเขตเสียงผิดพลาดมากยิ่งขึ้น ตำแหน่งที่ถูกเสนอเป็นตำแหน่งขอบเขตเสียงเป็นตำแหน่งของการเปลี่ยนแปลงค่าลักษณะสำคัญ ที่เกิดจากการเปลี่ยนแปลงค่าลักษณะสำคัญของ ขอบเขตเสียงอื่นที่อยู่ก่อนหน้าและขอบเขตเสียงถัดไป (คลาดเคลื่อนห่างไปจากกรอบตำแหน่งขอบเขตเสียงที่พิจารณา) ดังนั้นตำแหน่งที่ถูกเสนอเป็นคำตอบของระบบจึงทำให้เกิดความผิดพลาดของคำตอบ ส่งผลให้ผลการระบุตำแหน่งขอบเขตเสียงลดลง



รูปที่ 5.7 ความผิดพลาดที่ลดลงเมื่อใช้ขนาดกรอบพิจารณาหาตำแหน่งขอบเขตที่แตกต่างกัน

จากรูปที่ 5.7 เปรียบเทียบร้อยละความถูกต้องของการระบุตำแหน่งขอบเขตเสียงเมื่อใช้ขนาดกรอบพิจารณา 5 เฟรม 7 เฟรมและ 9 เฟรม ตามลำดับ ซึ่งแสดงผลที่ระดับความคลาดเคลื่อนที่ยอมรับแตกต่างกัน พบว่าการใช้กรอบพิจารณาขนาด 5 เฟรมมีค่าร้อยละความถูกต้องมากกว่าการใช้ขนาด 7 และ 9 เฟรม ซึ่งรูปนี้รายงานตามระดับความคลาดเคลื่อนเมื่อเทียบกับคำตอบจริง คือ ไม่คลาดเคลื่อนจากคำตอบจริงเลย (0 เฟรม) ความคลาดเคลื่อนที่  $\pm 1$   $\pm 2$  และ  $\pm 3$  เฟรมตามลำดับ และในตารางที่ 5.11 แสดงผลการระบุตำแหน่งขอบเขตหน่วยเสียงวิธีที่เสนอซึ่งใช้ขนาดกรอบพิจารณาต่างกัน โดยรายงานผลแต่ละชนิดขอบเขตเสียงเปรียบเทียบกับผลการระบุตำแหน่งขอบเขตหน่วยเสียงของแบบอ้างอิง

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.12 ผลการระบุตำแหน่งขอบเขตหน่วยเสียงวิธีที่เสนอ ซึ่งใช้ขนาดกรอบพิจารณาต่างกัน

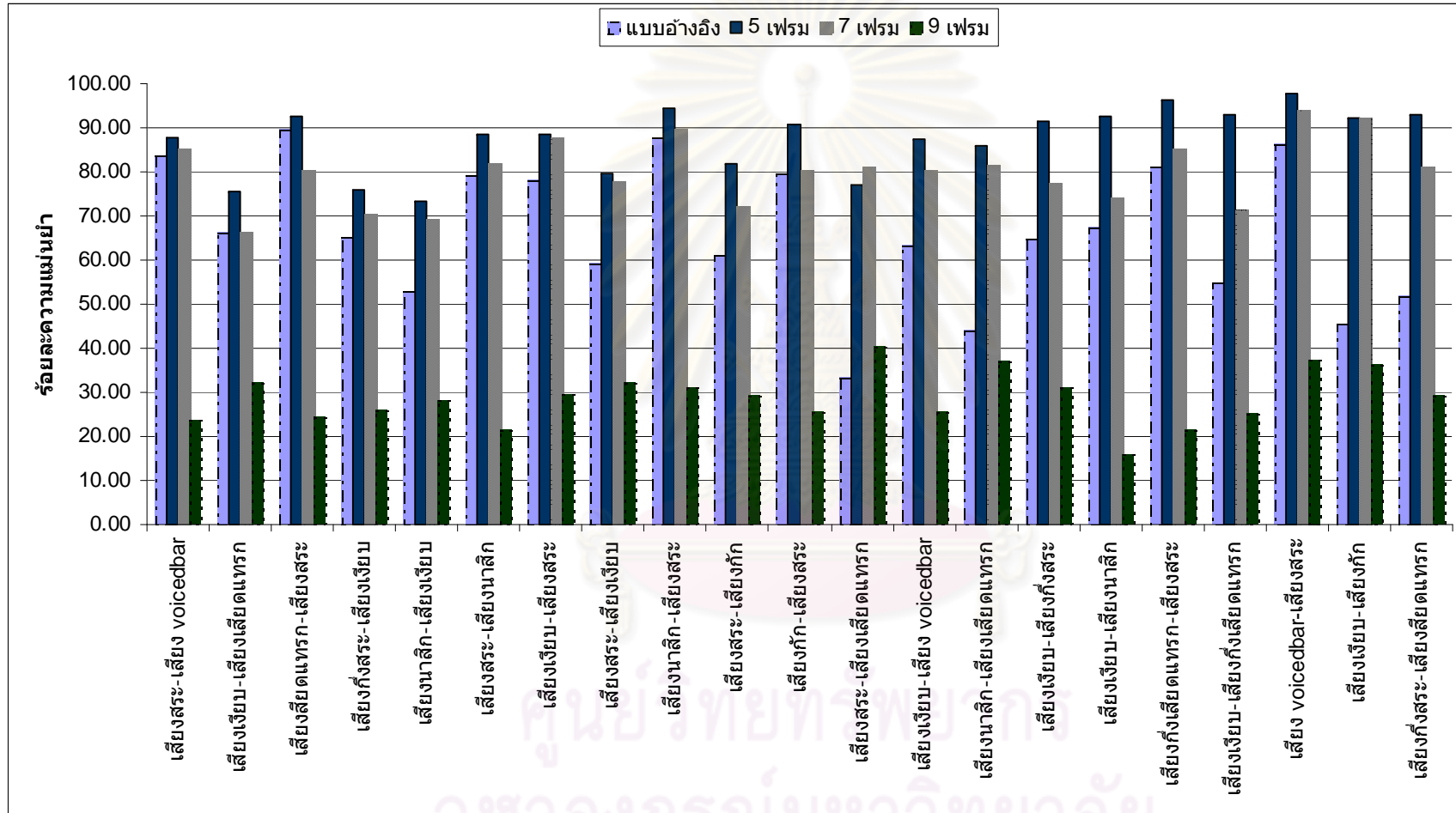
ชนิดขอบเขตหน่วยเสียง	ร้อยละความแม่นยำ			ความผิดพลาดลดลง			
	แบบ อ้างอิง	ขนาดกรอบพิจารณา			ขนาดกรอบพิจารณา		
		5 เฟรม	7 เฟรม	9 เฟรม	5 เฟรม	7 เฟรม	9 เฟรม
เสียงสระ-เสียงvb	83.74	87.80	85.37	23.58	25.00	10.00	-370.00
เสียงเจียบ-เสียงเสียดแทรก	66.36	75.45	66.36	32.27	27.03	0.01	-101.33
เสียงเสียดแทรก-เสียงสระ	89.62	92.63	80.30	24.51	29.01	-89.78	-627.25
เสียงกึ่งสระ-เสียงเจียบ	65.23	75.91	70.23	25.91	30.71	14.37	-113.09
เสียงนาสิก-เสียงเจียบ	52.79	73.15	69.10	28.06	43.13	34.54	-52.37
เสียงสระ-เสียงนาสิก	79.09	88.51	81.70	21.39	45.04	12.48	-275.94
เสียงเจียบ-เสียงสระ	78.26	88.70	87.83	29.57	48.00	44.00	-223.99
เสียงสระ-เสียงเจียบ	59.12	79.50	77.74	32.29	49.86	45.55	-65.64
เสียงนาสิก-เสียงสระ	87.87	94.32	89.55	30.97	53.20	13.84	-469.10
เสียงสระ-เสียงกัก	61.11	81.94	72.22	29.17	53.57	28.57	-82.14
เสียงกัก-เสียงสระ	79.55	90.66	80.19	25.58	54.35	3.15	-263.91
เสียงสระ-เสียงเสียดแทรก	33.20	77.08	81.03	40.32	65.68	71.60	10.65
เสียงเจียบ-เสียงvb	63.50	87.59	80.29	25.55	66.00	46.01	-103.98
เสียงนาสิก-เสียงเสียดแทรก	43.95	85.99	81.53	36.94	75.00	67.04	-12.50
เสียงเจียบ-เสียงกึ่งสระ	64.79	91.55	77.46	30.99	76.00	36.00	-96.01
เสียงเจียบ-เสียงนาสิก	67.44	92.56	73.95	15.81	77.14	20.00	-158.56
เสียงกึ่งเสียดแทรก-เสียงสระ	80.98	96.34	85.37	21.46	80.76	23.06	-312.92
เสียงเจียบ-เสียงกึ่งเสียดแทรก	54.66	92.93	71.38	25.08	84.40	36.88	-65.24
เสียงvb-เสียงสระ	86.22	97.88	93.99	37.46	84.61	56.41	-353.88
เสียงเจียบ-เสียงกัก	45.44	92.15	92.40	36.12	85.61	86.07	-17.08
เสียงกึ่งสระ-เสียงเสียดแทรก	51.72	93.10	81.03	29.31	85.72	60.72	-46.42
รวมทั้งหมด	65.00	80.23	75.74	35.37	43.51	30.69	-84.66

จากตารางที่ 5.12 รายงานผลการระบุตำแหน่งขอบเขตหน่วยเสียงของแต่ละชนิด ขอบเขตเสียงเมื่อใช้ขนาดกรอบหน้าต่างพิจารณาแตกต่างกัน โดยวัดผลตามระดับความคลาดเคลื่อนที่ยอมรับได้ 1 เฟรมเทียบกับคำตอบจริง พบว่าชนิดขอบเขตเสียงส่วนใหญ่เมื่อใช้ขนาดกรอบหน้าต่างพิจารณา 5 เฟรม ทำให้ผลการทดสอบมีผลการระบุขอบเขตเสียงถูกต้องมากกว่าการทดสอบที่ใช้ขนาดกรอบหน้าต่างพิจารณา 7 และ 9 เฟรม ซึ่งแสดงเป็นฮิสโตแกรมของร้อยละความถูกต้องของการหาคำแหน่งขอบเขตเสียงจากชนิดขอบเขตเสียงต่าง ๆ เมื่อใช้ขนาดกรอบพิจารณาแตกต่างกันโดยระดับความคลาดเคลื่อนที่ 1 เฟรมเทียบกับคำตอบที่แท้จริง ตามรูปที่ 5.8 พบว่าทุกชนิดขอบเขตเสียงมีแผนภูมิแท่งของผลการระบุหาคำแหน่งขอบเขตเสียงถูกต้องเมื่อใช้ขนาดกรอบพิจารณามีค่า 5 เฟรมสูงกว่าค่าแผนภูมิแท่งของผลการระบุหาคำแหน่งขอบเขตเสียงถูกต้องเมื่อใช้ขนาดกรอบพิจารณา 7 และ 9 เฟรม



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



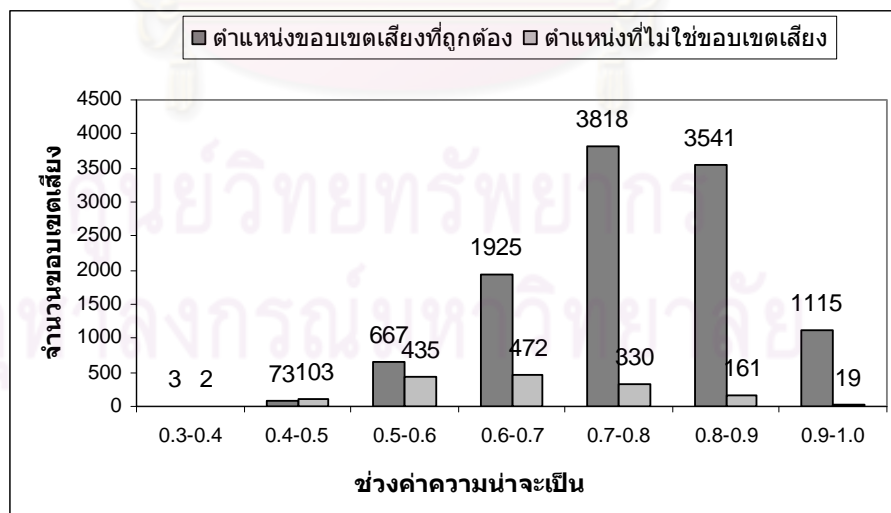


รูปที่ 5.8 สีสโตแกรมร้อยละความถูกต้องของการหาขอบเขตเสียงเมื่อใช้ขนาดกรอบพิจารณาเพื่อปรับหาดำแหน่งขอบเขตที่แตกต่างกัน

## 6. ผลการวิเคราะห์คะแนนในการพิจารณาคำแหน่งขอบเขตเสียง

วิทยานิพนธ์นี้เสนอการหาคำแหน่งขอบเขตเสียงโดยใช้ผลการจำแนกกลุ่ม ซึ่งผลจากจำแนกกลุ่มนี้สามารถนำมาคำนวณตามสมการที่ (2.12) เพื่อเป็นคะแนนที่ใช้ตัดสินใจเลือกตำแหน่งเฟรมที่เหมาะสมกับตำแหน่งขอบเขตเสียงที่แท้จริง การพิจารณาจากคะแนนของแต่ละเฟรมนั้น คะแนนของแต่ละเฟรมภายในกรอบพิจารณามีค่ามากหรือน้อยจะส่งผลกระทบต่อความมั่นใจของการเลือกเฟรมนั้นมาเป็นคำตอบของตำแหน่งขอบเขตเสียงที่แท้จริง และจากขั้นตอนการตัดสินใจเลือกตำแหน่งเฟรมนั้นพิจารณาจากตำแหน่งเฟรมที่มีค่าคะแนนที่สูงที่สุด ดังนั้นจากการวิเคราะห์การเลือกตำแหน่งเพื่อเป็นผลลัพธ์ของวิธีที่เสนอ นำมาวิเคราะห์ความสัมพันธ์ของคะแนนที่ตำแหน่งที่ได้รับเลือกมาตอบเป็นตำแหน่งขอบเขตเสียงและตำแหน่งที่เลือกตอบนั้นเป็นตำแหน่งขอบเขตเสียงที่ถูกต้อง

ผลการวิเคราะห์นี้พิจารณาคะแนนที่ถูกเลือกมาตอบเป็นตำแหน่งขอบเขตเสียง ถ้าคะแนนของตำแหน่งเฟรมที่พิจารณามีค่ามาก ตำแหน่งเฟรมนั้นจะมีโอกาสที่จะเป็นตำแหน่งขอบเขตเสียงที่แท้จริงมาก เช่นเดียวกันเมื่อคะแนนของตำแหน่งเฟรมที่พิจารณามีค่าน้อย ตำแหน่งเฟรมนั้นจะมีโอกาสที่จะเป็นตำแหน่งขอบเขตเสียงน้อย ดังนั้นรูปที่ 5.9 ต้องการแสดงความสัมพันธ์ของคะแนนที่ตำแหน่งที่ได้รับเลือกมาตอบเป็นตำแหน่งขอบเขตเสียงและตำแหน่งที่เลือกตอบนั้นไม่เป็นตำแหน่งขอบเขตเสียงที่แท้จริง ซึ่งการเลือกตัดสินใจให้เป็นคำตอบของวิธีที่เสนอ ใช้ขนาดกรอบพิจารณา 5 เฟรม

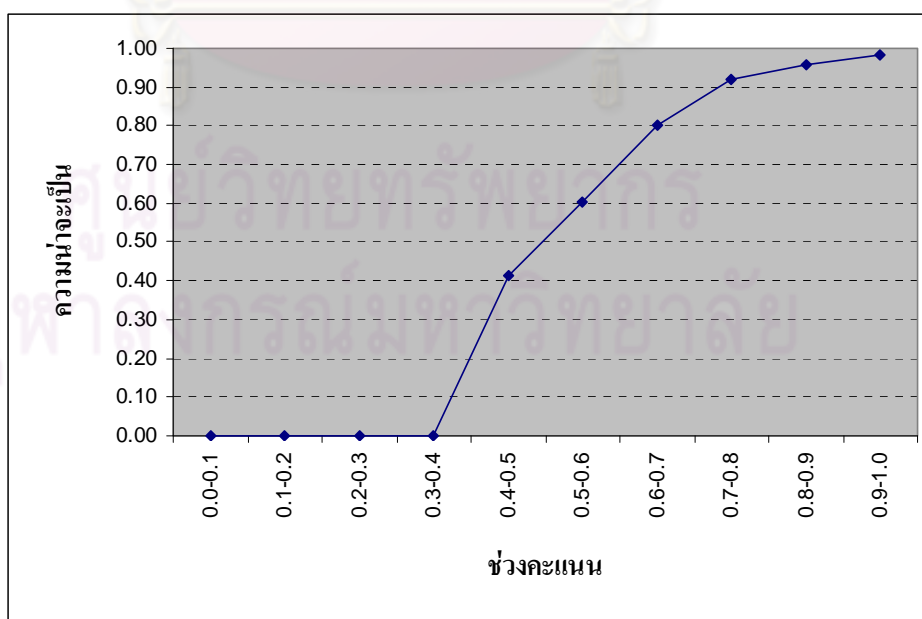


รูปที่ 5.9 ฮิสโตแกรมจำนวนขอบเขตเสียงที่ระบุขอบเขตได้ถูกต้องในช่วงคะแนนต่าง ๆ

จากรูปที่ 5.9 แสดงความถี่จำนวนขอบเขตเสียงที่ระบุขอบเขตได้ถูกต้อง จากขั้นตอนการเลือกตำแหน่งขอบเขตเสียงโดยใช้สถิติ โดยพิจารณาจากคะแนนที่ระบุว่าเป็นประเภท

กลุ่มขอบเขตเสี่ยงจากผลลัพธ์การจำแนกกลุ่ม พบว่าคะแนนที่แสดงโอกาสของแต่ละเฟรมที่จะถูกเลือกมาเป็นคำตอบ (ตำแหน่งขอบเขตเสี่ยง) นั้นมีความสัมพันธ์เป็นไปตามสมมติฐานที่เสนอไว้ นั่นคือ การพิจารณาโอกาสของเฟรมที่มีค่าคะแนนมากจะแสดงความมั่นใจในการเลือกเฟรมนั้น เป็นคำตอบตำแหน่งขอบเขตเสี่ยงมาก คะแนนส่วนใหญ่ที่ถูกเลือกมาตอบเป็นตำแหน่งขอบเขตเสี่ยงมีค่าอยู่ในช่วงที่มากกว่า 0.5 ขึ้นไป ซึ่งค่าเฉลี่ยของคะแนนที่ตอบเป็นตำแหน่งขอบเขตเสี่ยงที่ถูกต้องมีค่า 0.77 หมายถึง ตำแหน่งที่ถูกเลือกมีคะแนนเพื่อบอกความมั่นใจเฉลี่ย 0.77 ตำแหน่งที่ถูกเลือกนั้นมีจำนวนที่ตอบถูกต้องมาก ทำให้สรุปได้ว่าคะแนนที่ได้จากความน่าจะเป็นที่ระบุประเภทกลุ่มขอบเขตเสี่ยงจากการจำแนกนี้มีความน่าเชื่อถือ ทำให้ระบบการปรับหาขอบเขตเสี่ยงมีความแม่นยำ

เพื่อประเมินคะแนนที่ใช้ระบุความเป็นตำแหน่งขอบเขตเสี่ยง จึงเสนอการวิเคราะห์ค่าอัตราส่วนของจำนวนขอบเขตเสี่ยงเสี่ยงในช่วงคะแนนต่างซึ่งถูกพิจารณาเลือกตอบเป็นตำแหน่งขอบเขตเสี่ยงของวิธีที่เสนอ ที่ตอบเป็นตำแหน่งขอบเขตเสี่ยงที่ถูกต้องเทียบกับจำนวนขอบเขตเสี่ยงที่ถูกเลือกตอบ แต่ไม่ใช่ ตำแหน่งขอบเขตเสี่ยงที่แท้จริง หรือตอบไม่ถูกต้อง พบว่า ช่วงคะแนนระหว่าง 0.9-1.0 แสดงความน่าจะเป็นมากที่สุด ที่มีโอกาสการเลือกตำแหน่งที่มีคะแนนช่วงดังกล่าวเพื่อมาพิจารณาเป็นตำแหน่งขอบเขตเสี่ยงที่ถูกต้อง ในทางตรงข้ามตำแหน่งที่ถูกพิจารณาเสนอเป็นตำแหน่งขอบเขตเสี่ยงที่เหมาะสม จากช่วงคะแนนต่ำๆ พบว่า มีโอกาสความน่าจะเป็นของการเลือกเป็นตำแหน่งขอบเขตเสี่ยงที่ถูกต้องน้อยกว่า ซึ่งแสดงแนวโน้มโอกาสตอบเป็นตำแหน่งขอบเขตเสี่ยงถูกต้องของช่วงคะแนนต่างๆ ดังรูปที่ 5.10



รูปที่ 5.10 แสดงแนวโน้มช่วงคะแนนต่างๆที่สามารถระบุตำแหน่งขอบเขตเสี่ยงได้ถูกต้อง

ถูกต้อง

จากรูปที่ 5.10 ที่แสดงช่วงความน่าจะเป็นของตำแหน่งที่ถูกเลือกเป็นขอบเขตเสี่ยง นั้นตอบถูกต้องตรงกับตำแหน่งขอบเขตเสี่ยงที่แท้จริง ในช่วงคะแนนต่างๆ พบว่า แนวโน้มในช่วงคะแนนสูงๆ มีโอกาสตอบเป็นตำแหน่งขอบเขตเสี่ยงถูกต้องมากกว่า



ศูนย์วิทยพัทยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 6

### บทสรุปผลการวิจัยและข้อเสนอแนะ

#### สรุปผลการวิจัย

การหาตำแหน่งขอบเขตเสียงนี้วิทยานิพนธ์นี้เสนอวิธีหาตำแหน่งขอบเขตเสียงในเสียงพูดต่อเนื่องภาษาไทยจากค่าลักษณะสำคัญทางเสียงและคุณลักษณะทางสัทศาสตร์ซึ่งในวิทยานิพนธ์นี้ได้ศึกษาลักษณะและสมบัติของค่าลักษณะสำคัญทางเสียงต่าง ๆ จากผลวิเคราะห์สรุปได้ว่าลักษณะสำคัญทางเสียงและคุณลักษณะทางสัทศาสตร์ที่ใช้ใช้นั้น ควรเลือกค่าที่แตกต่างกันสำหรับแต่ละชนิดขอบเขตเสียงขึ้นอยู่กับตามชนิดขอบเขตเสียงโดยขั้นตอนการวิเคราะห์ลักษณะสำคัญทางเสียงและสรุปได้ว่าค่าสัมประสิทธิ์เมลฟรีเควินซีเคปสตรอลเป็นค่าที่เหมาะสมกับการใช้หาตำแหน่งการเปลี่ยนแปลงของรูปร่างของสเปกตรัมของเสียง ซึ่งตรงกับทฤษฎีกำเนิดเสียงและตัวกรองของกลไกการเกิดเสียงของมนุษย์ เมื่อเสียงผ่านตามช่องทางเดินเสียงในช่องปากมีรูปร่างของสเปกตรัมถูกเปลี่ยนแปลงรูปร่างไปตามตัวกรองและตัวกำเนิดเสียง ดังนั้นบริเวณรอยต่อระหว่างหน่วยเสียงจึงเกิดการเปลี่ยนแปลงของรูปร่างสเปกตรัมทันที สะท้อนให้ค่าลักษณะสำคัญทางเสียงที่สกัดได้จากสเปกตรัมมีค่าเปลี่ยนแปลงตามเช่นกัน วิทยานิพนธ์นี้ใช้ค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงมาเป็นข้อมูลเพื่อพิจารณาหาตำแหน่งขอบเขตเสียงจากเหตุผลของรอยต่อระหว่างสองหน่วยเสียงเกิดการเปลี่ยนแปลงทันที 1 ค่าเพื่อเป็นตัวแทนการเปลี่ยนแปลงเฟรมดังกล่าวแทนในเวกเตอร์เฉพาะที่ใช้ฝึกฝนและจำแนกนั้นเพียงพอและใช้เวลาในการประมวลผลน้อยที่สุด

วิทยานิพนธ์นี้ใช้การระบุหาตำแหน่งขอบเขตเสียงอัตโนมัติที่มีการทำงานเป็นสองส่วน ส่วนแรกจากการระบุตำแหน่งหน่วยเสียงแบบบังคับจากแบบจำลองอิดเด็นมาคอฟ และปรับหาตำแหน่งขอบเขตเสียงให้แม่นยำมากยิ่งขึ้นด้วยการพิจารณาจากผลการจำแนกกลุ่มจากการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น เนื่องจากการปรับหาตำแหน่งขอบเขตเสียงซึ่งเสนอขั้นตอนสร้างกฎตัดสินใจ ต้องศึกษาการเปลี่ยนแปลงของลักษณะสำคัญทางเสียงทีละค่า และพิจารณาจากการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงจากสัญญาณเสียงทีละสัญญาณ ทำให้ใช้เวลาในการพิจารณาอีกทั้งข้อผิดพลาดอาจเกิดขึ้นได้จากขั้นตอนการสร้างกฎที่เฉพาะเจาะจงกับข้อมูลเกินไปและข้อมูลที่ศึกษาไม่ใช่ตัวแทนของข้อมูลเสียงทั้งหมด นั่นคือ ในขั้นตอนการศึกษานั้นพิจารณาจากข้อมูลบางส่วนเท่านั้นไม่ครอบคลุมลักษณะที่เกิดขึ้นกับสัญญาณเสียงทั้งหมด อาจจะทำให้กฎที่สร้างขึ้นไม่เหมาะสมกับข้อมูลทั้งหมดได้ และเมื่อต้องการนำไปใช้จริงที่ต้องการเปลี่ยนแปลงฐานข้อมูลจะทำให้กฎที่ได้จะไม่เหมาะสมกับข้อมูลดังกล่าว ดังผลรายงานใน



ตารางที่ 5.6 และ 5.7 เป็นผลการทดลองที่ได้จากการระบุตำแหน่งขอบเขตเสียงเมื่อใช้ทดสอบกับฐานข้อมูลชุดที่ 2 ซึ่งเป็นข้อมูลใช้เพื่อพัฒนาระบบ และฐานข้อมูลชุดที่ 3 ซึ่งเป็นข้อมูลใช้เพื่อทดสอบระบบ ตามลำดับ โดยใช้ในขั้นตอนการสร้างกฎฐานใช้ข้อมูลชุดที่ 2 จากนั้นเมื่อนำกฎไปทดสอบกับฐานข้อมูลชุดที่ 3 พบว่า ค่าร้อยละความแม่นยำของการระบุเมื่อใช้ทดสอบกับฐานข้อมูลชุดทดสอบที่ 3 ให้ผลเป็นที่ยอมรับได้ แต่ค่าความผิดพลาดที่ลดลงเมื่อเทียบกับคำตอบที่แท้จริง พบว่ามีค่าน้อยกว่าการทดสอบกับฐานข้อมูลที่ใช้ในการสร้างกฎ และจากตารางที่ 5.8 ซึ่งแสดงผลการระบุตำแหน่งขอบเขตเสียงเมื่อใช้กฎตัดสิน พบว่า ในบางชนิดขอบเขตเสียงมีผลการระบุตำแหน่งขอบเขตเสียงต่ำกว่าแบบอ้างอิง ซึ่งขัดแย้งกับสมมติฐานในขั้นตอนการสร้างกฎที่ สร้างเมื่อทดสอบกับข้อมูลแล้วพบว่า กฎสามารถปรับหาตำแหน่งขอบเขตเสียงที่เหมาะสมกับชนิดขอบเขตเสียงนั้นได้ จึงสรุปได้ว่า กฎที่ได้อาจจะมีความเฉพาะเจาะจงกับข้อมูลที่ศึกษา

นอกจากนี้ผลความความแม่นยำของการระบุตำแหน่งขอบเขตเสียงที่ได้จากวิธีการปรับหาตำแหน่งขอบเขตเสียงจากวิธีใช้กฎตัดสินใจให้ความแม่นยำของตำแหน่งที่ระบุเปรียบเทียบกับคำตอบที่แท้จริง มีผลความแม่นยำน้อยกว่าผลที่ได้จากการปรับตำแหน่งขอบเขตเสียงเมื่อใช้วิธีสถิติซึ่งรายงานในบทที่ 5 ที่แสดงว่าวิธีการในการหาตำแหน่งขอบเขตเสียงซึ่งใช้การพิจารณาทางสถิติจากการจำแนกกลุ่มด้วยวิธีการวิเคราะห์สถิติสคริมิแนนต์เชิงเส้น ให้ผลความถูกต้องดีกว่า เมื่อใช้การจำแนกออกเป็นสองกลุ่มคือ กลุ่มที่เป็นขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง โดยแยกแต่ละชุดการจำแนกออกเป็นทั้งสิ้น 21 ชุด ตามชนิดขอบเขตเสียงและเลือกใช้ลักษณะสำคัญทางเสียงและคุณลักษณะทางสัทศาสตร์ที่เหมาะสมกับชนิดขอบเขตเสียงนั้น ๆ โดยใช้ตัวอย่างการฝึกฝนจากอัตราส่วนของจำนวนข้อมูลกลุ่มขอบเขตเสียงต่อจำนวนข้อมูลที่ไม่ใช่กลุ่มขอบเขตเสียงแบบ 3:4 มีประสิทธิภาพการตรวจหาตำแหน่งดีที่สุด มีร้อยละความถูกต้องของการระบุตำแหน่งขอบเขตเสียง 80.22 และ 89.52 เมื่อวัดผลตามระดับความคลาดเคลื่อนที่ยอมรับได้ 1 และ 2 เฟรมตามลำดับ ซึ่งวัดเป็นความผิดพลาดที่ลดลงเป็นร้อยละ 43.42 และ 20 เมื่อเทียบกับการระบุตำแหน่งขอบเขตหน่วยเสียงจากแบบอ้างอิงที่ใช้แบบจำลองฮิดเดินมาคอฟซึ่งมีค่าร้อยละความถูกต้องของการระบุตำแหน่งขอบเขตเสียงจากแบบจำลองฮิดเดินมาคอฟคือ 65.04 และ 86.90 เมื่อวัดผลตามระดับความคลาดเคลื่อนที่ยอมรับได้ 1 และ 2 เฟรมตามลำดับ ผลความถูกต้องของการระบุตำแหน่งขอบเขตเสียงมีความคลาดเคลื่อนเทียบกับการตอบจริงมีค่าเฉลี่ย 1 เฟรม (เฟรมมีขนาด 10 มิลลิวินาที) เปรียบเทียบกับชุดการทดลองอ้างอิงมีค่าเฉลี่ย 1.42 เฟรม พบว่าความถูกต้องมีผลความคลาดเคลื่อนเฉลี่ยดีขึ้นประมาณ 4.2 วินาทีเท่านั้น อาจจะเป็นค่าที่ไม่สามารถปรับปรุงและเพิ่มความแม่นยำในการนำไปใช้ประยุกต์กับระบบการเรียนรู้จำเสียง แต่เนื่องจากในขั้นตอนการเลือกหน่วยเสียง การสังเคราะห์เสียงนั้น ต้องการความแม่นยำสูงในการเลือกส่วนหน่วยเสียงที่ประกอบด้วยอะคูสติกเหมือนกันในหน่วยเสียงเดียวกัน เพื่อนำแต่ละหน่วยเสียงมาสังเคราะห์ให้เกิดเสียงที่ความ

เป็นธรรมชาติต่อไป ดังนั้นผลที่ได้จากการปรับหาขอบเขตเสียงในวิทยานิพนธ์นี้จึงให้ประโยชน์แก่การนำไปใช้สร้างฐานข้อมูลเสียงในได้ดี

### ข้อเสนอแนะ

การปรับปรุงประสิทธิภาพของการหาตำแหน่งขอบเขตเสียง สามารถปรับปรุงประสิทธิภาพของตัวจำแนกกลุ่มให้สามารถแบ่งแยกกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง ได้แก่ ปรับปรุงขั้นตอนการวิเคราะห์ลักษณะสำคัญทางเสียง โดยศึกษาลักษณะสำคัญทางเสียงอื่น ๆ ที่มีความสัมพันธ์กับสมบัติของกลุ่มเสียงเพิ่มเติม และสร้างวิธีการเลือกกลุ่มค่าลักษณะสำคัญตามสมบัติของกลุ่มเสียง และปรับปรุงในขั้นตอนการนำผลการจำแนกไปตรวจหาตำแหน่งขอบเขตเสียง เช่น การวิเคราะห์หาขนาดกรอบพิจารณาที่ยึดหยุ่นตามชนิดขอบเขตเสียงอาจทำให้สามารถตรวจหาตำแหน่งเฟรมที่มีโอกาสเป็นตำแหน่งที่เหมาะสมของการเกิดขอบเขตเสียง หรือการหาวิธีนำผลความน่าจะเป็นที่ได้จากการจำแนกเสียง ไปคิดกับวิธีการให้คะแนนความมั่นใจหรือการให้คะแนนจากการโหวตจากการตัดสินใจอื่นร่วมด้วย

การแบ่งกลุ่มเสียงที่เสนอในวิทยานิพนธ์นี้แบ่งตามการพิจารณาค่าอะคูสติคที่เกิดขึ้นบนหน่วยเสียง เนื่องจากการระบุตำแหน่งขอบเขตเสียงด้วยคนตัดสินใจพบว่าการแบ่งกลุ่มตามหลักภาษาศาสตร์นั้นทำให้ได้ส่วนหน่วยเสียงที่มีส่วนประกอบของส่วนที่แตกต่างกันเกิดขึ้นในส่วนหน่วยเสียง ไม่ตรงกับกับนิยามของส่วนหน่วยเสียงว่าภายในส่วนเดียวกันต้องมีค่าอะคูสติคเหมือนกัน ซึ่งจากการแบ่งตามอะคูสติคนี้ออกเป็นกลุ่มใหญ่ ๆ 9 กลุ่ม อาจจะไม่ละเอียดยังไม่สามารถตรวจหาส่วนหน่วยเสียงที่เกิดการเปลี่ยนแปลงคุณสมบัติภายในหน่วยเสียง เช่น กลุ่มเสียงกึ่งสระมีการเคลื่อนที่ของฟอร์แมนท์จากจุดเริ่มต้น ไปยังจุดสิ้นสุดหน่วยเสียงอย่างแตกต่างกันชัดเจน ขณะที่กลุ่มเสียงสระมีการเปลี่ยนแปลงของฟอร์แมนท์ในสระบางเสียงอย่างชัดเจนเช่นกัน ดังนั้นการรวมกลุ่มหรือแยกกลุ่มของกลุ่มเสียงกึ่งสระหรือกลุ่มสระจึงสามารถทดลองการจัดกลุ่ม (clustering) จากค่าลักษณะสำคัญทางเสียงร่วมด้วย ทำให้ได้กลุ่มที่เกิดขึ้นที่สมาชิกทุกตัวมีสมบัติคล้ายกัน แทนวิธีการแยกชนิดกลุ่มเสียงของหน่วยเสียงจากการพิจารณาตามหลักภาษาศาสตร์เท่านั้น

การประเมินประสิทธิภาพของระบบสามารถวัดประสิทธิภาพของตำแหน่งขอบเขตของเสียงที่ได้จากการหาตำแหน่งที่เหมาะสมที่สุดมาเป็นตำแหน่งขอบเขตเสียงระหว่างสองหน่วยเสียง เทียบกับคำตอบจริงที่ได้จากคน การวัดประสิทธิภาพสามารถวัดทางอ้อม เช่น วัดประสิทธิภาพของผลที่ได้จากการรู้จำเมื่อตำแหน่งของขอบเขตที่ได้จากระบบไปเป็นข้อมูลเกี่ยวกับเวลาสัญลักษณ์ที่นำไปฝึกฝนแบบจำลองในตัวรู้จำเสียง เทียบกับ ประสิทธิภาพของระบบรู้จำเสียงที่สองที่ฝึกฝนแบบจำลองจากการข้อมูลจากเพื่อวัดตำแหน่งขอบเขตของเสียงที่ผลได้จากการปรับหา

ตำแหน่งที่เหมาะสมที่สุดมาเป็นตำแหน่งขอบเขตเสียง ระหว่างสองหน่วยเสียง การวัดประสิทธิภาพของผล สามารถวัดทางอ้อม เช่น วัดประสิทธิภาพของผลที่ได้จากการรู้จำเมื่อใช้ข้อมูลเกี่ยวกับเวลาสี่ลักษณะที่นำไปฝึกฝน ในระบบรู้จำเสียงแรกจากแบบอ้างอิง เทียบกับ ตัวรู้จำที่สองจากการนำผลที่ได้จากการปรับตำแหน่ง



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## รายการอ้างอิง

- [1] Muller, J. The physiology of the senses, voice and muscular motion with the mental faculties. 1848.
- [2] Fosler-Lussier, E. Lectures notes in Foundations of Spoken Language Processing [Online]. Available from: <http://www.cse.ohio-state.edu/~fosler/794L/lectures/lecture3-handouts.pdf>
- [3] Wikipedia, the free encyclopedia [Online]. Available from: [http://en.wikipedia.org/wiki/Segment\\_%28linguistics%29](http://en.wikipedia.org/wiki/Segment_%28linguistics%29)
- [4] Crystal, D.. Dictionary of Linguistics & Phonetics, 2003
- [5] Kasuriya, S., Sornlertlamvanich, V., Cotsomrong, P., Kanokphara, S., and Thatphithakkul, N.. Thai Speech Corpus for Speech Recognition. Proceeding of the COCOSDA'03, (2003) :54-61.
- [6] Schmi, P.. Explicit N-Best Formant Features for Segment-Based Speech Recognition P octoral dissertation, Department of Electrical Engineering and Computer Science, University of Bern, Switzerland. 1990.
- [7] Wikipedia, the free encyclopedia [Online]. Available from: [http://en.wikipedia.org/wiki/International\\_Phonetic\\_Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet)
- [8] ณัฐกร ทับทอง. การรู้จำคำพูดภาษาไทยโดยใช้ลักษณะบ่งความต่างของหน่วยเสียง. วิทยานิพนธ์ปริญญาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย, 2538
- [9] K. Sjolander and J. Beskow. Wavesurfer - an open source speech tool”, In Prodeccings of ICSLP. (2000)
- [10] กัลยา วานิชย์บัญชา, หนังสือ การวิเคราะห์ข้อมูลด้วย SPSS for windows, 2548 พิมพ์ครั้งที่ 8 โรงพิมพ์ธรรมสาร
- [11] Yeon-Jun Kim, A. C. Automatic segmentation Combining an HMM-Based Approach and Spectral Boundary Correction. In Prodeccings of ICSLP. (2002)
- [12] Wilson, E., A. J. A. C. Segmentation of continuous speech using acoustic -phonetic parameter and statistical learning. In Proceedings of 9th International Conference on Neural Information Processing (2002): 726-730.

- [13] Wilson, E., A. J. A. C. Speech Segmentation using probabilistic phonetic feature hierarchy and Support Vector Machine. Proceedings of IJCNN(2003)
- [14] Tarun Pruthi, C. Y. E.-W. Automatic Classification of nasal and semivowels. Proceedings of 15th ICPhS Barcelon.
- [15] Suchato A.. Classification of Stop Consonant Place or Articulation. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. 2004.
- [16] Lin C.Y., Roger Jang J.S. and Ting H.. Automatic Segmentation and Labeling for Mandarin Chinese Speech Corpora for Concatenation-Based TTS. Chinese Journal of Computational Linguistics(2005): 145-166
- [17] Wang L.J., Zhao Y., Chu M., Zhou J.L., and Cao Z.G.. Refining segmental boundaries for TTS database using fine contextual-dependent boundary models. Proceedings of ICASSP (2004): 641-644
- [18] Sethy A. and Narayanam S.. Refined speech segmentation for concatenative speech synthesis. Proceedings of ICSLP (2002): 145-148
- [19] Toledano, D.T.. Neural network boundary refining for automatic speech segmentation. Proceedings of ICASSP (2000): 3438-3441
- [20] Wilson, E., C. Y.. Acoustic measures for linguistic features distinguishing the semivowel /wɹl/ in American English. Journal of the Acoustical Society of America (1992): 736-757.
- [21] Bitar, N.. Acoustic analysis and modeling of speech based on phonetic features. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Boston university. 1997
- [22] Sharlene, L.A.. Landmark detection for distinctive feature based speech recognition. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. 1995.
- [23] Lu,P.. Identifying Key Phoneme Features in Spectrograms. SUNFEST Report. Center for Sensor Technologies. Department of Electrical Engineering. University of Pennsylvania (1999)
- [24] Chen Y., M.. Nasal Detection Module for a knowledge-based speech recognition system. Proceedings of the ICSLP(2000): 636-639



- [25] Stevens, K. N.. Models of Phonetic Recognition II: An Approach to Feature based Recognition. Proceedings of Montreal Symposium on Speech Recognition. McGill University(1986)
- [26] Micallef, P. and Chilton, T.. Automatic identification of phoneme boundaries using a mixed parameter model. Proceedings of EUROSPEECH(1997): 485-488.
- [27] Zhijie, Z., Guangrui H., Iiao S.. A new method for the voiced/unvoiced decision based on patternclassification theory. Proceedings of Signal Processing 3rd International Conference, (1996)
- [28] Atal and Rabiner, L. R. A Pattern Recognition Approach to Voiced-Unvoiced-silence Classification with Application to Speech Recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing (1976): 24
- [29] Sarma, VVS. and Venugopal, D.. Studies on Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification. In Proceedings IEEE ICASSP(1978): 1-4.
- [30] Siegel et al. Voiced/Unvoiced/Mixed Excitation Classification of Speech, in IEEE Transactions on Acoustics, Speech, and Signal Processing(1982): 451-460.
- [31] Achan, K., Roweis, S.T., Aaron, H., Frey,B.. A Segment-Based Probabilistic Generative Model of Speech. IEEE ICASSP (2005)
- [32] Jan P. H. van Santen1 and Sproat R.. High-accuracy automatic segmentation. Proceedings of EUROSPEECH (1999)
- [33] Glass, J.R., and Zue,v.w.. Multi-Level Acoustic Segmentation of Continuous Speech Proceeding of ICASSP (1988): 429-432
- [34] Lee K.S., MLP-Based Phone Boundary Refining for a TTS Database. IEEE Trans Audio, Speech and Language Processing(2006): 981-989
- [35] Park E.-Y., Kim S.H. and Chung J.H., Automatic speech synthesis unit generation with MLP based postprocessor against auto-segmented phoneme errors. Proceedings of ICASSP(1999): 2985-2990
- [36] Lo, H.Y. and Wang, H.M.. Phonetic Boundary Refinement Using Support Vector Machine. to appear in IEEE Int. Conf. Acoustics, Speech, Signal processing ICASSP(2007)
- [37] Paulo S., and Oliveira L.C.. Improving the Accuracy of the Speech Synthesis Based Phonetic Alignment Using Multiple Acoustic Features. Proceedings of PROPOR(2003): 31-39

- [38] Little J. and Moler, C.. The MathWorks - MATLAB and Simulink for Technical Computing [Computer software]. Available from: [www.mathworks.com](http://www.mathworks.com)
- [39] Steve Young et.al, The HTK Book (for HTK Version 3.2), 2002
- [40] Boersma P. and Wennik D., Praat, a system for doing phonetics by computer [Online]. Available from: <http://www.praat.org> (1996)
- [41] Thongprasert R., Charoenporn T., Sinthupinyo W., and Sornlertlamvanich V.. Development of Very Large Corpora in Thailand. Proceedings of 6th Natural Language Processing Pacific Rim Symposium Post-Conference Workshop(2001)
- [42] Dusan S. and Rabiner L.. On the Relation between Maximum Spectral Transition Positions and phone Boundaries. Proceedings of Interspeech/ICSLP(2006)
- [43] Boonsuk S., Punyabukkana P., and Suchato S.. Precise Phone Boundary Detection using Selective Context-dependent Acoustic Refinement. Proceedings of the Fourth Annual Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology International Conference(2007)
- [44] Rabiner, L. R. and Juang, B.H.. Fundamentals of Speech Recognition. A. Oppenheim, Series Editor, Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [45] Boonsuk S., Punyabukkana P., and Suchato S.. Phone Boundary Detection using Selective Refinements and Context-dependent Acoustic Features. to be appear Proceedings of Interspeech/ICSLP(2007)

### ภาคผนวก ก

ตาราง หน่วยเสียงและสัญลักษณ์สากล

กลุ่มเสียง	หน่วยเสียง	สัญลักษณ์สากล
เสียงกัก	/pʰ/	[p <sup>h</sup> ]
	/p/	[p]
	/tʰ/	[t <sup>h</sup> ]
	/t/	[d]
	/kʰ/	[k <sup>h</sup> ]
	/k/	[k]
	/b/	[b]
	/d/	[d]
	/z/	[ʔ]
เสียงนาสิก	/m/	[m]
	/n/	[n]
	/ŋ/	[ŋ]
เสียงข้างลิ้น	/l/	[l]
เสียงกึ่งเสียดสี	/ç/	[tʃ]
	/çʰ/	[tʃ <sup>h</sup> ]
เสียงร้ว	/r/	[r]
เสียงเสียดแทรก	/s/	[s]
	/h/	[h]
	/f/	[f]
เสียงกึ่งสระ	/j/	[j]
	/w/	[w]

เสียงในภาษาไทย	หน่วยเสียง	สัญลักษณ์สากล
อะ อา	/a/, /aa/	[a], [a : ]
อี อี	/i/, /ii/	[i], [i : ]
อึ อึอ	/y/, /yy/	[ʏ], [ʏ : ]
อุ อุ	/u/, /uu/	[u], [u : ]
เอะ เอ	/e/, /ee/	[e], [e : ]
แอะ แอ	/ɛ/, /ɛɛ/	[ɛ], [ɛ : ]
โอะ โอ	/o/, /oo/	[o], [o : ]
เออะ เออ	/ɔ/, /ɔɔ/	[ɔ], [ɔ : ]
เอาะ ออ	/ɔ̄/, /ɔ̄ɔ̄/	[ɔ̄], [ɔ̄ : ]
เอียะ เอีย	/ia/, /iia/	[ia], [i : a]
เอือะ เอือ	/va/, /vva/	[ʋa], [ʋ : a]
อัวะ อัว	/ua/, /uua/	[ua], [u : a]

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## ภาคผนวก ข

ชนิดขอบเขตเสียง	จำนวนขอบเขต
เสียงสระ-เสียงนาสิก	8,483
เสียงเงิบ-เสียงกัก	8,072
เสียงกัก-เสียงสระ	7,737
เสียงสระ-เสียงเงิบ	7,272
เสียงนาสิก-เสียงเงิบ	4,069
เสียงสระ-เสียงกึ่งสระ	3,382
เสียงนาสิก-เสียงสระ	3,102
เสียงเสียดแทรก-เสียงสระ	2,724
เสียงร้ว-เสียงสระ	2,299
เสียงกึ่งสระ-เสียงเงิบ	1,693
เสียงข้างลิ้น-เสียงสระ	1,627
เสียงกึ่งเสียดแทรก-เสียงสระ	1,614
เสียงกึ่งสระ-เสียงสระ	1,474
เสียงเงิบ-เสียงกึ่งเสียดแทรก	1,198
เสียงกัก-เสียงร้ว	1,174
เสียงสระ-เสียงเสียดแทรก	1,066
เสียงvb-เสียงสระ	1,030
เสียงเงิบ-เสียงเสียดแทรก	852
เสียงเงิบ-เสียงนาสิก	820
เสียงนาสิก-เสียงเสียดแทรก	657
เสียงสระ-เสียงvb	556
เสียงเงิบ-เสียงvb	553
เสียงvb-เสียงกัก	501
เสียงเงิบ-เสียงข้างลิ้น	494
เสียงนาสิก-เสียงนาสิก	484
เสียงสระ-เสียงข้างลิ้น	451
เสียงเงิบ-เสียงร้ว	436
เสียงเงิบ-เสียงสระ	414



ชนิดขอบเขตเสียง	จำนวนขอบเขต
เสียงกัก-เสียงข้างลิ้น	381
เสียงกึ่งสระ-เสียงนาสิก	325
เสียงสระ-เสียงร้าว	321
เสียงนาสิก-เสียงvb	315
เสียงนาสิก-เสียงกึ่งสระ	306
เสียงเงียบ-เสียงกึ่งสระ	289
เสียงกัก-เสียงกึ่งสระ	260
เสียงสระ-เสียงกัก	256
เสียงกึ่งสระ-เสียงเสียดแทรก	244
เสียงสระ-เสียงกึ่งเสียดแทรก	234
เสียงนาสิก-เสียงร้าว	223
เสียงนาสิก-เสียงข้างลิ้น	200
เสียงสระ-เสียงสระ	135
เสียงนาสิก-เสียงกึ่งเสียดแทรก	133
เสียงกึ่งสระ-เสียงร้าว	128
เสียงกึ่งสระ-เสียงกัก	97
เสียงกึ่งสระ-เสียงกึ่งสระ	90
เสียงกึ่งสระ-เสียงข้างลิ้น	78
เสียงเสียดแทรก-เสียงเงียบ	76
เสียงvb-เสียงข้างลิ้น	21
เสียงvb-เสียงร้าว	12
เสียงเสียดแทรก-เสียงข้างลิ้น	10
เสียงข้างลิ้น-เสียงเงียบ	9
เสียงกึ่งเสียดแทรก-เสียงเงียบ	8
เสียงเสียดแทรก-เสียงเสียดแทรก	8
เสียงเสียดแทรก-เสียงร้าว	5
เสียงเสียดแทรก-เสียงนาสิก	2
เสียงเสียดแทรก-เสียงกัก	2
เสียงกึ่งเสียดแทรก-เสียงข้างลิ้น	1
เสียงกัก-เสียงเงียบ	1
รวมทั้งหมด	68,404

## ภาคผนวก ค

ผลการระบุขอบเขตเสี่ยงจากแบบอ้างอิงตามจำนวนรอบการฝึกฝนแบบจำลอง

จำนวนรอบHMM re-estimate	ร้อยละความถูกต้องตามระดับความคลาดเคลื่อนที่ยอมรับ			
	±0 เฟรม	±1 เฟรม	±2 เฟรม	±3 เฟรม
3	23.33	65.04	86.90	93.63
5	21.80	62.07	84.83	92.73
7	21.07	60.43	83.62	92.25
9	20.58	59.28	82.73	91.90
10	20.29	58.77	82.38	91.71

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## ภาคผนวก ง

## ผลทางสถิติที่ใช้วิเคราะห์ค่าลักษณะทางเสียง

ผลทางสถิติที่ใช้วิเคราะห์ค่าลักษณะทางเสียงแต่ละค่า สำหรับแต่ละชนิดขอบเขตเสียง ซึ่งตารางแสดงค่า F-ratio ที่ทำการปรับค่าปกติ (หรือนอร์มอลไลซ์) ค่า P จากการวิเคราะห์ความแปรปรวน ค่าร้อยละความถูกต้องการจำแนกกลุ่มจากการใช้ค่าลักษณะทางเสียงนั้นเป็นข้อมูลการจำแนก และค่าความผิดพลาดการจำแนกกลุ่มระหว่างกลุ่มขอบเขตเสียงและกลุ่มที่ไม่ใช่ขอบเขตเสียง ซึ่งค่าลักษณะทางเสียงต่าง ๆ ใช้สัญลักษณ์แทนค่าดังนี้ STM แทนค่าการเปลี่ยนแปลงรูปร่างของสเปกตรัม F1 F2 F3 แทนความถี่ฟอร์แมนท์ที่หนึ่ง สอง สามตามลำดับ F2F1 แทนความกว้างระหว่างความถี่ฟอร์แมนท์ที่หนึ่งและสอง E1, E2, E3, E4, E5 แทนพลังงานในช่วงความถี่ที่หนึ่ง สอง สาม สี่ ห้า ตามลำดับ ENSN แทนอัตราส่วนพลังงานสำหรับช่วง LOGEN แทนล็อกของพลังงาน TOTALENERGY แทนพลังงานเสียง CORR แทนความสัมพันธ์ระหว่างเฟรมที่สนใจและเฟรมที่อยู่ก่อนหน้า F0 แทนความถี่มูลฐาน ZCR แทนอัตราการตัดศูนย์

## เสียงกึ่งเสียงดแทรก-เสียงสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.32	0.00	0.147	85.95
En1	1.00	0.00	0.081	92.64
En2	0.44	0.00	0.162	80.40
En3	0.14	0.00	0.291	64.51
En4	0.02	0.00	0.426	63.59
En5	0.15	0.00	0.306	70.95
ENSN	0.03	0.00	0.394	68.72
LOGEN	0.49	0.00	0.174	80.93
TOTALENERGY	0.29	0.00	0.166	82.24
F1	0.35	0.00	0.163	80.33
F2	0.62	0.00	0.153	84.08
F2F1	0.47	0.00	0.160	84.25
F3	0.57	0.00	0.156	82.17
F0	0.38	0.00	0.186	76.54

STM	0.85	0.00	0.129	86.52
ZCR	0.51	0.00	0.175	87.08

## เสียงเสียดแทรก-เสียงสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.37	0.00	0.146	86.31
En1	0.96	0.00	0.102	91.59
En2	0.61	0.00	0.142	85.05
En3	0.22	0.00	0.255	71.73
En4	0.00	0.31	0.495	50.04
En5	0.10	0.00	0.345	65.08
ENSN	0.00	0.37	0.496	50.73
LOGEN	0.70	0.00	0.147	86.85
TOTALENERGY	0.37	0.00	0.139	88.24
F1	0.40	0.00	0.170	78.81
F2	0.50	0.00	0.194	78.46
F2F1	0.38	0.00	0.198	79.62
F3	0.53	0.00	0.186	80.47
F0	0.27	0.00	0.235	73.48
STM	1.00	0.00	0.128	88.92
ZCR	0.44	0.00	0.202	82.66

## เสียงนาสิก-เสียงเสียดแทรก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.13	0.00	0.272	76.79
En1	0.28	0.00	0.252	73.07
En2	0.01	0.00	0.472	58.04
En3	0.16	0.00	0.297	76.04
En4	0.41	0.00	0.189	81.85
En5	0.54	0.00	0.184	79.76
ENSN	0.58	0.00	0.161	87.50

LOGEN	0.09	0.00	0.387	63.10
TOTALENERGY	0.03	0.00	0.416	62.80
F1	0.24	0.00	0.190	75.22
F2	0.53	0.00	0.188	82.51
F2F1	0.67	0.00	0.172	84.90
F3	0.37	0.00	0.252	76.12
F0	0.39	0.00	0.231	75.45
STM	0.03	0.00	0.393	58.41
ZCR	1.00	0.00	0.133	84.08

## เสียงนาสิก-เสียงเจียบ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.21	0.00	0.212	76.63
En1	0.80	0.00	0.185	76.86
En2	0.13	0.00	0.352	61.68
En3	0.04	0.00	0.389	61.40
En4	0.00	0.07	0.491	50.46
En5	0.00	0.00	0.489	51.48
ENSN	0.36	0.00	0.322	74.16
LOGEN	0.80	0.00	0.246	71.73
TOTALENERGY	0.18	0.00	0.207	76.38
F1	0.21	0.00	0.237	71.38
F2	0.65	0.00	0.266	73.07
F2F1	0.95	0.00	0.187	78.83
F3	0.50	0.00	0.310	69.35
F0	0.56	0.00	0.269	71.97
STM	1.00	0.00	0.208	76.57
ZCR	0.17	0.00	0.273	75.09



## เสียงนาสิก-เสียงสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.14	0.00	0.320	70.08
En1	0.46	0.00	0.252	76.20
En2	0.76	0.00	0.196	78.05
En3	0.76	0.00	0.177	83.75
En4	0.62	0.00	0.188	79.97
En5	0.15	0.00	0.333	68.48
ENSN	0.74	0.00	0.176	82.60
LOGEN	0.28	0.00	0.296	65.88
TOTALENERGY	0.17	0.00	0.303	70.40
F1	0.13	0.00	0.321	69.28
F2	0.00	0.42	0.542	52.14
F2F1	0.09	0.00	0.334	67.33
F3	0.01	0.00	0.457	56.50
F0	0.00	0.01	0.500	49.95
STM	1.00	0.00	0.185	82.88
ZCR	0.43	0.00	0.264	72.47

## เสียงกึ่งสระ-เสียงเสียดแทรก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.48	0.00	0.218	81.13
En1	0.98	0.00	0.176	84.56
En2	0.16	0.00	0.390	61.76
En3	0.05	0.00	0.422	49.02
En4	0.20	0.00	0.364	70.10
En5	0.60	0.00	0.269	75.98
ENSN	0.14	0.00	0.324	72.55
LOGEN	0.46	0.00	0.298	67.40
TOTALENERGY	0.24	0.00	0.300	71.32
F1	0.51	0.00	0.192	76.47

F2	0.78	0.00	0.199	78.43
F2F1	1.00	0.00	0.166	86.27
F3	0.93	0.00	0.195	81.13
F0	0.86	0.00	0.199	76.72
STM	0.64	0.00	0.271	73.04
ZCR	0.94	0.00	0.209	80.88

## เสียงกึ่งสระ-เสียงเงียบ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.19	0.00	0.178	79.87
En1	0.49	0.00	0.145	81.85
En2	0.26	0.00	0.227	74.62
En3	0.19	0.00	0.269	72.28
En4	0.16	0.00	0.298	68.44
En5	0.07	0.00	0.350	61.52
ENSN	0.01	0.00	0.468	55.54
LOGEN	0.74	0.00	0.216	75.83
TOTALENERGY	0.15	0.00	0.177	80.86
F1	0.15	0.00	0.277	66.58
F2	0.31	0.00	0.283	68.18
F2F1	0.47	0.00	0.214	77.59
F3	0.37	0.00	0.272	73.82
F0	0.48	0.00	0.226	76.70
STM	1.00	0.00	0.151	82.39
ZCR	0.02	0.00	0.426	59.86

## เสียงเงียบ-เสียงกึ่งเสียงดแทรก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.00	0.00	0.483	53.56
En1	0.15	0.00	0.233	75.99
En2	0.29	0.00	0.141	87.59

En3	0.30	0.00	0.133	87.95
En4	0.62	0.00	0.124	87.59
En5	0.54	0.00	0.132	87.44
ENSN	0.19	0.00	0.234	77.42
LOGEN	0.56	0.00	0.185	77.72
TOTALENERGY	0.06	0.00	0.139	85.86
F1	0.00	0.09	0.484	50.81
F2	0.01	0.00	0.455	53.00
F2F1	0.00	0.00	0.460	55.34
F3	0.00	0.77	0.492	49.95
F0	0.02	0.00	0.468	52.49
STM	1.00	0.00	0.117	88.91
ZCR	0.50	0.00	0.210	78.84

## เสียงเจียบ-เสียงเสียดแทรก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.00	0.08	0.501	49.53
En1	0.18	0.00	0.311	63.63
En2	0.40	0.00	0.179	80.23
En3	0.47	0.00	0.146	82.57
En4	0.55	0.00	0.137	83.74
En5	0.52	0.00	0.188	81.35
ENSN	0.37	0.00	0.237	68.71
LOGEN	0.47	0.00	0.280	66.08
TOTALENERGY	0.07	0.00	0.251	70.00
F1	0.00	0.12	0.481	53.63
F2	0.01	0.00	0.472	55.73
F2F1	0.01	0.00	0.450	53.68
F3	0.00	0.03	0.472	51.87
F0	0.00	0.00	0.489	51.29
STM	1.00	0.00	0.186	81.35

ZCR	0.71	0.00	0.241	72.22
-----	------	------	-------	-------

## เสียงเจียบ-เสียงนาสิก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.10	0.00	0.165	84.33
En1	0.60	0.00	0.093	90.34
En2	0.20	0.00	0.179	78.87
En3	0.13	0.00	0.251	74.25
En4	0.07	0.00	0.323	66.03
En5	0.02	0.00	0.405	61.28
ENSN	0.11	0.00	0.350	78.21
LOGEN	0.81	0.00	0.148	84.63
TOTALENERGY	0.10	0.00	0.135	88.00
F1	0.17	0.00	0.208	72.75
F2	0.32	0.00	0.272	73.65
F2F1	0.51	0.00	0.165	83.97
F3	0.24	0.00	0.310	72.03
F0	0.57	0.00	0.134	88.06
STM	1.00	0.00	0.127	88.12
ZCR	0.12	0.00	0.256	78.09

## เสียงเจียบ-เสียงกึ่งสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.14	0.00	0.154	84.66
En1	0.54	0.00	0.106	87.06
En2	0.22	0.00	0.204	79.71
En3	0.15	0.00	0.257	73.96
En4	0.11	0.00	0.312	70.13
En5	0.04	0.00	0.408	59.90
ENSN	0.02	0.00	0.434	57.67
LOGEN	1.00	0.00	0.149	83.71

TOTALENERGY	0.14	0.00	0.109	86.74
F1	0.12	0.00	0.299	66.45
F2	0.26	0.00	0.325	64.86
F2F1	0.36	0.00	0.215	77.96
F3	0.27	0.00	0.310	67.89
F0	0.60	0.00	0.147	86.74
STM	0.99	0.00	0.134	85.78
ZCR	0.07	0.00	0.272	74.60

เทียบเทียบ-เทียบ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.04	0.00	0.290	70.84
En1	0.29	0.00	0.172	81.81
En2	0.28	0.00	0.175	82.09
En3	0.20	0.00	0.215	80.37
En4	0.23	0.00	0.224	76.91
En5	0.16	0.00	0.258	72.53
ENSN	0.11	0.00	0.322	67.91
LOGEN	0.54	0.00	0.191	77.45
TOTALENERGY	0.07	0.00	0.175	80.47
F1	0.02	0.00	0.421	57.44
F2	0.06	0.00	0.413	58.90
F2F1	0.03	0.00	0.414	57.13
F3	0.06	0.00	0.410	60.49
F0	0.03	0.00	0.444	55.84
STM	1.00	0.00	0.121	88.28
ZCR	0.09	0.00	0.370	62.62



## เสียงเงียบ-เสียงงvb

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.12	0.00	0.182	82.19
En1	0.66	0.00	0.123	87.60
En2	0.13	0.00	0.308	67.60
En3	0.04	0.00	0.391	61.15
En4	0.03	0.00	0.473	51.25
En5	0.03	0.00	0.500	52.08
ENSN	0.51	0.00	0.260	80.52
LOGEN	1.00	0.00	0.186	80.63
TOTALENERGY	0.13	0.00	0.148	86.15
F1	0.27	0.00	0.223	71.67
F2	0.50	0.00	0.304	72.60
F2F1	0.67	0.00	0.199	81.35
F3	0.32	0.00	0.342	69.27
F0	0.45	0.00	0.250	78.75
STM	0.96	0.00	0.164	84.69
ZCR	0.16	0.00	0.240	77.50

## เสียงเงียบ-เสียงสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.15	0.00	0.179	79.19
En1	0.43	0.00	0.114	85.60
En2	0.34	0.00	0.152	81.41
En3	0.23	0.00	0.193	81.41
En4	0.22	0.00	0.183	80.76
En5	0.11	0.00	0.258	73.17
ENSN	0.09	0.00	0.316	63.61
LOGEN	0.79	0.00	0.134	82.98
TOTALENERGY	0.16	0.00	0.141	83.12
F1	0.20	0.00	0.246	67.54

F2	0.25	0.00	0.283	72.38
F2F1	0.16	0.00	0.292	67.28
F3	0.26	0.00	0.262	72.77
F0	0.17	0.00	0.283	69.63
STM	1.00	0.00	0.107	88.87
ZCR	0.01	0.00	0.451	53.01

## เสียงกัก-เสียงสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.39	0.00	0.210	81.93
En1	1.00	0.00	0.161	85.90
En2	0.46	0.00	0.248	74.65
En3	0.16	0.00	0.349	64.95
En4	0.02	0.00	0.441	53.27
En5	0.00	0.00	0.469	51.37
ENSN	0.01	0.00	0.472	53.75
LOGEN	0.81	0.00	0.198	81.33
TOTALENERGY	0.44	0.00	0.188	82.96
F1	0.18	0.00	0.324	65.55
F2	0.31	0.00	0.305	67.73
F2F1	0.37	0.00	0.271	74.60
F3	0.35	0.00	0.284	71.31
F0	0.49	0.00	0.234	70.77
STM	0.76	0.00	0.215	79.70
ZCR	0.23	0.00	0.309	74.25

## เสียงvb-เสียงสระ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.20	0.00	0.146	86.30
En1	0.64	0.00	0.070	93.01
En2	0.52	0.00	0.092	87.40

En3	0.29	0.00	0.132	84.80
En4	0.21	0.00	0.163	80.06
En5	0.07	0.00	0.255	71.91
ENSN	0.32	0.00	0.118	83.58
LOGEN	0.50	0.00	0.113	89.08
TOTALENERGY	0.23	0.00	0.094	89.83
F1	0.28	0.00	0.142	84.45
F2	0.03	0.00	0.372	61.21
F2F1	0.04	0.00	0.253	70.06
F3	0.06	0.00	0.343	62.83
F0	0.02	0.00	0.432	58.96
STM	1.00	0.00	0.072	90.75
ZCR	0.15	0.00	0.222	75.32

## เสียงสระ-เสียงเสียดแทรก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.50	0.00	0.228	78.70
En1	0.87	0.00	0.195	80.91
En2	0.31	0.00	0.318	66.21
En3	0.05	0.00	0.431	52.97
En4	0.22	0.00	0.333	66.78
En5	0.65	0.00	0.245	73.89
ENSN	0.15	0.00	0.342	67.91
LOGEN	0.43	0.00	0.298	69.27
TOTALENERGY	0.27	0.00	0.297	71.39
F1	0.50	0.00	0.189	76.44
F2	0.94	0.00	0.194	79.88
F2F1	0.98	0.00	0.191	77.33
F3	0.90	0.00	0.205	77.71
F0	0.66	0.00	0.216	75.07
STM	0.66	0.00	0.261	73.28

ZCR	1.00	0.00	0.211	80.87
-----	------	------	-------	-------

## เสียงสระ-เสียงนาสิก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.22	0.00	0.275	72.28
En1	0.52	0.00	0.225	78.33
En2	0.69	0.00	0.183	80.34
En3	0.45	0.00	0.215	80.98
En4	0.34	0.00	0.226	76.26
En5	0.11	0.00	0.345	65.44
ENSN	0.60	0.00	0.180	84.03
LOGEN	0.31	0.00	0.287	67.92
TOTALENERGY	0.18	0.00	0.290	70.36
F1	0.12	0.00	0.317	68.72
F2	0.01	0.00	0.428	51.90
F2F1	0.08	0.00	0.340	68.30
F3	0.01	0.00	0.476	49.78
F0	0.00	0.00	0.496	50.22
STM	1.00	0.00	0.168	84.66
ZCR	0.26	0.00	0.294	70.53

## เสียงสระ-เสียงเย็บ

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.21	0.00	0.156	83.06
En1	0.50	0.00	0.129	84.96
En2	0.32	0.00	0.173	81.66
En3	0.22	0.00	0.216	78.83
En4	0.17	0.00	0.263	73.42
En5	0.07	0.00	0.331	65.31
ENSN	0.01	0.00	0.438	56.34
LOGEN	0.73	0.00	0.184	78.30

TOTALENERGY	0.16	0.00	0.161	82.12
F1	0.15	0.00	0.290	68.30
F2	0.31	0.00	0.281	70.42
F2F1	0.29	0.00	0.250	72.01
F3	0.30	0.00	0.290	69.82
F0	0.40	0.00	0.220	77.26
STM	1.00	0.00	0.138	85.92
ZCR	0.00	0.00	0.479	52.67

## เสียงสระ-เสียงกัก

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.64	0.00	0.254	76.58
En1	0.98	0.00	0.254	71.85
En2	0.31	0.00	0.374	62.39
En3	0.13	0.00	0.420	54.28
En4	0.01	0.16	0.534	53.38
En5	0.11	0.00	0.412	56.08
ENSN	0.17	0.00	0.381	59.23
LOGEN	0.66	0.00	0.342	64.19
TOTALENERGY	0.49	0.00	0.304	68.92
F1	0.26	0.00	0.286	73.42
F2	0.44	0.00	0.329	69.14
F2F1	1.00	0.00	0.269	73.65
F3	0.80	0.00	0.301	75.23
F0	0.98	0.00	0.225	75.45
STM	0.52	0.00	0.335	62.84
ZCR	0.73	0.00	0.288	68.69



## เสียงสระ-เสียงvb

	normalized F-ratio	P-value	ML Error	1st classification results
CORR	0.32	0.00	0.185	79.89
En1	0.64	0.00	0.146	85.71
En2	0.55	0.00	0.146	84.59
En3	0.42	0.00	0.153	84.02
En4	0.32	0.00	0.184	78.29
En5	0.15	0.00	0.254	71.62
ENSN	0.55	0.00	0.160	84.49
LOGEN	0.49	0.00	0.210	75.66
TOTALENERGY	0.32	0.00	0.173	83.93
F1	0.31	0.00	0.214	78.76
F2	0.04	0.00	0.396	57.80
F2F1	0.09	0.00	0.263	72.18
F3	0.07	0.00	0.385	61.65
F0	0.04	0.00	0.416	57.24
STM	1.00	0.00	0.132	86.94
ZCR	0.22	0.00	0.256	73.21

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศิรินุช บุญสุข เกิดเมื่อวันที่ 23 สิงหาคม พ.ศ. 2523 สำเร็จการศึกษาระดับปริญญาบัณฑิตในสาขาวิศวกรรมคอมพิวเตอร์จากคณะวิศวกรรมศาสตร์มหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2545



ศูนย์วิทยพัทยากร  
จุฬาลงกรณ์มหาวิทยาลัย