

การจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์

นางสาว บุศมาศ พลกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AUTOMATIC CLASSIFICATION OF THAI SYLLABLE-INITIAL STOP CONSONANTS WITH KNOWN  
PLACE OF ARTICULATION

Miss Budsamas Pholkul

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

Department of Computer Engineering

Faculty of Engineering

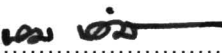
Chulalongkorn University

Academic Year 2007

Copyright of Chulalongkorn University

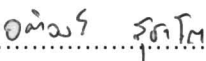
หัวข้อวิทยานิพนธ์	การจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดย
	อัตโนมัติเมื่อทราบฐานกรณ์
โดย	นางสาว บุศมาส พลกุล
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร. อติวงศ์ สุชาโต

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....  ..... คณบดีคณะวิศวกรรมศาสตร์  
(รองศาสตราจารย์ ดร. บุญสม เลิศศิริวงษ์)

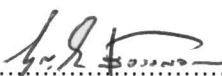
คณะกรรมการสอบวิทยานิพนธ์

.....  ..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร. ไปรดปราน บุญยพุกกณะ)

.....  ..... อาจารย์ที่ปรึกษา  
(อาจารย์ ดร. อติวงศ์ สุชาโต)

.....  ..... กรรมการภายนอกมหาวิทยาลัย  
(ผู้ช่วยศาสตราจารย์ ดร. ศรีวิมล มโนเชียวพินิจ)

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

.....  ..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ บุญชัย ไสววรรณนิชกุล)

บุคมาส พลกุล : การจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์. (AUTOMATIC CLASSIFICATION OF THAI SYLLABLE-INITIAL STOP CONSONANTS WITH KNOWN PLACE OF ARTICULATION) อ. ที่ปรึกษา : อ.ดร. อติวงศ์ สุชาติ, 94 หน้า.

วิทยานิพนธ์ฉบับนี้ได้ทำการศึกษาเกี่ยวกับการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ โดยอาศัยหลักการสวณสัทศาสตร์ มาหาค่าสวณลักษณะของพยัญชนะต้นเสียงกักภาษาไทย ซึ่งได้แก่ ความยาวช่วงกักลม ความยาวช่วงปล่อยลม ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำที่ 60 ถึง 300 เฮิรตซ์ ของสัญญาณเสียงช่วงกักลม ค่าพลังงานของช่วงปล่อยลม และจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลม โดยอาศัยวิธีการแบ่งเซกเมนต์เพื่อหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกักที่ถูกบันทึกในรูปแบบ ของ “สระ - พยัญชนะกัก - สระ” โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักต้นพยางค์ มีด้วยกันเป็น 8 เสียงหลัก บ ป พ ด ต ท ก ข และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ สระอะ อา อิ อี อุ อู แอะ แอ หลังจากนั้นจะนำสัญญาณเสียงช่วงกักลม และช่วงปล่อยลมที่ได้ไปสกัดค่าสวณลักษณะทั้ง 5 ค่า เพื่อใช้เป็นค่าคุณลักษณะของการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น ในการวิเคราะห์จะใช้วิธีวัดระยะห่างตำแหน่งระหว่างสัญญาณเสียงทดสอบกับสัญญาณเสียงต้นแบบตามวิธีของยุคลิด และผลที่ได้จากการจำแนกประเภทเสียงพยัญชนะกักแบ่งเป็น 3 ประเภท ได้แก่ เสียงโฆชะ เสียงอโฆชะแบบสถิต และเสียงอโฆชะแบบชนิด หลังจากนำสัญญาณเสียงของคนปกติไปทดสอบการจำแนกประเภทเสียงตามวิธีการที่ได้ศึกษา พบว่าความถูกต้องเฉลี่ยที่ได้มีค่าร้อยละ 83 89 และ 88 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ และเมื่อสัญญาณเสียงที่เข้ามาทดสอบการจำแนกประเภทเสียงเป็นเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ ผลการจำแนกประเภทเสียงที่ได้ตรงกับผลการจำแนกเสียงของผู้เชี่ยวชาญทางด้านสวณลักษณะถึงร้อยละ 71 จากผลดังกล่าวพบว่ามีค่าความถูกต้องมากกว่าผลการจำแนกประเภทเสียงของวิธีของอิดเดนมาร์คอฟถึงร้อยละ 18

ภาควิชา.....วิศวกรรมคอมพิวเตอร์..... ลายมือชื่อนิสิต..... บุคมาส พลกุล.....  
 สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์..... ลายมือชื่ออาจารย์ที่ปรึกษา..... อติวงศ์.....  
 ปีการศึกษา 2550

## 4971439521 : MAJOR COMPUTER SCIENCE

KEY WORD: ACOUSTIC MEASUREMENTS / AUTOMATIC CLASSIFICATION OF STOP CONSONANTS / ACOUSTIC PHONETICS / SPEECH PRODUCTION

BUDSAMAS PHOLKUL : AUTOMATIC CLASSIFICATION OF THAI SYLLABLE-INITIAL STOP CONSONANTS WITH KNOWN PLACE OF ARTICULATION. THESIS ADVISOR : ATIWONG SUCHATO, Ph.D., 94 pp.

The goal of this thesis is to create an automatic classification of Thai syllable-initial stop consonants with known place of articulation. The baseline algorithm makes use of acoustic-phonetic knowledge. This work focuses on acoustic discrimination among voiced, voiceless unaspirated and voiceless aspirated stop consonants with the same place of articulation. Region of interest associated with a stop consonant in the vowel-consonant-vowel ( $V_1C_sV_2$ ) where  $V_1$  was a schwa,  $C_s$  was one of the eight Thai stop consonants, /b/ /p/ /p<sup>h</sup>/ /d/ /t/ /t<sup>h</sup>/ /k/ and /k<sup>h</sup>/, and  $V_2$  was either one of eight Thai corner vowels. Since the interested region is in the middle of the speech, the segmentation algorithm is also applied for justifying the stop consonant part. Acoustic measurements based on the Voicing Onset Time (VOT), the Closure Duration, which is the duration from the point where the closure is formed to the point where it is released, the time-averaged Energy in 60-300 Hz frequency range of the speech signal in the closure interval ( $E_{ic}$ ), the Energy in the interval starting from the release to the point where voicing of  $V_2$  start ( $E_{re}$ ) and the Zero Crossing Rate in the Closure Duration were used to form the classification feature vectors of Linear Discriminant Analysis algorithm. Experiment reports were done on cleft lips and palates (CLP) utterances and non-CLP utterances. The classification accuracies of non-CLP tokens are 83%, 89% and 88% for labial, alveolar and velar cases respectively and the overall accuracy of CLP tokens is 71%, while the Hidden Markov Models algorithm can classify CLP data with only 53% which is 18% less than the classification accuracy from the acoustic-phonetic algorithm.

Department.....Computer Engineering..... Student's signature.....  
Field of study....Computer Science..... Advisor's signature.....  
Academic year 2007.....

## กิตติกรรมประกาศ

กิตติกรรมประกาศนี้ขออุทิศให้กับผู้ที่มีส่วนให้การช่วยเหลือจนสามารถทำวิทยานิพนธ์นี้สำเร็จไปได้ด้วยดี โดยเฉพาะอย่างยิ่งท่าน อาจารย์ ดร. อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ผู้ช่วยศาสตราจารย์ ดร. โปรตปราน บุญยพุกกณะ ที่ได้สละเวลาให้คำปรึกษาแนะนำแนวทางเกี่ยวกับงานวิจัยอย่างถี่ถ้วนมาจนเสร็จสมบูรณ์ และผู้วิจัยขอกราบขอบพระคุณท่านคณะกรรมการสอบวิทยานิพนธ์ที่ได้ให้คำแนะนำ ข้อคิดเห็น และแนวทางในการพัฒนางานวิจัยนี้

ขอขอบคุณ คุณ นันทนา ประชาฤทธิ์ภักดี นักรรณบำบัด จากโรงพยาบาลจุฬาลงกรณ์ ที่ให้คำแนะนำเกี่ยวกับลักษณะอาการปวดที่ผิดปกติของผู้ป่วยโรคปากแห้งเพดานโหว่ และช่วยเหลือในการบันทึกเสียงของผู้ป่วย ทำให้งานวิจัยนี้สำเร็จลุล่วงไปด้วยดี

ขอขอบคุณ นาย ปรีติ ประชาฤทธิ์ภักดี นาย ปริญา สงวนสัตย์ นางสาว ศิรินาถ ตั้งรวมทรัพย์ และ นาย ไพโรจน์ ลีลาภักดิ์ ที่ให้คำแนะนำและช่วยเหลือในการสร้างวิธีการจำแนกประเภทเสียงให้สำเร็จลุล่วงเป็นอย่างดี และขอขอบคุณ เพื่อนทุกคนที่ได้เสียสละเวลามาบันทึกเสียงพูด เพื่อใช้ในงานวิจัยฉบับนี้

ขอขอบคุณ จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ประสิทธิประสาทวิชาความรู้

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ที่คอยสนับสนุน และผลักดันให้ผู้วิจัยสามารถทำวิทยานิพนธ์นี้สำเร็จ

## สารบัญ

หน้า

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ .....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ .....	ญ
บทที่ 1 บทนำ.....	1
1.1. ความเป็นมาและความสำคัญของปัญหา.....	1
1.2. วัตถุประสงค์ของการวิจัย .....	2
1.3. ขอบเขตของการวิจัย .....	2
1.4. วิธีดำเนินการวิจัย .....	3
1.5. ประโยชน์ที่คาดว่าจะได้รับ .....	4
1.6. ผลงานที่ตีพิมพ์จากวิทยานิพนธ์.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	5
แนวคิดและทฤษฎี.....	5
2.1. ระบบเสียงของภาษาไทย .....	5
2.2. พยัญชนะกักตันพยางค์.....	7
2.3. สอนสัทศาสตร์.....	10
2.4. การแปลงฟูเรียร์อย่างรวดเร็ว .....	12
2.5. สเปกโตรแกรมของสัญญาณเสียงพูด .....	16
2.6. การแปลงข้อมูลให้เป็นบรรทัดฐาน .....	20
2.7. แบบจำลองฮิดเดนมาร์คคอฟ.....	20
2.8. วิธีการวิเคราะห์แบบดิสคริมีแนนต์เชิงเส้น .....	24
งานวิจัยที่เกี่ยวข้อง.....	30
บทที่ 3 การจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์เมื่อทราบฐาน กรณีโดยใช้สอนสัทศาสตร์.....	35
3.1. ขั้นตอนการนำเข้าสู่สัญญาณเสียง .....	35
3.2. ขั้นตอนหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกัก .....	36

3.3. ขั้นตอนการจำแนกประเภทเสียงโสมพะ อโสมพะ.....	47
บทที่ 4 การทดลองและผลการทดลอง.....	51
4.1. การประเมินความสามารถในการจำแนกประเภทเสียงพยัญชนะกักของค่าลักษณะ สำคัญทางสวณลักษณะที่ใช้.....	52
4.2. การทดลองวัดความคลาดเคลื่อนการหาช่วงกักลม และช่วงปล่อยลมของพยัญชนะ ต้นเสียงกักโดยอัตโนมัติ .....	59
4.3. การทดลองหาความถูกต้องของการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยใน ตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสวณสัทศาสตร์มา จำแนกประเภทเสียง.....	63
4.4. การทดลองหาความถูกต้องของการจำแนกประเภทเสียงพยัญชนะเสียงกักภาษาไทย ในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสวณสัทศาสตร์มา จำแนกประเภทเสียง โดยใช้สัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่เป็น สัญญาณทดสอบ .....	69
บทที่ 5 สรุปผลการวิจัย และข้อเสนอแนะ.....	74
5.1. สรุปผลการวิจัย.....	74
5.2. ข้อเสนอแนะ .....	78
รายการอ้างอิง.....	80
ภาคผนวก.....	82
ภาคผนวก ก ต้นแบบโปรแกรมการจำแนกเสียงพยัญชนะกักภาษาไทยโดยอัตโนมัติ.....	83
ภาคผนวก ข ผลงานตีพิมพ์.....	87
ประวัติผู้เขียนวิทยานิพนธ์.....	94



## สารบัญตาราง

ณ

ตาราง

หน้า

2.1	เสียงพยัญชนะกักตันพยางค์ในภาษาไทย .....	10
4.1	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานริมฝีปากเมื่อใช้ 4 ค่าลักษณะ สำคัญทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง .....	54
4.2	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานปุ่มเหงือกเมื่อใช้ 4 ค่า ลักษณะสำคัญทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง .....	55
4.3	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานเพดานอ่อนเมื่อใช้ 4 ค่า ลักษณะสำคัญทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง .....	56
4.4	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานริมฝีปากเมื่อใช้ 5 ค่าลักษณะ สำคัญทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง .....	57
4.5	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานปุ่มเหงือกเมื่อใช้ 5 ค่า ลักษณะสำคัญทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง .....	57
4.6	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานเพดานอ่อนเมื่อใช้ 5 ค่า ลักษณะสำคัญทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง .....	58
4.7	ค่าคลาดเคลื่อนเฉลี่ย และค่าคลาดเคลื่อนมากที่สุดของช่วงกักลม จากชุดข้อมูลสอน และชุดข้อมูลทดสอบ .....	61
4.8	ค่าคลาดเคลื่อนเฉลี่ย และค่าคลาดเคลื่อนมากที่สุดของช่วงปล่อยลม จากชุดข้อมูล สอน และชุดข้อมูลทดสอบ .....	62
4.9	ผลการจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อ ทราบฐานกรณ์ เมื่อทดสอบกับชุดข้อมูลสอน .....	67
4.10	ผลการจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อ ทราบฐานกรณ์ เมื่อทดสอบกับชุดข้อมูลทดสอบ .....	68
4.11	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยโดยอัตโนมัติเมื่อทราบฐานกรณ์ ตามวิธีของหลักการสวณศาสตร์ เมื่อใช้ชุดทดสอบจากสัญญาณเสียงของผู้ป่วย ในรูปแบบตารางค่าความสับสน .....	71
4.12	ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยเมื่อทราบฐานกรณ์ตามวิธีของฮิด เดนมาร์คอฟ เมื่อใช้ชุดทดสอบจากสัญญาณเสียงของผู้ป่วย ในรูปแบบตารางค่า ความสับสน .....	72

ภาพประกอบ	หน้า
2.1 ภาพฐานที่เกิดของเสียงพยัญชนะกักตันพยางค์.....	8
2.2 ภาพการสั่นของเส้นเสียงขณะออกเสียงพยัญชนะกักตันพยางค์แบบโฆชะ.....	9
2.3 ภาพการเปรียบเทียบการสั่นของฐานเส้นเสียง.....	9
2.4 ภาพแบบจำลองของแหล่งกำเนิดเสียงตัวกรองของกระบวนการสร้างเสียงพูดของมนุษย์.....	11
2.5 ภาพเปรียบเทียบระหว่างสัญญาณเสียงในโดเมนเวลา และสัญญาณเสียงในโดเมนความถี่.....	13
2.6 สเปกโตรแกรมของสัญญาณเสียงพูด อะบี.....	17
2.7 สเปกโตรแกรมแสดงตำแหน่งเซกเมนต์ของสัญญาณพยัญชนะเสียงกัก อะพะ.....	18
2.8 สเปกโตรแกรมแสดงความถี่สั้นพ้องของเซกเมนต์สระ.....	18
2.9 ภาพการสร้างแบบจำลองมาร์คอฟ.....	22
2.10 ตัวอย่างขั้นตอนการรู้จำของแบบจำลองฮิดเดนมาร์คอฟ.....	24
2.11 ภาพตัวอย่างการวิเคราะห์ข้อมูลตามหลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น.....	25
2.12 ภาพทิศทางการเวกเตอร์คุณลักษณะที่เกิดจากการเปลี่ยนรูปคลาสแบบไม่อิสระ.....	29
2.13 ภาพทิศทางการเวกเตอร์คุณลักษณะที่เกิดจากการเปลี่ยนรูปคลาสแบบอิสระ.....	30
3.1 ภาพขั้นตอนการจำแนกประเภทเสียงโฆชะ เสียงอโฆชะ แบบสถิติ และ เสียงอโฆชะแบบชนิดของพยัญชนะกักโดยอัตโนมัติ.....	35
3.2 ภาพขั้นตอนการหาดำแหน่งช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์.....	37
3.3 ภาพเซกเมนต์ $V_1C_sV_2$ โดยที่แผนภาพด้านบนเป็นสเปกโตรแกรม และแผนภาพด้านล่างเป็นคลื่นของพลังงาน.....	38
3.4 ภาพเปรียบเทียบสัญญาณเสียง กับสัญญาณค่าผลรวม.....	39
3.5 ภาพเปรียบเทียบสัญญาณค่าผลรวม กับสัญญาณค่าความแตกต่าง.....	39
3.6 ภาพเปรียบเทียบสัญญาณค่าความแตกต่าง กับสัญญาณจุดยอด.....	40
3.7 ภาพส่วนที่ตัดได้ของพยัญชนะกักตันพยางค์.....	40
3.8 ภาพขั้นตอนการพิจารณาตำแหน่งเซกเมนต์ตามสัญญาณจุดยอด.....	41
3.9 ภาพตำแหน่งบริเวณจุดยอดที่เป็นขอบของเซกเมนต์พยัญชนะกัก.....	42

ภาพประกอบ	หน้า
3.10 ภาพเปรียบเทียบพลังงานของ absLowLeft absHighLeft absLowRight และ absHighRight ของสัญญาณเสียงพยัญชนะกัก อะพะ.....	46
3.11 ภาพส่วนประกอบของขั้นตอนการจำแนกประเภทเสียงโฆษะ อโฆษะ .....	47
5.1 สเปกโตรแกรมสัญญาณเสียงพยัญชนะกัก อะบุ ของผู้ป่วยที่เป็นโรคปากแห้งเพดาน โห่ว.....	78
ก-1 หน้าแรกของโปรแกรมต้นแบบการฝึกพูดเสียงพยัญชนะกักภาษาไทย.....	83
ก-2 หน้าลงทะเบียนผู้ใช้งานใหม่ของโปรแกรมต้นแบบ .....	84
ก-3 หน้าเลือกประเภทการออกเสียงของโปรแกรมต้นแบบ .....	84
ก-4 หน้าเลือกการออกเสียงพยัญชนะกักในตำแหน่งต้นพยางค์.....	85
ก-5 หน้าฝึกการออกเสียงพยัญชนะกักฐานริมฝีปากของโปรแกรมต้นแบบ .....	85
ก-6 หน้าทดสอบการออกเสียงพยัญชนะกักฐานริมฝีปากของโปรแกรมต้นแบบ.....	86

# บทที่ 1

## บทนำ

### 1.1. ความเป็นมาและความสำคัญของปัญหา

เมื่อมนุษย์ต้องการติดต่อสื่อสารระหว่างกันสามารถทำได้หลายวิธี โดยวิธีที่สำคัญที่สุดวิธีหนึ่งคือ การใช้ภาษาพูด ทำให้มนุษย์เกิดความสนใจที่จะขัดเกลาเสียงพูด สำเนียงพูดเพื่อให้สามารถติดต่อสื่อสารได้ถูกต้อง และเพื่อความก้าวหน้าทางสังคมของตน หากสามารถนำความรู้ทางด้าน การจำแนกประเภทเสียงโดยอัตโนมัติมาช่วยพัฒนาโปรแกรมคอมพิวเตอร์ สำหรับฝึกพูด เพื่อให้ผู้ที่มีปัญหาทางด้าน การออกเสียงใช้งาน จะช่วยให้เกิดประโยชน์มหาศาลทั้งต่อตัวผู้ที่มีปัญหา และต่อสังคมที่ใช้ภาษาพูดในการติดต่อสื่อสาร

ในการฝึกพูดส่วนมากจะเริ่มฝึกตั้งแต่หน่วยเสียง ซึ่งประกอบไปด้วยอย่างน้อย เสียงพยัญชนะต้นพยางค์ และเสียงสระ โดยการฝึกเสียงพยัญชนะต้นพยางค์จะฝึกไปที่ละประเภทของอวัยวะที่ใช้ในการออกเสียง (ฐานกรณ์) และลักษณะการออกเสียง เพื่อให้เกิดทักษะการใช้ฐานกรณ์ในการพูด

ปัจจุบันนี้การจำแนกประเภทเสียงพยัญชนะต้นพยางค์โดยอัตโนมัติได้ถูกศึกษาวิจัยและพัฒนาอย่างต่อเนื่อง โดยเฉพาะเสียงพยัญชนะกักที่เป็นเสียงส่วนมากของเสียงพยัญชนะ แต่จากการศึกษาพบว่างานวิจัยส่วนใหญ่เป็นการจำแนกประเภทเสียงพยัญชนะกักภาษาต่างประเทศ โดยเฉพาะภาษาอังกฤษ [1] ซึ่งเป็นภาษาที่เน้นเสียงหนักเบา จังหวะเสียง และทำนองเสียง ยังมีได้ครอบคลุมเสียงพยัญชนะกักภาษาไทย ซึ่งเป็นภาษาที่มีเสียงวรรณยุกต์และไม่นั้นเสียงหนักเบา จังหวะเสียง และทำนองเสียง มากนัก ทำให้ยังไม่มีโปรแกรมคอมพิวเตอร์ฝึกพูดเสียงพยัญชนะกักภาษาไทย

ดังนั้นวิทยานิพนธ์นี้จึงศึกษาวิธีการจำแนกประเภทเสียงพยัญชนะกักภาษาไทย ในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เพื่อสร้างแนวทางในการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับฝึกพูด เพื่อให้ผู้ที่มีปัญหาในการออกเสียงพยัญชนะกักภาษาไทยใช้งาน โดยเฉพาะอย่างยิ่งกับเด็กที่ป่วยเป็นโรคปากแหว่งเพดานโหว่ในประเทศไทย ซึ่งโดยส่วนมากจะมีความผิดปกติในการแทนที่กัน ของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณ์เดียวกัน นอกจากนี้ยังสร้างแนวทางในการศึกษาความรู้ทางสัทศาสตร์ของผู้ป่วยโรคปากแหว่งเพดานโหว่ โดยนำเอาวิธีการจำแนกประเภทเสียงที่ศึกษาไปใช้กับเสียงของผู้ป่วย

## 1.2. วัตถุประสงค์ของการวิจัย

เพื่อศึกษาวิธีการจำแนกประเภทเสียง และนำเสนอต้นแบบขั้นตอนวิธีการจำแนกประเภทเสียงอย่างอัตโนมัติของสัญญาณเสียงโสมพะิม โสมพะิมแบบสถิต และโสมพะิมแบบอนิต ของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณ์เดียวกัน เพื่อใช้เป็นแนวทางในการพัฒนาโปรแกรมคอมพิวเตอร์ฝึกพูดการออกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์

## 1.3. ขอบเขตของการวิจัย

1.3.1 เนื่องจากงานวิจัยนี้มุ่งเน้นในการจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ หากสัญญาณเสียงที่เข้ามาเป็น พยัญชนะกักภาษาอื่น หรือเป็นพยัญชนะกักภาษาไทยแต่ไม่ได้อยู่ตำแหน่งต้นพยางค์ ค่าการจำแนกประเภทที่ได้ อาจมีความคลาดเคลื่อน

1.3.2 งานวิจัยนี้ไม่ได้ครอบคลุมงานหาค่าความคลาดเคลื่อนเมื่อสัญญาณเสียงที่เข้ามาไม่ได้อยู่ในกลุ่มของพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ หรือเป็นเสียงพยัญชนะกักต้นพยางค์แต่เปลี่ยนฐานกรณ์ของเสียงให้ไม่ตรงกับเงื่อนไขที่ระบุก่อนทำงานทดสอบการจำแนกประเภท

1.3.3 รูปแบบของสัญญาณเสียงที่เข้าจะอยู่ในรูป  $V_1C_sV_2$  กล่าวคือ  $V_1$  เป็นเสียงสระอะไม่เน้นพยางค์  $C_s$  เป็นเสียงของพยัญชนะกักต้นพยางค์ และ  $V_2$  เป็นเสียงของสระที่สามารถเป็นได้ 8 เสียง อันได้แก่ อะ อา อี อี อู อู แอะ แอ ถ้าหากสัญญาณเสียงที่เข้ามามีเกิน 3 หน่วยเสียงที่ต่อเนื่องกันการจำแนกประเภทสัญญาณเสียงนั้นอาจเกิดความคลาดเคลื่อนได้

1.3.4 ถ้าหากในส่วน  $C_s$  ของสัญญาณเกิดมีการเปลี่ยนแปลงเป็นจากเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์เป็นสัญญาณเสียงแบบอื่น เช่น จากเสียง บ เป็นเสียง ม ซึ่งเป็นเสียงที่ขึ้นจมูก ถ้าเกิดการเปลี่ยนแปลงของสัญญาณเสียงของพยัญชนะกักต้นพยางค์เป็นเสียงพยัญชนะต้นกลุ่มอื่น หรือเป็นเสียงพยัญชนะกักต้นพยางค์แต่ใช้ฐานกรณ์ของเสียงต่างไป การจำแนกประเภทจะยังสามารถทำได้ แต่ค่าการจำแนกประเภทที่ได้ อาจมีความคลาดเคลื่อน

1.3.5 การบันทึกสัญญาณเสียงที่ใช้เป็นฐานความรู้สวลลักษณะ (Acoustic Feature) และเป็นสัญญาณเสียงเข้าของขั้นตอนการจำแนกประเภทแบบอัตโนมัติจะต้องใช้วิธีการ

เดียวกัน นั่นคือบันทึกที่ความถี่ 44,100 เฮิรตซ์ และใช้ไมโครโฟนที่ติดหูฟังเพื่อให้ระยะห่างระหว่างปากกับไมโครโฟนน้อยที่สุด

1.3.6 สภาพแวดล้อมขณะบันทึกเสียง จะต้องเป็นสภาพแวดล้อมที่เงียบสงบ ปราศจากเสียงรบกวน เนื่องจากส่วนในการจำแนกประเภทเสียงใช้วิธีการจำแนกตามฐานความรู้ของสวณลักษณะที่ซึ่งสภาพแวดล้อมสามารถมีผลต่อค่าของสวณลักษณะได้ ดังนั้นถ้าหากเปลี่ยนจากสถานที่ที่เงียบสงบไปเป็นสถานที่ที่มีเสียงดังมารบกวน อาจทำให้เกิดความคลาดเคลื่อนของการจำแนกประเภทเสียงได้

1.3.7 งานวิจัยนี้เน้นเพื่อหาแนวทางการจำแนกประเภทเสียงแบบอัตโนมัติของเสียงโห่ระ อโห่ระแบบชนิด และ อโห่ระแบบสถิลของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณของเสียงเดียวกัน เพื่อใช้เป็นแนวทางช่วยเหลือผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่ หากไปประยุกต์ใช้กับผู้ป่วยประเภทอื่นเช่นผู้พิการทางสมอง ที่ไม่สามารถพูดต่อเนื่องได้ หรือผู้ที่พูดได้ไม่ต่อเนื่อง ติดขัด เป็นต้น จะทำให้ประสิทธิภาพในการจำแนกคลาดเคลื่อนได้

#### 1.4. วิธีดำเนินการวิจัย

1.4.1 สร้างชุดข้อมูลเสียงเพื่อใช้เป็นฐานความรู้ในการจำแนกประเภทเสียง

1.4.2 ศึกษาวิธีในการตัดสัญญาณเสียงออกเป็นเสียงช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์

1.4.3 ศึกษาวิธีในการจำแนกประเภทสัญญาณเสียงแบบอัตโนมัติของเสียงโห่ระ อโห่ระแบบชนิด และ อโห่ระแบบสถิลของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ ที่ใช้ฐานกรณเดียวกัน

1.4.4 ออกแบบขั้นตอนวิธีการจำแนกประเภทสัญญาณเสียงแบบอัตโนมัติ โดยนำความรู้ที่ได้จากการศึกษาการตัดสัญญาณเสียง และการจำแนกประเภทสัญญาณเสียงมาประยุกต์เป็นขั้นตอนการจำแนกประเภทสัญญาณเสียงแบบอัตโนมัติ

1.4.5 พัฒนาขั้นตอนการจำแนกประเภทสัญญาณเสียงแบบอัตโนมัติ ตามที่ได้ออกแบบ

1.4.6 ทดสอบความถูกต้องของขั้นตอนการจำแนกประเภทแบบอัตโนมัติ

1.4.7 พัฒนาต้นแบบการจำแนกประเภทเสียงอัตโนมัติของสัญญาณเสียง โห้ชะ อโห้ชะแบบสถิต และอโห้ชะแบบจลิต ของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้น พยางค์ ที่ใช้ฐานกรณเดียวกัน

1.4.8 สรุปผล และวิจารณ์ผลที่ได้

1.4.9 จัดทำวิทยานิพนธ์

## 1.5. ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 วิธีการตัดสัญญาณเสียงอัตโนมัติที่เข้ามาในรูปแบบ  $V_1C_sV_2$  กล่าวคือ  $V_1$  เป็นเสียงสระอะไม่เน้นพยางค์  $C_s$  เป็นเสียงของพยัญชนะกักต้นพยางค์ และ  $V_2$  เป็นเสียงของสระ ที่สามารถเป็นได้ 8 เสียง ดังที่ได้กล่าวไว้ข้างต้น

1.5.2 วิธีการจำแนกประเภทเสียงอัตโนมัติของสัญญาณเสียง โห้ชะ อโห้ชะ แบบสถิต และอโห้ชะแบบจลิต ของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ ที่ใช้ฐาน กรณเดียวกัน

1.5.3 ต้นแบบขั้นตอนการจำแนกประเภทเสียงอัตโนมัติของสัญญาณเสียง โห้ชะ อโห้ชะแบบสถิต และอโห้ชะแบบจลิต ของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้น พยางค์ ที่ใช้ฐานกรณเดียวกัน

1.5.4 แนวทางในการพัฒนาโปรแกรมคอมพิวเตอร์ฝึกพูดเสียงพยัญชนะกัก ภาษาไทยในตำแหน่งต้นพยางค์

## 1.6. ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตอบรับให้นำเสนอเป็นบทความวิชาการใน หัวข้อเรื่อง “Stop Consonant Voicing Classification for Computer-assisted Speech Training of Patients with Cleft Lips and Palates” โดยบุศมาส พลกุล โปรดปราน บุญยพุกกณะ อติวงศ์ สุขชาติ ในงานประชุมวิชาการ “International Convention for Rehabilitation Engineering & Assistive Technology (i-Create 2007)” ณ ประเทศ สิงคโปร์ ในระหว่างวันที่ 23-26 เมษายน 2550

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้จะนำเสนอทฤษฎีพื้นฐานและแนวคิดที่เกี่ยวข้องกับการพัฒนาวิธีการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์เมื่อทราบฐานกรณ์ ซึ่งเป็นทฤษฎีที่ศึกษาทางด้านของระบบเสียงของภาษาไทย พยัญชนะกักต้นพยางค์ และสวณศาสตร์ (Acoustic Phonetics)

จากนั้นจะนำเสนอทฤษฎีการวิเคราะห์เสียงพูดในโดเมนความถี่ด้วยการแปลงฟูเรียร์อย่างรวดเร็ว (Fast Fourier Transform) และสเปกโตรแกรม (Spectrogram) ของสัญญาณเสียงพูด รวมถึงวิธีการแปลงข้อมูลให้เป็นบรรทัดฐาน (Data Normalization) แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model) และวิธีการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น (Linear Discriminant Analysis) รวมทั้งเสนองานวิจัยต่างๆที่เกี่ยวข้องกับวิธีการแบ่งเสียงพยัญชนะต้นและวิธีการจำแนกประเภทเสียงพยัญชนะกัก

#### แนวคิดและทฤษฎี

##### 2.1. ระบบเสียงของภาษาไทย

เสียงในภาษาไทย ประกอบด้วย เสียงพยัญชนะ 21 หน่วยเสียง เสียงพยัญชนะควบกล้ำ 12 หน่วยเสียง เสียงสระเดี่ยว 18 หน่วยเสียง เสียงสระประสม 6 หน่วยเสียง และเสียงวรรณยุกต์ 5 หน่วยเสียง เมื่อนำหน่วยเสียงพื้นฐานทั้งหมดมาประสมพยางค์ตามหลักไวยากรณ์ จะได้ 26,928 พยางค์ [2], [3]

##### 2.1.1. เสียงพยัญชนะต้น

ภาษาไทยมีเสียงพยัญชนะต้นทั้งหมด 33 หน่วยเสียง ได้แก่ เสียงพยัญชนะต้นเดี่ยว 21 หน่วยเสียง และเสียงพยัญชนะควบกล้ำ 12 หน่วยเสียง ในบรรดา 21 หน่วยเสียง สามารถแบ่งกลุ่มได้ตามฐานกรณ์ของเสียง และลักษณะการออกเสียง

##### 2.1.1.1. ฐานกรณ์ของเสียง (Place of Articulation)

แบ่งเป็น 5 ฐาน ได้แก่ ฐานริมฝีปาก (Labial) ฐานปุ่มเหงือก (Alveolar) ฐานเพดานแข็ง (Palatal) ฐานเพดานอ่อน (Velar) และฐานเส้นเสียง (Glottal) เป็นต้น



2.1.1.2. **ลักษณะการออกเสียง** แบ่งเป็น 2 เสียง ได้แก่ เสียงกักลม และเสียงไม่กักลม ซึ่งเสียงกักลมสามารถแบ่งย่อยเป็น เสียงโฆษะ เสียงอโฆษะแบบ สติลิต และ เสียงอโฆษะแบบชนิด ส่วนเสียงไม่กักลมสามารถแบ่งได้เป็น เสียงนาสิก (Nasals) เสียงเสียดแทรก (Fricatives) เสียงร้ว (Trill) เสียงข้างลิ้น (Lateral) และเสียงเปิด (Approximants)

### 2.1.2. เสียงตัวสะกด

เสียงตัวสะกดในภาษาไทยประกอบด้วย 8 หน่วยเสียง มี 3 หน่วยเสียง เป็นเสียงกักลม ได้แก่ เสียง ป ท และ ก ส่วนอีก 5 หน่วยเสียงเป็นเสียงไม่กักลม ได้แก่ เสียงนาสิก 3 หน่วยเสียง ม น และ ง เสียงเปิด 2 หน่วยเสียง ย และ ว

### 2.1.3. เสียงสระ

เสียงสระในภาษาไทยประกอบด้วย เสียงสระเดี่ยว 18 หน่วยเสียง เสียงสระประสม 6 หน่วยเสียง

2.1.3.1. **สระเดี่ยว** (Monophthongs) เป็นสระเสียงแท้ มีทั้งสิ้น 18 หน่วยเสียง เป็นสระเสียงสั้น 9 หน่วยเสียง และสระเสียงยาว 9 หน่วยเสียง

2.1.3.2. **สระประสม** (Diphthongs) เป็นสระที่เกิดจากการออกเสียงประสมกันของสระแท้ มีทั้งสิ้น 6 หน่วยเสียง เป็นสระเสียงสั้น 3 หน่วยเสียง และสระเสียงยาว 3 หน่วยเสียง

### 2.1.4. เสียงวรรณยุกต์

เสียงวรรณยุกต์ คือเสียงที่เกิดจากการเปลี่ยนแปลงระดับเสียงของ พยางค์ ภาษาไทยเป็นภาษาวรรณยุกต์ (Tonal Language) เมื่อเปลี่ยนเสียงวรรณยุกต์แล้วทำให้ ความหมายของพยางค์หรือกลุ่มคำเปลี่ยนไป ถึงแม้พยางค์หรือกลุ่มคำจะมีพยัญชนะ และ สระ เหมือนกัน ซึ่งจะแตกต่างไปจากบางภาษา เช่น ภาษาอังกฤษ ที่ถึงแม้จะพูดเปลี่ยนเสียงวรรณยุกต์ ไป แต่ความหมายของพยางค์หรือกลุ่มคำยังคงไม่เปลี่ยนแปลง เสียงวรรณยุกต์ในภาษาไทยมี 5 หน่วยเสียง แบ่งได้เป็น 2 ประเภทตามลักษณะการเปลี่ยนแปลงของระดับเสียง

2.1.4.1. **เสียงวรรณยุกต์ที่มีระดับความถี่ค่อนข้างคงที่ตลอดพยางค์** (Static Tone) แบ่งเป็น เสียงวรรณยุกต์สามัญ (Mid Tone) เสียงวรรณยุกต์เอก (Low Tone) และเสียงวรรณยุกต์ตรี (High Tone)

2.1.4.2. **เสียงวรรณยุกต์เปลี่ยนระดับ** (Dynamic Tone) แบ่งเป็น เสียงวรรณยุกต์โท (Falling Tone) และเสียงวรรณยุกต์จัตวา (Rising Tone)

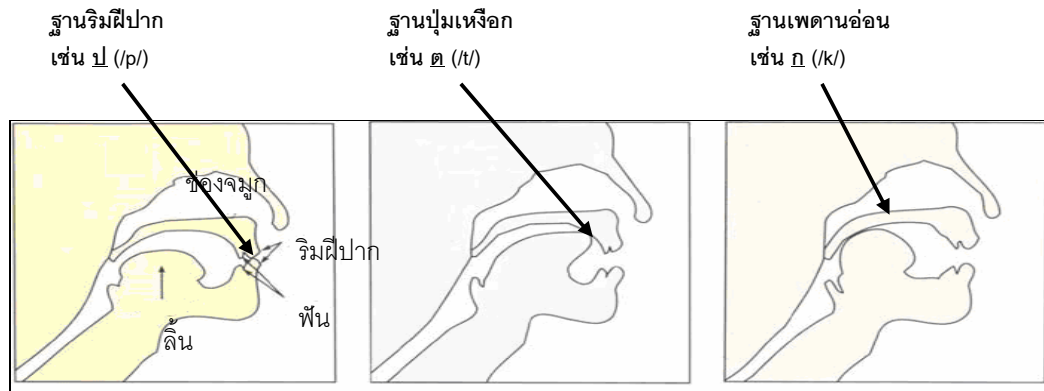
## 2.2. พยัญชนะกักตันพยางค์ (Stop Consonant)

พยัญชนะกักตันพยางค์ เกิดจากการกักลมไว้ ณ จุดใดจุดหนึ่งในช่องปากและขณะเดียวกันเพดานออกก็ยกตัวขึ้น ทำให้มีการปิดกั้นลมทางช่องจมูกด้วย เมื่ออวัยวะในการออกเสียงเคลื่อนที่ออกจากกัน ลมก็จะพุ่งออก และเกิดเป็นเสียงพยัญชนะกักขึ้นมา เพื่อให้เข้าใจถึงวิธีการออกเสียงพยัญชนะเสียงกัก ในหัวข้อนี้จะนำเสนอหลักสวณศาสตร์ของพยัญชนะกักตันพยางค์ และประเภทของเสียงพยัญชนะกักภาษาไทยตำแหน่งตันพยางค์

### 2.2.1. สวณศาสตร์ของพยัญชนะกักตันพยางค์

สวณศาสตร์ของพยัญชนะกักตันพยางค์ คือการศึกษาเกี่ยวกับลักษณะธรรมชาติของเสียงพยัญชนะกักตันพยางค์ รวมถึงการศึกษาในด้านการทำงานของอวัยวะต่างๆในการออกเสียงพยัญชนะกักตันพยางค์ โดยปกติแล้วการออกเสียงพยัญชนะกักตันพยางค์จะมีสวณลักษณะที่สำคัญดังต่อไปนี้

2.2.1.1. **ฐานกรณ์ที่ทำให้เกิดพยัญชนะกักตันพยางค์** (Place of Articulation) คือจุดที่ลมถูกกักหรือบีบก่อนที่จะถูกปล่อยให้ผ่านออกมาทางปากหรือช่องจมูก และช่องดังกล่าวนี้เกิดจากการที่กรณ หรืออวัยวะที่เคลื่อนที่ได้เคลื่อนที่ไปสัมผัสกับฐานหรืออวัยวะที่ใช้ในการออกเสียงไม่เคลื่อนที่ ดังรูปที่ 2.1 แสดงฐานกรณ์ของเสียงพยัญชนะกักตันพยางค์



รูปที่ 2.1 ภาพฐานที่เกิดของเสียงพยัญชนะกักตันพยางค์

[ดัดแปลงจาก [http://www.ling.upenn.edu/courses/Summer\\_2004/ling001/lecture2.html](http://www.ling.upenn.edu/courses/Summer_2004/ling001/lecture2.html) ]

ซึ่งจำแนกประเภทของฐานได้ดังต่อไปนี้

- **ฐานริมฝีปาก (Labial)**  
 ฐาน : ริมฝีปากบน  
 กรณ : ริมฝีปากล่าง  
 เช่นเสียง **บ** (/b/) **ป** (/p/) และ **พ** (/p<sup>h</sup>)
- **ฐานปุ่มเหงือก (Alveolar)**  
 ฐาน : ปุ่มเหงือก  
 กรณ : ปลายสุดลิ้น หรือ ลิ้นส่วนปลาย  
 เช่นเสียง **ด** (/d/) **ต** (/t/) และ **ท** (/t<sup>h</sup>)
- **ฐานเพดานอ่อน (Velar)**  
 ฐาน : เพดานอ่อน  
 กรณ : ลิ้นส่วนหลัง  
 เช่นเสียง **ก** (/k/) และ **ข** (/k<sup>h</sup>)

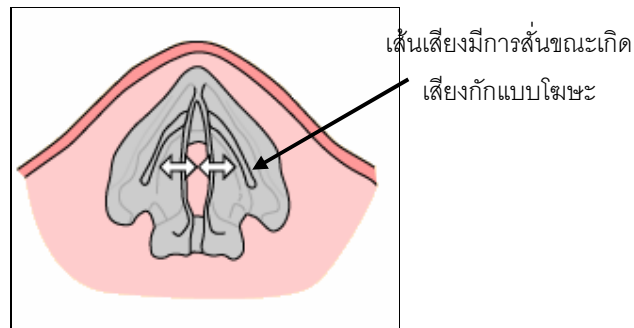
#### 2.2.1.2. ลักษณะการออกเสียงพยัญชนะกักตัน

พยางค์ แบ่งเป็น 3 ประเภท คือ

##### 2.2.1.2.1. พยัญชนะกักตันพยางค์

แบบโฆษะ (Voiced Stop Consonant Sound) คือพยัญชนะกักตันพยางค์ที่ ณ เวลาก็กลมไว้

เกิดการสั่นของเส้นเสียง (Vocal Fold) ทำให้มีพลังงานในช่วงความถี่ต่ำเกิดขึ้น ดังรูปที่ 2.2 พยัญชนะต้นเสียงกักแบบโฆชะ ได้แก่ บ ด เป็นต้น



รูปที่ 2.2 ภาพการสั่นของเส้นเสียงขณะออกเสียงพยัญชนะกักต้นพยางค์แบบโฆชะ

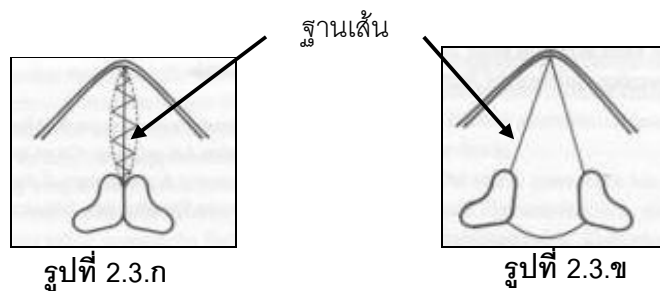
[ดัดแปลงจาก <http://hyperphysics.phy-astr.gsu.edu/hbase/music/voice.html> ]

#### 2.2.1.2.2. พยัญชนะกักต้นพยางค์

**แบบโฆชะแบบชนิด** (Voiceless Aspirated Stop Consonant Sound) คือพยัญชนะต้นเสียงกักที่ออกเสียงแล้วมีกลุ่มลมตามออกมา ดังรูปที่ 2.3.ก พยัญชนะกักต้นพยางค์แบบโฆชะแบบชนิด ได้แก่ พ ท ค เป็นต้น

#### 2.2.1.2.3. พยัญชนะกักต้นพยางค์

**แบบโฆชะแบบชนิด** (Voiceless Unaspirated Stop Consonant Sound) คือพยัญชนะต้นเสียงกักที่ออกเสียงแล้วไม่มีกลุ่มลมตามออกมา ดังรูปที่ 2.3.ข พยัญชนะกักต้นพยางค์แบบโฆชะแบบชนิด ได้แก่ ป ต ก เป็นต้น



รูปที่ 2.3 ภาพการเปรียบเทียบการสั่นของฐานเส้นเสียง (Glottis) เมื่อปล่อยลมออกของพยัญชนะกักต้นพยางค์โฆชะแบบชนิด (2.3.ก) และ พยัญชนะกักต้นพยางค์โฆชะแบบชนิด (2.3.ข)

## 2.2.2. ประเภทของเสียงพยัญชนะกักภาษาไทยต้นพยางค์

ภาษาไทยมีเสียงพยัญชนะกักต้นพยางค์ 8 หน่วยเสียง ดังตารางที่ 2.1 แสดงพยัญชนะกักต้นพยางค์ภาษาไทย

ตารางที่ 2.1 เสียงพยัญชนะกักต้นพยางค์ในภาษาไทย

ฐานกรณ์ของเสียง (Place Of Articulation)	ลักษณะการออกเสียง (Voicing Characteristic)		
	เสียงโฆษะ (Voiced)	เสียงอโฆษะแบบสลิกล (Voiceless Unaspirated)	เสียงอโฆษะแบบธนิต (Voiceless Aspirated)
ฐานริมฝีปาก (Labial)	<u>บ</u> (/b/)	<u>ป</u> (/p/)	<u>พ</u> (/p <sup>h</sup> /)
ฐานปุ่มเหงือก (Alveolar)	<u>ด</u> (/d/)	<u>ต</u> (/t/)	<u>ท</u> (/t <sup>h</sup> /)
ฐานเพดานอ่อน (Velar)		<u>ก</u> (/k/)	<u>ข</u> (/k <sup>h</sup> /)

จากตารางที่ 2.1 มีการแบ่งพยัญชนะกักภาษาไทยต้นพยางค์ตาม ฐานกรณ์ของเสียง และลักษณะการออกเสียง ได้ดังนี้

2.2.2.1. **ฐานริมฝีปาก (Labial)** สามารถแบ่งตามลักษณะการออกเสียงได้เป็น 3 หน่วยเสียง ได้แก่ บ (/b/) ป (/p/) และ พ (/p<sup>h</sup>/)

2.2.2.2. **ฐานปุ่มเหงือก (Alveolar)** สามารถแบ่งตามลักษณะการออกเสียงได้เป็น 3 หน่วยเสียง ได้แก่ ด (/d/) ต (/t/) และ ท (/t<sup>h</sup>/)

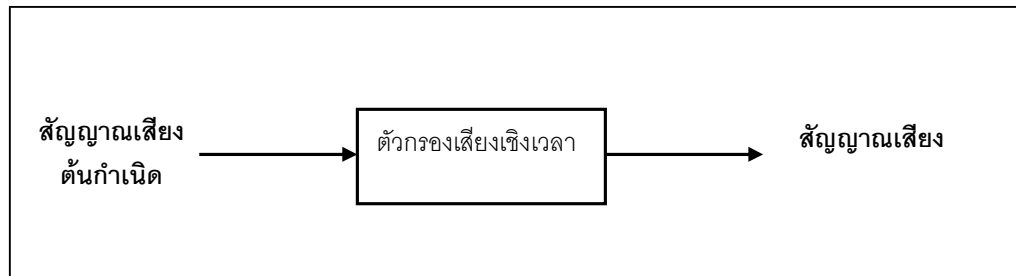
2.2.2.3. **ฐานเพดานอ่อน (Velar)** สามารถแบ่งตามลักษณะการออกเสียงได้เป็น 2 หน่วยเสียง ได้แก่ ก (/k/) และ ข (/k<sup>h</sup>/)

## 2.3. สอนสัทศาสตร์ (Acoustic-Phonetic Feature)

### 2.3.1. กระบวนการสร้างเสียงพูดของมนุษย์ (Human Speech Production)

เป็นที่ยอมรับกันดีว่าแบบจำลองของกระบวนการสร้างเสียงพูดของมนุษย์นั้นเป็นแบบจำลองของแหล่งกำเนิดเสียงตัวกรอง (Source-filter model) อันเกิดมาจากสัญญาณเสียงของมนุษย์นั้นไม่ได้ถูกปล่อยออกมาได้โดยตรง ก่อนที่สัญญาณจะปล่อยมาภายนอกนั้น สัญญาณจะต้องผ่านมาทางช่องปาก ที่เปรียบได้เสมือนเป็นช่องกรองสัญญาณ

เพื่อให้ได้สัญญาณตามความถี่ที่ฟังที่ต้องการ และเนื่องจากขณะที่มนุษย์ออกเสียงพูดนั้น อวัยวะตั้งแต่ปากจนถึงคอมีการเคลื่อนที่เปลี่ยนแปลงตามเวลาตลอดเวลา ฉะนั้นการกรองสัญญาณเสียงของมนุษย์จึงเป็นแบบการกรองเชิงเวลา ดังรูปที่ 2.4 แสดงภาพแบบจำลองของแหล่งกำเนิดเสียงตัวกรองของกระบวนการสร้างเสียงพูดของมนุษย์



รูปที่ 2.4 ภาพแบบจำลองของแหล่งกำเนิดเสียงตัวกรองของกระบวนการสร้างเสียงพูดของมนุษย์

ต้นกำเนิดของสัญญาณเสียงมนุษย์มีสองแบบ แบบแรกเป็นต้นกำเนิดแบบสัญญาณกึ่งคาบ (Quasi-periodic Signal) เกิดจากการสั่นของเส้นเสียงในขณะที่สัญญาณวิ่งผ่าน กับแบบหลังเป็นต้นกำเนิดแบบสัญญาณรบกวน เกิดขึ้นเมื่อสัญญาณเสียงวิ่งผ่านช่องกรองขนาดเล็กอย่างรวดเร็ว เสียงที่ได้จากช่องทางเสียงจึงมีลักษณะเป็นสัญญาณรบกวนเช่นเดียวกับต้นกำเนิด แต่ขนาดของสเปกตรัมที่ความถี่สูงจะถูกขยายขึ้นมาก รูปร่างของสเปกตรัมนี้สามารถนำมาวิเคราะห์เพื่อประมาณเหตุการณ์ของช่องเสียงผู้พูดได้

### 2.3.2. ลักษณะสำคัญทางสวณลักษณะ (Acoustic Feature)

ลักษณะสำคัญทางสวณลักษณะโดยทั่วไปเกิดจากการวิเคราะห์คุณสมบัติความถี่ของเสียงตามลักษณะของแหล่งกำเนิดเสียง (Source Characteristics) ลักษณะการออกเสียง (Manner of Articulation) และตำแหน่งของฐานกรณ์ในการออกเสียง (Place of Articulation)

#### 2.3.2.1. คุณสมบัติความถี่ของเสียง เสียงพูดของ

มนุษย์เมื่ออากาศถูกดันออกมาจากปอดด้วยแรงดันที่มากจะทำให้เส้นเสียงสั่น ทำให้สัญญาณที่เกิดจากแหล่งกำเนิดเสียงมีลักษณะเป็นคาบ ส่วนสัญญาณเสียงที่ไม่ได้เกิดจากการสั่นของเส้นเสียงจะมีลักษณะไม่เป็นคาบ

### 2.3.2.2. ลักษณะการออกเสียง สามารถพิจารณาจาก

ลักษณะของช่องทางเดินเสียงว่ามีการเปิด ปิดอย่างไร มีการกักเสียงไว้มากน้อยแค่ไหน และการออกเสียงนั้นอากาศไหลผ่านบริเวณช่องปากหรือผ่านไปในช่องโพรงจมูกบ้างหรือไม่ เสียงที่ผ่านช่องปากไปโดยไม่ได้ถูกกักเอาไว้เพียงพอต่อการสร้างเสียงรบกวนหรือกักการไหลของอากาศจะเรียกว่าเสียงที่มีเสียงสั้น (Sonorant) ซึ่งประกอบไปด้วยเสียงสระ (Vowels) เสียงกึ่งสระ (Semi-vowels) และเสียงนาสิก (Nasals) ส่วนเสียงที่ไม่มีคุณสมบัติความเป็นเสียงสั้น ได้แก่ เสียงพยัญชนะกัก (Stop Consonants) และเสียงเสียดแทรก (Fricatives)

### 2.3.2.3. ตำแหน่งของฐานในการออกเสียง ในกรณี

ของเสียงพยัญชนะเสียงกักและเสียงพยัญชนะเสียดแทรกจะพิจารณาจากตำแหน่งฐานที่ใช้เช่น ฟัน ลิ้น หรือริมฝีปากในการกั้นไม่ให้อากาศไหลผ่านช่องปากออกไปได้โดยตรง ในกรณีของเสียงสระจะพิจารณาจากตำแหน่งของลิ้นในขณะที่อากาศเดินทางผ่านช่องปากออกไป

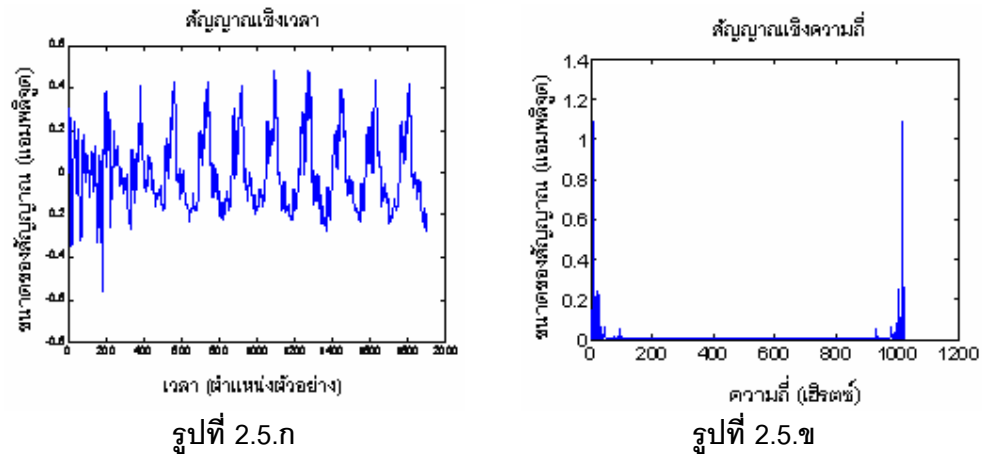
## 2.4. การแปลงฟูรีเยร์อย่างรวดเร็ว (Fast Fourier Transform)

เสียงเมื่อเดินทางผ่านอากาศมาที่อวัยวะรับการได้ยินของมนุษย์ คือ หู ซึ่งเป็นระบบเปิด ที่มีความไวต่อความดันอากาศมาก มนุษย์จะรู้สึกได้ยินก็ต่อเมื่อ มีการนำสัญญาณนั้นผ่านกระบวนการได้ยินในหูชั้นต่างๆ สามารถอธิบายเสียงที่ได้ยินในเชิงสมการคณิตศาสตร์ได้สองลักษณะหลัก คือ ฟังก์ชันทางเวลา (Time Domain) และฟังก์ชันทางความถี่ (Frequency Domain)

ฟังก์ชันทางเวลา สามารถอธิบายในลักษณะความดังของเสียง ในช่วงเวลาที่เปลี่ยนแปลงไปเท่านั้น แต่ไม่สามารถบอกได้ในเชิงความถี่ว่ามีความถี่ต่ำหรือสูง ในขณะที่ฟังก์ชันทางความถี่สามารถอธิบายลักษณะของเสียงว่ามีลักษณะสูงหรือต่ำได้ ซึ่งเป็นข้อมูลที่สำคัญในการเข้าถึงธรรมชาติเสียงนั้น

โดยธรรมชาติ หูของมนุษย์จะทำการแปลงสัญญาณเสียงให้อยู่ในรูปของแถบความถี่เสียง นั่นคือ แอมพลิจูด (Amplitude) และความถี่ โดยการแปลงแบบนี้สามารถทำได้ โดยอาศัยหลักการทางคณิตศาสตร์ที่เรียกว่าการแปลงแบบฟูรีเยร์ (Fourier Transform) ก็คือการแปลงสัญญาณในโดเมนเวลาให้อยู่ในโดเมนของความถี่นั่นเอง และสัญญาณเสียงที่บันทึกทั่วไปนั้นจะอยู่ในรูปของโดเมนเวลา (Time Domain) โดยตัวอย่างของการแปลงฟูรีเยร์เป็นไปตาม

รูปที่ 2.5 แสดงภาพเปรียบเทียบระหว่างสัญญาณเสียงในโดเมนเวลา และสัญญาณเสียงในโดเมนความถี่



รูปที่ 2.5 ภาพเปรียบเทียบระหว่างสัญญาณเสียงในโดเมนเวลา (2.5.ก) และสัญญาณเสียงในโดเมนความถี่ (2.5.ข)

การแปลงแบบฟูเรียร์นั้นประกอบด้วย

- การแปลงฟูเรียร์แบบต่อเนื่อง (Fourier Transform) ใช้วิธีการคำนวณโดยการอินทิเกรต
- การแปลงฟูเรียร์แบบวิฤต (Discrete Fourier Transform) ใช้วิธีการหาผลบวกแทนการอินทิเกรต

การแปลงฟูเรียร์อย่างรวดเร็ว (Fast Fourier Transform หรือ FFT) นั้นเป็นขั้นตอนวิธีหนึ่งของการแปลงฟูเรียร์แบบวิฤตที่ใช้เวลาน้อย ซึ่งถูกคิดค้นขึ้นโดย James W. Cooley และ John W. Turkey [4] เมื่อปี ค.ศ. 1965 การแปลงฟูเรียร์อย่างรวดเร็วนี้มีขั้นตอนการแปลงเริ่มจากนิยามการแปลงฟูเรียร์แบบวิฤตซึ่งกำหนดให้

$$x[k] = \sum_{n=0}^{N-1} x[n]W_N^{kn} \quad (2.1)$$

เมื่อ  $k = 0, 1, \dots, N-1$



โดย  $W_n = e^{-j(2\pi/N)}$  และ  $N$  คือขนาดจำนวนข้อมูลสัญญาณที่ถูกสุ่ม

แยกพจน์ทางขวามือของฟังก์ชัน (2.1) ออกเป็น 2 ส่วน

$$x[k] = \sum_{n=0}^{(N/2)-1} x[n]W_N^{kn} + \sum_{n=N/2}^{N-1} x[n]W_N^{kn} \quad (2.2)$$

เมื่อ  $k = 0, 1, \dots, N-1$

เปลี่ยนตัวแปร  $n$  ของพจน์ที่สองของฟังก์ชัน (2.2) ใหม่จะได้ว่า

$$x[k] = \sum_{n=0}^{(N/2)-1} x[n]W_N^{kn} + \sum_{n=0}^{(N/2)-1} x\left[n + \frac{N}{2}\right]W_N^{k\left(n + \frac{N}{2}\right)} \quad (2.3)$$

เมื่อ  $k = 0, 1, \dots, N-1$

จัดรูปฟังก์ชัน (2.3) ใหม่ได้เป็น

$$x[k] = \sum_{n=0}^{(N/2)-1} x[n]W_N^{kn} + \sum_{n=0}^{(N/2)-1} x\left[n + \frac{N}{2}\right]W_N^{kn}W_N^{kN/2} \quad (2.4)$$

เนื่องจากพจน์  $W_n^{kN/2} = e^{-j(2\pi k/N)N/2} = e^{-j\pi k} = (-1)^k$

ดังนั้นจะได้ว่า

$$x[k] = \sum_{n=0}^{(N/2)-1} \left\{ x[n] + (-1)^k x\left[n + \frac{N}{2}\right] \right\} W_N^{kn} \quad (2.5)$$

เมื่อ  $k = 0, 1, \dots, N-1$

จากนั้นทำการแบ่งแยกฟังก์ชันออกเป็น 2 ส่วนคือ ฟังก์ชันที่เป็นจำนวนคู่ และฟังก์ชันที่เป็นจำนวนคี่ ได้ฟังก์ชันใหม่เป็น

$$x[2k] = \sum_{n=0}^{(N/2)-1} \left\{ x[n] + x\left[n + \frac{N}{2}\right] \right\} W_N^{2kn} \quad (2.6)$$

เมื่อ  $k = 0, 1, \dots, \frac{N}{2} - 1$  และ

$$x[2k + 1] = \sum_{n=0}^{(N/2)-1} \left\{ x[n] - x\left[n + \frac{N}{2}\right] \right\} W_N^{(2k+1)n} \quad (2.7)$$

เมื่อ  $k = 0, 1, \dots, \frac{N}{2} - 1$

$$\begin{aligned} \text{เนื่องจาก } W_N^2 &= W_{N/2} \text{ เพราะ } W_N^2 = e^{-2j(2\pi/N)} \\ &= e^{-j2\pi/(N/2)} = W_{N/2} \end{aligned}$$

ดังนั้นเมื่อแทนค่า  $W_N^2 = W_{N/2}$  ในฟังก์ชัน (2.6) และ (2.7) จะจัดรูปใหม่ได้เป็น

$$x[2k] = \sum_{n=0}^{(N/2)-1} \left\{ x[n] + x\left[n + \frac{N}{2}\right] \right\} W_{N/2}^{kn} \quad (2.8)$$

เมื่อ  $k = 0, 1, \dots, \frac{N}{2} - 1$  และ

$$x[2k + 1] = \sum_{n=0}^{(N/2)-1} \left\{ x[n] - x\left[n + \frac{N}{2}\right] \right\} W_N^n W_{N/2}^{kn} \quad (2.9)$$

เมื่อ  $k = 0, 1, \dots, \frac{N}{2} - 1$

จะเห็นได้ว่า การแปลงข้อมูล  $N$  จำนวน สามารถทำได้โดยคำนวณเพียง  $N/2$  จำนวน ของพจน์ที่เป็นจำนวนคู่ และจำนวนคี่ จากนั้นนำผลของทั้งสองพจน์มารวมเข้าด้วยกัน ทำให้ได้ค่าข้อมูลการแปลง  $N$  จำนวน

หาก  $N$  เป็นเลขยกกำลังของ 2 ก็ทำซ้ำไปเรื่อยๆ จนได้ถึงค่าฐาน ทำการรวมไปเรื่อยๆ แบบ Recursive สุดท้ายจะได้ผลลัพธ์ของการแปลง จากหลักการดังกล่าวพบว่ามีวิธีการ

คล้ายกับขั้นตอนของวิธีแบ่งต่อสู้และเอาชนะ (Divide and Conquer) ซึ่งจะใช้เวลาในการคำนวณเพียง  $O(N \log N)$  มีค่าน้อยกว่าขั้นตอนวิธีแบบธรรมดา ที่ใช้เวลาถึง  $O(N^2)$

เนื่องจากวิทยานิพนธ์นี้ต้องการจำแนกประเภทเสียงระหว่าง เสียงโหมะะ เสียงอโหมะะแบบสถิต และแบบธิตของพยัญชนะกักตันพยางค์ภาษาไทย จากผลการทดลองของวิทยานิพนธ์นี้พบว่า ณ ช่วงกักลม (Closure Duration) ของเสียงโหมะะ มีพลังงานที่ความถี่ต่ำ ในขณะที่เสียงอโหมะะ ไม่มี หรือมีพลังงานน้อยมากในช่วงกักลมดังกล่าว ทำให้เลือกใช้วิธีการแปลงฟูเรียร์อย่างรวดเร็วมาเป็นวิธีการหนึ่งในการหาลักษณะสำคัญ เพื่อช่วยจำแนกเสียงโหมะะของพยัญชนะกักตันพยางค์

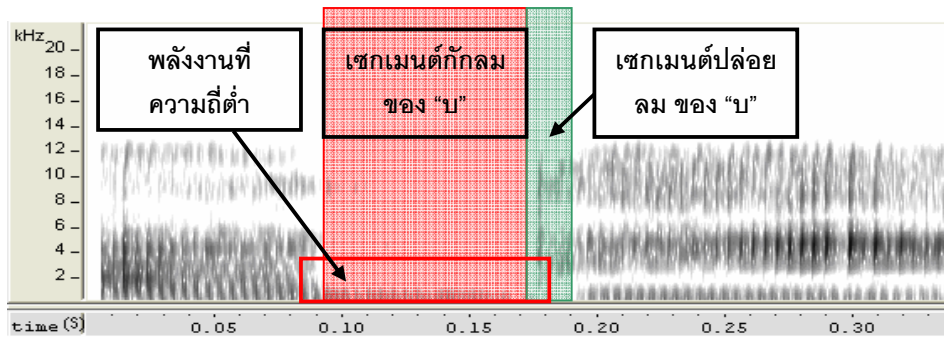
## 2.5. สเปกโตรแกรมของสัญญาณเสียงพูด [5]

สัญญาณเสียงมีลักษณะรูปร่างเป็นคลื่นที่สั้นแกว่งไปมา ดังนั้นการอ่านหน่วยเสียงในรูปแบบของคลื่นในโดเมนเวลาจึงเป็นเรื่องที่ทำได้ยาก ทำให้เกิดการวิเคราะห์หน่วยเสียงในโดเมนความถี่แทน หลังจากวิเคราะห์ในโดเมนความถี่จะได้สเปกโตรแกรมซึ่งสามารถนำไปสกัดหาค่าสวนลักษณะของพยัญชนะกักตันพยางค์ได้

ช่วงความถี่ที่มนุษย์สามารถได้ยินจะอยู่ระหว่าง 20 – 20,000 เฮิรตซ์ (20 กิโลเฮิรตซ์) มนุษย์ไม่สามารถได้ยินความถี่สูงเกินซึ่งเกิดขึ้นที่ความถี่ต่ำกว่า 20 ครั้งต่อวินาที และไม่สามารถรับรู้ความถี่ที่สูงกว่า 20 กิโลเฮิรตซ์ เสียงคำพูดจะประกอบด้วยพลังงานที่ระดับความถี่ต่างๆในช่วงที่มนุษย์สามารถได้ยินได้ โดยที่เสียงทั้งหมดจะอยู่ที่ระดับต่ำกว่า 8,000 เฮิรตซ์

การนำเทคนิคทางคณิตศาสตร์ที่เรียกว่าการวิเคราะห์ฟูเรียร์มาใช้กับรูปแบบคลื่นคำพูดที่มีความถี่ใดที่เกิดขึ้นในเวลาต่างกันในสัญญาณเสียงพูด ผลจากการทำการวิเคราะห์ฟูเรียร์จะได้ สเปกตรัม (Spectrum) หลังจากคำนวณสเปกตรัมสำหรับช่วงเวลาสั้นๆ (5-20 มิลลิวินาที) ของคำพูดต่อไปเรื่อยๆจนสิ้นสุดรูปแบบของคลื่นโดยทั่วไป สเปกตรัมที่อยู่ติดกันจะเปลี่ยนแปลงอย่างช้าๆและราบรื่น ซึ่งให้เห็นถึง การเคลื่อนไหวอย่างช้าๆของเสียงเทียบกับช่วงเวลาทีวิเคราะห์

การวิเคราะห์ดังกล่าวเรียกว่า การวิเคราะห์ฟูเรียร์ในช่วงเวลาสั้น (Short-time Fourier analysis) ซึ่งชุดของสเปกตรัมที่ได้มาจากช่วงเวลาต่างๆ นั้น สามารถนำไปแสดงผลได้ในรูปแบบของสเปกโตรแกรม ดังตัวอย่างตามรูปที่ 2.6 สเปกโตรแกรมของสัญญาณเสียงพูด อะบิ /abi/



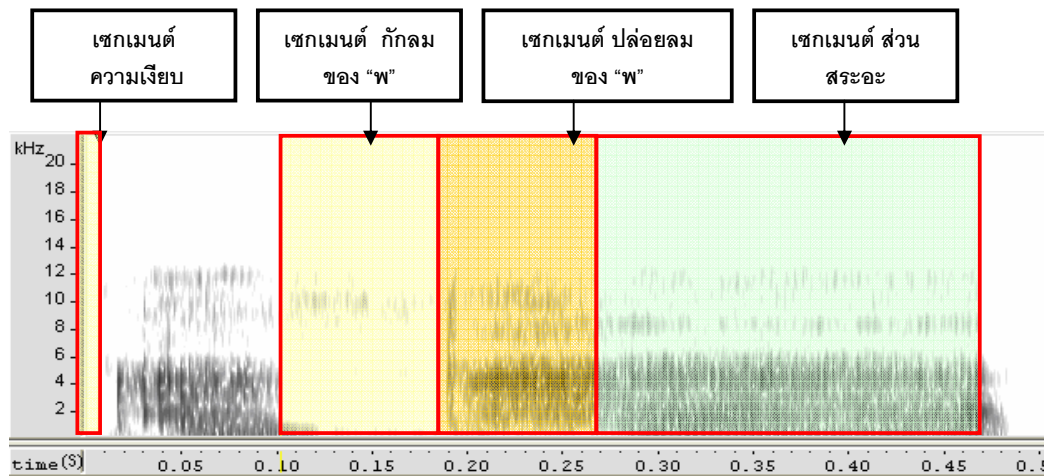
รูปที่ 2.6 สเปกโตรแกรมของสัญญาณเสียงพูด อะบี /abi/

สเปกโตรแกรมคำนวณจากรูปแบบคลื่นของคำพูด แกนตั้งของสเปกโตรแกรมแสดงความถี่โดยที่ระดับฐานจะเท่ากับ 0 เฮิรตซ์ ช่วงความถี่ที่แสดงจะเพิ่มขึ้นทีละ 2,000 เฮิรตซ์ แกนนอนแสดงถึงเวลา การเลื่อนไปทางขวาตามแกนนอน แสดงถึงสเปกตรัมตามเวลาที่เพิ่มขึ้น สเปกโตรแกรมจะถูกคำนวณค่าพลังงานของเสียงและเก็บไว้ในอาเรย์ขนาดสองมิติ สำหรับสเปกโตรแกรม  $S$  ใดๆ ความแรงของสัญญาณความถี่  $f$  ที่เวลา  $t$  ในสัญญาณเสียงพูดจะแสดงโดยความเข้มหรือสีในกราฟที่จุด  $S(t, f)$

การอ่านสเปกโตรแกรมจะใช้พื้นฐานความรู้ทางด้านการออกสัญญาณเสียงพูด เพื่อแบ่งแยกลำดับของสัญญาณเสียงที่ส่งเข้ามา โดยวิเคราะห์จากสเปกโตรแกรมของสัญญาณเสียงพูด โดยขั้นตอนการแปลความหมายของสเปกโตรแกรมมีดังนี้

2.5.1. พิจารณาขอบเขต (Boundary) ของสัญญาณเสียงตามที่เห็นในสเปกโตรแกรม ตามตำแหน่งทางเวลาที่เกิดความไม่ต่อเนื่องนั้นๆ

2.5.2. พิจารณาหา เซกเมนต์ (Segment) ย่อยของสัญญาณเสียงที่ได้จากขอบเขต โดยทั่วไป 1 เซกเมนต์จะอยู่ระหว่างขอบเขต 2 ขอบเขตที่อยู่ติดกัน ยกเว้นเซกเมนต์ที่เกิดจากเสียงพยัญชนะกัก และเสียงกึ่งเสียดแทรก (Affricate) ที่ 1 เซกเมนต์นั้น มักจะมีขอบเขตอีก 1 ขอบเขตอยู่ในระหว่างเซกเมนต์ ขอบเขตนี้สอดคล้องกับความไม่ต่อเนื่องจากการปล่อยช่องปิด ในกระบวนการกำเนิดเสียงทั้งสองประเภท ดังตัวอย่างตามรูปที่ 2.7 สเปกโตรแกรมแสดงตำแหน่งขอบเขต และส่วนของเซกเมนต์ของสัญญาณเสียงพยัญชนะเสียงกัก อะพะ /apha/

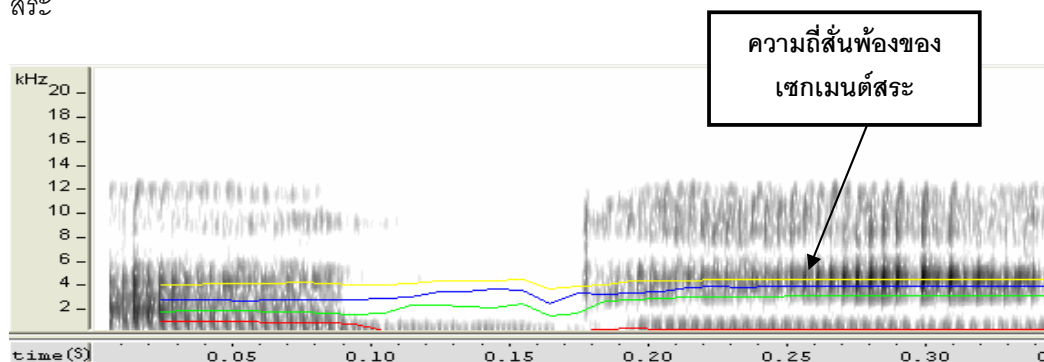


รูปที่ 2.7 สเปกโตรแกรมแสดงตำแหน่งเซกเมนต์ของสัญญาณพยัญชนะเสียงกัก อะพะ /apha/

2.5.3. พิจารณาประเภทเซกเมนต์ จากเซกเมนต์ที่ได้ โดยวิธีการกำหนดประเภทของเซกเมนต์ มีดังนี้

2.5.3.1. เซกเมนต์ความเงียบ ไม่ได้เป็นสัญญาณเสียงโดยตรง แต่เป็นส่วนหนึ่งของทุกๆสัญญาณเสียงพูด ไม่มีพลังงานปรากฏให้เห็นใน สเปกโตรแกรมตลอดช่วงเวลาที่ทั้งหมดดังรูปที่ 2.7

2.5.3.2. เซกเมนต์สระ จะมีพลังงานสูง มีการไหลของกระแสลมที่ทำให้เกิดสัญญาณเสียงมากที่สุด อีกทั้งจะเห็นโครงสร้างของกลุ่มความถี่สั้นพ้อง (Formant) ได้ชัดเจน และเป็นสัญญาณที่มีลักษณะเป็นคาบ (Periodic) สามารถสังเกตได้จากรูปร่างของคลื่นเสียงทางเวลา ดังรูปที่ 2.8 แสดงสเปกโตรแกรมแสดงความถี่สั้นพ้องของเซกเมนต์สระ



รูปที่ 2.8 สเปกโตรแกรมแสดงความถี่สั้นพ้องของเซกเมนต์สระ

2.5.3.3. **เซกเมนต์พญูชนะเสียงกัก** ดังรูปที่ 2.6 จะประกอบไปด้วยเซกเมนต์ย่อย 2 เซกเมนต์ คือ

2.5.3.3.1. **เซกเมนต์กักลม** ซึ่งจะเป็นบริเวณที่มีพลังงานหายไป เนื่องจากการสร้างช่องปิด แต่อาจมีพลังงานที่ความถี่ต่ำอันเกิดจากการสั่นของเส้นเสียงในขณะที่เกิดช่องปิดในกรณีของเสียงพญูชนะกักแบบโฆชะ โดยพลังงานที่ส่งออกมาในอากาศนี้ ผ่านออกมาจากการแผ่รังสีจากกระพุ้งแก้ม มิได้เกิดจากช่องปาก

2.5.3.3.2. **เซกเมนต์ปล่อยลม** ในขณะที่ช่องปิดถูกปล่อยออกอย่างรวดเร็ว อาจเกิดเสียงรบกวนสั้นๆ ทำให้เกิดพลังงานที่มีรูปร่างยาวออกไปทางแนวตั้งในสเปกโตรแกรม หรือก็คือแถบพลังงานในช่วงเซกเมนต์ปล่อยลม

2.5.3.4. **เซกเมนต์พญูชนะเสียงเสียดแทรก** ลักษณะของสัญญาณจะเป็นเสียงรบกวนที่เกิดจากกระแสลมถูกขับผ่านช่องแคบ พลังงานของเสียงรบกวนนี้จะหนาแน่นที่ช่วงความถี่ใดนั้น ขึ้นอยู่กับตำแหน่งของช่องแคบที่ใช้ในการสร้างเสียงเสียดแทรกนั้นๆ

2.5.3.5. **เซกเมนต์พญูชนะเสียงกึ่งเสียดแทรก** จะมีลักษณะในสเปกโตรแกรมเช่นเดียวกับ ลักษณะที่เกิดจากการสร้างช่องปิด เพื่อเตรียมเปล่งเสียงพญูชนะเสียงกัก แล้วตามด้วยเสียงเสียดแทรก หลังจากที่ช่องปิดถูกปล่อยออก

2.5.3.6. **เซกเมนต์พญูชนะเสียงนาสิก** พลังงานในช่วงความถี่กลางถึงสูงจะลดต่ำลงจากระดับพลังงานของสระที่อยู่ใกล้เคียงเล็กน้อย เนื่องจากการสูญเสียพลังงานในโพรงจมุก

2.5.3.7. **เซกเมนต์เสียงกึ่งสระ** มีการเปลี่ยนแปลงของลักษณะโครงสร้างของความถี่สั้นพ้องที่รวดเร็วและมากกว่าการเปลี่ยนแปลงที่เกิดในสระ เนื่องจากการสร้างช่องแคบที่แคบกว่าแต่มักจะไม่มีความต่อเนื่องของความถี่สั้นพ้องต่างๆ

2.5.4. หลังจากแบ่งเสียงเป็นเซกเมนต์ ต่างๆ พร้อมทั้งกำหนดประเภทของเซกเมนต์ของเสียงเรียบร้อยแล้ว จะต้องมีการคาดคะเนตำแหน่งของลิ้น หรือช่องแคบช่องปิดต่างๆ ให้สอดคล้องกับประเภทของเสียงนั้นๆ ถ้าเป็นเสียงของพญูชนะเสียงกัก พญูชนะเสียงเสียดแทรก หรือเสียงกึ่งเสียดแทรก จะต้องพิจารณาการสั่นของเส้นเสียงควบคู่ไปด้วย

2.5.5. ขั้นตอนสุดท้ายในการอ่านสเปกโตรแกรมของเสียงพูด คือ การรวมเสียงต่างๆที่ได้พิจารณาจากขั้นตอนข้างต้นมารวมให้เป็นค่า การรวมเสียงเป็นค่านั้นจะต้องใช้ความหมายของประโยคเข้ามาช่วยในการพิจารณา การใช้ความหมาย (Semantic) เข้ามาช่วยจะทำให้ตัดสนใจเกี่ยวกับเสียงบางเสียงที่ยังไม่แน่ใจจากการพิจารณาจากสัญญาณได้ดีขึ้น

## 2.6. การแปลงข้อมูลให้เป็นบรรทัดฐาน (Data Normalization)

เนื่องจากสัญญาณเสียงพยางค์จะแตกต่างกันพยางค์ที่ใช้ทำการทดลองมีความหลากหลายของค่าพลังงาน ทั้งนี้อาจเนื่องมาจากปัจจัยภายนอกที่เป็นสภาพแวดล้อมขณะบันทึกเสียง หรือปัจจัยภายในของผู้พูดเสียงพยางค์จะแตกต่างกันพยางค์ เช่น ผู้พูดเปล่งเสียงค่อย ดัง, อายุของผู้พูด และเพศของผู้พูด เพื่อลดความแปรปรวนของค่าพลังงานสัญญาณเสียงพยางค์จะแตกต่างกัน จึงจำเป็นที่จะต้องทำการแปลงข้อมูลให้เป็นบรรทัดฐานเดียวกันก่อน เพื่อให้การจำแนกเสียงมีความถูกต้องมากยิ่งขึ้น

การแปลงข้อมูลให้เป็นบรรทัดฐาน เป็นการปรับค่าของข้อมูลให้มีขอบเขตอยู่ในช่วงเล็กกลาง เช่น อยู่ในช่วง  $-1.0$  ถึง  $1.0$  หรือ ช่วง  $0.0$  ถึง  $1.0$  ซึ่งวิธีที่นิยมใช้กันอย่างแพร่หลาย เช่น การแปลงตามค่าต่ำสุด-สูงสุด (Min-Max Normalization) การแปลงตามค่าคะแนนมาตรฐานซี (z-score Normalization) และการปรับมาตราทศนิยม (Decimal Scaling) เป็นต้น

ในงานวิจัยนี้ผู้วิจัยเลือกใช้วิธีการแปลงตามค่าต่ำสุด-สูงสุด (Min-Max Normalization) โดยเลือกใช้ค่าสูงสุดที่เปอร์เซ็นต์ไทล์ที่ 95

## 2.7. แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model)

ระบบแบบจำลองฮิดเดนมาร์คอฟ [6] เป็นแบบจำลองมาร์คอฟระบบหนึ่ง ซึ่งไม่สามารถระบุได้ชัดเจนว่า ในขณะหนึ่งระบบอยู่ในสถานะใด หากแต่สามารถสังเกตปรากฏการณ์อื่นๆ เพื่อนำมาทำนายความน่าจะเป็นของสถานะในแบบจำลองมาร์คอฟที่เวลาใดๆได้ เหตุผลที่แบบจำลองฮิดเดนมาร์คอฟเป็นที่นิยมมีด้วยกันสองประการ คือ

ประการแรก แบบจำลองนี้อาศัยโครงสร้างทางคณิตศาสตร์ และสามารถเปลี่ยนแปลงทฤษฎีพื้นฐานเพื่อประยุกต์ใช้ได้อย่างกว้างขวาง

ประการที่สอง แบบจำลองนี้สามารถทำงานได้เป็นอย่างดีเมื่อประยุกต์ใช้อย่างเหมาะสม โดยในการวิจัยนี้จะพิจารณาถึงหลักการรู้จำคำโดดของแบบจำลองฮิดเดนมาร์คอฟ ซึ่งมีรายละเอียดดังนี้

ในการรู้จำคำโดดของแบบจำลองฮิดเดนมาร์คอฟนั้นจะแทนคำโดดแต่ละคำที่เข้ามาจำด้วยเวกเตอร์ลำดับการพูด หรือเรียกอีกอย่างว่า ค่าสังเกต (Observation)  $O$  โดยกำหนดให้

$$O = o_1, o_2, \dots, o_T \quad (2.10)$$

เมื่อ  $o_t$  คือค่าสังเกตในเวลา  $t$  และผลการจำแนกคำโดดที่ได้จากแบบจำลองฮิดเดนมาร์คอฟสามารถเขียนให้อยู่ในสัญลักษณ์  $\arg \max_i \{ P(w_i | O) \}$  โดยที่  $w_i$  เป็นคำโดดลำดับที่  $i$  ที่แบบจำลองฮิดเดนมาร์คอฟได้เรียนรู้ ส่วน  $P(w_i | O)$  เป็นค่าความน่าจะเป็นที่จะเป็นคำโดด  $w_i$  ของสัญญาณเสียงในรูปแบบค่าสังเกต  $O$  แต่การคำนวณหาค่า  $P(w_i | O)$  นั้นไม่สามารถคำนวณได้โดยตรง จะต้องอาศัยกฎของเบย์ (Bayes' Rule) มาช่วยหาค่าความน่าจะเป็นดังกล่าว โดยมีสมการการหาค่าเป็นดังนี้

$$P(w_i | O) = \frac{P(O | w_i) P(w_i)}{P(O)} \quad (2.11)$$

โดยที่จะทราบค่าความน่าจะเป็น  $P(w_i)$  ของแต่ละคำโดดในฐานข้อมูลการเรียนรู้ของแบบจำลองฮิดเดนมาร์คอฟอยู่ก่อนแล้ว ดังนั้นการหาค่าความน่าจะเป็นที่จะเป็นคำโดด  $w_i$  เมื่อมีสัญญาณเสียงในรูปแบบค่าสังเกต  $O$  นั้นจะขึ้นตรงกับค่าความน่าจะเป็น  $P(O | w_i)$  หรือ  $P(o_1, o_2, \dots | w_i)$

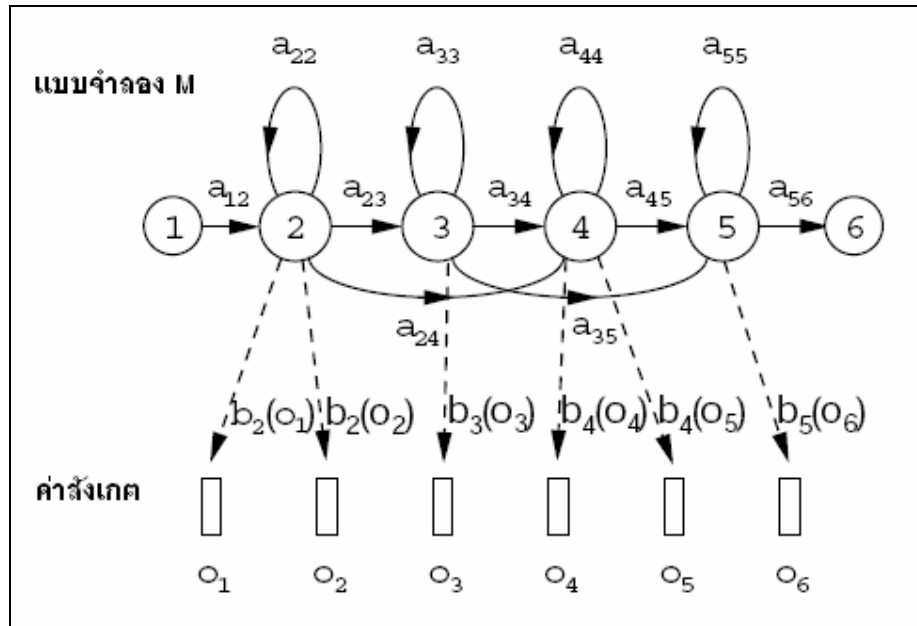
ในการรู้จำของแบบจำลองฮิดเดนมาร์คอฟนั้นจะสมมติให้เวกเตอร์ลำดับการพูดถูกสร้างจากแบบจำลองมาร์คอฟดังรูปที่ 2.9 โดยที่แบบจำลองมาร์คอฟนั้นจะเป็นเครื่องสถานะจำกัด (Finite State Machine) ที่เปลี่ยนสถานะทุกครั้งเมื่อเข้าไปสู่สถานะใหม่ และค่า  $o_t$  ของสัญญาณเสียงที่สถานะ  $j$  ในเวลา  $t$  จะถูกคำนวณจากค่าความหนาแน่นของความน่าจะเป็น  $b_j(o_t)$  ยิ่งกว่านั้นการเปลี่ยนสถานะจากสถานะที่  $i$  เป็นสถานะที่  $j$  จะถูกกำหนดจากความน่าจะเป็นเชิงวิฤต (Discrete Probability)  $a_{ij}$  จากรูปที่ 2.9 จะพบว่า มีสถานะของ



แบบจำลองอยู่หกสถานะ

และมีลำดับการเปลี่ยนสถานะของสัญญาณเสียงเป็น

$X = 1, 2, 2, 3, 4, 4, 5, 6$  ซึ่งจะถูกแปลงเป็นเวกเตอร์ลำดับการพูด  $o_1$  to  $o_6$



รูปที่ 2.9 ภาพการสร้างแบบจำลองมาร์คอฟ [ดัดแปลงจาก [6]]

การหาค่าความน่าจะเป็นร่วมของค่าสังเกต  $O$  (Joint Probability) ที่ถูกสร้างจากแบบจำลองมาร์คอฟ  $M$  ที่มีการเปลี่ยนสถานะตามรูปแบบของ  $X$  สามารถหาได้จากผลคูณของความน่าจะเป็น ตามสมการดังต่อไปนี้

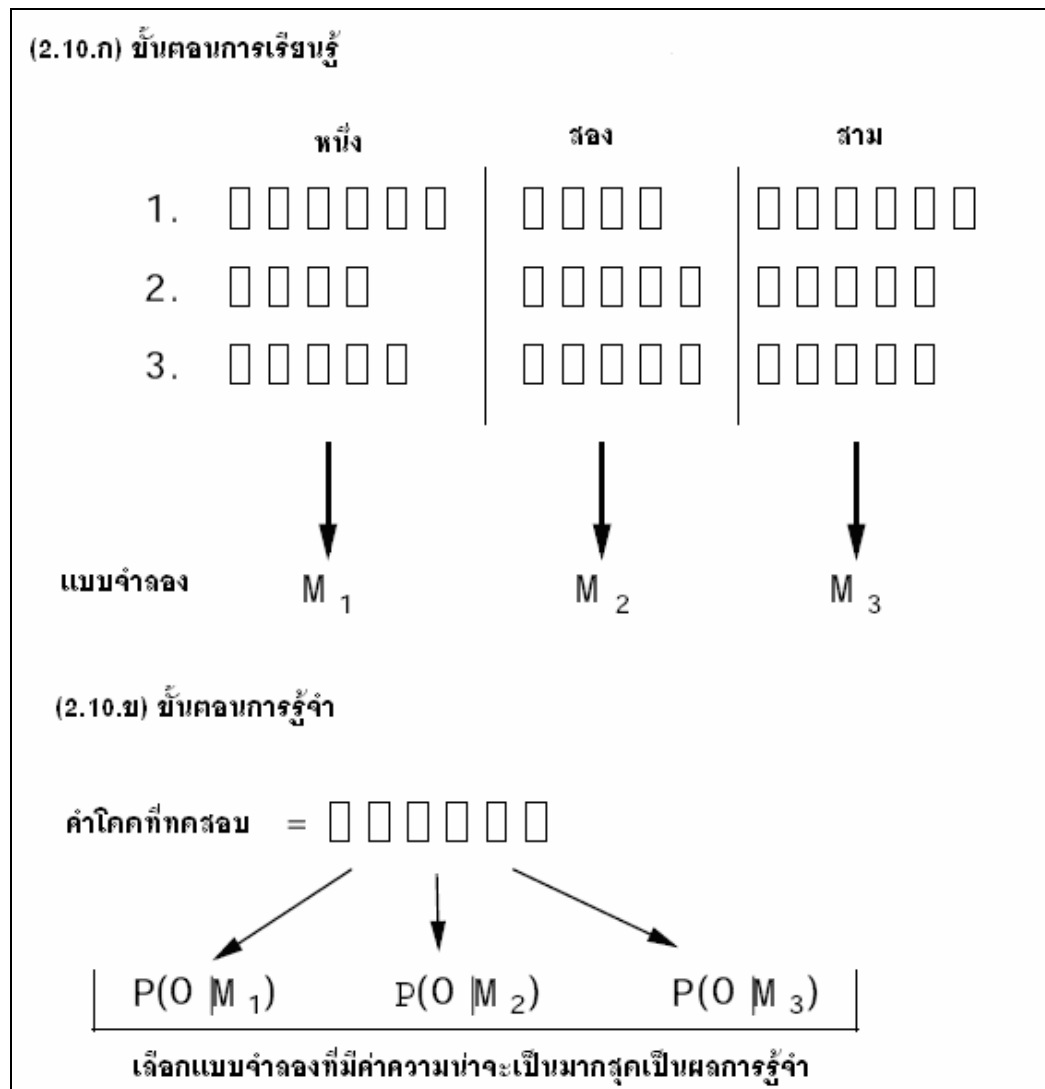
$$P(O, X|M) = a_{12}b_2(o_1)a_{22}b_2(o_2)a_{23}b_3(o_3) \quad (2.12)$$

โดยที่จะทราบเพียงค่าสังเกต  $O$  เท่านั้นแต่จะไม่สามารถทราบลำดับที่แท้จริงของการเปลี่ยนสถานะ  $X$  ได้ ซึ่งเป็นที่มาของชื่อแบบจำลองฮิดเดนมาร์คอฟนั่นเอง เนื่องจากไม่ทราบสถานะ  $X$  ดังนั้นในการหาค่า  $P(O, X|M)$  จะใช้ค่า  $P(O|M)$  แทน โดยจะใช้วิธีการประมาณหาค่ามากที่สุดมาช่วยในการคำนวณค่า  $P(O|M)$  ดังสมการต่อไปนี้

$$P(O|M) = \max_x \left\{ a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(o_t) a_{x(t)x(t+1)} \right\} \quad (2.13)$$

จากสมการที่ (2.13) พบว่าสามารถนำหลักการเวียนบังเกิดมาช่วยแก้สมการได้ ซึ่งทำให้การคำนวณใช้เวลาลดลง เมื่อได้ค่า  $P(O|M_i)$  จะสมมติให้ มีค่าเท่ากับ  $P(O|w_i)$  นอกจากนี้แต่ระบบจำลองมาร์คอฟ  $M_i$  จะทราบค่าความน่าจะเป็นเชิงวิฤต  $\{a_{ij}\}$  และค่าความหนาแน่นของความน่าจะเป็น  $\{b_j(o_t)\}$  อยู่ก่อนแล้ว ทำให้สามารถคำนวณหาค่าความน่าจะเป็นค่าโดดของแบบจำลองฮิดเดนมาร์คอฟ  $P(w_i|O)$  ได้ในที่สุด

นอกจากนี้เพื่อให้ง่ายต่อการทำความเข้าใจ จะสรุปการทำงานของกรรผู้จำค่าโดดของแบบจำลองฮิดเดนมาร์คอฟดังรูปที่ 2.10 โดยในแบบจำลองฮิดเดนมาร์คอฟนั้นจะต้องเรียนรู้จำแต่ละค่าโดดจากฐานข้อมูลการเรียนรู้ก่อน พร้อมทั้งสร้างแบบจำลองของแต่ละค่าโดดออกมา จากตัวอย่างในรูปที่ 2.10 พบว่ามีค่าที่จะถูกเรียนรู้อยู่สามค่าได้แก่ “หนึ่ง สอง สาม” หลังจากนั้นจึงค่อยให้แบบจำลองฮิดเดนมาร์คอฟทดสอบการรู้จำค่าโดดที่ไม่เคยเห็นมาก่อน โดยค่าโดดที่ไม่เคยเห็นมาก่อนจะถูกเปลี่ยนเป็นค่าสังเกต จากนั้นนำค่าสังเกตที่ได้ไปวิเคราะห์หาความน่าจะเป็นที่จะเป็นค่าโดดจากฐานข้อมูลการเรียนรู้ โดยผลการรู้จำที่ได้จะเลือกค่าโดดที่มีความน่าจะเป็นมากที่สุดเมื่อเปรียบเทียบกับค่าสังเกตของสัญญาณเสียงทดสอบ



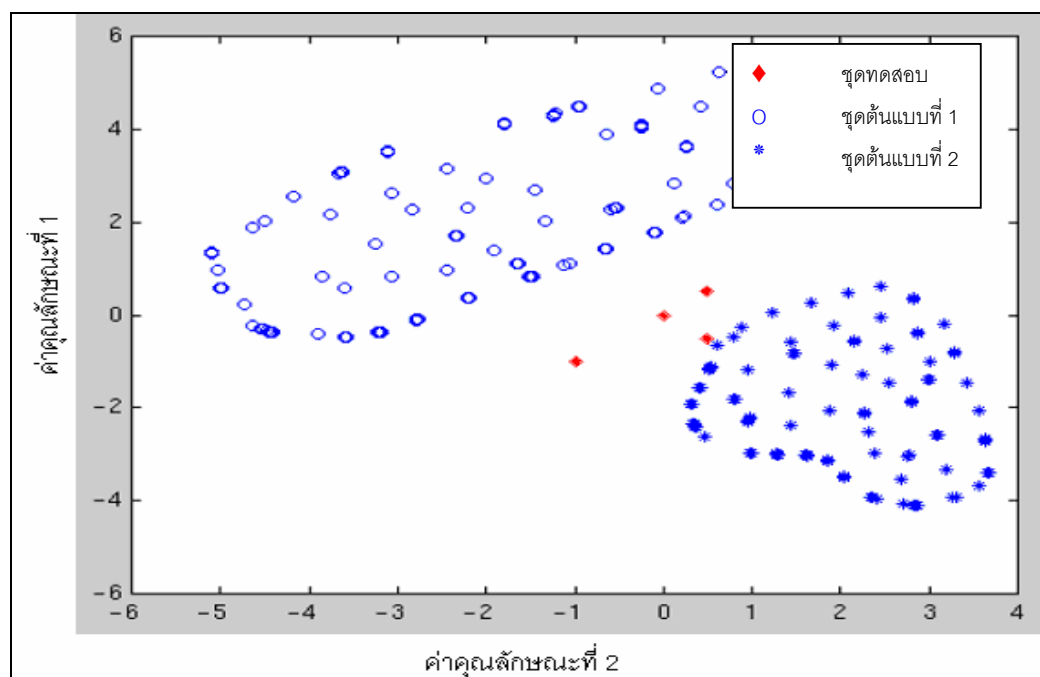
รูปที่ 2.10 ตัวอย่างขั้นตอนการรู้จำของแบบจำลองฮิดเดนมาร์คอฟ [ดัดแปลงจาก [6]]

## 2.8. วิธีการวิเคราะห์แบบดิสคริมีแนนต์เชิงเส้น (Linear Discriminant Analysis) [7]

ในปัจจุบันมีวิธีการจำแนกประเภทอยู่หลากหลายวิธี วิธีการวิเคราะห์แบบดิสคริมีแนนต์เชิงเส้นเป็นอีกวิธีหนึ่งที่ยอมรับในปัจจุบัน วิธีการวิเคราะห์แบบดิสคริมีแนนต์เชิงเส้นนี้จะอาศัยการจัดการอย่างง่ายกับกรณีของชุดข้อมูลที่มีความถี่ภายในแต่ละคลาส (Within-class frequency) ไม่เท่ากัน จากนั้นจึงวิเคราะห์ชุดข้อมูลทดสอบ โดยจะสำรวจค่าความถี่ของชุดข้อมูลทดสอบอย่างอิสระ ในระหว่างสำรวจจะมีการพยายามเพิ่มขนาดของค่าอัตราส่วนความแปรปรวน

ระหว่างคลาส (Between-class variance) กับความแปรปรวนภายในคลาส (Within-class variance) ในแต่ละชุดของข้อมูล เพื่อเป็นการรับประกันค่าความแตกต่างของแต่ละคลาสมากยิ่งขึ้น ทำให้สามารถจำแนกประเภทคลาสของชุดข้อมูลทดสอบได้แม่นยำมากขึ้น ส่งผลให้วิธีการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นนี้ได้ถูกนำมาใช้ในการแก้ปัญหาในการจำแนกประเภทของเสียง และการรู้จำเสียง

วิธีการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นจะไม่มี การเปลี่ยนตำแหน่งจุดของชุดข้อมูล แต่อย่างไรก็ตามในระหว่างที่ทำการวิเคราะห์ แต่จะพยายามจัดหาวิธีแยกแยะความแตกต่างระหว่างคลาส เพื่อให้สามารถกำหนดบริเวณของตำแหน่งแต่ละคลาสได้ชัดเจน ส่งผลให้การตัดสินใจถูกต้องแม่นยำมากขึ้น นอกจากนี้ยังช่วยให้เข้าใจถึงการกระจายของค่าคุณลักษณะของข้อมูลมากยิ่งขึ้นด้วย ดังตัวอย่างรูปที่ 2.11 แสดงภาพตัวอย่างการวิเคราะห์ข้อมูลตามหลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น



รูปที่ 2.11 ภาพตัวอย่างการวิเคราะห์ข้อมูลตามหลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น [ดัดแปลงจาก [7]]

### 2.8.1. ประเภทของการเปลี่ยนรูปในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น

ในการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นจะมีการเปลี่ยนรูปชุดข้อมูลต้นแบบของแต่ละคลาสและชุดข้อมูลทดสอบให้ไปอยู่ในปริภูมิใหม่เสียก่อน จากนั้นจึงทำการจำแนกประเภทของชุดข้อมูลทดสอบ โดยวิธีการเปลี่ยนรูปของชุดข้อมูลแต่ละคลาสนั้น มีอยู่ด้วยกัน 2 แบบ

2.8.1.1. การเปลี่ยนรูปคลาสแบบไม่อิสระ (Class-dependent transformation) ในการเปลี่ยนรูปคลาสแบบไม่อิสระนี้ จะใช้วิธีการขยายขนาดอัตราส่วนความแปรปรวนระหว่างคลาสกับความแปรปรวนภายในคลาส การขยายขนาดอัตราส่วนนี้จะช่วยให้สามารถแยกประเภทของคลาสได้ชัดเจนมากขึ้น

2.8.1.2. การเปลี่ยนรูปคลาสแบบอิสระ (Class-independent transformation) ในการเปลี่ยนรูปคลาสแบบอิสระนี้ จะใช้วิธีขยายค่าอัตราส่วนความแปรปรวนทั้งหมดกับความแปรปรวนภายในคลาส การเปลี่ยนรูปคลาสแบบอิสระนี้จะวิเคราะห์ค่าความแตกต่างภายในคลาสเปรียบเทียบกับค่าความแตกต่างของคลาสอื่นในชุดข้อมูลต้นแบบ

### 2.8.2. วิธีการคำนวณการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น

ในส่วนนี้จะอธิบายวิธีการคำนวณการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น โดยจะขอสมมติให้มีชุดข้อมูลต้นแบบตามรูปที่ 2.11 คือมีคลาส 2 คลาส แต่ละคลาสจะอยู่ในชุดข้อมูลต้นแบบแต่ละชุด ซึ่งแต่ละชุดจะมีข้อมูล 100 ข้อมูล โดยการคำนวณที่จะกล่าวต่อไปนี้ได้อิงตามขั้นตอนการวิเคราะห์ในแมทแลป (Matlab)

2.8.2.1. สร้างเซตของชุดข้อมูลต้นแบบ และชุดข้อมูลทดสอบ ซึ่งค่าที่เก็บในเซตจะเป็นค่าคุณลักษณะดั้งเดิมของข้อมูลต้นแบบ เพื่อให้ง่ายต่อการทำความเข้าใจ จะขอนำเมทริกซ์มาเก็บข้อมูลของแต่ละเซต ดังสมการต่อไปนี้

$$set1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \dots & \dots \\ \dots & \dots \\ a_{m1} & a_{m2} \end{bmatrix} \quad set2 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \dots & \dots \\ \dots & \dots \\ b_{m1} & b_{m2} \end{bmatrix} \quad (2.14)$$

2.8.2.2. คำนวณหาค่าเฉลี่ยของแต่ละชุดข้อมูลต้นแบบ และค่าเฉลี่ยของข้อมูลทั้งหมด โดยให้  $\mu_1$  และ  $\mu_2$  แทนค่าเฉลี่ยของ  $set1$  และ  $set2$  ตามลำดับ และให้  $\mu_3$  แทนค่าเฉลี่ยของข้อมูลทั้งหมด โดยสามารถคำนวณได้จากการรวมข้อมูล จาก  $set1$  และ  $set2$  ดังสมการ (2.15)

$$\mu_3 = p_1 \times \mu_1 + p_2 \times \mu_2 \quad (2.15)$$

โดยที่  $p_1$  และ  $p_2$  เป็นความน่าจะเป็นของแต่ละคลาสจากความรู้ก่อนหน้า (The apriori probabilities of the classes) ในเบื้องต้นกำหนดให้  $p_1$  และ  $p_2$  มีค่าเป็น 0.5

2.8.2.3. ในการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น จะใช้การกระจายค่า (scatter) ภายในคลาส และระหว่างคลาส มาสร้างเกณฑ์ในการจำแนกความแตกต่าง โดยการกระจายภายในคลาส (Within-class scatter) จะคำนวณหาค่าความแปรปรวนภายในคลาส จากนั้นเมื่อได้ค่าความแปรปรวนภายในคลาสแล้ว จึงจะสามารถคำนวณหาค่าความแปรปรวนระหว่างคลาสได้ ให้  $S_w$  แทนค่าการกระจายภายในคลาส จะสามารถคำนวณได้ตามสมการ (2.16) และ (2.17)

$$S_w = \sum_j p_j \times (\text{cov}_j) \quad (2.16)$$

จากชุดข้อมูลตามรูปที่ 2.11 ซึ่งมี 2 คลาส จะได้ว่า

$$S_w = 0.5 \times \text{cov}_1 + 0.5 \times \text{cov}_2 \quad (2.17)$$

โดยที่  $\text{cov}_1$  และ  $\text{cov}_2$  เป็นเมทริกซ์ความแปรปรวนร่วมเกี่ยว (Covariance Matrices) ที่สมมาตรกัน และเป็นของชุดข้อมูลจาก *set1* และ *set2* ตามลำดับ ค่าเมทริกซ์ความแปรปรวนร่วมเกี่ยวสามารถคำนวณได้ตามสมการ (2.18)

$$\text{cov}_j = (x_j - \mu_j)(x_j - \mu_j)^T \quad (2.18)$$

ส่วนค่าการกระจายระหว่างคลาส (Between-class scatter) สามารถคำนวณได้ตามสมการ (2.19) โดยในที่นี้ให้  $S_b$  แทนค่าการกระจายระหว่างคลาส

$$S_b = \sum_j (\mu_j - \mu_3)(\mu_j - \mu_3)^T \quad (2.19)$$

จะสังเกตเห็นว่า  $S_b$  จะมีวิธีการคำนวณคล้ายกับการคำนวณค่าความแปรปรวนร่วมเกี่ยว เพียงแต่ค่า  $S_b$  นั้นจะมีสมาชิกภายในเป็นค่าเฉลี่ย  $\mu_j$  แทนค่าข้อมูลดั้งเดิม

จากที่ทราบมาก่อนแล้วข้างต้นว่าการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น จะมีการเปลี่ยนรูปคลาสก่อนทำการวิเคราะห์ ซึ่งถ้าเป็นการเปลี่ยนรูปคลาสแบบไม่อิสระนั้นก็พยายามขยายขนาดอัตราส่วนความแปรปรวนระหว่างคลาส หรือก็คือค่าการกระจายระหว่างคลาส กับความแปรปรวนภายในคลาส หรือก็คือค่าการกระจายภายในคลาส และการเปลี่ยนรูปคลาสแบบอิสระจะใช้การขยายขนาดอัตราส่วนความแปรปรวนทั้งหมดกับความแปรปรวนภายในคลาสแทน

โดยเกณฑ์ที่เหมาะสมของประเภทการเปลี่ยนรูปคลาสแบบไม่อิสระสามารถเขียนเป็นสมการได้ดังสมการ (2.20) และของการเปลี่ยนรูปคลาสแบบอิสระสามารถเขียนได้ดังสมการ (2.21)

$$\text{criterion}_j = \text{inverse}(\text{cov}_j) \times S_b \quad (2.20)$$

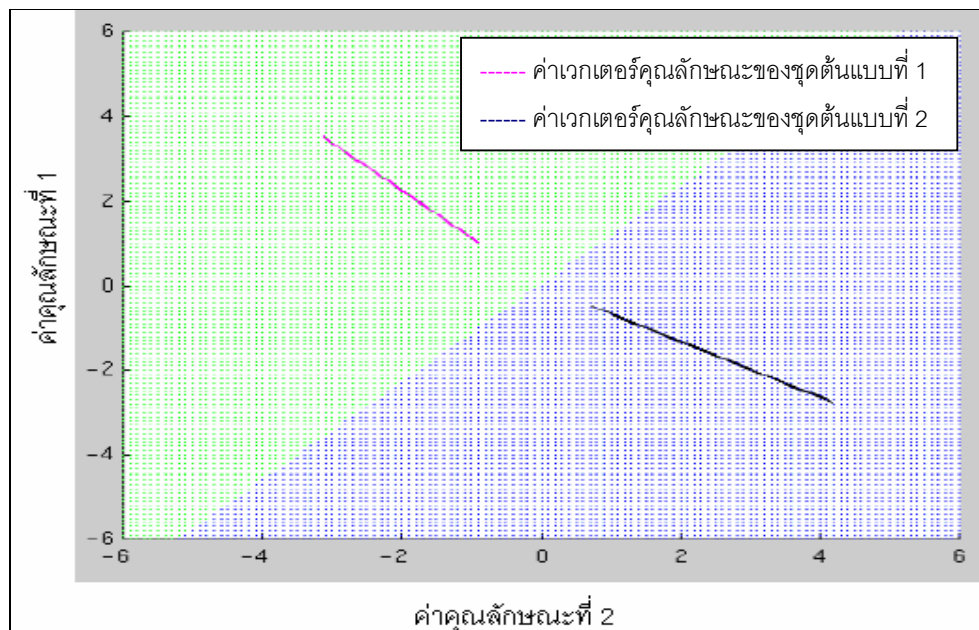
$$\text{criterion}_j = \text{inverse}(S_w) \times S_b \quad (2.21)$$

2.8.2.4. ค่าเวกเตอร์ลักษณะเฉพาะ (Eigen Vector) จะเป็นเสมือนตัวแทนการเปลี่ยนรูปของปริภูมีย่อยหนึ่งมิติ จากปริภูมิเวกเตอร์ทั้งหมดของชุดข้อมูล เมื่อคำนวณค่าเวกเตอร์ลักษณะเฉพาะของทุกเวกเตอร์ในปริภูมิแล้วจะได้จะเซตของเวกเตอร์ลักษณะเฉพาะขึ้นมา โดยที่สมาชิกในเซตจะต้องไม่มีค่าเป็นศูนย์ และมีการจัดเรียงกันอย่างอิสระเชิงเส้น พร้อมทั้งจะไม่มีเปลี่ยนแปลงตำแหน่งในปริภูมิเวกเตอร์อีกภายใต้การเปลี่ยนรูปครั้ง

ต่อไป ดังนั้นจะเห็นได้ว่าปริภูมิเวกเตอร์นี้สามารถแทนได้ใหม่ในรูปแบบของการรวมกลุ่มของค่าเวกเตอร์ลักษณะเฉพาะแบบเชิงเส้น

2.8.2.5. สำหรับปัญหาการจำแนกประเภท  $L$  คลาสใดๆ จะพบว่าจำเป็นต้องคำนวณหาค่าเวกเตอร์ลักษณะเฉพาะ  $L - 1$  ค่า และจะต้องไม่มีค่าเป็นศูนย์อีกด้วย ดังจะเห็นได้จากตัวอย่างสมการที่ (2.15)

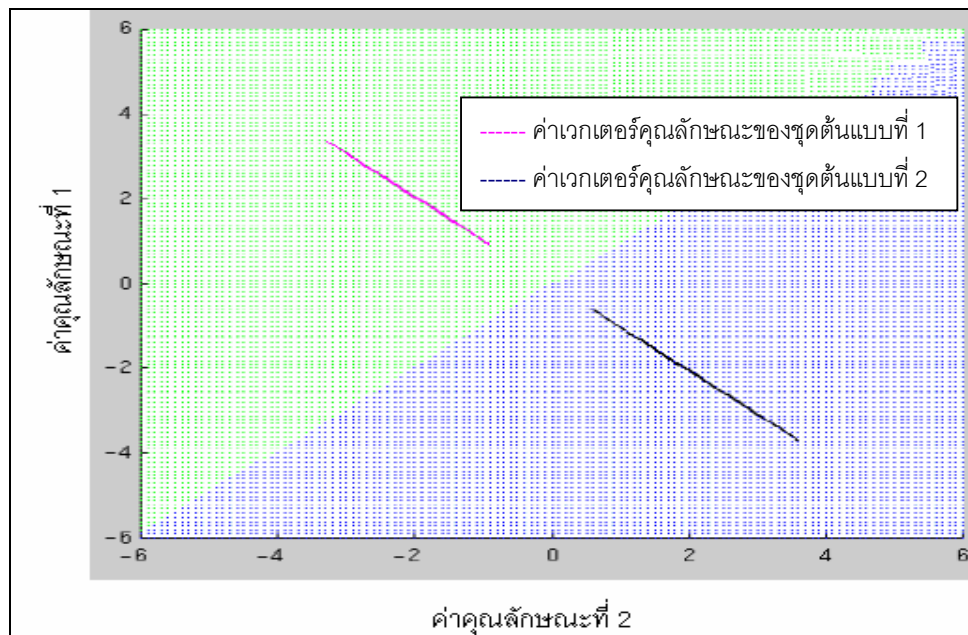
2.8.2.6. จากตัวอย่างรูปที่ 2.11 ซึ่งจะมีคลาส 2 คลาส จะเห็นได้ว่าเมื่อคำนวณหาค่าเวกเตอร์ลักษณะเฉพาะเรียบร้อยแล้วนำไปหาตำแหน่งในปริภูมิเวกเตอร์จะเห็นทิศทางของค่าเวกเตอร์ลักษณะเฉพาะที่มีนัยสำคัญ ดังรูปที่ 2.12 และ 2.13



รูปที่ 2.12 ภาพทิศทางค่าเวกเตอร์คุณลักษณะที่เกิดจากการเปลี่ยนรูปคลาสแบบไม่อิสระ

[ดัดแปลงจาก [7]]





รูปที่ 2.13 ภาพทิศทางค่าเวกเตอร์คุณลักษณะที่เกิดจากการเปลี่ยนรูปคลาสแบบอิสระ

[ดัดแปลงจาก [7]]

จากรูปที่ 2.12 และ 2.13 จะเห็นว่าเมื่อเปลี่ยนรูปคลาสข้อมูลทั้งหมดของชุดต้นแบบแล้ว พร้อมกับพิจารณาทิศทางของค่าเวกเตอร์คุณลักษณะ พบว่าสามารถใช้แกนเพียงแกนเดียวเท่านั้นในการกำหนดขอบเขตของแต่ละคลาส หรือแต่ละชุดได้ และเมื่อต้องการพิจารณาประเภทของคลาสของชุดข้อมูลทดสอบ ก็จะทำในลักษณะเดียวกันกับของชุดข้อมูลต้นแบบ กล่าวคือจะมีการเปลี่ยนรูปคลาส จากนั้นวิเคราะห์ประเภทของคลาสโดยใช้หลักการการหาระยะห่างระหว่างตำแหน่งข้อมูลทดสอบกับตำแหน่งข้อมูลต้นแบบตามวิธีของยูคลิด (Euclidean Distance)

### งานวิจัยที่เกี่ยวข้อง

เนื่องจากในปัจจุบันการใช้เครื่องคอมพิวเตอร์เป็นที่แพร่หลาย ผนวกกับเกิดการคิดค้นกลวิธีการรู้จำคำพูดแบบอัตโนมัติ ทำให้เกิดโปรแกรมคอมพิวเตอร์ฝึกสอนการพูดขึ้นมากมาย ตัวอย่างเช่น สปีชวิวเวอร์ (Speech Viewer) จากการวิจัยของ Adams และคณะ [8] พบว่าสปีชวิวเวอร์เป็นเครื่องมือฝึกสอนการพูดที่เหมาะสมสำหรับผู้ที่มีความบกพร่องหลายแบบ เช่น ความบกพร่องทางการพูด ความบกพร่องทางช่วงพัฒนาการ การได้รับความกระทบกระเทือนทางสมอง หรือการเป็นอัมพาตสมองใหญ่ โดยสปีชวิวเวอร์จะรับเสียงของผู้ใช้งานไปตรวจสอบการรู้จำ

แล้วแสดงผลเป็นรูปภาพ นอกจากนี้สปีชวิเวอร์ยังพบว่ามีการประมวลผลคอมพิวเตอร์จำนวนมากถูกสร้างขึ้นเพื่อช่วยฝึกสอนการแปลงเสียง และหน่วยเสียงในหลากหลายภาษา ทั้งภาษาทางฝั่งตะวันตก และภาษาทางฝั่งเอเชีย ผู้ใช้ส่วนมากมักเป็นผู้มีความผิดปกติทางการฟัง และกลวิธีการรู้จำในโปรแกรมคอมพิวเตอร์ส่วนใหญ่ใช้กลวิธีการรู้จำแบบ (Pattern Matching Technique) ที่อาศัยการเปรียบเทียบโครงสร้างสเปกตรัมของสัญญาณเสียง แบบจำลองพื้นฐานที่ใช้ในกลวิธีการรู้จำแบบ อาทิเช่น แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model) แบบจำลองไดนามิกไทม์วอร์ปิ้ง (Dynamic Time Warping) เป็นต้น

จะพบว่าโปรแกรมคอมพิวเตอร์ฝึกสอนการพูดโดยส่วนใหญ่จะมีฟังก์ชันการทำงานที่เน้นทางด้านเปรียบเทียบโครงสร้างสเปกตรัม ไม่ได้เน้นไปตรวจสอบถึงลักษณะการออกเสียงของพยัญชนะต้นที่แท้จริง ซึ่งจะเป็นประโยชน์ในการฝึกสอนการพูดได้ ดังนั้นโปรแกรมคอมพิวเตอร์ฝึกสอนการพูดโดยทั่วไปยังไม่สามารถตอบสนองความต้องการในการฝึกลักษณะการออกเสียงของเสียงพยัญชนะต้น โดยเฉพาะเสียงพยัญชนะกักได้โดยตรง การหาแนวคิดในการจำแนกเสียงตามลักษณะการออกเสียงของพยัญชนะกักต้นพยางค์จึงเป็นสิ่งจำเป็น

การจำแนกลักษณะการออกเสียงสามารถใช้วิธีสวสัทศาสตร์ มาศึกษาคุณลักษณะเด่นทางสัทศาสตร์ของแต่ละประเภทเสียง เพื่อใช้ในการแยกประเภทของเสียงที่ต้องการได้ โดยในเบื้องต้นต้องทำการแยกส่วนของหน่วยเสียงก่อน เพื่อให้สามารถศึกษาคุณลักษณะของเสียงได้ จากงานวิจัยของ Liu [9] เสนอหลักการตัดสัญญาณเสียง โดยอาศัยวิธีการค้นหาจุดหลักที่สัญญาณเสียงมีการเปลี่ยนแปลงชัดเจน (Landmark of Distinctive Features) จุดหลักนี้จะระบุถึงจุดตำแหน่งของเวลาที่สัญญาณเสียงมีการเปลี่ยนแปลงลักษณะอย่างชัดเจน วิธีการค้นหาจะใช้การแยกแยะส่วนประกอบของสัญญาณเสียงออกมาในรูปแบบของการเปลี่ยนแปลงพลังงานอย่างรวดเร็วของช่วงความถี่ 5 ช่วง แล้วหาตำแหน่งที่มีการเปลี่ยนแปลงของพลังงานมากที่สุดจากทุกความถี่

จุดหลักที่หาได้มีหลายประเภท ได้แก่ จุดหลักที่เกิดจากการสั่นของช่องเส้นเสียง (Glottis) จุดหลักที่เกิดจากซอนอรันต์ (Sonorant) และจุดหลักพลังลมฉับพลัน (Burst) การประเมินค่าความแม่นยำ จะใช้วิธีการหาค่าความคลาดเคลื่อนจุดหลักที่ได้จากการค้นหาอย่างอัตโนมัติ กับจากการใช้มือคนกำหนด ค่าความคลาดเคลื่อนที่ได้จากงานวิจัยนี้ไม่เกินร้อยละ 10

นอกจากวิธีการหาจุดหลักเพื่อใช้ในการแบ่งส่วนของสัญญาณเสียงแล้ว Leelaphattarakij และคณะ [10] ได้นำเสนอวิธีการหาขอบเขตของหน่วยเสียงที่ไม่ต่อเนื่องกันโดยใช้วิธีการสองขั้นตอนในการหา (Two-staged Approach) ก่อนทำการหาขอบเขตจะต้องรู้จำนวนขอบเขตของสัญญาณเสียงที่ต้องการ จากนั้นใช้วิธีการสองขั้นตอนในการวางตำแหน่งของขอบเขตให้กับสัญญาณเสียง ซึ่งวิธีการวางตำแหน่งของขอบเขตแบ่งออกเป็นสองขั้นตอน ขั้นตอนแรกจะ

หาตำแหน่งของเสียงที่มีความน่าจะเป็นตำแหน่งของขอบเขตสูงออกมาทั้งหมดก่อน โดยอาศัยสมมติฐานว่าบริเวณที่เป็นขอบเขตจะต้องเป็นบริเวณที่มีด้านข้างทั้งสองด้านมีสวณลักษณะที่แตกต่างกัน หรือในทางกลับกันบริเวณในขอบเขตเดียวกันจะมีสวณลักษณะใกล้เคียงกัน จากนั้นขั้นตอนสองจะตัดสินใจเลือกตำแหน่งของขอบเขตจากขั้นตอนแรกให้ตรงกับจำนวนขอบเขตที่ต้องการ โดยอาศัยกำหนดการพลวัตมาช่วยในการตัดสินใจหาตำแหน่งขอบเขตที่ดีที่สุด

สำหรับสวณลักษณะในขั้นตอนแรก จะใช้สองค่าลักษณะสำคัญทางสวณลักษณะ ได้แก่ ค่าพลังงานคลื่นสัญญาณเสียง แบ่งตาม 5 ช่วงความถี่ ได้แก่ 0 - 500 เฮิรตซ์ 500 - 1,000 เฮิรตซ์ 1,000 - 2,000 เฮิรตซ์ 2,000 - 4,000 เฮิรตซ์ และ 4,000 - 8,000 เฮิรตซ์ ในการคำนวณค่าพลังงานจะคำนวณที่ละกรอบของเวลา และการคำนวณจะอยู่ในรูปของพลังงานบรรทัดฐานเมื่อเปรียบเทียบกับกรอบของเวลาถัดไป โดยพลังงานบรรทัดฐานที่ได้จะถูกนำมายกกำลังสองอีกครั้งหนึ่ง ส่วนค่าลักษณะสำคัญทางสวณลักษณะที่สอง จะใช้สัมประสิทธิ์เซปสตรัมสเกลเมล (Mel Frequency Cepstral Coefficients: MFCC) โดยคำนวณหาที่ละกรอบของเวลา จากนั้นนำไปหาระยะห่างแบบยุคลิดของสัมประสิทธิ์เซปสตรัมสเกลเมลของกรอบเวลาปัจจุบันกับของกรอบเวลาถัดไป เมื่อได้ค่าลักษณะสำคัญทางสวณลักษณะที่ต้องการในแต่ละกรอบเวลา จะหาตำแหน่งของเวลาที่มีค่าลักษณะสำคัญทางสวณลักษณะมากที่สุดเมื่อเปรียบเทียบกับบริเวณใกล้เคียงให้เป็นตำแหน่งของเสียงที่มีความน่าจะเป็นตำแหน่งของขอบเขตสูง จากนั้นส่งให้ขั้นตอนที่สองต่อไป จากผลการทดลองของ Leelaphattarakij พบว่าค่าความแม่นยำสูงสุดที่ได้เป็น ร้อยละ 79.6 เมื่อยอมให้ตำแหน่งของขอบเขตเสียงที่เหมาะสมที่สุดโดย มีความคลาดเคลื่อนไม่เกิน 40 มิลลิวินาที ความแม่นยำจะมากขึ้นถ้าสามารถหาค่าลักษณะสำคัญทางสวณลักษณะที่บ่งบอกความเป็นขอบเขตได้เพิ่มเติม รวมทั้งสามารถหาความสัมพันธ์ของค่าลักษณะสำคัญทางสวณลักษณะระหว่างขอบเขตได้มากขึ้น วิธีการซัพพอร์ตเวกเตอร์แมชชีนเป็นอีกวิธีการที่น่าจะช่วยให้ได้ ต่อมา Leelaphattarakij [5] ได้นำเสนอวิธีการใหม่ที่อาศัยวิธีการซัพพอร์ตเวกเตอร์แมชชีนมาช่วยทำให้สามารถแบ่งเสียงพูดเป็นเซกเมนต์ที่มีประสิทธิภาพและทำงานได้รวดเร็วกว่าเดิม วิธีการแบ่งเสียงมีขั้นตอนการทำงาน 2 ขั้นตอนดังนี้ คือขั้นตอนการหาขอบเขตของหน่วยเสียงจากตำแหน่งที่มีการเปลี่ยนแปลงตามลักษณะการออกเสียง โดยอาศัยค่าลักษณะสำคัญทางสวณลักษณะ และวิธีการซัพพอร์ตเวกเตอร์แมชชีนมาจำแนกเสียงพูดตามลักษณะการออกเสียง ขั้นตอนต่อมาจะสร้างกราฟของเซกเมนต์โดยใช้วิธีการสร้างกราฟแบบหลายระดับ รวมถึงมีการคิดคะแนนให้กับขอบเขตของหน่วยเสียงที่หามาได้จากการเปลี่ยนแปลงสเปกตรัม โดยเมื่อยอมให้ขอบเขตของหน่วยเสียงที่หาได้คลาดเคลื่อนไปจากขอบเขตของหน่วยเสียงอ้างอิงไม่เกิน 20 มิลลิวินาที ในงานวิจัยนี้สามารถค้นหาขอบเขตของหน่วยเสียงได้ความแม่นยำและความครอบคลุมเพิ่มขึ้นร้อยละ 8.3 (จากร้อยละ 68.0 เป็นร้อยละ 76.3) และร้อยละ 5.1 (จากร้อยละ 82.1 เป็นร้อยละ 87.2)

ตามลำดับ และสามารถลดขนาดกราฟของเซกเมนต์ได้ประมาณ 14 เท่าโดยที่ยังรักษาระดับความครอบคลุมไว้ได้ที่ร้อยละ 77.4 เมื่อเปรียบกับวิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด

เมื่อสามารถแยกส่วนของสัญญาณเสียงพยัญชนะต้น โดยเฉพาะเสียงพยัญชนะกักได้ จะนำส่วนของสัญญาณที่ได้ไปศึกษาสัทศาสตร์ ในปัจจุบันพบว่ามิงงานวิจัยหลายงานที่ศึกษาสัทศาสตร์ของเสียงพยัญชนะกักต้นพยางค์เพื่อใช้ในการจำแนกประเภทเสียงของเสียงพยัญชนะกักต้นพยางค์ ตัวอย่างเช่น งานวิจัยของ Ali และคณะ [11] ได้นำเสนอค่าลักษณะสำคัญทางสวณลักษณะชุดใหม่ของเสียงพยัญชนะกักภาษาอังกฤษแบบอเมริกัน ได้แก่ ความถี่ของช่วงปล่อยลมของเสียงพยัญชนะกัก (Burst Frequency) ความถี่สั้นพ้องลำดับที่สองของสระที่ตามหลังเสียงพยัญชนะกัก (Second Formant (F2)) ความชันมากที่สุดของสเปกตรัม ความถี่สั้นพ้องที่เด่นสุด รูปแบบการเปลี่ยนสถานะของความถี่สั้นพ้องทั้งก่อนและหลังเสียงพยัญชนะกัก รวมทั้งพิจารณาลักษณะการออกเสียงของเสียงพยัญชนะกักประกอบเข้าด้วยกัน เพื่อใช้ในการจำแนกประเภทเสียงพยัญชนะกักภาษาอังกฤษแบบอเมริกัน ในการจำแนกประเภทจะหาค่าลักษณะสำคัญทางสวณลักษณะดังกล่าวข้างต้นก่อน จากนั้นจะอาศัยหลักการต้นไม้ตัดสินใจมาช่วยจำแนกประเภทของเสียงพยัญชนะกักเป็น /t/ /d/ /k/ /g/ /p/ /b/ โดยผลลัพธ์ของการจำแนกประเภทเสียงที่ได้มีค่าความถูกต้องถึงร้อยละ 86

ต่อมา Suchato [12] ได้เสนอลักษณะสำคัญทางสวณลักษณะที่ใช้ในการจำแนกประเภทฐานกรณ์ของการออกเสียงพยัญชนะกักภาษาอังกฤษโดยอาศัยหลักการศึกษาของสวณสัทศาสตร์ โดยลักษณะสำคัญทางสวณลักษณะที่ได้จะเกี่ยวข้องการสกัดค่าแอมพลิจูด พลังงานและความถี่สั้นพ้องในช่วงความถี่สูง และความถี่กลางของสัญญาณเสียง พร้อมทั้งนำเสนออิทธิพลของแต่ละลักษณะสำคัญทางสวณลักษณะที่มีผลต่อการจำแนกประเภทฐานกรณ์ของเสียงพยัญชนะกัก เนื่องจากค่าลักษณะสำคัญทางสวณลักษณะที่ได้มีจำนวนมาก Suchato และ คณะ [13] ได้ใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) และการวิเคราะห์ดิสคริมิแนนต์ (Discriminant Analysis) จัดกลุ่มค่าลักษณะสำคัญทางสวณลักษณะที่ได้เป็นองค์ประกอบใหม่เพื่อใช้ในการจำแนกประเภทฐานกรณ์ของเสียงพยัญชนะกักแทนการวิเคราะห์จากค่าลักษณะสำคัญทางสวณลักษณะจำนวนมาก โดยองค์ประกอบที่นำเสนอ คือ ขนาดของบรรทัดฐานสัญญาณเสียงของช่วงปล่อยลม (Normalized Burst Amplitude) รูปร่างสเปกตรัมของช่วงปล่อยลม (Burst Shape) ความถี่สั้นพ้อง (Formant Frequency) รูปแบบการเปลี่ยนสถานะของความถี่สั้นพ้อง (Formant Transition) สององค์ประกอบแรกจะมีอิทธิพลในการจำแนกประเภทฐานกรณ์ของเสียง เมื่อปราศจากข้อมูลส่วนหน้าของสระที่ตามมา แต่ถ้ามีข้อมูลส่วนหน้าของสระสององค์ประกอบหลัง จะมีอิทธิพลมากกว่า

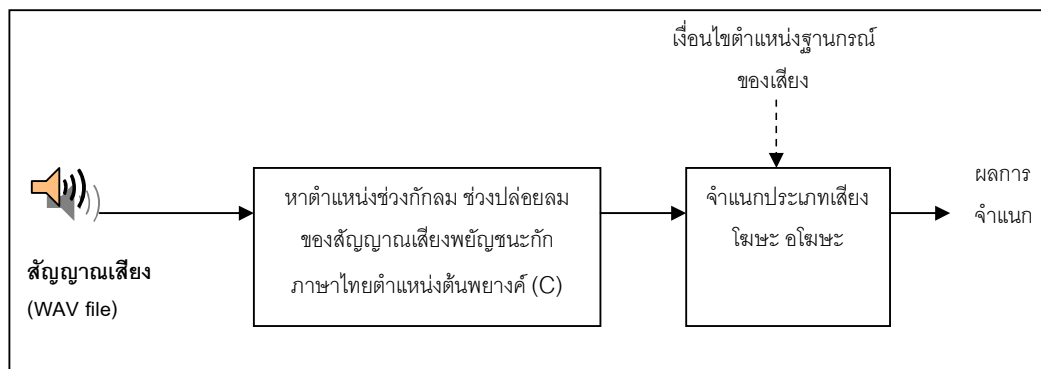
นอกจากงานวิจัยที่เกี่ยวข้องกับเสียงพยัญชนะกัก ยังมีงานวิจัยของ Dareyoah [14] ที่นำเสนอวิธีการตรวจสอบลักษณะการออกเสียงแบบโห้ชะ อโห้ชะของเสียงพูดขณะอยู่ในสภาพห้องปกติ (Room-level SNR) โดยอาศัยสวนลักษณะ จำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินศูนย์ (Zero Crossing Rate: ZCR) และค่าพลังงานที่ความถี่ต่ำ จะพบว่าถ้าสัญญาณเสียงที่ใช้ในการจำแนกประเภทเป็นสัญญาณเสียงแบบอโห้ชะ จะมีค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ที่สูงกว่าสัญญาณเสียงแบบโห้ชะ แต่สัญญาณเสียงแบบโห้ชะจะมีค่าพลังงานที่ความถี่ต่ำสูงกว่าสัญญาณเสียงแบบอโห้ชะ ในการทดลองได้ใช้ข้อมูลสัญญาณเสียง สองแบบ แบบแรกเป็นสัญญาณเสียงอ่านขณะอยู่ในห้องปิด (Clean Speech) แบบที่สองเป็นสัญญาณเสียงสนทนาขณะอยู่ในห้องปกติ (Room-level Speech) วิธีการวิเคราะห์ใช้การแบ่งเสียงพูดเป็นกรอบเวลาที่เท่ากัน จากนั้นพิจารณาทีละกรอบของเวลาคำนวณหาค่าสวนลักษณะทั้งสองค่า นำค่าที่ได้ไปพิจารณาการโห้ชะ อโห้ชะของเสียง จากผลการทดลองพบว่าที่ข้อมูลสัญญาณเสียงอ่านขณะอยู่ในห้องปิด ค่าความถูกต้องที่ได้สูงถึงร้อยละ 95 แต่กับข้อมูลสัญญาณเสียงสนทนาขณะอยู่ในห้องปกติ ค่าความถูกต้องที่ได้จะลดลง

งานวิจัยส่วนใหญ่มักจะมุ่งเน้นในการจำแนกประเภทเสียงพยัญชนะกักต้นพยางค์ การหาตำแหน่งของอวัยวะที่เกิดของเสียงพยัญชนะกักต้นพยางค์ หรือแม้กระทั่งการจำแนกประเภทลักษณะการออกเสียงในภาษาอังกฤษ แต่ยังมีได้มีงานวิจัยที่หาวิธีจำแนกประเภทเสียงตามลักษณะการออกเสียงของเสียงพยัญชนะกักต้นพยางค์ที่ใช้ฐานกรณ์เดียวกันในภาษาไทยโดยตรง ดังนั้นการศึกษารูปแบบการจำแนกประเภทเสียงตามลักษณะการออกเสียงของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณ์เดียวกันจึงเป็นสิ่งจำเป็น

### บทที่ 3

## การจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์เมื่อทราบ ฐานกรณ์โดยใช้สวณศาสตร์

การจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ดังกล่าวนี้ จะจำแนกตามลักษณะการออกเสียง ที่ประกอบไปด้วย เสียงโฆชะ กับ เสียงอโฆชะ แบบสถิต และ เสียงอโฆชะ แบบอนิต โดยใช้ฐานกรณ์เดียวกัน ซึ่งมีขั้นตอนหลักสองขั้นตอนด้วยกัน ดังรูปที่ 3.1 แสดงภาพขั้นตอนการจำแนกประเภทสัญญาณเสียงโดยอัตโนมัติของเสียงโฆชะ กับ เสียงอโฆชะ แบบสถิต และ เสียงอโฆชะ แบบอนิต



รูปที่ 3.1 ภาพขั้นตอนการจำแนกประเภทเสียงโฆชะ เสียงอโฆชะ แบบสถิต และ เสียงอโฆชะ แบบอนิตของพยัญชนะกักโดยอัตโนมัติ

จากรูปจะพบว่า มีขั้นตอนหลักสองขั้นตอน คือ ขั้นตอนการหาตำแหน่งช่วงกักลม และช่วงปล่อยลม และขั้นตอนการจำแนกประเภทเสียงโฆชะ และอโฆชะ โดยก่อนที่จะผ่านการหาตำแหน่งช่วงกักลม และช่วงปล่อยลมจะต้องผ่านขั้นตอนการเตรียมสัญญาณเสียงก่อน จากนั้นจึงค่อยนำเข้าสู่สัญญาณเสียงส่งต่อไปกับขั้นตอนการหาตำแหน่งช่วงกักลม และช่วงปล่อยลม ซึ่งมีรายละเอียดของแต่ละขั้นตอนดังนี้

#### 3.1. ขั้นตอนการนำเข้าสู่สัญญาณเสียง

สัญญาณเสียงที่ป้อนเข้าสู่ระบบอยู่ในรูป สระ (Vowel  $V_1$ ) พยัญชนะกัก (Stop Consonant  $C_s$ ) สระ (Vowel  $V_2$ ) จะเขียนย่อด้วยสัญลักษณ์  $V_1C_sV_2$  โดยที่สระตัวหน้าเป็นสระไม่เน้นพยางค์ และพยัญชนะเป็นพยัญชนะกักต้นพยางค์ ได้แก่ บ ป พ ต ต ฑ ก ข ต่อด้วยเสียงสระ

ในภาษาไทย 8 หน่วยเสียง เช่น สระอะ อา อี อี อู อู แอะ แอ และสัญญาณเสียงที่เข้ามาจะบันทึกด้วยความถี่ 44100 เฮิรตซ์

สาเหตุในการกำหนดรูปแบบของสัญญาณเสียงให้อยู่ในรูปค่าโดดแบบ  $V_1C_SV_2$  เนื่องจาก

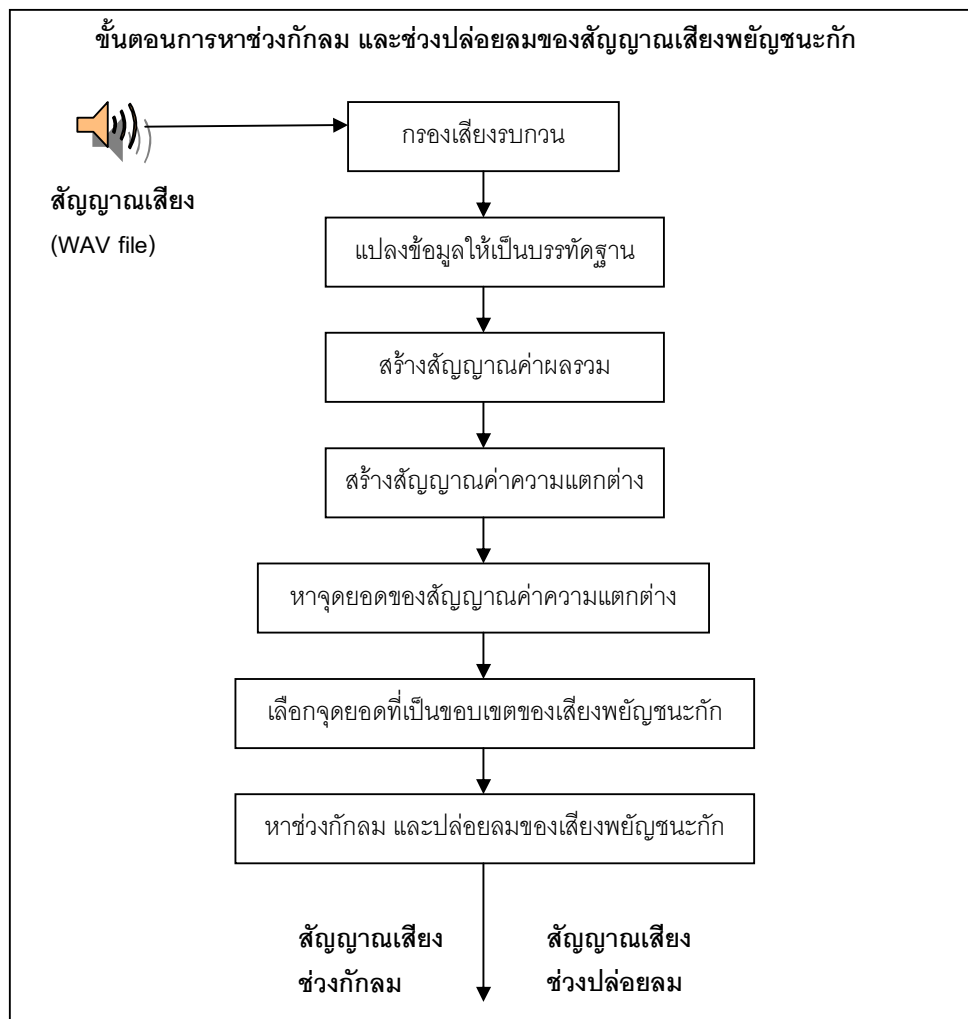
3.1.1. การฝึกการออกเสียงพยัญชนะกักตันพยางค์ในขั้นพื้นฐานจะเริ่มฝึกที่กลุ่มของค่าโดด

3.1.2. การให้ก่อนเริ่มออกเสียงพยัญชนะกักตันพยางค์มีการออกเสียงนำหน้าด้วยสระอะแบบไม่เน้นพยางค์ ( $V_1$ ) จะช่วยให้ผู้ฝึกสามารถสร้างช่องปิดได้สมบูรณ์แบบส่งผลให้การพิจารณาในสัทศาสตร์ในช่วงของการกักลมสามารถทำได้ง่ายขึ้น

3.1.3. เนื่องจากในการเปล่งเสียงค่าโดดแต่ละครั้งนอกจากมีเสียงพยัญชนะต้นแล้ว สิ่งที่เราไม่ได้คือสระที่ตามมา ดังนั้นเพื่อให้ตรงกับสภาพความเป็นจริง ต้องพิจารณาถึงส่วนของสระที่ตามมาด้วย ในที่นี้จึงเน้นสระที่สำคัญในภาษาไทยทั้ง 8 หน่วยเสียง เพื่อให้เกิดครอบคลุมกับความเป็นจริงมากที่สุด

### 3.2. ขั้นตอนหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกัก

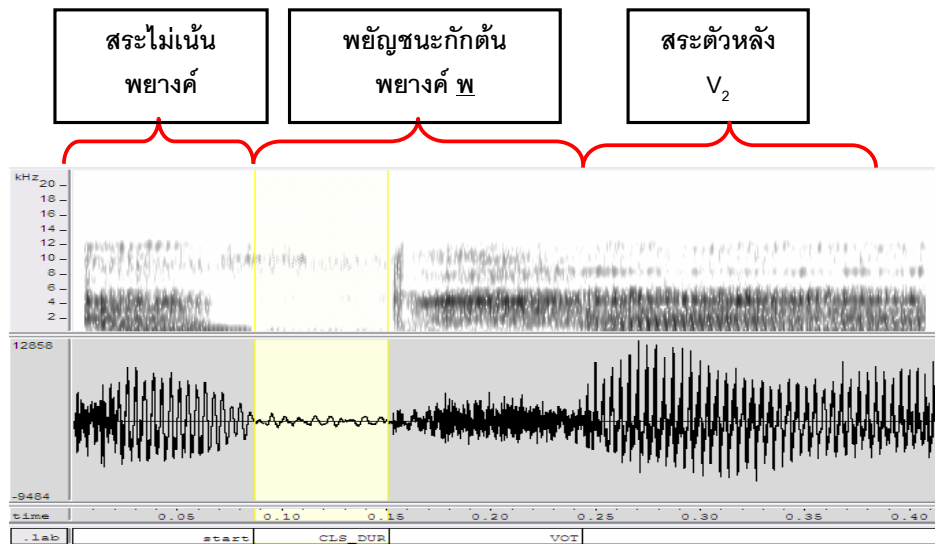
ในขั้นตอนนี้จะรับสัญญาณเสียงจากขั้นตอนของการนำเข้าสัญญาณเสียงเข้ามาเพื่อหาตำแหน่งช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกัก แล้วส่งต่อไปให้ขั้นตอนการจำแนกประเภทเสียงโม่ชะ อโม่ชะ ดังรูปที่ 3.2 แสดงภาพขั้นตอนการหาตำแหน่งช่วงกักลมช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกักภาษาไทยตำแหน่งต้นพยางค์



**รูปที่ 3.2** ภาพขั้นตอนการหาตำแหน่งช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะ  
กักภาษาไทยในตำแหน่งต้นพยางค์

ในการหาช่วงกักลมและช่วงปล่อยลมนั้นจะอาศัยหลักการแบ่งเสียงพูดเป็น  
เซกเมนต์มาช่วยในการตัดคำ ทั้งนี้เพราะว่าข้อมูลที่ใช้ในการเรียนรู้จำและเรียนรู้การตัดคำนั้นมี  
จำกัด ทำให้การใช้วิธีแบบเซกเมนต์จะเหมาะสมกว่าการใช้วิธีการแบ่งตามกรอบของเวลา อีกทั้ง  
เมื่อพิจารณารูปคลื่นของพลังงาน (Energy Waveform) ในแต่ละครั้งที่เข้ามาแล้ว พบว่ามีขอบเขต  
ของ  $V_1C_SV_2$  ที่ค่อนข้างชัดเจน ดังจะเห็นได้จากรูปที่ 3.3 แสดงส่วนของเซกเมนต์ของ  $V_1C_SV_2$  ทำ  
ให้การเลือกใช้วิธีแบบเซกเมนต์น่าจะเหมาะสมกับการตัดสัญญาณเสียงที่เข้าในระบบ





รูปที่ 3.3 ภาพเซกเมนต์  $V_1C_5V_2$  โดยที่แผนภาพด้านบนเป็นสเปกโตรแกรม (Spectrogram) และแผนภาพด้านล่างเป็นคลื่นของพลังงาน (Energy Waveform)

โดยมีรายละเอียดของขั้นตอนการแบ่งเสียงพูดแบบวิธีเซกเมนต์ดังนี้

3.2.1. กรองเสียงรบกวน (Noise) ออกจากสัญญาณเสียงที่เข้ามาในระบบ โดยใช้วิธีจำกัดช่วงความถี่ (Band Pass)

3.2.2. เมื่อผ่านการกรองเสียงรบกวน จะทำการแปลงข้อมูลเสียงที่ได้ให้เป็นบรรทัดฐาน ตามที่ได้กล่าวถึงในทฤษฎีที่หัวข้อ 2.6 โดยใช้วิธีการแปลงค่าสูงสุดที่เปอร์เซ็นต์ไทล์ที่ 95

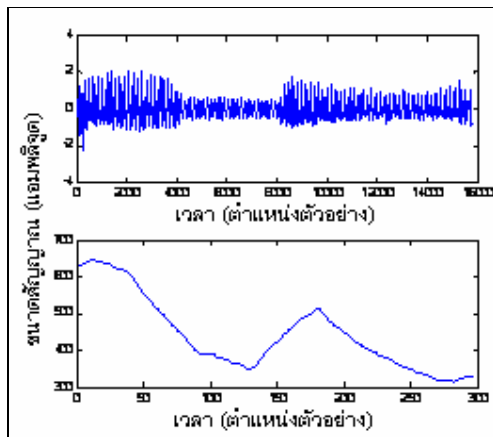
3.2.3. สร้างกรอบหน้าต่างของเวลาเพื่อใช้เป็นกรอบในการพิจารณาสัญญาณเสียงทีละส่วน จากนั้นหาผลรวมของพลังงานของสัญญาณเสียงในกรอบหน้าต่างนั้น เมื่อหาผลรวมเสร็จจะเลื่อนหน้าต่างไปยังส่วนถัดไปของสัญญาณเสียง ทำไปจนกระทั่งสิ้นสุดสัญญาณเสียงที่เข้ามา จะได้สัญญาณของค่าผลรวมเป็นสัญญาณใหม่ ดังรูปที่ 3.4 แสดงแผนภาพเปรียบเทียบสัญญาณเสียงกับสัญญาณของค่าผลรวม

3.2.4. ปรับความเรียบให้กับสัญญาณของค่าผลรวม เพื่อลดค่าการเปลี่ยนแปลงที่ไม่มีนัยสำคัญลง

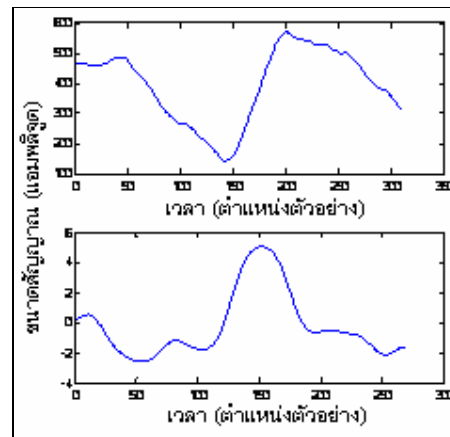
3.2.5. หาค่าความแตกต่างของสัญญาณของค่าผลรวม แล้วสร้างเป็นสัญญาณของค่าความแตกต่าง ดังรูปที่ 3.5 แสดงแผนภาพเปรียบเทียบระหว่างสัญญาณของค่าผลรวมกับสัญญาณของค่าความแตกต่าง

3.2.6. จากนั้นหาจุดยอดของสัญญาณค่าความแตกต่าง จะพบว่าหากขนาดของจุดยอดมีค่าเป็นบวกแสดงว่าสัญญาณเสียงในช่วงนั้นมีค่าพลังงานเพิ่มขึ้นจนถึงจุดสูงสุดในขณะนั้น ในทางกลับกันถ้าขนาดของจุดยอดมีค่าเป็นลบแสดงว่าสัญญาณเสียงมีค่าพลังงานลดลงจนถึงจุดต่ำสุดในช่วงนั้น ถ้าหากไม่ใช่ส่วนที่เป็นจุดยอดของพลังงานของสัญญาณเสียงจะให้ค่าเป็น 0 ดังรูปที่ 3.6 แสดงแผนภาพเปรียบเทียบสัญญาณค่าความแตกต่างกับสัญญาณจุดยอด

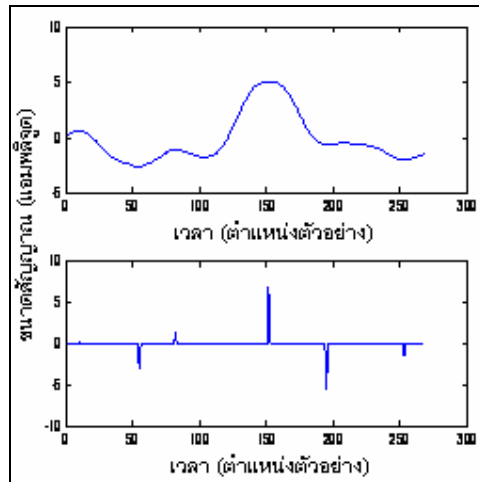
3.2.7. ลบตำแหน่งของค่าจุดยอดที่มีโอกาสเป็นขอบของเซกเมนต์น้อยทิ้งไป เพื่อให้สามารถตัดสินใจเลือกจุดยอดได้แม่นยำขึ้น โดยจะลบตำแหน่งของจุดยอดที่มีทิศทางของค่า (บวก, ลบ) เหมือนกับตำแหน่งของค่าจุดยอดก่อนหน้า และมีระยะห่างน้อยกว่าระยะห่างค่าเฉลี่ยทั้งหมดของตำแหน่งจุดยอดทิ้งไป เนื่องจากตำแหน่งดังกล่าวมักจะเป็นบริเวณรอยหยักของกราฟสัญญาณค่าความแตกต่างมากกว่าจะเป็นจุดยอดของสัญญาณค่าความแตกต่าง



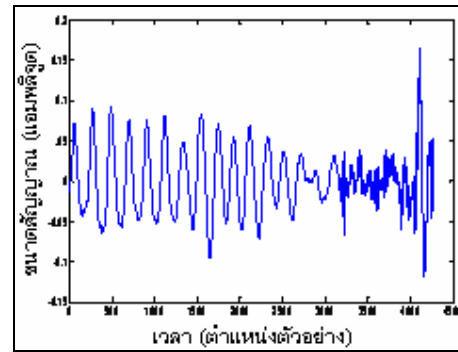
รูปที่ 3.4 ภาพเปรียบเทียบสัญญาณเสียง (ภาพบน) กับสัญญาณค่าผลรวม (ภาพล่าง)



รูปที่ 3.5 ภาพเปรียบเทียบสัญญาณค่าผลรวม (ภาพบน) กับสัญญาณค่าความแตกต่าง (ภาพล่าง)

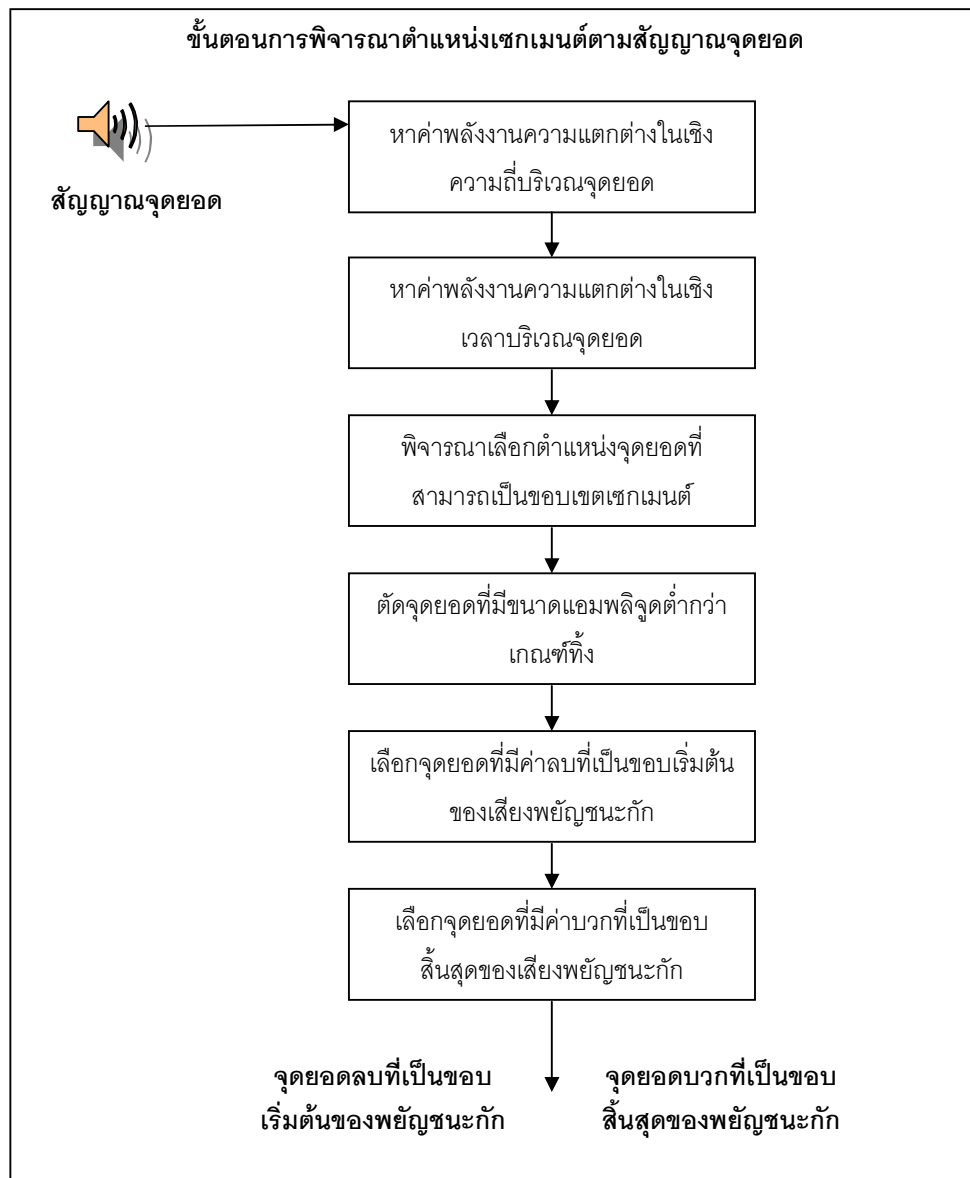


รูปที่ 3.6 ภาพเปรียบเทียบสัญญาณค่าความแตกต่าง (ภาพบน) กับสัญญาณจุดยอด (ภาพล่าง)



รูปที่ 3.7 ภาพส่วนที่ตัดได้ของพัญชนะกักต้นพยางค์

3.2.8. นำส่วนของจุดยอดที่ได้มาพิจารณาหาตำแหน่งที่แท้จริงของสัญญาณเสียงเพื่อให้ได้ส่วนของเสียงพัญชนะกักต้นพยางค์ สำหรับวิธีการพิจารณาดำเนินการตามจุดยอดนั้นมีวิธีการพิจารณา ดังรูปที่ 3.8 ต่อไปนี้

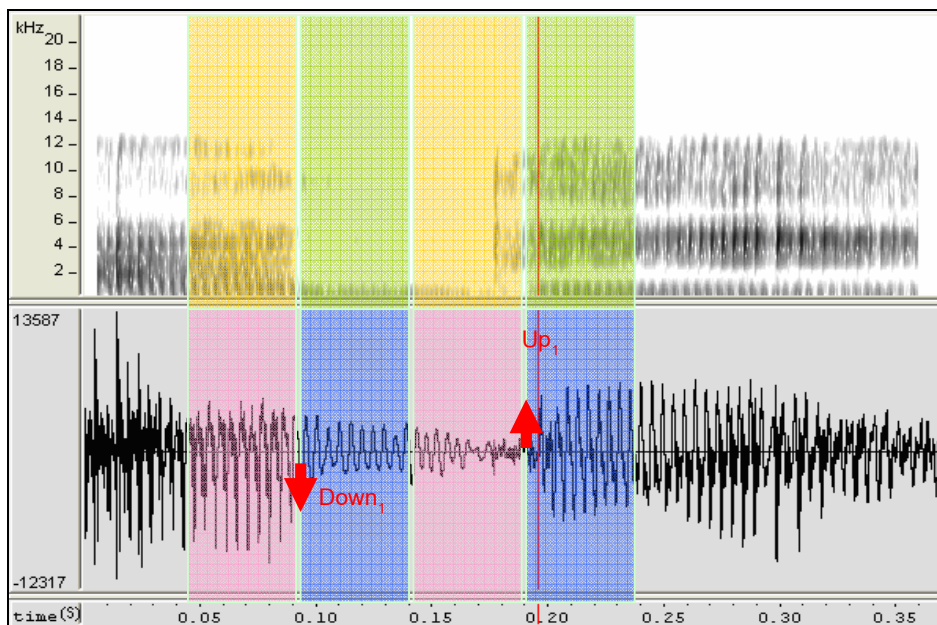


**รูปที่ 3.8** ภาพขั้นตอนการพิจารณาตำแหน่งเซกเมนต์ตามสัญญาณจุดยอด

**3.2.8.1.** หาค่าพลังงานความแตกต่างในเชิงความถี่ของพลังงานบริเวณก่อนหน้าและหลังจากตำแหน่งจุดยอด (ความกว้างประมาณ 50 มิลลิวินาที) โดยค่าพลังงานที่ได้จะหามาจากการแปลงฟูเรียร์อย่างรวดเร็ว จากนั้นหาผลรวมของพลังงานตั้งแต่ความถี่ที่ 1 ถึง 22,000 เฮิรตซ์ ดังตัวอย่างจากรูปที่ 3.9 แสดงภาพตำแหน่งบริเวณจุดยอดที่เป็นขอบของเซกเมนต์พัลส์ชณะกัก

### 3.2.8.2. หาค่าพลังงานความแตกต่างในเชิงเวลาของบริเวณ

ก่อนหน้าและหลังตำแหน่งจุดยอด โดยค่าความแตกต่างที่ได้จะหามาจากการลบกันระหว่างค่าพลังงานผลรวมในเชิงเวลาของบริเวณก่อนหน้าและหลังตำแหน่งจุดยอด (ความกว้างประมาณ 50 มิลลิวินาที) ดังตัวอย่างจากรูปที่ 3.9 แสดงภาพตำแหน่งบริเวณจุดยอดที่เป็นขอบของเซกเมนต์พยัญชนะเสียงกัก



รูปที่ 3.9 ภาพตำแหน่งบริเวณจุดยอดที่เป็นขอบของเซกเมนต์พยัญชนะกัก (บริเวณที่เป็นสีเหลืองและสีเขียว เป็นบริเวณหน้าและหลังจุดยอดที่ใช้คำนวณพลังงานในเชิงความถี่ และบริเวณสีชมพูและสีน้ำเงิน เป็นบริเวณหน้าและหลังจุดยอดที่ใช้คำนวณพลังงานในเชิงเวลา)

### 3.2.8.3. นำค่าพลังงานความแตกต่างทั้งสองมา

พิจารณาเลือกตำแหน่งจุดยอดที่มีโอกาสเป็นขอบของเซกเมนต์พยัญชนะกัก

โดยถ้าเป็นขอบเริ่มต้นของพยัญชนะกัก จะเป็นบริเวณที่มีค่าจุดยอดเป็นลบ มีค่าพลังงานความแตกต่างในเชิงความถี่ หรือในเชิงเวลาเป็นบวก หากมีบริเวณจุดยอดที่เป็นลบ แต่ไม่ได้มีค่าพลังงานความแตกต่างเป็นบวกจะไม่นำมาพิจารณาเป็นขอบของเซกเมนต์

แต่ถ้าหากเป็นขอบสิ้นสุดของพยัญชนะกัก จะเป็นบริเวณที่มีค่าจุดยอดเป็นบวก และมีค่าพลังงานความแตกต่างในเชิงความถี่ หรือในเชิงเวลาเป็น

ลบ หากมีบริเวณจุดยอดที่เป็นบวก แต่ไม่ได้มีค่าพลังงานความแตกต่างเป็นลบจะไม่นำมาพิจารณาเป็นขอบสิ้นสุดของเซกเมนต์พยัญชนะกัก

3.2.8.4. กรองบริเวณจุดยอดที่ขนาดแอมพลิจูดของจุดยอดน้อยกว่าเกณฑ์ (Ratio) ทิ้งไป โดยเกณฑ์ที่ใช้สามารถคำนวณได้จากสมการ (3.1) และค่าเกณฑ์ที่จะใช้จะแยกเป็น 2 ค่า เป็นเกณฑ์ของจุดยอดที่เป็นบวก และเกณฑ์ของจุดยอดเป็นลบ โดยการกรองนี้จะทำหลังจากการพิจารณาตัดจุดยอดโดยการใช้ค่าพลังงานความแตกต่างในเชิงความถี่และเวลาจากข้างต้น และจะพิจารณาต่อเมื่อมีจุดยอดบวก หรือจุดยอดลบมากกว่า 2 จุดยอด

$$Ratio = ((Max - Min) \times 0.3) + Min \quad (3.1)$$

โดยที่  $Max$ ,  $Min$  คือค่าสัมบูรณ์มากที่สุด และน้อยสุดของจุดยอดในเชิงเวลา ตามลำดับ

3.2.8.5. เมื่อได้จุดยอดที่มีโอกาสเป็นขอบของเซกเมนต์พยัญชนะต้นเสียงกักมาแล้ว จะทำการแยกพิจารณาเพื่อหาขอบเริ่มต้น โดยพิจารณาบริเวณจุดยอดเป็นค่าลบ และพิจารณาเพื่อหาขอบสิ้นสุด โดยพิจารณาบริเวณจุดยอดมีค่าเป็นบวก

ในการพิจารณาจะเลือกบริเวณจุดยอดที่มีค่าเป็นบวก และมีค่าเป็นลบ ที่มีขนาดแอมพลิจูดสูงสุดอย่างละ 2 อันดับแรก มาวิเคราะห์ โดยที่บริเวณจุดยอดที่มีขนาดแอมพลิจูดมากที่สุดและมีค่าเป็นลบ ( $Down_1$ ) จะต้องปรากฏก่อนบริเวณของจุดยอดที่มีค่าแอมพลิจูดมากที่สุดและมีค่าเป็นบวก ( $Up_1$ ) ทั้งนี้เนื่องมาจากว่าบริเวณ  $Down_1$  จะเป็นบริเวณขอบเริ่มต้นของพยัญชนะกัก และบริเวณ  $Up_1$  จะเป็นบริเวณขอบสิ้นสุดของพยัญชนะต้นเสียงกักตามที่ได้กล่าวแล้วข้างต้น ดังรูปที่ 3.9 แสดงภาพตำแหน่งบริเวณจุดยอดที่เป็นขอบของเซกเมนต์พยัญชนะเสียงกัก แต่ถ้าหากพบว่าบริเวณ  $Down_1$  อยู่ไกลกว่าบริเวณ  $Up_1$  จะทำการพิจารณาบริเวณจุดยอดลบใหม่แทน โดยการเลื่อนตำแหน่งไปพิจารณาที่จุดยอดเป็นลบที่มีขนาดแอมพลิจูดมากเป็นอันดับที่สองแทน

3.2.8.6. เมื่อสามารถเลือกบริเวณที่เป็นขอบเริ่มต้นของพยัญชนะเสียงกัก ขั้นตอนต่อมาจะพิจารณาค่าจุดยอดบวกที่น่าจะเป็นขอบสิ้นสุดของพยัญชนะต้นเสียงกักมากที่สุด เนื่องจากพยัญชนะเสียงกักเป็นพยัญชนะที่ประกอบไปด้วย สองเซกเมนต์ คือเซกเมนต์กักลม และเซกเมนต์ปล่อยลม ในช่วงเริ่มต้นของเซกเมนต์ปล่อยลมถ้ามีการเปลี่ยนแปลงชัดเจนในเชิงเวลา หรือในเชิงความถี่ จะปรากฏจุดยอดที่เป็นบวกขึ้นมาหนึ่งจุดยอดเพิ่มเติมจากบริเวณที่เป็นขอบสิ้นสุดของเซกเมนต์พยัญชนะเสียงกัก โดยขนาดของแอมพลิจูดของ

ทั้ง 2 จุดยอดบวกมักจะมีค่ามากที่สุดเป็น อันดับหนึ่ง หรือ อันดับสอง แล้วแต่สัญญาณเสียงที่เข้ามา ถ้าจุดยอดบวกที่มีแอมพลิจูดมากเป็นอันดับหนึ่ง อยู่ไกลกว่าจุดยอดที่มีแอมพลิจูดเป็นอันดับสองจะเลือกให้จุดยอดบวกอันดับหนึ่ง นั้นเป็นขอบของเซกเมนต์พัญชนะเสียงกัก แต่ถ้าพบว่าจุดยอดบวกที่มีขนาดแอมพลิจูดมากเป็นอันดับหนึ่ง นั้นอยู่ในตำแหน่งที่ถึงก่อนจุดยอดบวกที่มีค่าแอมพลิจูดเป็นอันดับสอง จะไม่สามารถตัดสินใจได้ทันทีว่าจุดยอดบวกใดเป็นขอบของเซกเมนต์พัญชนะเสียงกัก ทำให้ในการพิจารณาเลือกบริเวณที่เป็นขอบสิ้นสุดของเซกเมนต์นั้นจำเป็นต้องประยุกต์ใช้วิธีการแก้ปัญหาแบบศึกษาสำนึก (Heuristic Approach) จากฐานข้อมูลเสียงที่มีอยู่มาเป็นเครื่องมือช่วยในการตัดสินใจเลือกจุดยอดที่เป็นบวก

### 3.2.8.7. วิธีการแก้ปัญหาแบบศึกษาสำนึก

(Heuristic Approach) ที่ใช้เป็นเครื่องมือในการเลือกบริเวณจุดยอดบวกที่น่าจะเป็นขอบสิ้นสุดของเซกเมนต์พัญชนะเสียงกัก มีรายละเอียดดังต่อไปนี้

- ขั้นตอนที่ 1 หาตำแหน่งจุดยอดบวกที่มีขนาดแอมพลิจูดมากเป็นอันดับที่ 1 ( $Up_1$ ) และตำแหน่งจุดยอดบวกที่มีแอมพลิจูดมากเป็นอันดับที่ 2 ( $Up_2$ ) จากนั้นพิจารณาระหว่าง  $Up_1$  กับ  $Up_2$  หากบริเวณ  $Up_2$  ปรากฏขึ้นมาก่อน จะให้บริเวณ  $Up_1$  นั้นเป็นตำแหน่งขอบเริ่มต้นของเซกเมนต์ช่วงปล่อยลม หากบริเวณ  $Up_2$  ปรากฏขึ้นมาทีหลัง โดยจะถือว่าจุดยอด  $Up_1$  นี้เป็นจุดยอดทางซ้าย (Left) และจุดยอด  $Up_2$  ที่ปรากฏทีหลังเป็นจุดยอดทางขวา (Right) โดยจุดยอด Right นี้มีความน่าจะเป็นขอบของเซกเมนต์พัญชนะต้นเสียงกักมากกว่า Left

- ขั้นตอนที่ 2 หาค่าความต่างของพลังงานรวมในเชิงความถี่ที่ความถี่ต่ำของบริเวณก่อนหน้าและหลังของจากตำแหน่ง Left ไป 50 มิลลิวินาที และพลังงานที่จะรวมนั้นรวมจากช่วงความถี่ 1 ถึง 100 เฮิร์ตซ์ จะขอแทนค่าความต่างที่ได้ด้วย  $absLowLeft$

- ขั้นตอนที่ 3 หาค่าความต่างของพลังงานรวมในเชิงความถี่ที่ความถี่ต่ำของบริเวณก่อนหน้าและหลังของจากตำแหน่ง Right ไป 50 มิลลิวินาที และพลังงานที่จะรวมนั้นรวมจากช่วงความถี่ 1 ถึง 100 เฮิร์ตซ์ จะขอแทนค่าความต่างที่ได้ด้วย  $absLowRight$

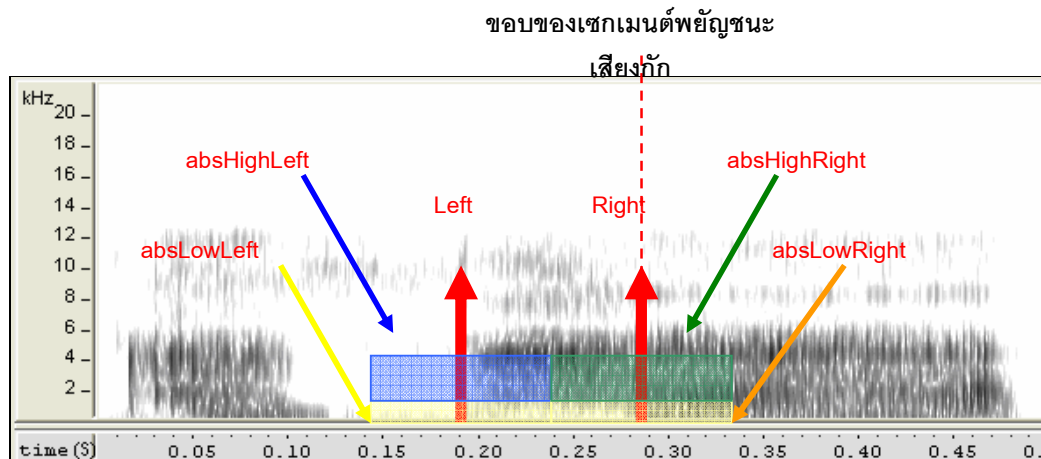
- ขั้นตอนที่ 4 หาค่าความต่างของพลังงานรวมในเชิงความถี่ที่ความถี่สูงของบริเวณก่อนหน้าและหลังของจากตำแหน่ง Left ไป 50 มิลลิวินาที และพลังงานที่จะรวมนั้นรวมจากช่วงความถี่ 2,000 ถึง 4,000 เฮิร์ตซ์ จะขอแทนค่าความต่างที่ได้ด้วย  $absHighLeft$

- ขั้นตอนที่ 5 หาค่าความต่างของพลังงานรวมในเชิงความถี่ที่ความถี่สูงของบริเวณก่อนหน้าและหลังของจากตำแหน่ง Right ไป 50 มิลลิวินาที และพลังงานที่จะรวมนั้นรวมจากช่วงความถี่ 2,000 ถึง 4,000 เฮิรตซ์ จะขอแทนค่าความต่างที่ได้ด้วย  $absHighRight$

- ขั้นตอนที่ 6 เมื่อได้ค่า  $absLowLeft$ ,  $absLowRight$ ,  $absHighLeft$ ,  $absHighRight$  มาเรียบร้อยแล้ว จะเริ่มพิจารณาที่  $absLowRight$  ก่อน โดยที่ถ้าจุดยอด Right เป็นขอบของเซกเมนต์พัลซึนระดับเสียงกักจริง มักจะต้องมีค่า  $absLowRight$  มาก (จากฐานข้อมูลเสียงที่มีอยู่พบว่า ถ้า  $absLowRight$  มีค่ามากกว่า 25) ให้ถือว่าจุดยอด Right เป็นขอบของเซกเมนต์พัลซึนระดับเสียงกักได้ทันที ทั้งนี้เนื่องจากสังเกตเห็นว่าบริเวณที่เป็นช่วงปล่อยลมนั้นจะไม่มีพลังงานที่ความถี่ต่ำ ส่วนบริเวณที่เป็นช่วงกักลมนั้นมักจะมีพลังงานที่ความถี่ต่ำ โดยเฉพาะอย่างยิ่งเสียงพัลซึนระดับเสียงกักแบบโหมชะ ส่วนบริเวณที่เป็นช่วงของสระจะมีพลังงานที่ความถี่ต่ำ ทำให้บริเวณที่เป็นขอบระหว่างพัลซึนระดับเสียงกักและสระ นั้นมีความน่าจะเป็นที่จะมีความแตกต่างของพลังงานที่ความถี่ต่ำสูง ดังรูปที่ 3.10 แสดงภาพเปรียบเทียบพลังงานของ  $absLowLeft$ ,  $absHighLeft$ ,  $absLowRight$  และ  $absHighRight$  แต่ถ้าค่าของ  $absLowRight$  มีค่าน้อย ความน่าจะเป็นที่จะจุดยอด Left เป็นขอบของพัลซึนระดับเสียงกักจะมีค่ามากกว่า ทั้งนี้จะนำค่า  $absHighLeft$  และ  $absHighRight$  มาพิจารณาควบคู่ในการตัดสินใจ

- ขั้นตอนที่ 7 ในกรณีที่ยังไม่สามารถตัดสินใจได้จากค่า  $absLowRight$  จะนำค่า  $absHighLeft$  และ  $absHighRight$  มาพิจารณา ในการตัดสินใจ โดยถ้าจุดยอด Right เป็นขอบจริง จะมีค่า  $absHighRight$  น้อยกว่า  $absHighLeft$  และมีค่า  $absLowLeft$  มาก แต่ถ้าหากค่า  $absLowRight$  กับค่า  $absHighRight$  มีค่าน้อยมาก แสดงว่าจุดยอด Right นั้นน่าจะอยู่ในช่วงของเซกเมนต์สระ ทำให้จุดยอด Left เป็นขอบของพัลซึนระดับเสียงกักแทน





รูปที่ 3.10 ภาพเปรียบเทียบพลังงานของ absLowLeft absHighLeft absLowRight และ absHighRight ของสัญญาณเสียงพยัญชนะกัก อะพะ /a/

เมื่อพิจารณาเลือกจุดยอดลบและจุดยอดบวกที่เป็นขอบของเซกเมนต์สัญญาณเสียงพยัญชนะกักเรียบร้อยแล้วจะได้สัญญาณเสียงพยัญชนะกักออกมาดังรูปที่ 3.7 แสดงแผนภาพส่วนของเสียงพยัญชนะกักต้นพยางค์ที่ได้จากการทำเซกเมนต์

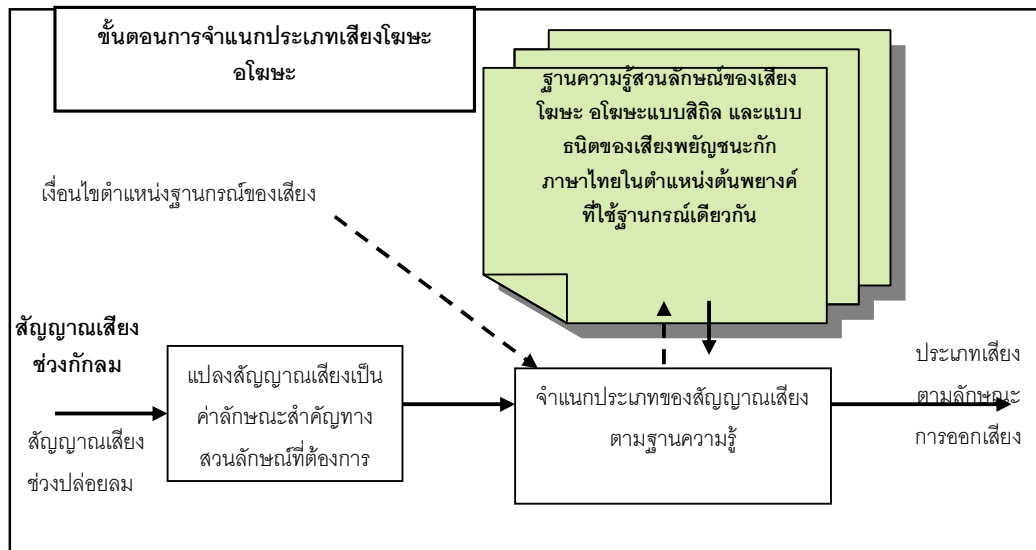
3.2.9. หาดำแหน่งเริ่มต้นช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกักต้นพยางค์หลังจากที่ได้ช่วงของสัญญาณเสียงพยัญชนะกัก ในการหาบริเวณเริ่มต้นของช่วงปล่อยลมนั้นจะใช้ค่าพลังงานในเชิงความถี่มาเป็นตัวกรองหาบริเวณที่น่าจะเป็นตำแหน่งเริ่มต้นของช่วงปล่อยลม โดยจะสร้างกรอบหน้าต่างของเวลาขึ้นมาแล้วหาพลังงานรวมที่ความถี่ต่างๆ เลื่อนกรอบหน้าต่างไปเรื่อยๆจนกระทั่งสิ้นสุดสัญญาณเสียง หลังจากได้ค่าพลังงานรวมแต่ละความถี่ในแต่ละกรอบหน้าต่าง จะนำค่าที่ได้ไปกรองความถี่ หากพบว่าบริเวณใดที่เริ่มมีพลังงานสูงทั้งช่วงความถี่ต่ำ กลาง และสูง บริเวณนั้นจะเป็นตำแหน่งเริ่มต้นของช่วงปล่อยลม

เกณฑ์ที่ใช้ในการกรองหาบริเวณตำแหน่งเริ่มต้นช่วงปล่อยลมนั้นจะแบ่งออกเป็น 3 ช่วงได้แก่ ช่วงความถี่ต่ำ (0 ถึง 1,000 เฮิรตซ์) ความถี่กลาง (2,000 ถึง 6,000 เฮิรตซ์) และความถี่สูง (8,000 ถึง 10,000 เฮิรตซ์) โดยที่ช่วงความถี่ต่ำมีค่าพลังงานประมาณ -25 ถึง -10 เดซิเบล ที่ช่วงความถี่กลางมีค่าพลังงานประมาณ -30 ถึง -10 เดซิเบล และที่ช่วงความถี่สูงมีค่าพลังงานประมาณ -55 ถึง -35 เดซิเบล ทั้งนี้ขึ้นอยู่กับฐานกรณ์ของสัญญาณเสียง

เมื่อหาบริเวณเริ่มต้นของช่วงปล่อยลมได้ จะนำตำแหน่งดังกล่าว พร้อม ทั้งจุดยอดกลับที่เป็นขอบเริ่มต้น และจุดยอดบวกที่เป็นขอบสิ้นสุดของพยัญชนะกัก มาหาส่วนของ ช่วงกักลม และช่วงปล่อยลมพยัญชนะต้นเสียงกัก เพื่อส่งต่อไปกับขั้นตอนการจำแนกประเภท เสียงโฆชะ อโฆชะต่อไป

### 3.3. ขั้นตอนการจำแนกประเภทเสียงโฆชะ อโฆชะ

ในขั้นตอนนี้จะรับสัญญาณเสียงสองส่วนมาจากขั้นตอนการหาช่วงกักลม ช่วง ปล่อยลมของสัญญาณเสียงพยัญชนะกัก ได้แก่ส่วนของช่วงกักลม และส่วนของช่วงปล่อยลมมา ใช้ในการจำแนกประเภท และรับเงื่อนไขตำแหน่งฐานกรณ์ของเสียงมาเป็นเงื่อนไขการทำงานของ ส่วนจำแนกประเภทของสัญญาณเสียงตามลักษณะการออกเสียงโฆชะ อโฆชะแบบสติดิล และ แบบชนิด ในการจำแนกประเภทประกอบไปด้วยหลายส่วน ดังรูปที่ 3.11 แสดงภาพส่วนประกอบ ของขั้นตอนการจำแนกประเภทเสียงโฆชะ อโฆชะ มีรายละเอียดดังนี้



รูปที่ 3.11 ภาพส่วนประกอบของขั้นตอนการจำแนกประเภทเสียงโฆชะ อโฆชะ

3.3.1. จัดเตรียมฐานความรู้สวณลักษณะของเสียงโหระ อโหระ แบบสถิต และแบบชนิดของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณ์เดียวกัน ฐานความรู้สวณลักษณะนี้ได้มาจากการนำสัญญาณเสียงพยัญชนะกักไปสกัดหาค่าลักษณะสำคัญทางสวณลักษณะ จากนั้นนำไปใช้เป็นชุดข้อมูลสอนเพื่อเรียนรู้ลักษณะที่แตกต่างกันของเสียงโหระ อโหระแบบสถิต และอโหระแบบชนิด โดยสัญญาณเสียงพยัญชนะกักต้นพยางค์ที่ใช้จะเก็บจาก **กลุ่มของคนปกติ** มาเป็นฐานความรู้สวณลักษณะ

สัญญาณเสียงกลุ่มของคนปกติ ประกอบไปด้วยเสียงของคนปกติที่พูดภาษาไทยได้ จำนวน 10 คน เป็นเพศชาย 5 คน และเพศหญิง 5 คน การเก็บเสียงของคนปกติ จะช่วยให้ศึกษาและค้นคว้าหาข้อแตกต่างของเสียงโหระ เสียงอโหระแบบสถิต และเสียงอโหระแบบชนิดในแบบที่ถูกต้อง การบันทึกข้อมูลจะใช้ความถี่ที่ 44,100 เฮิร์ตซ์ และกำหนดให้สัญญาณเสียงที่เข้ามาอยู่ในรูปแบบ **สระ** (Vowel) **พยัญชนะ** (Stop Consonant) **สระ** (Vowel) ( $V_1-C_s-V_2$ ) โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักต้นพยางค์ มีด้วยกันเป็น 8 เสียงหลัก **บ ป พ ต ท ก ข** และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ **สระอะ อา อิ อี อู อุ แอะ แอ** รวมเสียงที่จัดเก็บทั้งสิ้น 640 เสียง

3.3.2. แปลงสัญญาณเสียงเป็นค่าลักษณะสำคัญทางสวณลักษณะ **ที่ต้องการ** ในงานวิจัยนี้จะใช้ค่าลักษณะสำคัญทางสวณลักษณะทั้งหมด 5 ค่าด้วยกัน คือ ความยาวช่วงกักลม ความยาวช่วงปล่อยลม ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำของสัญญาณเสียงช่วงกักลม และ ค่าพลังงานของช่วงปล่อยลม ทั้ง 4 ค่านี้ได้มาจากผลงานวิจัยของผู้ทำวิจัยที่ได้ตีพิมพ์ [15] นอกจากนี้ยังเพิ่มเติมค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ (Zero Crossing Rate: ZCR) ในช่วงกักลมมาเป็นค่าลักษณะสำคัญทางสวณลักษณะที่ 5 เพื่อช่วยให้การจำแนกเสียงโหระ แบบสถิตแม่นยำมากยิ่งขึ้น

- การหาความยาวช่วงกักลม และช่วงปล่อยลม จะวัดจากการหาค่าผลรวมของจำนวนตัวอย่าง (Sampling) ที่ได้จากสัญญาณเสียงในเชิงเวลา

- ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำของสัญญาณเสียง **ในช่วงกักลม** สามารถหาได้ตามฟังก์ชันต่อไปนี้

$$E_{Ave} = \left\{ \sum_{i=1}^N PowerY_{low}(x_i) \right\} / length(x) \quad (3.2)$$

โดยที่  $E_{Ave}$  คือค่าพลังงานเฉลี่ยที่ความถี่ต่ำของช่วงกักลมของ สัญญาณเสียง  $x$  ส่วน  $length(x)$  คือความยาวช่วงกักลม และ  $i, N$  คือตำแหน่งตัวอย่าง (Sampling) ของสัญญาณเสียงช่วงกักลมเชิงเวลา โดยที่  $N$  คือตำแหน่งตัวอย่างสุดท้ายของ สัญญาณเสียงช่วงกักลมที่พิจารณา

$$PowerY_{low}(x_i) = Y_i * conj(Y_i) / 1024 \quad (3.3)$$

โดยที่  $PowerY_{low}(x_i)$  คือค่าพลังงานรวมในเชิงความถี่ที่ 60 ถึง 300 เฮิรตซ์ของสัญญาณเสียงช่วง  $x_i$  และ  $conj(Y_i)$  คือค่าสังยุค (Conjugate) ของ  $Y_i$

สำหรับ  $Y_i$  นั้นสามารถหาได้ดังนี้

$$Y_i = fft(x_i) \quad (3.4)$$

โดยที่  $Y_i$  คือสัญญาณเสียงในเชิงความถี่จากการแปลงฟูเรียร์อย่างรวดเร็ว  $fft(x_i)$  สัญญาณเสียงเชิงเวลาช่วง  $x_i$

- **ค่าพลังงานของช่วงปล่อยลม** จะวัดจากค่าพลังงานของ สัญญาณเสียงช่วงปล่อยลม สามารถหาได้ดังต่อไปนี้

$$E_{Sum} = \sum_{i=1}^N PowerY_{total}(x_i) \quad (3.5)$$

โดยที่  $E_{Sum}$  คือพลังงานรวมของช่วงปล่อยลม ส่วน  $PowerY_{total}(x_i)$  คือค่าพลังงานรวมในเชิงความถี่ของทุกความถี่ของสัญญาณเสียงเชิงเวลาช่วง  $x_i$  และ  $i, N$  คือ ตำแหน่งตัวอย่าง (Sampling) ของสัญญาณเสียงช่วงปล่อยลม โดยที่  $N$  คือตำแหน่งตัวอย่าง สุดท้ายของช่วงสัญญาณเสียงที่พิจารณา

- **ค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกิน ค่าศูนย์ของช่วงกักลม** สามารถหาได้ดังต่อไปนี้

$$S_{num} = \sum_{i=1}^N Zcr(x_i) \quad (3.6)$$

โดยที่  $S_{num}$  คือค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ของช่วงกักลม  $x_i$  คือสัญญาณเสียงช่วงกักลมตำแหน่งตัวอย่างที่  $i$  ส่วน  $i, N$  คือตำแหน่งตัวอย่าง (Sampling) ของสัญญาณเสียงช่วงกักลม โดยที่  $N$  คือตำแหน่งตัวอย่างสุดท้ายของช่วงสัญญาณเสียงที่พิจารณา และ  $Zcr(x_i)$  สามารถหาได้ดังนี้

$$Zcr(x_i) = \{ 1 \mid (x_i * x_{i-1}) < 0 \} \quad (3.7)$$

หรือ

$$Zcr(x_i) = \{ 0 \mid (x_i * x_{i-1}) > 0 \} \quad (3.8)$$

**3.3.3. จำแนกประเภทของสัญญาณเสียงตามฐานความรู้** เมื่อหาค่าลักษณะสำคัญทางสวณลักษณะของสัญญาณเสียงเรียบร้อยแล้ว จะส่งมายังขั้นตอนจำแนกประเภทของสัญญาณเสียง ในขั้นตอนนี้จะรับค่าลักษณะสำคัญทางสวณลักษณะที่ได้ไปจัดกลุ่มเปรียบเทียบกับข้อมูลค่าลักษณะสำคัญทางสวณลักษณะที่มีอยู่ในฐานความรู้สวณลักษณะของเสียงโหระหะ โหระหะแบบสติล และโหระหะแบบชนิดของเสียงพยัญชนะกักตันพยางค์ที่ใช้ฐานกรณเดียวกัน

ในการทำงานขั้นตอนนี้จะรับเงื่อนไขตำแหน่งฐานกรณของเสียงพยัญชนะกักตันพยางค์ภาษาไทยจากภายนอกมาเป็นเงื่อนไขในการจำแนกประเภทเสียงด้วย โดยเงื่อนไขที่ได้จะเป็นกุญแจในการดึงข้อมูลจากฐานความรู้สวณลักษณะ ยกตัวอย่างเช่น เลือกเงื่อนไขการออกเสียงเป็นฐานริมฝีปาก ขั้นตอนจำแนกประเภทจะดึงชุดข้อมูลค่าลักษณะสำคัญทางสวณลักษณะจากกลุ่มฐานริมฝีปากมาเป็นข้อมูลในการสวณลักษณะรูปแบบของเสียงโหระหะ โหระหะแบบสติล และโหระหะแบบชนิดของสัญญาณเสียงฐานริมฝีปาก

สำหรับการจำแนกประเภทสัญญาณเสียงนั้นจะนำหลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นตามที่ได้อธิบายไว้แล้วในหัวข้อ 2.8 มาใช้ควบคู่ไปกับค่าลักษณะสำคัญทางสวณลักษณะที่ได้มาจากหลักการสวณสัทศาสตร์ เพื่อช่วยในการจำแนกประเภทเสียงโหระหะ หรือโหระหะแบบสติล หรือโหระหะแบบชนิดของพยัญชนะต้นเสียงกักที่ใช้ฐานกรณเดียวกัน

## บทที่ 4

### การทดลองและผลการทดลอง

บทนี้จะนำเสนอเกี่ยวกับการทดลอง ผลการทดลอง และการวิเคราะห์ผลการทดลองของการจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ ซึ่งเป็นการทดลองเพื่อหาค่าความถูกต้องของการจำแนกเสียง โดยการทดลองนั้นจะแบ่งออกเป็น 4 ชุด ดังนี้

4.1. การประเมินความสามารถในการจำแนกประเภทเสียงพยัญชนะกักของค่าลักษณะสำคัญทางสัทศาสตร์ที่ใช้

4.2. การทดลองวัดความคลาดเคลื่อนการหาช่วงกักลม และช่วงปล่อยลมของพยัญชนะต้นเสียงกักโดยอัตโนมัติ

4.3. การทดลองหาความถูกต้องของการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสัทศาสตร์มาจำแนกประเภทเสียง

4.4. การทดลองหาความถูกต้องของการจำแนกประเภทเสียงพยัญชนะเสียงกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสัทศาสตร์มาจำแนกประเภทเสียง โดยใช้สัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่เป็นสัญญาณทดสอบ

โดยในการทดลองชุดที่ 4.1 และ 4.2 จะเปรียบเทียบกับผลการจำแนกประเภทและวัดของผู้ทำวิจัย ส่วนการทดลองชุดที่ 4.3 จะนำผลการจำแนกประเภทที่ได้เปรียบเทียบกับผลการจำแนกประเภทเสียงของผู้ทำวิจัย และจากวิธีของฮิดเดนมาร์คอฟ ส่วนในการทดลองชุดที่ 4.4 ผลการจำแนกประเภทที่ได้ไปเปรียบเทียบกับผลการจำแนกของผู้เชี่ยวชาญทางด้านสัทศาสตร์จำนวน 3 ท่าน และจากวิธีของฮิดเดนมาร์คอฟ รายละเอียดของวิธีการทดลอง ชุดข้อมูลที่ใช้ในการทดลอง ผลการทดลองของ และผลการวิเคราะห์การทดลอง ในแต่ละชุดการทดลองมีดังนี้

#### 4.1. การประเมินความสามารถในการจำแนกประเภทเสียงพยัญชนะกักของค่าลักษณะสำคัญทางสวณลักษณะที่ใช้

การประเมินความสามารถในการจำแนกประเภทเสียงพยัญชนะกักของค่าลักษณะสำคัญทางสวณลักษณะนั้น จะนำสัญญาณเสียงที่บันทึกไว้มาผ่านการตัดแบ่งช่วงกักลม และช่วงปล่อยลมโดยผู้ทำวิจัย จากนั้นนำไปผ่านขั้นตอนการจำแนกประเภทเสียง ซึ่งมีรายละเอียดดังนี้

##### 4.1.1. การเตรียมการทดลอง

4.1.1.1. บันทึกสัญญาณเสียงในรูปแบบ **สระ** (Vowel) **พยัญชนะ** (Stop Consonant) **สระ** (Vowel) ( $V_1-C_s-V_2$ ) โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักต้นพยางค์ มีด้วยกันเป็น 8 เสียงหลัก **บ ป พ ด ต ท ก ข** และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ สระอะ **อา อิ อี อุ ู แอะ แอ** ในการบันทึกจะบันทึกจากเสียงของคนปกติ เพศหญิง 5 คน และเพศชาย 5 คน รวมทั้งสิ้น 640 เสียง

4.1.1.2. นำสัญญาณเสียงทั้ง 640 เสียง มาตัดแบ่งเป็นส่วนของเสียงพยัญชนะกัก จากนั้นหาช่วงกักลม และช่วงปล่อยลมจากเสียงพยัญชนะกักที่ได้

##### 4.1.2. การทดลอง

4.1.2.1. นำช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักที่ได้จากขั้นตอนการเตรียมการทดลองมาสกัดหาค่าลักษณะสำคัญทางสวณลักษณะ 4 ค่าได้แก่ ความยาวช่วงกักลม ความยาวช่วงปล่อยลม ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำของสัญญาณเสียงช่วงกักลม และ ค่าพลังงานของช่วงปล่อยลม สำหรับวิธีการสกัดค่าลักษณะสำคัญทางสวณลักษณะนั้นได้อธิบายไว้แล้วในหัวข้อ 3.3.2

4.1.2.2. ค่าลักษณะสำคัญทางสวณลักษณะที่สกัดได้ เปรียบเสมือนเป็นค่าลักษณะประจำของแต่ละสัญญาณเสียง การทดสอบการจำแนกประเภทเสียงพยัญชนะกักนั้นจะใช้หลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นมาช่วยในการจำแนกประเภทเสียง โดยการให้ค่าลักษณะสำคัญทางสวณลักษณะของเสียง **บ** และเสียง **ด** มาสร้างเวกเตอร์ลักษณะเฉพาะของเสียงโฆชะ ของฐานริมฝีปาก และฐานปุ่มเหงือก ส่วนเสียงโฆชะแบบสถิตินั้นจะใช้เสียง **ป ต** และเสียง **ก** มาสร้างเวกเตอร์ลักษณะเฉพาะของฐานริมฝีปาก ฐานปุ่มเหงือก และฐานเพดานอ่อน สำหรับเสียงโฆชะแบบอนิตจะใช้เสียง **ท ท** และเสียง **ข** มาสร้างเวกเตอร์ลักษณะเฉพาะของฐานริมฝีปาก ฐานปุ่มเหงือก และฐานเพดานอ่อน

4.1.2.3. ในการทดสอบนั้นใช้หลักการสุ่มเลือกออกหนึ่งเพื่อตรวจสอบ (Leave-One-Out Cross Validation) มาช่วยในการทดสอบ โดยจะเลือกหยิบสัญญาณเสียงออกมาทีละสัญญาณเสียงมาเป็นเสียงทดสอบของการจำแนกประเภทเสียง จากนั้นจะให้สัญญาณเสียงที่เหลือ เป็นเสียงของชุดข้อมูลสอน ชุดข้อมูลสอนนั้นจะช่วยให้การวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นได้เรียนรู้ถึงลักษณะประจำของแต่ละสัญญาณเสียง โดยในการเรียนรู้นั้นจะใช้ค่าลักษณะสำคัญทางสวณลักษณะมาเป็นค่าลักษณะประจำของแต่ละสัญญาณเสียง สำหรับการทดสอบนี้จะแบ่งออกเป็น 3 ฐานเสียง ได้แก่ฐานริมฝีปาก ฐานปุ่มเหงือก และฐานเพดานอ่อน แต่ละฐานจะมีชุดข้อมูลสอนแยกจากกัน กล่าวคือในฐานริมฝีปากมีสัญญาณเสียง 240 เสียง ฐานปุ่มเหงือกมีสัญญาณเสียง 240 เสียง และฐานเพดานอ่อนมีสัญญาณเสียง 160 เสียง และในการสุ่มเลือกหยิบจะสุ่มเลือกหยิบภายในกลุ่มของฐานเดียวกัน

4.1.2.4. สุ่มเลือกหยิบสัญญาณเสียงมาทดสอบจนครบ จากนั้นบันทึกผลการทดลองที่ได้ ในการบันทึกนั้นจะเปรียบเทียบผลการทดลองที่ได้กับผลการจำแนกประเภทเสียงของผู้ทำวิจัย โดยในที่นี้จะถือว่าผลการจำแนกของผู้ทำวิจัยมีความถูกต้องเชื่อถือได้ เพื่อให้สามารถประเมินประสิทธิภาพในการจำแนกประเภทเสียงจากค่าลักษณะสำคัญทางสวณลักษณะทั้ง 4 ค่าได้

4.1.2.5. ผลการทดลองที่บันทึกจะถูกบันทึกในรูปแบบของตารางค่าความสับสน (Confusion Matrices) โดยที่จะให้แถวแนวนอนเป็นประเภทการจำแนกประเภทเสียงจากผู้ทำวิจัย ส่วนคอลัมน์แนวตั้งเป็นประเภทการจำแนกเสียงจากการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น ดังตารางที่ 4.1 4.2 และ 4.3

4.1.2.6. นอกเหนือไปจากค่าลักษณะสำคัญทางสวณลักษณะทั้ง 4 ค่า ยังมีค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแวก์เกินค่าศูนย์ในช่วงกักลม ที่น่าจะมีส่วนช่วยเพิ่มความสามารถในการจำแนกประเภทเสียงโดยการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น ดังนั้นในขั้นตอนนี้จะทำในลักษณะเดียวกับขั้นตอน 4.1.2.1 เพียงแต่จะเพิ่มให้มีการสกัด ค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแวก์เกินค่าศูนย์ในช่วงกักลม มาเป็นค่าลักษณะสำคัญทางสวณลักษณะอีกค่าหนึ่ง

4.1.2.7. นำค่าลักษณะสำคัญทางสวณลักษณะในข้อ 4.1.2.6 ทำการทดลองซ้ำตามขั้นตอน 4.1.2.2 ถึง 4.1.2.5 ผลการทดลองที่ได้จะแสดง ดังตารางที่ 4.4 4.5 และ 4.6



#### 4.1.3. ผลการทดลอง

ในส่วนนี้จะนำเสนอผลการทดลองที่ได้เมื่อใช้ค่าลักษณะสำคัญของทางสวณลักษณะ 4 ค่า ได้แก่ ความยาวช่วงกักลม ความยาวช่วงปล่อยลม ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำของสัญญาณเสียงช่วงกักลม และ ค่าพลังงานของช่วงปล่อยลม และผลการทดลองเมื่อใช้ค่าลักษณะสำคัญของทางสวณลักษณะ 5 ค่า โดยเพิ่มจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมเป็นค่าลักษณะสำคัญของทางสวณลักษณะอีกค่าหนึ่ง

ผลการทดลองที่ได้จะแสดงในรูปแบบของตารางค่าความสับสน (Confusion Matrices) โดยจะให้แถวแนวนอนเป็นผลการจำแนกประเภทเสียงจากผู้ทำวิจัย คอลัมน์แนวตั้งเป็นผลการจำแนกประเภทเสียงของการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น ในแนวเส้นทแยงมุมจะเป็นบริเวณที่ผลการจำแนกจากทั้งสองแบบมีค่าตรงกัน หรือคือเป็นจำนวนสัญญาณเสียงที่การวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นสามารถจำแนกประเภทเสียงได้ตรงกับการจำแนกของผู้ทำวิจัย ความสามารถที่ได้จะคำนวณในรูปแบบของร้อยละของความถูกต้อง โดยคำนวณจากจำนวนสัญญาณเสียงในแนวเส้นทแยงมุมส่วนด้วยจำนวนเสียงทั้งหมดในแต่ละฐาน (240 เสียง) คุณด้วยหนึ่งร้อย จะได้เป็นร้อยละความถูกต้อง

**ตารางที่ 4.1** ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานริมฝีปากเมื่อใช้ 4 ค่า ลักษณะสำคัญของทางสวณลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง

		ประเภทเสียงพยัญชนะกักได้จากการจำแนก โดยใช้ค่าลักษณะสำคัญของทางสวณลักษณะ 4 ค่า (ประเภทเสียงที่ทดสอบได้)			ร้อยละความ ถูกต้องในแต่ละ ประเภท เสียง (ร้อยละ)	จำนวนเสียง ที่ใช้ทดสอบ ในแต่ละ ประเภทเสียง (เสียง)
		โสมพะ (เสียง)	อโสมพะแบบ สิถิล (เสียง)	อโสมพะแบบ ธนิต (เสียง)		
ประเภทเสียงพยัญชนะ กักได้จากผู้ทำวิจัย (ประเภทเสียงข้างจริง)	โสมพะ	<u>62</u>	18	0	77.5	80
	อโสมพะแบบ สิถิล	20	<u>60</u>	0	75	80
	อโสมพะแบบ ธนิต	0	2	<u>78</u>	97.5	80
	ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง				<u>83.3</u>	
		จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)				<u>240</u>

จากตารางที่ 4.1 พบว่าการจำแนกประเภทเสียงอโหะระแบบธนิตของพยัญชนะต้นเสียงกักฐานริมฝีปากมีความถูกต้องมากกว่าประเภทเสียงอื่น โดยจะพบว่าผลการจำแนกจากชุดข้อมูลทดสอบนั้นมีค่าใกล้เคียงกับผลการจำแนกจากประเภทเสียงอ้างอิงถึง 78 เสียงจากสัญญาณเสียงอโหะระแบบธนิตที่ใช้ทดสอบจำนวน 80 เสียง ทั้งนี้เป็นเพราะเสียงอโหะระแบบธนิตของเสียงพยัญชนะกักในฐานริมฝีปากนั้นมีช่วงปล่อยลมที่ยาวกว่าประเภทเสียงอื่น ทำให้สามารถแยกประเภทเสียงโดยใช้ค่าความยาวช่วงปล่อยลมออกมาได้ง่ายกว่าประเภทเสียงอื่น ส่วนประเภทเสียงโหะระ และอโหะระแบบสถิตินั้นมีร้อยละความถูกต้องของการจำแนกประเภทเสียงใกล้เคียงกัน คือที่ร้อยละ 77.5 และ 75 ตามลำดับ ทั้งนี้จะสังเกตเห็นว่าสัญญาณเสียงที่คลาดเคลื่อนนั้นจะสลับกันระหว่างเสียงอโหะระแบบสถิตินั้น และเสียงโหะระ สาเหตุมาจากทั้งสองเสียงดังกล่าวมีความยาวช่วงปล่อยลมใกล้เคียงกัน ทำให้ค่าลักษณะสำคัญทางสวันลักษณะที่เป็นความยาวช่วงปล่อยลมนั้นไม่มีอิทธิพลในการจำแนกระหว่างสองเสียงนี้มากเท่าใดนัก การจำแนกที่ได้จะอาศัยค่าลักษณะสำคัญทางสวันลักษณะในช่วงกักลมเป็นหลักซึ่งจะพบว่าการจำแนกที่ได้สามารถมีความคลาดเคลื่อนได้ ทั้งนี้เพราะพลังงานที่ความถี่ต่ำช่วงกักลมนั้นสามารถถูกรบกวนจากสัญญาณเสียงอื่นได้ง่าย

**ตารางที่ 4.2** ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานปุ่มเหงือกเมื่อใช้ 4 ค่าลักษณะสำคัญทางสวันลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง

		ประเภทเสียงที่ได้จากการจำแนก โดยใช้ค่าลักษณะสำคัญทางสวันลักษณะ 4 ค่า (ประเภทเสียงที่ทดสอบได้)			ร้อยละความถูกต้องในแต่ละประเภทเสียง (ร้อยละ)	จำนวนเสียงที่ใช้ทดสอบในแต่ละประเภทเสียง (เสียง)
		โหะระ (เสียง)	อโหะระแบบสถิตินั้น (เสียง)	อโหะระแบบธนิต (เสียง)		
ประเภทเสียงที่ได้จาก ผู้ทำวิจัย (ประเภทเสียงอ้างอิง)	โหะระ	<u>70</u>	10	0	87.5	80
	อโหะระแบบสถิตินั้น	15	<u>65</u>	0	81.25	80
	อโหะระแบบธนิต	3	1	<u>76</u>	95	80
	ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง				<u>87.9</u>	
จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)						<u>240</u>

จากตารางที่ 4.2 พบว่าผลการจำแนกประเภทเสียงพยัญชนะกักของฐานปุ่มเหงือกนั้นมีแนวโน้มใกล้เคียงกับผลการจำแนกประเภทเสียงพยัญชนะกักของฐานริมฝีปาก กล่าวคือเสียงอโหชนะแบบชนิดจะมีความถูกต้องมากที่สุด ซึ่งในที่นี้มีมากถึงร้อยละ 87.9 ส่วนเสียงอโหชนะ และอโหชนะแบบสติดนั้นมีความถูกต้องของการจำแนกประเภทเสียงใกล้เคียงกัน โดยมีค่าความถูกต้องของการจำแนกประเภทเสียงที่ร้อยละ 87.5 และ 81.25 ตามลำดับ

**ตารางที่ 4.3** ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานเพดานอ่อนเมื่อใช้ 4 ค่า ลักษณะสำคัญทางสวลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง

		ประเภทเสียงที่ได้จากการจำแนก โดยใช้ค่าลักษณะสำคัญทางสวลักษณะ 4 ค่า (ประเภทเสียงที่ทดสอบได้)		ร้อยละความถูกต้องในแต่ละประเภทเสียง (ร้อยละ)	จำนวนเสียงที่ใช้ทดสอบในแต่ละประเภทเสียง (เสียง)
		อโหชนะแบบสติด (เสียง)	อโหชนะแบบชนิด (เสียง)		
ประเภทเสียงที่ได้จากผู้ทักวิจัย (ประเภทเสียงอ้างอิง)	อโหชนะแบบสติด	79	1	98.75	80
	อโหชนะแบบชนิด	3	77	96.25	80
		ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง		97.5	
		จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)			160

จากตารางที่ 4.3 พบว่าการจำแนกเสียงพยัญชนะกักในฐานเพดานอ่อนมีค่ามากที่สุดเมื่อเปรียบเทียบกับฐานริมฝีปาก และฐานปุ่มเหงือก โดยมีความถูกต้องของเสียงอโหชนะแบบสติดอยู่ที่ร้อยละ 98.75 และของเสียงอโหชนะแบบชนิดอยู่ที่ร้อยละ 96.25 สาเหตุที่การจำแนกเสียงพยัญชนะกักฐานเพดานอ่อนนั้นมีความถูกต้องมากที่สุด เนื่องมาจากในฐานดังกล่าวมีเพียงสองเสียงเท่านั้น คือเสียงอโหชนะแบบสติด และเสียงอโหชนะแบบชนิด ซึ่งพบค่าความยาวของช่วงปล่อยลมมีส่วนช่วยให้การจำแนกทั้งสองสามารถทำได้ถูกต้องแม่นยำ

ผลการทดลองถัดมาจะเป็นผลการทดลองเมื่อใช้ค่าลักษณะสำคัญทางสวลลักษณะจำนวน 5 ค่า มาช่วยในการจำแนกประเภทเสียงพยัญชนะกักต้นพยางค์ในแต่ละฐานกรณ

ตารางที่ 4.4 ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานริมฝีปากเมื่อใช้ 5 ค่า  
ลักษณะสำคัญทางสวันลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง

		ประเภทเสียงพยัญชนะกักได้จากการจำแนก โดยใช้ค่าลักษณะสำคัญทางสวันลักษณะ 4 ค่า (ประเภทเสียงที่ทดสอบได้)			ร้อยละความ ถูกต้องในแต่ละ ประเภท เสียง (ร้อยละ)	จำนวนเสียง ที่ใช้ทดสอบ ในแต่ละ ประเภทเสียง (เสียง)
		โสมพะ (เสียง)	อโสมพะแบบ สีถิล (เสียง)	อโสมพะแบบ ธนิต (เสียง)		
ประเภทเสียงพยัญชนะ กักได้จากผู้ทักวิจัย (ประเภทเสียงข้างอิง)	โสมพะ	<u>67</u>	13	0	83.75	80
	อโสมพะแบบ สีถิล	22	<u>58</u>	0	72.5	80
	อโสมพะแบบ ธนิต	0	1	<u>79</u>	98.75	80
	ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง				<u>85</u>	
จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)						<u>240</u>

ตารางที่ 4.5 ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานปุ่มเหงือกเมื่อใช้ 5 ค่า  
ลักษณะสำคัญทางสวันลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง

		ประเภทเสียงพยัญชนะกักได้จากการจำแนก โดยใช้ค่าลักษณะสำคัญทางสวันลักษณะ 4 ค่า (ประเภทเสียงที่ทดสอบได้)			ร้อยละความ ถูกต้องในแต่ละ ประเภท เสียง (ร้อยละ)	จำนวนเสียง ที่ใช้ทดสอบ ในแต่ละ ประเภทเสียง (เสียง)
		โสมพะ (เสียง)	อโสมพะแบบ สีถิล (เสียง)	อโสมพะแบบ ธนิต (เสียง)		
ประเภทเสียงพยัญชนะ กักได้จากผู้ทักวิจัย (ประเภทเสียงข้างอิง)	โสมพะ	<u>73</u>	7	0	91.25	80
	อโสมพะแบบ สีถิล	14	<u>66</u>	0	82.5	80
	อโสมพะแบบ ธนิต	3	1	<u>76</u>	95	80
	ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง				<u>89.5</u>	
จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)						<u>240</u>

**ตารางที่ 4.6** ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยฐานพีดานอ่อนเมื่อใช้ 5 ค่า  
ลักษณะสำคัญทางสวันลักษณะเป็นค่าลักษณะประจำของการจำแนกประเภทเสียง

		ประเภทเสียงที่ได้จากการจำแนก โดยใช้ค่า ลักษณะสำคัญทางสวันลักษณะ 4 ค่า (ประเภทเสียงที่ทดสอบได้)		ร้อยละความ ถูกต้องในแต่ ละประเภท เสียง (ร้อยละ)	จำนวนเสียง ที่ใช้ทดสอบ ในแต่ละ ประเภทเสียง (เสียง)
		อโหชนะแบบ สถิต (เสียง)	อโหชนะแบบ ธนิต (เสียง)		
ประเภทเสียงที่ได้จาก ผู้ทำวิจัย (ประเภทเสียงอ้างอิง)	อโหชนะแบบ สถิต	79	1	98.75	80
	อโหชนะแบบ ธนิต	2	78	97.5	80
		ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง		98	
		จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)			160

จากตารางที่ 4.4 4.5 และ 4.6 พบว่าเมื่อเพิ่มค่าลักษณะสำคัญทางสวันลักษณะ  
จาก 4 ค่าอันได้แก่ ความยาวช่วงกักลม ความยาวช่วงปล่อยลม ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำ  
ของสัญญาณเสียงช่วงกักลม และ ค่าพลังงานของช่วงปล่อยลม เป็น 5 ค่า โดยเพิ่มค่าจำนวนครั้ง  
ของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมเข้าไปช่วยในการจำแนกประเภทเสียง  
จะช่วยให้การจำแนกเสียงโหชนะของฐานริมฝีปากมีความถูกต้องเพิ่มจากร้อยละ 77.5 เป็น 83.75  
และเสียงโหชนะของฐานปุ่มเหงือกมีความถูกต้องเพิ่มจากร้อยละ 87.5 เป็น 91.25 รวมทั้งเสียง  
อโหชนะแบบสถิตของฐานปุ่มเหงือกซึ่งมีความถูกต้องเพิ่มจากร้อยละ 81.25 เป็น 82.5 ถึงแม้ว่า  
ความถูกต้องในการจำแนกประเภทเสียงอโหชนะแบบสถิตจะมีค่าลดลงในฐานริมฝีปาก จากร้อยละ  
75 เป็น 72.5 ในฐานริมฝีปาก แต่ความถูกต้องโดยรวมของฐานริมฝีปาก และฐานปุ่มเหงือกยังคง  
มีค่าเพิ่มขึ้นจากร้อยละ 83.3 เป็น 85 และร้อยละ 87.9 เป็น 89.5 ตามลำดับ ทั้งนี้เป็นเพราะว่า  
จำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมนั้นจะมีส่วนช่วยในการ  
จำแนกประเภทเสียงโหชนะ และอโหชนะ [14] ทำให้เมื่อนำมาใช้เป็นค่าลักษณะสำคัญทางสวัน  
ลักษณะจะส่งผลการการจำแนกประเภทเสียงโหชนะ ออกจากเสียงอโหชนะแบบสถิตมีประสิทธิภาพ  
มากขึ้น ทั้งนี้ค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมของเสียง  
โหชนะจะมีค่าน้อยกว่าเสียงอโหชนะแบบสถิต เนื่องจากว่าเสียงโหชนะนั้นจะมีพลังงานที่ความถี่ต่ำ

ในช่วงกักลม ทำให้บริเวณช่วงกักลมดังกล่าวมีลักษณะของการแกว่งน้อยลง หรือมีค่าแอมพลิจูดสูงกว่าของเสียงอโหพระแบบสถิต

จากผลการทดลองในฐานเพดานอ่อนนั้นพบว่า การเพิ่มจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมมาเป็นค่าลักษณะสำคัญทางสวณลักษณะ ไม่มีส่วนช่วยเพิ่มประสิทธิภาพในการจำแนกเสียงมากนัก ทั้งนี้เพราะว่าประเภทเสียงในฐานเพดานอ่อนนั้นมีเพียง เสียงอโหพระแบบสถิต และเสียงอโหพระแบบชนิดเท่านั้น มิได้มีเสียงอโหพระ หรือกล่าวอีกนัยหนึ่งคือ จำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมจะมีส่วนช่วยในการจำแนกประเภทเสียงอโหพระ กับอโหพระเท่านั้น แต่มิได้ช่วยให้การจำแนกประเภทเสียงระหว่าง เสียงอโหพระแบบสถิต และเสียงอโหพระแบบชนิดมากนัก

#### 4.2. การทดลองวัดความคลาดเคลื่อนการหาช่วงกักลม และช่วงปล่อยลมของพยัญชนะต้นเสียงกักโดยอัตโนมัติ

การทดลองนี้จะแบ่งชุดข้อมูลที่ใช้ทดสอบออกเป็น 2 ชุดย่อย คือ ชุดข้อมูลสอน และชุดข้อมูลทดสอบ และในการทดลองจะตรวจสอบความยาวของช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักที่ได้จากการหาโดยอัตโนมัติกับการหาโดยผู้ทำวิจัย สำหรับผลความคลาดเคลื่อนที่ได้จะแสดงในรูปแบบของค่าคลาดเคลื่อนเฉลี่ย ค่าคลาดเคลื่อนมากที่สุดของแต่ละชุดข้อมูลที่ใช้ทดสอบ โดยมีรายละเอียดต่างๆ ดังต่อไปนี้

##### 4.2.1. การเตรียมการทดลอง

4.2.1.1. ในการทดลองชุดนี้จะแบ่งชุดข้อมูลออกเป็น 2 ชุดย่อย คือ ชุดข้อมูลสอน และชุดข้อมูลทดสอบ โดยชุดข้อมูลสอนนั้นมาจากสัญญาณเสียงของคนปกติ 10 คนที่ออกเสียงพูดในรูปแบบ **สระ (Vowel) พยัญชนะ (Stop Consonant) สระ (Vowel) (V<sub>1</sub>-C<sub>s</sub>-V<sub>2</sub>)** โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักต้นพยางค์ มีด้วยกันเป็น 8 เสียงหลัก **บ ป พ ด ต ท ก ข** และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ สระ **อะ อา อิ อี อุ อู แอะ แอ** รวมทั้งสิ้น 640 เสียง ส่วนชุดข้อมูลทดสอบนั้นจะถูกบันทึกในลักษณะเดียวกันกับชุดข้อมูลสอน เพียงแต่เป็นสัญญาณเสียงจากกลุ่มคนปกติกลุ่มใหม่จำนวน 4 คน มีสัญญาณเสียงทั้งสิ้น 256 เสียง

4.2.1.2. นำสัญญาณเสียงทั้ง 2 ชุด ไปทดสอบการหาช่วงกักลม และช่วงปล่อยของพยัญชนะต้นเสียงกัก โดยในการทดลองชุดแรกจะใช้ชุดข้อมูลสอน และในการทดลองที่สองจะใช้กับชุดข้อมูลทดสอบ การทดลองกับชุดข้อมูลสอนนั้นถูกสร้างขึ้นเพื่อใช้ในการปรับกระบวนการหาช่วงอัตโนมัติให้มีความแม่นยำมากที่สุด เมื่อได้กระบวนการที่มีความแม่นยำมากที่สุดแล้ว จะนำกระบวนการที่ได้ไปทดสอบกับชุดข้อมูลทดสอบ เพื่อทดสอบว่ากระบวนการที่ได้มีได้ขึ้นอยู่กับชุดข้อมูลสอนเท่านั้น แต่สามารถใช้ได้กับสัญญาณเสียงทั่วไป

#### 4.2.2. การทดลอง

4.2.2.1. นำสัญญาณเสียงทั้ง 640 เสียงจากชุดข้อมูลสอน ไปทดสอบการหาช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักโดยอัตโนมัติ

4.2.2.2. หาความยาวช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงทั้ง 640 เสียงจากชุดข้อมูลสอนโดยใช้การพิจารณาจากผู้ทำวิจัย

4.2.2.3. นำความยาวช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักในชุดข้อมูลสอน ที่ได้จากวิธีการหาโดยอัตโนมัติ และโดยการพิจารณาจากผู้ทำวิจัยมาเปรียบเทียบกัน เพื่อตรวจสอบหาความคลาดเคลื่อน

4.2.2.4. นำสัญญาณเสียงของชุดข้อมูลทดสอบ 256 เสียง ไปทดสอบการหาช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักโดยอัตโนมัติ

4.2.2.5. หาความยาวช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงทั้ง 256 เสียงจากชุดข้อมูลทดสอบ โดยใช้การพิจารณาจากผู้ทำวิจัย

4.2.2.6. นำความยาวช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักในชุดข้อมูลทดสอบ ที่ได้จากวิธีการหาโดยอัตโนมัติ และโดยการพิจารณาจากผู้ทำวิจัยมาเปรียบเทียบกัน เพื่อตรวจสอบหาความคลาดเคลื่อน

#### 4.2.3. ผลการทดลอง

ในส่วนนี้จะนำเสนอผลการทดลองที่ได้จากการหาช่วงกักลม และช่วงปล่อยลมของเสียงพญานะกักตันพยางค์จากชุดข้อมูลสอน และชุดข้อมูลทดสอบโดยอัตโนมัติ โดยค่าที่ได้จะแสดงในรูปแบบของค่าคลาดเคลื่อนเฉลี่ย และค่าคลาดเคลื่อนมากที่สุดของแต่ละชุดข้อมูล เมื่อเปรียบเทียบกับค่าที่ได้จากวิธีการพิจารณาของผู้ทำวิจัย

ตารางที่ 4.7 ค่าคลาดเคลื่อนเฉลี่ย และค่าคลาดเคลื่อนมากที่สุดของช่วงกักลม จากชุดข้อมูลสอน และชุดข้อมูลทดสอบ

ประเภทเสียงของ ช่วงกักลม	ชุดข้อมูลสอน		ชุดข้อมูลทดสอบ	
	ค่าคลาดเคลื่อน เฉลี่ย (มิลลิวินาที)	ค่าคลาดเคลื่อน มากที่สุด (มิลลิวินาที)	ค่าคลาดเคลื่อน เฉลี่ย (มิลลิวินาที)	ค่า คลาดเคลื่อน มากที่สุด (มิลลิวินาที)
<b>พญานะเสียงกักภาษาไทยที่ใช้ฐานริมฝีปาก</b>				
เสียงโฆชะ	14.10	108.62	8.40	42.38
เสียงอโฆชะแบบสติด	20.36	87.07	16.23	124.19
เสียงอโฆชะแบบชนิด	16.15	68.95	26.50	106.07
<b>พญานะเสียงกักภาษาไทยที่ใช้ฐานเพดานอ่อน</b>				
เสียงโฆชะ	13.12	48.30	10.09	28.63
เสียงอโฆชะแบบสติด	23.88	75.78	8.62	31.24
เสียงอโฆชะแบบชนิด	15.93	66.08	15.76	43.76
<b>พญานะเสียงกักภาษาไทยที่ใช้ฐานปุ่มเหงือก</b>				
เสียงอโฆชะแบบสติด	23.88	75.78	24.43	63.26
เสียงอโฆชะแบบชนิด	15.93	66.10	29.56	106.67



ตารางที่ 4.8 ค่าคลาดเคลื่อนเฉลี่ย และค่าคลาดเคลื่อนมากที่สุดของช่วงปล่อยลม จากชุดข้อมูล สอน และชุดข้อมูลทดสอบ

ประเภทเสียงของ ช่วงปล่อยลม	ชุดข้อมูลสอน		ชุดข้อมูลทดสอบ	
	ค่าคลาดเคลื่อน เฉลี่ย (มิลลิวินาที)	ค่าคลาดเคลื่อน มากที่สุด (มิลลิวินาที)	ค่าคลาดเคลื่อน เฉลี่ย (มิลลิวินาที)	ค่า คลาดเคลื่อน มากที่สุด (มิลลิวินาที)
<b>พยัญชนะเสียงกักภาษาไทยที่ใช้ฐานริมฝีปาก</b>				
เสียงโหระ	7.64	23.99	6.73	20.54
เสียงอโหระแบบสถิต	9.43	81.54	19.39	251.24
เสียงอโหระแบบชนิด	15.06	85.94	22.51	100.65
<b>พยัญชนะเสียงกักภาษาไทยที่ใช้ฐานเพดานอ่อน</b>				
เสียงโหระ	9.91	105.62	8.30	17.86
เสียงอโหระแบบสถิต	10.24	186.28	3.48	10.06
เสียงอโหระแบบชนิด	16.00	47.64	26.49	166.82
<b>พยัญชนะเสียงกักภาษาไทยที่ใช้ฐานปุ่มเหงือก</b>				
เสียงอโหระแบบสถิต	10.24	186.28	16.97	196.00
เสียงอโหระแบบชนิด	16.00	47.64	26.75	272.36

จากตารางที่ 4.7 จะพบว่าค่าคลาดเคลื่อนเฉลี่ยของการตัดช่วงกักลมพยัญชนะ กักโดยอัตโนมัติของชุดข้อมูลสอน มีค่าคลาดเคลื่อนเฉลี่ยประมาณ 13 ถึง 24 มิลลิวินาที เมื่อนำ ช่วงกักลมที่หาได้ไปเปรียบเทียบกับความยาวทั้งหมดของสัญญาณเสียง พบว่ามีค่าคลาดเคลื่อน เฉลี่ยร้อยละ 3 ถึง 5 ในส่วนของชุดข้อมูลทดสอบ นั้นมีค่าคลาดเคลื่อนเฉลี่ยของการหาช่วงกักลม โดยอัตโนมัติอยู่ที่ประมาณ 8 ถึง 30 มิลลิวินาที เมื่อนำช่วงกักลมที่หาได้ไปเปรียบเทียบกับความ

ยาวทั้งหมดของสัญญาณเสียง พบว่ามีค่าคลาดเคลื่อนเฉลี่ยร้อยละ 2 ถึง 5 จะเห็นได้ว่าค่าคลาดเคลื่อนช่วงกักลมที่ได้จากทั้ง 2 ชุดทดลองมีค่าใกล้เคียงกัน

จากตารางที่ 4.8 จะพบว่าค่าคลาดเคลื่อนเฉลี่ยของการตัดช่วงปล่อยลมพัลซุญชนะกักโดยอัตโนมัติของชุดข้อมูลสอน มีค่าคลาดเคลื่อนเฉลี่ยประมาณ 8 ถึง 16 มิลลิวินาที เมื่อนำช่วงปล่อยลมที่หาได้ไปเปรียบเทียบกับความยาวทั้งหมดของสัญญาณเสียง พบว่ามีค่าคลาดเคลื่อนเฉลี่ยร้อยละ 2 ถึง 3 ในส่วนของชุดข้อมูลทดสอบ นั้นมีค่าคลาดเคลื่อนเฉลี่ยของการหาช่วงปล่อยลม โดยอัตโนมัติอยู่ที่ประมาณ 7 ถึง 27 มิลลิวินาที เมื่อนำช่วงปล่อยลมที่หาได้ไปเปรียบเทียบกับความยาวทั้งหมดของสัญญาณเสียง พบว่ามีค่าคลาดเคลื่อนเฉลี่ยร้อยละ 1 ถึง 5 จะเห็นได้ว่าค่าคลาดเคลื่อนช่วงปล่อยลมที่ได้จากทั้ง 2 ชุดทดลองมีค่าใกล้เคียงกัน สามารถเชื่อถือได้ระดับหนึ่งว่าวิธีการหาช่วงกักลม และช่วงปล่อยลมของวิทยานิพนธ์นี้สามารถนำช่วยหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพัลซุญชนะกักต้นพยางค์โดยอัตโนมัติได้

#### 4.3. การทดลองหาความถูกต้องของการจำแนกประเภทเสียงพัลซุญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสวนศาสตร์มาจำแนกประเภทเสียง

ในการทดลองนี้แบ่งออกเป็นสองชุดย่อย คือชุดข้อมูลสอน และชุดข้อมูลทดสอบ โดยข้อมูลแต่ละชุดจะถูกนำมาทดสอบการจำแนกประเภทเสียงโดยอัตโนมัติจากการใช้หลักการสวนศาสตร์ และหลักการฮิดเดนมาร์คอฟ เพื่อเปรียบเทียบความถูกต้องของการจำแนกประเภทเสียงได้จากทั้งสองหลักการ

จุดประสงค์ของการทดสอบกับชุดข้อมูลสอนนั้นเพื่อใช้สร้างวิธีการจำแนกประเภทเสียงอัตโนมัติที่ได้จากหลักการสวนศาสตร์ ส่วนจุดประสงค์ของการทดสอบกับชุดทดสอบนั้นเพื่อทำให้เกิดความมั่นใจว่าวิธีการจำแนกโดยอัตโนมัติที่ได้ มิได้ขึ้นอยู่กับชุดข้อมูลสอนแต่เพียงอย่างเดียว แต่สามารถใช้ได้กับชุดข้อมูลทดสอบด้วย โดยมีรายละเอียดของการทดลองดังต่อไปนี้

#### 4.3.1. การเตรียมการทดลอง

4.3.1.1. ในการทดลองชุดนี้จะแบ่งชุดข้อมูลออกเป็น 2 ชุดย่อย คือ ชุดข้อมูลสอน และชุดข้อมูลทดสอบ โดยชุดข้อมูลสอนนั้นมาจากสัญญาณเสียงของคนปกติ 10 คนที่ออกเสียงพูดในรูปแบบ **สระ (Vowel) พยัญชนะ (Stop Consonant) สระ (Vowel)** ( $V_1-C_s-V_2$ ) โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักตันพยางค์ มีด้วยกันเป็น 8 เสียงหลัก **บ ป พ ด ต ท ก ข** และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ **สระอะ อา อี อี้ อุ อู แอะ แอ** รวมทั้งสิ้น 640 เสียง ส่วนชุดข้อมูลทดสอบนั้นจะบันทึกในลักษณะเดียวกันกับชุดข้อมูลสอน เพียงแต่เป็นสัญญาณเสียงจากกลุ่มคนปกติกลุ่มใหม่จำนวน 4 คน มีสัญญาณเสียงทั้งสิ้น 256 เสียง

4.3.1.2. นำสัญญาณเสียงทั้ง 2 ชุด ไปทดสอบการจำแนกประเภทเสียงโดยอัตโนมัติตามหลักการสวนสัทศาสตร์ และตามหลักการของฮิดเดนมาร์คอฟ

#### 4.3.2. การทดลอง

4.3.2.1. นำสัญญาณเสียงจากชุดข้อมูลสอนทั้ง 640 เสียงไปผ่านการจำแนกประเภทเสียงโดยอัตโนมัติตามหลักการสวนสัทศาสตร์ โดยเริ่มจากการหาช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักโดยอัตโนมัติ จากนั้นสกัดค่าลักษณะสำคัญทางสวนลักษณะทั้ง 5 ค่า จากช่วงกักลม และช่วงปล่อยลมดังกล่าว ในการจำแนกประเภทเสียงจะใช้หลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นมาช่วยในการจำแนกประเภทเสียง

4.3.2.2. เนื่องจากชุดข้อมูลที่ใช้มีจำกัด ในการทดสอบนี้จะใช้วิธีการสุ่มเลือกออกหนึ่งเพื่อตรวจสอบ (Leave-One-Out Cross Validation) มาช่วยในการทดสอบ โดยจะเลือกหยิบสัญญาณเสียงออกมาทีละสัญญาณเสียงมาเป็นเสียงทดสอบการจำแนกประเภทเสียง จากนั้นจะให้สัญญาณเสียงที่เหลือ เป็นเสียงของชุดข้อมูลสอน โดยในการทดสอบจะแบ่งออกเป็น 3 ฐานเสียง ได้แก่ฐานริมฝีปาก ฐานปุ่มเหงือก และฐานเพดานอ่อน แต่ละฐานจะมีชุดข้อมูลแยกจากกัน กล่าวคือในฐานริมฝีปากมีสัญญาณเสียง 240 เสียง ฐานปุ่มเหงือกมีสัญญาณเสียง 240 เสียง และฐานเพดานอ่อนมีสัญญาณเสียง 160 เสียง และในการสุ่มเลือกหยิบจะสุ่มเลือกหยิบภายในกลุ่มของฐานเดียวกัน

4.3.2.3. นำผลการจำแนกประเภทเสียงโดยอัตโนมัติที่ได้จากหลักการ  
สวณศาสตร์ไปเปรียบเทียบกับผลการจำแนกประเภทเสียงของผู้ทำวิจัย

4.3.2.4. หลังจากที่ได้ทราบถึงความถูกต้องของการจำแนกประเภทเสียง  
โดยอัตโนมัติตามหลักการสวณศาสตร์ นำสัญญาณเสียงชุดข้อมูลสอนดังกล่าวไปทดสอบการ  
จำแนกประเภทเสียงโดยอัตโนมัติตามหลักการฮิดเดนมาร์คอฟ ซึ่งในที่นี้ได้อาศัยโปรแกรม  
Hidden markov Toolkit – HTK [6] มาช่วยในการทดสอบ โดยมีรายละเอียดการใช้งานดังต่อไปนี้

4.3.2.4.1. แปลงสัญญาณเสียงที่จะใช้เป็นชุดข้อมูลสอน  
เป็นไฟล์ค่าสัมประสิทธิ์เซปตริ่มสเกลเมล โดยใช้ HCopy โปรแกรม จากชุดเครื่องมือ HTK โดยจะ  
คำนวณหาสัมประสิทธิ์เซปตริ่มของทุกกรอบเวลา 25 มิลลิวินาที และแต่ละกรอบของเวลาจะมี  
ระยะห่างกัน 10 มิลลิวินาที ได้ออกมาเป็นเวกเตอร์ลักษณะสำคัญที่มีขนาด 39 มิติ และความถี่  
ของสัญญาณเสียงอยู่ที่ 44100 เฮิรตซ์

4.3.2.4.2. เตรียมไฟล์พจนานุกรม เพื่อแสดงรายการของ  
คำศัพท์การรู้จำเสียง

4.3.2.4.3. เตรียมไฟล์รายการหน่วยเสียง

4.3.2.4.4. เตรียมไฟล์ต้นแบบทอพอโลยีของแบบจำลอง  
เสียง (HMM Prototype File) โดยเลือกใช้แบบจำลองฮิดเดนมาร์คอฟที่มีจำนวนสถานะ 5  
สถานะ (proto5s)

4.3.2.4.5. สร้างต้นแบบเสียงเดี่ยว (Monophone Model)  
โดยใช้โปรแกรม HCompV จากชุดเครื่องมือ HTK

4.3.2.4.6. ประมาณค่าพารามิเตอร์ของต้นแบบเสียงเดี่ยว  
จำนวน 3 รอบ โดยใช้โปรแกรม HERest จากชุดเครื่องมือ HTK

4.3.2.4.7. สร้างต้นแบบสามเสียง (Triphone Model)  
จากต้นแบบเสียงเดี่ยว โดยใช้โปรแกรม HLed จากชุดเครื่องมือ HTK

4.3.2.4.8. ประมาณค่าพารามิเตอร์ของต้นแบบสามเสียง  
จำนวน 3 รอบ โดยใช้โปรแกรม HERest จากชุดเครื่องมือ HTK

4.3.2.4.9. นำสัญญาณเสียงชุดทดสอบมาทดสอบการ  
จำแนกเสียงโดยใช้โปรแกรม HVite จากชุดเครื่องมือ HTK

4.3.2.5. นำสัญญาณเสียงจากชุดข้อมูลทดสอบทั้ง 256 เสียงไปผ่านการ  
ทดสอบในลักษณะเดียวกับวิธีการทดสอบของชุดข้อมูลสอน เพียงแต่ในการทดสอบนี้จะไม่อาศัย  
วิธีการสุ่มเลือกออกหนึ่งเพื่อตรวจสอบ (Leave-One-Out Cross Validation) มาช่วยในการ  
ทดสอบแต่จะใช้ชุดข้อมูลสอนจากข้างต้นมาเป็นชุดข้อมูลสอนของการวิเคราะห์แบบดิสคริมิแนนต์  
เชิงเส้น และให้ชุดข้อมูลทดสอบทั้ง 256 เสียงมาเป็นสัญญาณทดสอบ จะเห็นได้ว่าการแบ่งแยก  
ชุดข้อมูลสอน และชุดข้อมูลทดสอบออกจากกันโดยชัดเจน

4.3.2.6. หลังจากนำชุดข้อมูลทดสอบไปทดสอบการจำแนกประเภทเสียง  
ตามหลักการสวนศาสตร์ และหลักการฮิดเดนมาร์คอฟแล้ว เพื่อให้เกิดความมั่นใจมากขึ้น จะ  
ทำการทดสอบซ้ำจำนวน 3 รอบ โดยชุดข้อมูลทดสอบที่ได้จะถูกสุ่มเลือกขึ้นมาจากเสียงคนปกติ 4  
คน รวมทั้งสิ้น 256 เสียง และให้สัญญาณเสียงของ คนปกติ 10 คนที่เหลือเป็นสัญญาณเสียงชุด  
ข้อมูลสอน

4.3.2.7. บันทึกผลที่ได้ พร้อมทั้งแสดงผลการทดลองในรูปแบบของค่า  
ความถูกต้องโดยเฉลี่ยจากผลการทดลองทั้ง 3 รอบดังกล่าว

### 4.3.3. ผลการทดลอง

ในส่วนนี้จะนำเสนอผลการทดลองการจำแนกประเภทเสียงพยัญชนะกัก  
ภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อทดสอบกับชุดข้อมูลสอน  
และชุดข้อมูลทดสอบ โดยใช้วิธีการจำแนกประเภทเสียงตามหลักการสวนศาสตร์ และตาม  
หลักการฮิดเดนมาร์คอฟ ซึ่งผลการจำแนกประเภทเสียงที่ได้จากแต่ละหลักการนั้นจะถูกนำไป  
เปรียบเทียบกับผลการจำแนกประเภทเสียงของผู้ทำวิจัย แล้วแสดงผลในรูปแบบของร้อยละค่า  
ความถูกต้องของแต่ละหลักการ

ตารางที่ 4.9 ผลการจำแนกเสียงพยางค์ของนักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อทดสอบกับชุดข้อมูลสอน

ประเภทของเสียงพยางค์	ค่าความถูกต้องที่ได้จากวิธี ของหลักการสัทศาสตร์ (ร้อยละ)	ค่าความถูกต้องที่ได้จากวิธี ของฮิดเดนมาร์คอฟ (ร้อยละ)
ฐานริมฝีปาก	86	65
ฐานเพดานอ่อน	83	80
ฐานปุ่มเหงือก	91	85

จากตารางที่ 4.9 พบว่าความถูกต้องที่ได้จากการการจำแนกเสียงพยางค์ของนักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์จากหลักการสัทศาสตร์ เมื่อใช้กับชุดข้อมูลสอน มีความถูกต้องมากกว่าการจำแนกประเภทเสียงพยางค์ของนักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์จากหลักการของฮิดเดนมาร์คอฟ ทั้งนี้พบว่าการทดลองชุดข้อมูลสอนนั้นเมื่อถูกจำแนกประเภทเสียงตามหลักการสัทศาสตร์ มีความถูกต้องของการจำแนกประเภทเสียงมากถึงร้อยละ 86 83 และ 91 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ ในขณะที่ความถูกต้องของการจำแนกประเภทเสียงตามวิธีฮิดเดนมาร์คอฟมีค่าเพียงร้อยละ 65 80 และ 85 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ สาเหตุที่การจำแนกประเภทเสียงพยางค์จากชุดข้อมูลสอนตามหลักการสัทศาสตร์มีความถูกต้องมากกว่าตามหลักการฮิดเดนมาร์คอฟนั้น อาจเนื่องมาจากการผูกติดกันระหว่างลักษณะจำเพาะของข้อมูลในชุดข้อมูลสอน ที่เป็นชุดข้อมูลช่วยสร้าง กับขั้นตอนการตัดสินใจของกระบวนการวิเคราะห์ตามหลักการสัทศาสตร์

ตารางที่ 4.10 ผลการจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อทดสอบกับชุดข้อมูลทดสอบ

ประเภทของเสียงพยัญชนะ	ค่าความถูกต้องที่ได้จากวิธีของ หลักการสวณศาสตร์ (ร้อยละ)	ค่าความถูกต้องที่ได้จากวิธีของ ฮิดเดนมาร์คอฟ (ร้อยละ)
<b>ความถูกต้องเฉลี่ยจากการทดสอบจำนวน 3 รอบ</b>		
ฐานริมฝีปาก	83	81
ฐานเพดานอ่อน	89.33	88.33
ฐานปุ่มเหงือก	88	85.33

จากตารางที่ 4.10 พบว่าความถูกต้องที่ได้จากการจำแนกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์จากวิธีของหลักการสวณศาสตร์ มีความถูกต้องค่อนข้างมากกว่าผลการจำแนกเสียงจากวิธีของฮิดเดนมาร์คอฟ ทั้งนี้พบว่าความถูกต้องเฉลี่ยที่ได้มีค่าร้อยละเป็น 83 89.33 และ 88 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ ส่วนความถูกต้องที่ได้จากวิธีฮิดเดนมาร์คอฟมีค่าร้อยละเป็น 81 88.33 และ 85.33 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ

จากผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ของหลักการสวณศาสตร์ และผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ของหลักการฮิดเดนมาร์คอฟ แสดงให้เห็นวิธีการของหลักการสวณศาสตร์ค่อนข้างมีความถูกต้องมากกว่าการจำแนกประเภทเสียงพยัญชนะกักของวิธีฮิดเดนมาร์คอฟ ถึงแม้ว่าร้อยละความถูกต้องเฉลี่ยที่ได้จะมากกว่าเพียงร้อยละ 1 ถึง 2 เท่านั้น แต่ประโยชน์ที่ได้จากการจำแนกประเภทเสียงของหลักการสวณศาสตร์นั้นจะมากกว่าในกรณีที่สำคัญเสียงที่เข้ามาทดสอบการจำแนกประเภทนั้นเป็นสัญญาณเสียงที่ออกเสียงไม่ถูกต้อง เพราะหากการผิดพลาดนั้นมาจากการเปล่งเสียงไม่ตรงตามหลักการสวณศาสตร์ของการเปล่งเสียงพยัญชนะกักภาษาไทย ค่าลักษณะสำคัญทางสวณลักษณะที่ได้จากการวิเคราะห์ในช่วงสัญญาณเสียงพยัญชนะกักของการจำแนกประเภทนั้น อาจช่วยบ่งบอกถึงที่มาของปัญหาจากหลักการสวณศาสตร์ได้ โดยจะเห็นได้จากผลการทดลองชุดถัดไป

4.4. การทดลองหาความถูกต้องของการจำแนกประเภทเสียงพยัญชนะเสียงกักภาษาไทย ในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสัทศาสตร์มา จำแนกประเภทเสียง โดยใช้สัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่เป็น สัญญาณทดสอบ

ในการทดลองนี้จะใช้สัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ จำนวน 3 ราย มาเป็นชุดข้อมูลทดสอบ โดยชุดข้อมูลทดสอบจะถูกนำมาทดสอบการจำแนก ประเภทเสียงโดยอัตโนมัติของหลักการสัทศาสตร์ และหลักการฮิดเดนมาร์คคอฟ โดยผลการ จำแนกเสียงที่ได้ของแต่ละหลักการจะถูกเปรียบเทียบกับผลการจำแนกประเภทเสียงของ ผู้เชี่ยวชาญทางด้านสัทศาสตร์จำนวน 3 ท่าน ซึ่งรายละเอียดของการทดลอง มีดังต่อไปนี้

#### 4.4.1. การเตรียมการทดลอง

4.4.1.1. บันทึกสัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ จำนวน 3 รายในรูปแบบ **สระ** (Vowel) **พยัญชนะ** (Stop Consonant) **สระ** (Vowel) ( $V_1-C_s-V_2$ ) โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักต้นพยางค์ มี ด้วยกันเป็น 8 เสียงหลัก **บ ป พ ด ต ท ก ข** และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ **สระอะ อา อี อี้ อุ อู แอะ แอ** รวมทั้งสิ้น 192 เสียง

4.4.1.2. ผู้ป่วยที่ถูกบันทึกเสียงนั้น เป็นผู้ป่วยที่ทำการผ่าตัดช่องเพดาน และปากเรียบร้อยแล้ว เป็นเพศชายอายุประมาณ 5 ปี 2 คน และเป็นเพศหญิงอายุประมาณ 30 ปี 1 คน ในการทดสอบนี้จะใช้ข้อมูลเสียงของผู้ป่วยเป็นชุดข้อมูลทดสอบ

4.4.1.3. สำหรับชุดข้อมูลสอนนั้นจะนำมาจาก การทดลอง 4.3 ที่เก็บมา จากคนปกติจำนวน 10 คน รวมทั้งสิ้น 640 เสียง

4.4.1.4. การจัดเก็บของชุดข้อมูลทดสอบ และชุดข้อมูลสอนจะเก็บแยก ตามฐานกรณ์ของพยัญชนะเสียงกัก กล่าวคือ แยกเป็นฐานริมฝีปาก ฐานปุ่มเหงือก และฐาน เพดานอ่อน

4.4.1.5. การทดสอบจะแยกชุดข้อมูลสอน และชุดข้อมูลทดสอบอย่าง ชัดเจน ชุดข้อมูลทดสอบที่ได้จะถูกนำไปจำแนกประเภทเสียงตามหลักการสัทศาสตร์ และ หลักการฮิดเดนมาร์คคอฟ เพื่อเปรียบเทียบหาค่าความถูกต้องของทั้งสองหลักการ



#### 4.4.2. การทดลอง

4.4.2.1. นำสัญญาณเสียงของผู้ป่วยจากชุดข้อมูลทดสอบทั้ง 192 เสียง ไปผ่านการจำแนกประเภทเสียงโดยอัตโนมัติตามหลักการสวนศาสตร์ โดยเริ่มจากการหาช่วงกักลม และช่วงปล่อยลมของเสียงพยัญชนะกักโดยอัตโนมัติ จากนั้นสกัดค่าลักษณะสำคัญทางสวนลักษณะทั้ง 5 ค่า จากช่วงกักลม และช่วงปล่อยลมดังกล่าว ใช้ชุดข้อมูลสอนจากคนปกติ 10 คนมาเป็นชุดข้อมูลสอนลักษณะสำคัญทางสวนลักษณะให้กับการจำแนกประเภทเสียง เมื่อเตรียมข้อมูลทดสอบและชุดข้อมูลสอนเรียบร้อยแล้ว จะใช้หลักการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นมาช่วยตัดสินใจประเภทการออกเสียงของแต่ละเสียงทดสอบ

4.4.2.2. ในการทดสอบจะแบ่งสัญญาณเสียงออกเป็น 3 ฐานเสียง คือ ฐานริมฝีปาก ฐานปุ่มเหงือก และฐานเพดานอ่อน และผลการทดสอบที่ได้จะออกมาเป็นเสียงโสมพะ อโสมพะแบบสถิต และแบบธนิต เปรียบเทียบกับผลการจำแนกประเภทของเสียงของผู้เชี่ยวชาญทางด้านสวนลักษณะจำนวน 3 ท่าน

4.4.2.3. บันทึกผลการทดลองที่ได้ในรูปแบบของตารางค่าความสับสน (Confusion Matrices) โดยที่จะให้แถวแนวนอนเป็นประเภทการจำแนกประเสียงจากผู้เชี่ยวชาญทางสวนลักษณะ ส่วนคอลัมน์แนวตั้งเป็นประเภทการจำแนกเสียงจากการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้น ในแนวเส้นทแยงมุมจะเป็นบริเวณที่ผลการจำแนกจากทั้งสองแบบมีค่าตรงกัน หรือคือเป็นจำนวนสัญญาณเสียงที่การวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นสามารถจำแนกประเภทเสียงได้ตรงกับการจำแนกของผู้เชี่ยวชาญทางสวนลักษณะ ความสามารถที่ได้จะคำนวณในรูปแบบของร้อยละของความถูกต้อง โดยคำนวณจากจำนวนสัญญาณเสียงในแนวเส้นทแยงมุมส่วนด้วยจำนวนเสียงทั้งหมด คูณด้วยหนึ่งร้อย จะได้เป็นร้อยละความถูกต้อง

4.4.2.4. นำสัญญาณเสียงชุดทดสอบดังกล่าวไปทดสอบการจำแนกประเภทเสียงตามหลักการฮิดเดนมาร์คอฟ โดยขั้นตอนของฮิดเดนมาร์คอฟที่ใช้จะทำเหมือนกับการทดลอง 4.3

4.4.2.5. เปรียบเทียบผลการจำแนกประเภทเสียงที่ได้จากวิธีของฮิดเดนมาร์คอฟกับผลการจำแนกประเภทเสียงของผู้เชี่ยวชาญทางด้านสวนลักษณะจำนวน 3 ท่าน จากนั้นบันทึกผลการทดลองที่ได้ในรูปแบบตารางค่าความสับสน

#### 4.4.3. ผลการทดลอง

ในส่วนนี้จะนำเสนอผลการทดลองการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อชุดข้อมูลทดสอบเป็นสัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ โดยใช้วิธีการจำแนกประเภทเสียงตามหลักการสัทศาสตร์ และตามหลักการฮิดเดนมาร์คอฟ ซึ่งผลการจำแนกประเภทเสียงที่ได้จากแต่ละหลักการนั้นจะถูกนำไปเปรียบเทียบกับผลการจำแนกประเภทเสียงของผู้เชี่ยวชาญทางด้านสัทศาสตร์จำนวน 3 ท่าน ซึ่งมีรายละเอียดดังต่อไปนี้

**ตารางที่ 4.11** ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยโดยอัตโนมัติเมื่อทราบฐานกรณ์ตามวิธีของหลักการสัทศาสตร์ เมื่อใช้ชุดทดสอบจากสัญญาณเสียงของผู้ป่วย ในรูปแบบตารางค่าความสับสน

		ประเภทเสียงที่ได้จากการจำแนกเสียงตามวิธีของหลักการสัทศาสตร์ (ประเภทเสียงที่ทดสอบได้)			ร้อยละความถูกต้องในแต่ละประเภทเสียง (ร้อยละ)	จำนวนเสียงที่ใช้ทดสอบในแต่ละประเภทเสียง (เสียง)
		โสมพะ	อโสมพะแบบ สถิต	อโสมพะแบบ ธนิต		
ประเภทเสียงได้จาก ผู้เชี่ยวชาญ (ประเภทเสียงอ้างอิง)	โสมพะ	7	6	5	38	18
	อโสมพะแบบ สถิต	1	90	29	75	120
	อโสมพะแบบ ธนิต	0	13	33	71	46
	ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง			70.7		
		จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)			184	

จากตารางที่ 4.11 พบว่ามีร้อยละความถูกต้องโดยรวมของการจำแนกประเภทเสียงจากหลักการสัทศาสตร์ถึง 70.7 ซึ่งประเภทเสียงอโสมพะแบบสถิตเป็นประเภทเสียงที่ถูกจำแนกได้ถูกต้องมากที่สุด กล่าวคือถูกต้องใกล้เคียงกับการจำแนกของผู้เชี่ยวชาญทางด้านสัทศาสตร์ถึงร้อยละ 75 จากการสังเกตพบว่าเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ส่วนใหญ่จะมีลมออกในช่วงกักลมทำให้เกิดพลังงานที่ความถี่สูง ซึ่งส่งผลให้ค่าจำนวนครั้งของเวลาที่

สัญญาณคลื่นเสียงแกว่งเกินค่าศูนย์ในช่วงกักลมมีค่ามากตามไปด้วย ยิ่งมีค่าจำนวนครั้งมากเท่าใดโอกาสที่จะเป็นเสียงอโหะระจะมีมากขึ้น ดังนั้นการจำแนกเสียงของผู้ป่วยโดยใช้หลักการสวนสัทศาสตร์จะจำแนกออกมาเป็นเสียงอโหะระโดยส่วนใหญ่ ประกอบกับผู้ป่วยส่วนใหญ่มักจะออกแต่เสียงอโหะระ ดังนั้นการจำแนกเสียงที่ได้จะได้เป็นเสียงอโหะระ โดยเฉพาะอโหะระแบบสถิต ทำให้การจำแนกเสียงอโหะระแบบสถิตมีค่าความถูกต้องมากที่สุด ในทางกลับกันการจำแนกเสียงอโหะระมีความถูกต้องลดลง ทั้งนี้สืบเนื่องจากพลังงานที่ความถี่สูง ทำให้เสียงอโหะระมีค่าจำนวนครั้งของการแกว่งเกินค่าศูนย์ในช่วงกักลมเพิ่มมากขึ้น ส่งผลให้การจำแนกประเภทเสียงอโหะระมีแนวโน้มเป็นเสียงอโหะระแทน ทำให้ความถูกต้องมีค่าลดลง ในส่วนถัดมาขอเสนอผลการจำแนกประเภทเสียงตามวิธีของฮิตเดนมาร์คอฟในรูปแบบตารางค่าความสับสนดังตารางที่ 4.12

**ตารางที่ 4.12** ผลการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยเมื่อทราบฐานกรณ์ตามวิธีของฮิตเดนมาร์คอฟ เมื่อใช้ชุดทดสอบจากสัญญาณเสียงของผู้ป่วย ในรูปแบบตารางค่าความสับสน

		ประเภทเสียงที่ได้จากการจำแนกเสียงตามวิธีของฮิตเดนมาร์คอฟ (ประเภทเสียงที่ทดสอบได้)			ร้อยละความถูกต้องในแต่ละประเภทเสียง (ร้อยละ)	จำนวนเสียงที่ใช้ทดสอบในแต่ละประเภทเสียง (เสียง)
		โหะระ (เสียง)	อโหะระแบบสถิต (เสียง)	อโหะระแบบธนิต (เสียง)		
ประเภทเสียงได้จาก ผู้เชี่ยวชาญ (ประเภทเสียงอ้างอิง)	โหะระ	9	5	4	50	18
	อโหะระแบบสถิต	30	43	47	35.8	120
	อโหะระแบบธนิต	0	1	45	97.8	46
	ร้อยละความถูกต้องโดยรวมทุกประเภทเสียง			52.7		
		จำนวนเสียงที่ใช้ทดสอบทั้งหมด (เสียง)			184	

จากตารางที่ 4.12 พบว่ามีร้อยละความถูกต้องโดยรวมของการจำแนกประเภทเสียงตามวิธีของฮิตเดนมาร์คอฟเพียง 52.7 ทั้งนี้เห็นได้ว่าประเภทเสียงอโหะระแบบสถิตมีร้อยละค่าความถูกต้องน้อยที่สุด ในขณะที่ประเภทเสียงอโหะระแบบธนิตกลับมีร้อยละค่าความถูกต้องของการจำแนกประเภทเสียงมากที่สุดถึง 97.8 จะเห็นได้ว่าการจำแนกประเภทเสียงตามวิธีของฮิต

เดนมาร์คอฟจะมีความถูกต้องมากในกรณีที่เป็นเสียงอโหิษะแบบชนิด ซึ่ง เป็นประเภทเสียงที่มี ลักษณะรูปแบบของคลื่นสัญญาณเสียงค่อนข้างชัดเจนและแตกต่างจากเสียงเสียงอโหิษะ และ อโหิษะแบบสติดิล ทั้งนี้เพราะสัญญาณเสียงประเภทเสียงอโหิษะแบบชนิดนั้นจะมีพลังงานมาก ในช่วงปล่อยลมส่งผลให้รูปแบบของสัญญาณเสียงพยัญชนะกักสามารถสังเกตได้ง่าย ทำให้การ จำแนกเสียงอโหิษะแบบชนิดนี้มีความถูกต้องแม่นยำกว่าการจำแนกประเภทเสียงอื่น

จากผลการทดลองของตารางที่ 4.11 และ 4.12 พบว่าการจำแนกประเภทเสียง พยัญชนะกักโดยอัตโนมัติตามหลักการสวณส์ศาสตร์ มีค่าความถูกต้องมากกว่าการจำแนก ประเภทเสียงพยัญชนะกักตามหลักการฮิดเดนมาร์คอฟ ในกรณีที่สัญญาณเสียงทดสอบเป็น สัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ โดยพบว่าวิธีของฮิดเดนมาร์คอฟนั้นจะมี ความถูกต้องมากกว่ากรณีที่สัญญาณเสียงมีรูปแบบของคลื่นเสียงใกล้เคียงกับของชุดข้อมูลสอน หรือ มีการออกเสียงใกล้เคียงกับเสียงคนปกติ ในขณะที่การจำแนกประเภทเสียงพยัญชนะกักโดยอัตโนมัติ ตามหลักการสวณส์ศาสตร์จะมีความถูกต้องมากกว่า ถ้าสัญญาณเสียงที่เข้ามาทดสอบเป็น สัญญาณเสียงที่ไม่ถูกต้อง หรือเป็นสัญญาณเสียงที่มีรูปแบบไม่ถูกต้องเมื่อเปรียบเทียบกับ สัญญาณเสียงแบบปกติ ทั้งนี้เป็นเพราะว่าการจำแนกพยัญชนะเสียงกักตามหลักการสวณ ส์ศาสตร์มิได้ขึ้นอยู่กับรูปแบบของสัญญาณเสียงโดยตรงนัก แต่จะขึ้นอยู่กับค่าลักษณะสำคัญ ทางสวณส์ศาสตร์ที่สกัดได้มากกว่า ทำให้การจำแนกประเภทเสียงพยัญชนะกักที่ใช้สัญญาณเสียง ของผู้ป่วยตามหลักการสวณส์ศาสตร์มีความถูกต้องมากกว่าการจำแนกเสียงตามหลักการของ ฮิดเดนมาร์คอฟ

## บทที่ 5

### สรุปผลการวิจัย และข้อเสนอแนะ

ในบทนี้จะขอนำเสนอข้อสรุปผลการวิจัย และข้อเสนอแนะของงานวิจัยเรื่องวิธีการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ โดยใช้หลักการสวณศาสตร์มาช่วยในการจำแนกประเภทของเสียง มีรายละเอียดดังนี้

#### 5.1. สรุปผลการวิจัย

วิทยานิพนธ์นี้นำเสนอวิธีการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ โดยใช้หลักการสวณศาสตร์ ซึ่งประกอบไปด้วยสองขั้นตอนคือ ขั้นตอนหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกัก และขั้นตอนการจำแนกประเภทเสียงโหระ อโหระ เพื่อใช้เป็นแนวทางในการพัฒนาโปรแกรมคอมพิวเตอร์ที่ออกเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ โดยเปรียบเทียบประสิทธิภาพกับวิธีการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ โดยใช้หลักการฮิดเดนมาร์คอฟ ด้วยความหวังว่าวิธีการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ที่นำเสนอจะ สามารถเป็นต้นแบบขั้นตอนวิธีการจำแนกประเภทเสียงอย่างอัตโนมัติของสัญญาณเสียงโหระ อโหระแบบสถิต และอโหระแบบชนิด ของเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณ์เดียวกันได้ การจำแนกเสียงพยัญชนะกักดังกล่าวมีรายละเอียดดังนี้

##### 5.1.1. ขั้นตอนหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกัก

กระบวนการหาช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกักตามวิธีที่เสนอนี้ได้อาศัยหลักการของการแบ่งเสียงพูดเป็นเซกเมนต์มาช่วยหาตำแหน่งเริ่มต้น และสิ้นสุดของสัญญาณเสียงพยัญชนะกักที่อยู่ในรูปแบบ **สระ** (Vowel) **พยัญชนะ** (Stop Consonant) **สระ** (Vowel) ( $V_1-C_s-V_2$ ) โดยที่ สระตัวแรกเป็นสระอะแบบไม่เน้นพยางค์ พยัญชนะเป็นเสียงของพยัญชนะกักต้นพยางค์ มีด้วยกันเป็น 8 เสียงหลัก **บ ป พ ต ท ก ข** และสระตัวหลังเป็นเสียงสระ 8 เสียง คือ **สระอะ อา อิ อี อุ อู แอะ แอ**

กระบวนการแบ่งเสียงพูดเป็นเซกเมนต์ดังกล่าว จะเริ่มจากการแปลงสัญญาณเสียงให้เป็นสัญญาณค่าพลังงานรวม จากนั้นสร้างสัญญาณค่าความแตกต่างจากสัญญาณค่าพลังงานรวม นำสัญญาณค่าความแตกต่างไปหาจุดยอดของความแตกต่าง เมื่อได้จุดยอดจากสัญญาณค่าความแตกต่างแล้ว ให้นำจุดยอดดังกล่าวไปผ่านการตัดสินใจหาจุดยอดที่มีโอกาสเป็นขอบของสัญญาณเสียงพยัญชนะกัก โดยบริเวณของจุดยอดที่มีแอมพลิจูดของการเปลี่ยนแปลงพลังงานในเชิงเวลาและในเชิงความถี่ลดลงมากที่สุด จะมีโอกาสเป็นขอบเริ่มต้นของสัญญาณเสียงพยัญชนะกัก ส่วนบริเวณของจุดยอดที่มีแอมพลิจูดของการเปลี่ยนแปลงพลังงานในเชิงเวลาและในเชิงความถี่เพิ่มมากที่สุด จะมีโอกาสเป็นขอบสิ้นสุดของสัญญาณเสียงพยัญชนะกัก เมื่อสามารถหาขอบเริ่มต้นและสิ้นสุดของสัญญาณเสียงพยัญชนะกักได้ จะนำสัญญาณเสียงที่ได้ผ่านการกรองพลังงานในเชิงความถี่ เพื่อหาตำแหน่งเริ่มต้นของช่วงปล่อยลมที่อยู่ภายในสัญญาณเสียงพยัญชนะกัก ซึ่งบริเวณที่เป็นจุดเริ่มต้นของช่วงปล่อยลมนั้นจะมีพลังงานสูง กล่าวคือมีค่าพลังงานประมาณ -25 ถึง -10 เดซิเบล ที่ช่วงความถี่ต่ำ (0 ถึง 1,000 เฮิรตซ์) และมีค่าพลังงานประมาณ -30 ถึง -10 เดซิเบล ที่ช่วงความถี่กลาง (2,000 ถึง 6,000 เฮิรตซ์) นอกจากนี้ที่ช่วงความถี่สูง (8,000 ถึง 10,000 เฮิรตซ์) มีค่าพลังงานประมาณ -55 ถึง -35 เดซิเบล ทั้งนี้ขึ้นอยู่กับฐานกรณ์ของสัญญาณเสียง หลังจากนั้นให้นำตำแหน่งเริ่มต้นของช่วงปล่อยลมไปช่วยในการแบ่งสัญญาณเสียงพยัญชนะกัก ทำให้สามารถแบ่งสัญญาณเสียงของพยัญชนะกักออกมาเป็นสองช่วง คือช่วงกักลม และช่วงปล่อยลมได้

จากผลการทดลองเพื่อทดสอบวัดความคลาดเคลื่อนการหาช่วงกักลม และช่วงปล่อยลมของพยัญชนะต้นเสียงกัก พบว่าค่าความคลาดเคลื่อนเฉลี่ยเมื่อเปรียบเทียบกับการตัดโดยผู้ทำวิจัยมีค่าประมาณ 7 ถึง 27 มิลลิวินาทีจากสัญญาณเสียงทั้งหมด 896 สัญญาณเสียง หรือคิดเป็นร้อยละ 1 ถึง 5 ต่อความยาวทั้งหมดของสัญญาณเสียง จะเห็นได้ว่าวิธีการแบ่งเสียงพูดเป็นเซกเมนต์นี้ สามารถใช้หาช่วงกักลม และช่วงปล่อยลมของพยัญชนะเสียงกักได้อย่างมีประสิทธิภาพ สามารถนำมาช่วยในการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ได้

### 5.1.2. ขั้นตอนการจำแนกประเภทเสียงโฆษะ อโฆษะ

กระบวนการจำแนกประเภทเสียงโฆษะ และอโฆษะนั้น เป็นกระบวนการหลักของการจำแนกประเภทเสียงตามหลักการสวนสัทศาสตร์ โดยมีขั้นตอนเริ่มจาก นำสัญญาณเสียงช่วงกักลม และช่วงปล่อยลมของสัญญาณเสียงพยัญชนะกักไปสกัดหาค่าลักษณะสำคัญของสวนลักษณะ ซึ่งในค่าลักษณะสำคัญทางสวนลักษณะที่นำเสนอในวิทยานิพนธ์นี้มีอยู่ด้วยกันห้าค่า ได้แก่ ความยาวช่วงกักลม ความยาวช่วงปล่อยลม ค่าพลังงานเฉลี่ยในช่วงความถี่ต่ำของสัญญาณเสียงช่วงกักลม ค่าพลังงานของช่วงปล่อยลม และค่าจำนวนครั้งของเวลาที่สัญญาณคลื่นเสียงเกินค่าศูนย์ในช่วงกักลม โดยค่าลักษณะสำคัญทางสวนลักษณะที่ได้จะถือเป็นลักษณะประจำของสัญญาณเสียง เพื่อให้การวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นนำมาตัดสินใจประเภทการออกเสียงของเสียงดังกล่าวได้ โดยในการตัดสินใจนั้นจะรับเงื่อนไขฐานกรณ์ของการออกเสียงพยัญชนะกักต้นพยางค์มาเพื่อช่วยเลือกข้อมูลค่าลักษณะสำคัญทางสวนลักษณะจากชุดข้อมูลสอนที่มีฐานกรณ์การออกเสียงเดียวกันมาสร้างเป็นเวกเตอร์ลักษณะเฉพาะ สำหรับช่วยสอนให้การวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นได้เรียนรู้ลักษณะเฉพาะของการออกแบบเสียงโฆษะ อโฆษะแบบสติดิล และอโฆษะแบบธนิตของพยัญชนะต้นเสียงกักที่ใช้ฐานกรณ์เดียวกัน หลังจากนั้นการวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นจะตัดสินใจเลือกประเภทการออกเสียงให้กับสัญญาณเสียงทดสอบที่เข้ามาทดสอบ โดยประเภทการออกเสียงที่ได้มีอยู่ด้วยกันสามประเภท คือประเภทเสียงโฆษะ อโฆษะแบบสติดิล และอโฆษะแบบธนิต

จากการทดลองเพื่อหาความถูกต้องของการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ เมื่อใช้หลักการสวนสัทศาสตร์มาจำแนกประเภทเสียง เมื่อสัญญาณเสียงที่ทดสอบเป็นสัญญาณเสียงของคนปกติ พบว่าการทดสอบที่ใช้ชุดข้อมูลสอนเป็นชุดทดสอบมีความถูกต้องอยู่ที่ร้อยละ 86 83 และ 91 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ ส่วนการทดลองโดยใช้ชุดข้อมูลทดสอบมาทดสอบนั้นมีความถูกต้องของการจำแนกประเภทเสียงอยู่ที่ร้อยละ 83 89.33 และ 88 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ พบว่าการทดลองทั้งสองชุดมีความถูกต้องใกล้เคียงกัน และเมื่อเปรียบเทียบกับการจำแนกประเภทเสียงตามหลักการฮิดเดนมาร์คอฟที่มีค่าความถูกต้องของการจำแนกประเภทเสียงในชุดข้อมูลสอนอยู่ที่ร้อยละ 65 80 และ 85 และในชุดข้อมูลทดสอบอยู่ที่ร้อยละ 81 88.33 และ 85.33 ตามฐานริมฝีปาก ฐานเพดานอ่อน และฐานปุ่มเหงือก ตามลำดับ จะเห็นได้ว่าการจำแนกประเภทเสียงตามหลักการสวนสัทศาสตร์ของวิทยานิพนธ์นี้ ค่อนข้างมีความแม่นยำมากกว่าการจำแนกประเภทเสียงพยัญชนะกักตามหลักการฮิดเดนมาร์คอฟ ถึงแม้ว่า จะมีค่าความถูกต้องเฉลี่ยมากกว่าเพียงร้อยละ 1 ถึง 2 ในชุดข้อมูล

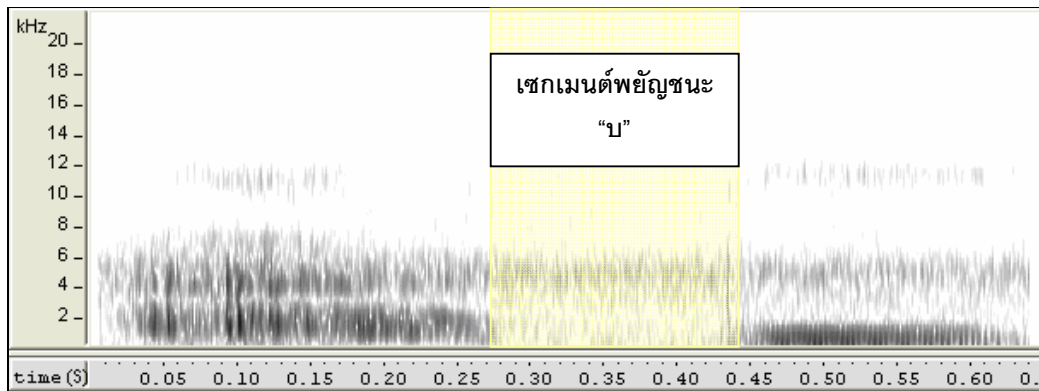
ทดสอบ แต่ประโยชน์ที่ได้จากการจำแนกประเภทเสียงตามหลักการสัทศาสตร์นั้นจะมากกว่า ในกรณีที่ใช้สัญญาณเสียงที่เข้ามาทดสอบการจำแนกประเภทเสียงนั้นเป็นสัญญาณเสียงที่ออกเสียง ไม่ถูกต้องตามหลักสัทศาสตร์ เพราะการจำแนกประเภทเสียงโดยใช้หลักการสัทศาสตร์จะ ใช้การสกัดค่าลักษณะสำคัญทางสวณลักษณะมาช่วยในการวิเคราะห์ ทำให้เราสามารถนำผลค่า ลักษณะสำคัญทางสวณลักษณะที่ได้ไปวิเคราะห์ถึงสาเหตุการผิดพลาดทางสัทศาสตร์ได้

นอกจากนี้ยังพบว่าถ้าเสียงที่เข้ามาทดสอบการจำแนกประเภทเสียงเป็นเสียงของ คนที่ออกเสียงไม่ชัดเจน หรือของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ ประสิทธิภาพของการ จำแนกเสียงพยัญชนะกักตันพยางค์โดยการใช้นหลักการสัทศาสตร์จะมีมากกว่าของหลักการ ฮิดเดนมาร์คอฟมาก หรือมีความสามารถในการจำแนกประเภทเสียงได้ใกล้เคียงกับของ ผู้เชี่ยวชาญด้านสวณลักษณะมากกว่าการจำแนกประเภทเสียงตามหลักการฮิดเดนมาร์คอฟ จากผล การทดลองที่ใช้สัญญาณเสียงของผู้ป่วยโรคปากแหว่งเพดานโหว่เป็นสัญญาณทดสอบ พบว่าการ จำแนกประเภทเสียงตามหลักการสัทศาสตร์มีค่าความถูกต้องโดยรวมอยู่ที่ร้อยละ 70.7 ในขณะที่การจำแนกประเภทเสียงตามหลักการของฮิดเดนมาร์คอฟ มีค่าความถูกต้องโดยรวม เพียงร้อยละ 52.7 จะเห็นได้ว่าวิธีการจำแนกประเภทเสียงพยัญชนะกักตันพยางค์ตามหลักการ สวณศาสตร์ของวิทยานิพนธ์นี้ สามารถจำแนกประเภทเสียงได้แม่นยำกว่าวิธีฮิดเดนมาร์คอฟถึง ร้อยละ 18 จากผลการทดลองที่ได้สะท้อนให้เห็นประสิทธิภาพของการจำแนกประเภทเสียงของ วิทยานิพนธ์นี้ ว่าน่าจะสามารถนำไปเป็นต้นแบบของขั้นตอนวิธีการจำแนกประเภทเสียงอย่าง อัตโนมัติของสัญญาณเสียงโหระหะ อโหระหะแบบสถิต และอโหระหะแบบชนิด ของเสียงพยัญชนะกัก ภาษาไทยในตำแหน่งต้นพยางค์ที่ใช้ฐานกรณ์เดียวกันได้ อีกทั้งยังสามารถนำไปใช้เป็นแนวทางใน การพัฒนาโปรแกรมคอมพิวเตอร์ฝึกรับรู้เสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์ สำหรับผู้ที่มีความผิดปกติในการออกเสียง โดยเฉพาะอย่างยิ่งผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ ได้ ดังจะเห็นได้จากผลการทดลองการจำแนกประเภทเสียงเมื่อใช้สัญญาณเสียงของผู้ป่วยที่เป็น โรคปากแหว่งเพดานโหว่มาเป็นสัญญาณเสียงทดสอบ อีกทั้งจะพบประโยชน์ที่ได้เพิ่มเติมจากการ จำแนกประเภทเสียงตามหลักการสัทศาสตร์ คือสามารถที่จะใช้ค่าลักษณะสำคัญทางสวณ ลักษณะที่ได้มาช่วยในการพิจารณาหาที่มาของการออกเสียงผิดพลาด หรือหาสาเหตุของการออก เสียงไม่ตรงตามหลักสัทศาสตร์ได้ ดังตัวอย่างต่อไปนี้

ในกรณีของผู้ป่วยที่เป็นโรคปากแหว่งเพดานโหว่ต้องการออกเสียงโหระหะ ฐานริม ฝีปาก (/บ/) แต่ออกเสียงผิดเป็นเสียงอโหระหะแบบสถิตแทน วิธีการจำแนกประเภทเสียงตาม หลักการสวณศาสตร์ของวิทยานิพนธ์นี้ จะสามารถจำแนกเป็นอโหระหะแบบสถิตได้ตรงตามที่



ออกเสียงผิด อีกทั้งยังสามารถระบุถึงสาเหตุข้อผิดพลาดตามหลักสัทศาสตร์ได้ ดังรูปที่ 5.1 สเปกโตรแกรมของสัญญาณเสียง อะบุ /abu/ ของผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่



รูปที่ 5.1 สเปกโตรแกรมสัญญาณเสียงพยัญชนะกัก อะบุ /abu/ ของผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่

จากรูปที่ 5.1 หากสังเกตโดยใช้สเปกโตรแกรมจะพบว่าในช่วงกักลมของพยัญชนะ “บ” นั้นไม่มีพลังงานที่ความถี่ต่ำ และเมื่อผ่านการจำแนกประเภทตามวิธีของวิทยานิพนธ์จะพบว่าค่าพลังงานเฉลี่ยในช่วงถี่ความต่ำของสัญญาณเสียงนี้มีเพียง -38 เดซิเบล เมื่อเปรียบเทียบกับค่าพลังงานเฉลี่ยที่ได้จากฐานข้อมูลสวณลักษณะมีมากถึง -20 เดซิเบล ทำให้เห็นที่มาว่าการออกสัญญาณเสียงนี้อาจมีข้อผิดพลาดจากการสร้างช่องปิดได้ไม่สมบูรณ์ทำให้ไม่เกิดการสั่นของเส้นเสียงที่เป็นต้นกำเนิดพลังงานที่ความถี่ต่ำ

## 5.2. ข้อเสนอแนะ

5.2.1. จากการทดสอบการจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติกับสัญญาณเสียงของผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่ พบว่าความผิดพลาดในการออกเสียงของผู้ป่วยนอกจากมีการแทนที่กันระหว่างเสียงโฆษะ และอโฆษะแล้วยังมีการแทนที่กันระหว่างฐานกรณ์ในการออกเสียง หรือการมีเสียงขึ้นจมูกตลอดเวลาที่ออกเสียงพยัญชนะกัก ดังนั้นหากสามารถศึกษาวิธีการจำแนกเสียงพยัญชนะกักภาษาไทยโดยมีทราบบรรณก่อนการจำแนก หรือจำแนกเสียงพยัญชนะเสียงนาสิก จากนั้นนำแนวทางการศึกษาที่ได้มา

สร้างโปรแกรมฝึกพูดการออกเสียงดังกล่าว อาจจะช่วยฝึกทักษะการออกเสียงให้ผู้ป่วยที่เป็นโรคปากแห้งเพดานโหว่ได้มากขึ้น

5.2.2. การจำแนกประเภทเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์โดยอัตโนมัติเมื่อทราบฐานกรณ์ของวิทยานิพนธ์นี้อาจพัฒนาให้มีความแม่นยำเพิ่มมากขึ้นได้ ถ้าสามารถศึกษาค่าลักษณะสำคัญทางสวณลักษณะที่มีผลต่อการจำแนกเพิ่มขึ้น ทั้งนี้ขึ้นอยู่กับความหลากหลายของฐานข้อมูลสัญญาณเสียงที่ใช้เป็นฐานความรู้ในการหาสวณลักษณะ

5.2.3. ในกรณีที่สัญญาณเสียงที่ถูกจำแนกมีความหลากหลายมากขึ้น หรือมีการกระจายตัวอยู่ปริภูมิที่แตกต่างกันมากขึ้นการใช้วิธีวิเคราะห์แบบดิสคริมิแนนต์เชิงเส้นอาจไม่ครอบคลุมได้หมด อาจพัฒนาให้วิธีการจำแนกตามแบบสวณศาสตร์นี้ใช้วิธีวิเคราะห์ค่าลักษณะสำคัญทางสวณลักษณะโดยวิธีเครือข่ายประสาท (Neural Network) แทน

5.2.4. จากผลการทดลองการหาช่วงกักลม และช่วงปล่อยลมของพยัญชนะเสียงกักพบว่ายังมีความคลาดเคลื่อนอยู่ที่ร้อยละ 1 ถึง 5 หากสามารถขยายฐานข้อมูลสัญญาณเสียงเพิ่มขึ้น และเปลี่ยนจากวิธีแบ่งเซกเมนต์เป็นวิธีการแบ่งตามกรอบของเวลา อาจทำให้ช่วงสัญญาณเสียงที่ได้คลาดเคลื่อนน้อยลง ส่งผลให้มีความแม่นยำในการจำแนกมากขึ้น

5.2.5. เนื่องจากสัญญาณเสียงมีการเปลี่ยนแปลงได้ตลอดตามเวลา อีกทั้งสามารถถูกรบกวนจากเสียงรอบข้างได้ทุกเมื่อ ฉะนั้นหากต้องการให้ผลการจำแนกถูกต้องแม่นยำมากยิ่งขึ้น จึงควรบันทึกเสียงที่ต้องการทดสอบในห้องปิดเพื่อตัดปัญหาการถูกรบกวนดังกล่าว

## รายการอ้างอิง

- [1] ดร. นันทนา วัฒนเกียรติ. สัทศาสตร์ ภาคทฤษฎีและภาคปฏิบัติ. สำนักพิมพ์ มหาวิทยาลัยธรรมศาสตร์, 2548.
- [2] Luksaneeyanawin S. A Three Dimensional Phonology: A Historical Implication. In: Proceedings of the 3rd International Symposium on Language and Linguistics. pp. 75-90, 1992.
- [3] Luksaneeyanawin S. Linguistics Research and Thai Speech Technology. In: Proceeding of the 5th International Conference on Thai Studies. School of Oriental and African Studies, University of London. pp. 1-29, 1993.
- [4] I. W. Cooley, J. W. Turkey. An algorithm for the machine calculation of complex Fourier series. Math. Of Comput., vol. 19, pp. 297-301, April 1965.
- [5] ไพโรจน์ ลีลาภักดิ์. การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยอาศัย เซกเมนต์โดยใช้สารสนเทศสวณศาสตร์. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2549.
- [6] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland. HTK Book (for HTK Version 3.3). Cambridge University Engineering Department. April 2005.
- [7] S. Balakrishnama, A. Ganapathiraju. Linear Discriminant Analysis – A Brief Tutorial. Institute for Signal and Information Processing. Department of Electrical and Computer Engineering. Mississippi State University.
- [8] F. R. Adams, H. Crepy, D. Jameson, J. Thatcher. International Business Machines Corporation. IBM Products for Persons with Disabilities. In: Proceedings of. IEEE. pp 27.5.1 – 27.5.5, 1989.
- [9] S. A. Liu. Landmark detection for distinctive feature-based speech recognition, J. Acoust. Soc. Am. 100(5). pp. 3417-3430, 1996.
- [10] P. Leelaphattarakij, P. Punyabukkana, A. Suchato. Locating Phone Boundaries from Acoustic Discontinuities using a Two-staged Approach. The Ninth International Conference on Spoken Language Processing: Interspeech 2006. Pittsburgh, Pennsylvania. pp. 673-676, 2006.

- [11] A.M.A. Ali, J. Van der Spiegel, and P. Mueller. Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants. In: Proceedings of. IEEE Trans. Speech Audio Processing. vol. 9. no. 8, 2001.
- [12] A. Suchato. Classification of Stop Consonant Place of Articulation. Ph.D. *Thesis*, MIT.
- [13] A. Suchato, P. Punyabukkana. Factors in Classification of Stop Consonant Place of Articulation. In: Proceeding of. Interspeech 2005, Lisbon. Portugal. pp 2969-2972, 2005.
- [14] P. Dareyoah, A. Suchato, P. Punyabukkana. A Study of Acoustic Measurements for Voicing Detection in Speech with Room-level SNR. The Sixth Symposium on Natural Language Processing 2005: SNLP2005. Chiang Rai, Thailand, 2005.
- [15] B. Pholkul, P. Punyabukkana, A. Suchato. Stop Consonant Voicing Classification for Computer-assisted Speech Training of Patients with Cleft Lips and Palates. International Convention for Rehabilitation Engineering & Assistive Technology: i-Create 2007. National Library Board Building & Hotel Intercontinental, Singapore. 2007.

ภาคผนวก

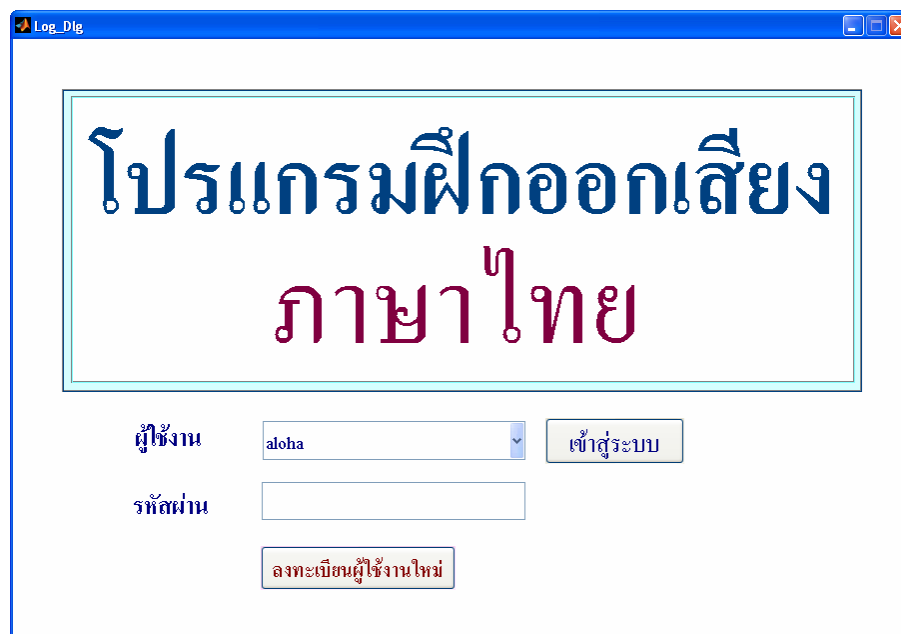
## ภาคผนวก ก

### ต้นแบบโปรแกรมการจำแนกเสียงพยัญชนะกักภาษาไทยโดยอัตโนมัติ

ในส่วนนี้จะขอแนะนำเสนอต้นแบบของโปรแกรมเพื่อใช้ในการฝึกพูดสัญญาณเสียงพยัญชนะกักภาษาไทยในตำแหน่งต้นพยางค์

#### ก.1. หน้าแรกของโปรแกรม

ผู้ใช้งานจะต้องลงทะเบียนผู้ใช้งานก่อนเข้าใช้งานโปรแกรมฝึกออกเสียง เพื่อให้โปรแกรมสามารถเก็บบันทึกผลการฝึกของผู้ใช้งาน แล้วผู้เชี่ยวชาญทางสวณลักษณะสามารถนำผลที่ได้ไปวิเคราะห์ภายหลัง



Log\_Dig

โปรแกรมฝึกออกเสียง  
ภาษาไทย

ผู้ใช้งาน

รหัสผ่าน

รูปที่ ก-1 หน้าแรกของโปรแกรมต้นแบบการฝึกพูดเสียงพยัญชนะกักภาษาไทย

### ก.2. หน้าลงทะเบียนผู้ใช้งานใหม่

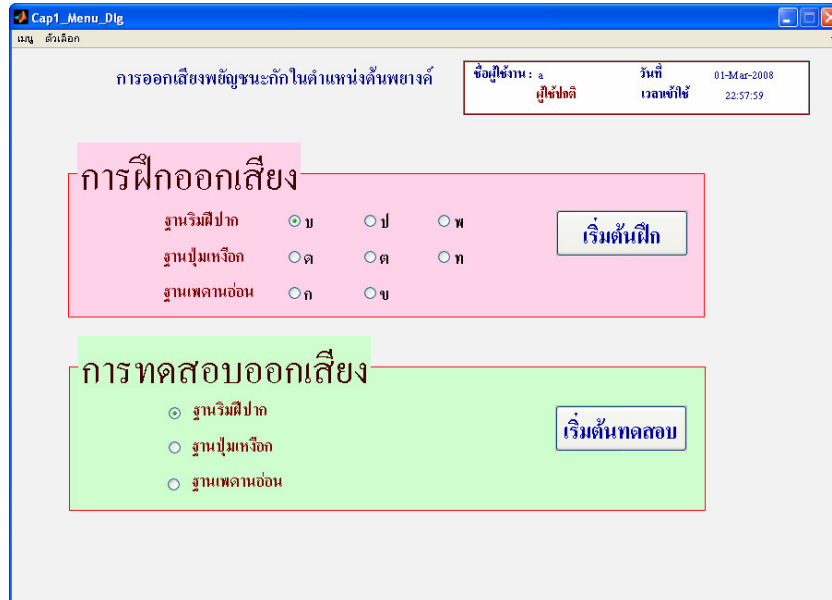
ในกรณีที่เป็นผู้ใช้งานใหม่ จะต้องมาลงทะเบียนก่อนเข้าใช้งานโปรแกรม โดยผู้ใช้งานมีได้ 2 สถานะ คือ สถานะผู้ใช้งานปกติ และผู้ใช้งานที่เป็นเจ้าหน้าที่

รูปที่ ก-2 หน้าลงทะเบียนผู้ใช้งานใหม่ของโปรแกรมต้นแบบ

### ก.3. หน้าประเภทการออกเสียงที่ต้องการฝึกออกเสียง

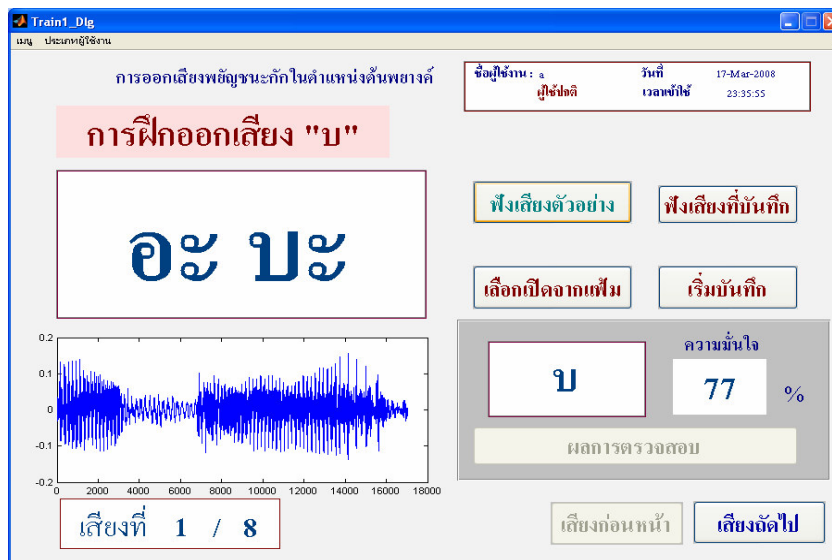
รูปที่ ก-3 หน้าเลือกประเภทการออกเสียงของโปรแกรมต้นแบบ

ก.4. หน้าแสดงประเภทการฝึก และการทดสอบการออกเสียงพยัญชนะเสียงัก



รูปที่ ก-4 หน้าเลือกการออกเสียงพยัญชนะกักในตำแหน่งต้นพยางค์

ก.5. หน้าตัวอย่างการฝึกการออกเสียงพยัญชนะกักฐานริมฝีปาก



รูปที่ ก-5 หน้าฝึกการออกเสียงพยัญชนะกักฐานริมฝีปากของโปรแกรมต้นแบบ



ก.6. หน้าตัวอย่างการทดสอบการออกเสียงพยัญชนะกัฐานริมฝีปาก

การออกเสียงพยัญชนะกัในตำแหน่งคั่นพยางค์

ชื่อผู้ใช้งาน : a      วันที่ : 17-Mar-2008  
ผู้จัดทำ :      เวลาที่ใช้ : 23:35:55

การทดสอบออกเสียง  
ฐานริมฝีปาก ("บ", "ป", "ท")

อะ แพะ

เสียงทดสอบที่ 1 / 20

ฟังเสียงตัวอย่าง

เริ่มบันทึก

เสียงถัดไป

รูปที่ ก-6 หน้าทดสอบการออกเสียงพยัญชนะกัฐานริมฝีปากของโปรแกรมต้นแบบ

## ภาคผนวก ข

### ผลงานตีพิมพ์

งานประชุมวิชาการ “International Convention for Rehabilitation Engineering & Assistive Technology (i-Create 2007)” ระหว่างวันที่ 23 ถึง 26 เมษายน 2550 ณ ประเทศสิงคโปร์ ในบทความเรื่อง Stop Consonant Voicing Classification for Computer-assisted Speech Training of Patients with Cleft Lips and Palates

# Stop Consonant Voicing Classification for Computer-assisted Speech Training of Patients with Cleft Lips and Palates

Budsamas Pholkul	Proadpran Punyabukkana	Atiwong Suchato
Spoken Language Systems Research Group	Spoken Language Systems Research Group	Spoken Language Systems Research Group
Department of Computer Engineering Chulalongkorn University Bangkok, THAILAND 66-2-648-9852	Department of Computer Engineering Chulalongkorn University Bangkok, THAILAND 66-2-218-9656	Department of Computer Engineering Chulalongkorn University Bangkok, THAILAND 66-2-218-9656
Budsamas.P@student.chula.ac.th	Proadpran.P@chula.ac.th	Atiwong.S@chula.ac.th

## ABSTRACT

Cleft lips and palates (CLP) may cause functional disorders even after adequate surgical treatments, speech disorders being one of them. Automatic algorithms utilizing acoustic-phonetic knowledge are needed in developing computer-based tools for assisting the speech training of CLP patients. This work focuses on acoustic discrimination among voiced, voiceless unaspirated, and voiceless aspirated stop consonants with the same place of articulation and aims at revealing a set of acoustic measurements capable of discriminating a CLP patients' speech. Acoustic measurements based on duration and signal energy are proposed and studied. Analysis of variance and classification experiments demonstrate high potentials in using these acoustic measurements in developing automatic voicing classification algorithms for speech training tools. The overall classification accuracy of 92% is achieved in classifying non-CLP data, in which the best result obtained is 99% for the alveolar case. The proposed measurements can classify data from CLP patients with almost 90% accuracy even when the classifier is trained only on the non-CLP data.

## Categories and Subject Descriptors

I.5.2 [Pattern Recognition]: Design Methodology – classifier design and evaluation, feature evaluation and selection and pattern analysis; K.4.2 [Computer and Society]: Social Issues – assistive technology for persons with disabilities

## General Terms

Measurement, Experimentation.

## Keywords

Computer-assisted speech training, phonological and acoustical analysis of Thai language, acoustic-phonetic features.

## 1. INTRODUCTION

Cleft lips and palates (CLP) are usually formed before birth. They can result in morphological and functional disorders. They affect not only primary disorders such as swallowing, breathing and mimic disorders but also speech development. CLP is in the top four in birth disorder ranking. Statistically, one in seven hundred infants have cleft lips and/or palates. Thai CLP children are typically found in remote inland areas while clinics treating such a disorder are situated in big cities. This renders being trained by a speech therapist unaffordable for those children as well as other CLP patients. Moreover, the lack of speech therapists throughout the country is another major obstacle in providing the CLP patients with proper training in speech articulation.

Stop consonants are sounds which are formed using the greatest degree of obstruction and complexity of movements in the vocal tract. The articulators used in the production of stop consonants form occlusions (closures) in the oral cavity of the vocal tract. Due to continuous airflow and the presence of closures, pressure is built up in the cavity behind the closures. After the closure phase comes the release phase. In this phase, the oral occlusion is broken. This immediately reduces the pressure and allows the air to resume its flow. This complex articulation causes difficulties in uttering stop consonants for CLP patients.

The most common speech disorder for Thai CLP patients is the substitution of stop consonants in syllable-initial positions. A cause of such substitutions arises from the necessity of the speaker making a complete closure in his/her oral cavity while maintaining air flow from the lungs. The closure is then promptly released. Although patients have received the required reconstructive surgical therapies, they still lack the ability to perform the required gesture in a good enough manner. Clinical observations indicate that one of the most frequently occurring substitutions is among stop consonants with the same place of articulation. In the Thai language, there are three places of articulation, namely labial, alveolar and velar. Such substitutions

are among the voiced, voiceless, and voiceless aspirated stop consonants within each place of articulation.

In this work, we present an acoustic-phonetic study whose results serve as a foundation to further develop a computer-assisted speech training program aiming at CLP patients. We investigate the acoustic-phonetic characteristics of stop consonants for detecting stop consonant voicing and apply these characteristics as classification features of an automatic stop consonant voicing classification. Stop consonants of subjects with and without CLP are analyzed and used for testing the automatic classification. The classification results in the case of CLP patients are then compared to experts' judgments in order to evaluate the performance of our acoustic measurements in doing the task.

First, we start with a broad look at how computers are involved in the speech training issue and describe the motivation of the need of subtle acoustic-phonetic analysis. Then, we introduce the acoustic measurements used in the classification experiments along with the analyses of their values obtained from different voicing characters. Classification results of patients with and without CLP are then discussed.

## 2. BACKGROUND

### 2.1 Computer-assisted Speech Training

Combined with the widespread usage of personal computers, advances in automatic speech recognition techniques, and speech-related sciences including physiology of human speech production and acoustic-phonetics of human speech lead to many exciting developments in computer-assisted speech training tools. An example of such a tool is a commercially available speech and language product called *SpeechViewer*, which is designed to help people of all ages who have a variety of disabilities, such as speech or language impairments, cerebral palsy, developmental delay, traumatic brain injury and speech disorders resulting from a stroke. A number of computer programs were also created to help people practice uttering sounds as well as phonemes in many languages, including many western languages [1], and some Asian languages, such as Chinese [2]. The majority of users are among people with hearing disorders. In most programs pattern matching techniques are utilized for verifying users' speech against targeted sounds by comparing the spectral features of their surface speech signals. Pattern matching based on Hidden Markov Models (HMMs) and Dynamic Time Warping (DTW) is usually the enabler of such verification of functionality.

As mentioned earlier, the most common speech disorder for the Thai CLP patients is substitution of stop consonants in syllable-initial positions. The substitution occurs within stop consonants with the same place of articulation. Therefore, automatically identifying the three types of stop consonant with the same place of articulation is an important functionality for a computer-assisted speech training program for CLP patients. However, such functionality requires subtle acoustic analysis that pattern matching techniques relying on gross spectral features such as HMMs and DTW cannot provide well enough. To discriminate among these subtle differences, using acoustic measurements based on acoustic-phonetics knowledge is unavoidable although selecting suitable sets of acoustic measurements requires intense knowledge engineering. It has been shown in a number of works that classification features based on acoustic measurements can

lead to successful classification tasks while valuable insights on acoustic nature of the classified sounds are obtained in the process [3],[4],[5],[6].

### 2.2 Phonological Analysis of Thai Language

In order to investigate acoustic cues significant to stop consonant voicing classification, we first explore the phonological aspect of Thai language in general. Thai language is known for being a tonal language. Here, we elaborate the five lexical tones and their distinctive linguistic features. There are 21 consonantal phonemes, 12 consonant clusters, 18 monophthongs, 6 diphthongs and 5 tones in the Thai language [7], [8]. Places of articulation of various sounds in the Thai sound system can be labial, alveolar, palatal, velar or glottal, while manners of articulation are classified into two major groups: stops and non-stops. The stops consist of stop consonants which are subcategorized into voiced, voiceless (unaspirated), and voiceless aspirated ones. The non-stops are subcategorized into nasals, fricatives, a trill, a lateral and approximants. There are 8 different stop consonants in the Thai language. Their places of articulation and voicing characteristics are shown in Table 1.

Table 1. Thai stop consonants

Place of Articulation	Voicing Characteristic		
	Voiced	Voiceless Unaspirated	Voiceless Aspirated
Labial	/b/	/p/	/p <sup>h</sup> /
Alveolar	/d/	/t/	/t <sup>h</sup> /
Velar	-	/k/	/k <sup>h</sup> /

### 2.3 Acoustic-phonetic Theory of Human Speech Production

A widely-accepted model of human speech production is the source-filter model where speech signals coming out of a speaker's mouth are the result of filtering some source signal with a time-varying filter, as shown in Figure 1. There are two types of source signals. The first one is the glottal source which is a quasi-periodic signal occurring at the speaker's vocal folds due to their vibration. During their vibration cycle, the vocal folds allow airflow to pass through the glottis when the vocal folds are abducted, and they block all airflow from the lungs when they are abducted. The other source is the noise source which occurs when a rapid airflow rushes through a very narrow constriction. The time-varying filter amplifies (or attenuates) different frequency components of the source signal differently based on the vocal tract configuration.

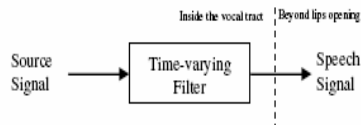


Figure 1. The source-filter model of speech production

In the production of stop consonants with the same place of articulation, the vocal tract configurations are the same. Therefore, the effect of the time-varying filter cannot make the output speech distinguishable. However, for voiced stop consonants, vocal folds vibrate and radiate some low-frequency signal through the wall of the mouth. For voiceless stop consonants, no such vibration occurs. Therefore, signal energy is

absent during the closure. In the syllable-initial position, voiceless stops can be aspirated or unaspirated. The glottis is spread during production of an aspirated voiceless stop consonant. There is a noise source caused by rapid airflow rushing through this spread glottis after the release of the closure and before the vocal folds start vibrating again as a part of the production of the following vowel. This noise is called an aspiration noise; it excites the vocal tract filter and manifests itself in the associated speech signal.

### 3. EXPERIMENTAL DETAILS

#### 3.1 Speech Data

Thai speech tokens in the vowel-consonant-vowel ( $V_1C_3V_2$ ) formation were used in the study. Attention was paid to acoustic events during the closure of the stop consonant ( $C_3$ ), and during the interval after the stop consonant was released and before the voicing of the vowel, followed that stops ( $V_2$ ), started.

For non-CLP speech data, utterances were recorded from 10 native Thai speakers with a sampling rate of 44,100 kHz. Each speaker was asked to say speech tokens in the  $V_1C_3V_2$  formation, where  $V_1$  was a schwa,  $C_3$  was one of the eight Thai stop consonants, /b/ /pʰ/ /d/ /t/ /k/ and /kʰ/, and  $V_2$  was either one of eight Thai corner vowels, including both tense and lax vowels.

For CLP speech data, utterances were recorded from two CLP children with ages of four to five years old and one 31 year-old CLP adult. Parents of the CLP children were informed prior to the examination. The patients were regular out patients and all of them are native Thai speakers. The CLP patients were asked to speak  $V_1C_3V_2$  tokens in the same fashion as when non-CLP speech data were collected. The speech samples were recorded with a close-talking microphone on a headset at a sampling frequency of 44100 kHz and quantized with 16 bit.

#### 3.2 Acoustic Measurements

This section describes the acoustic measurements selected as classification features. They are based on acoustic cues that have potential in discriminating the voiced, voiceless, and voiceless aspirated stop consonants with the same place of articulation.

##### 3.2.1 Durations

Regions of interest associated with a stop consonant in  $V_1C_3V_2$  formation consist of a closure interval, a release (transient, friction and aspiration) interval and transition intervals from the release towards  $V_2$ 's nucleus as well as from  $V_1$ 's nucleus towards the closure. Durations of these intervals were investigated by many researchers [3], [4], [6],[9]. The interval from the release to the point where the vocal folds start vibrating in the production of  $V_2$  is called the Voicing Onset Time (VOT). Voiceless stop consonants have an average VOT that is longer than the average VOT of the voiced ones. For aspirated stop consonants, their average VOT is clearly greater than the VOT of the unaspirated ones due to the presence of aspiration noise [10]. It has been shown that VOT could play a major role in the distinguishing stop consonant voicing, even though it is not as useful in the place of articulation classification [3],[4]. The closure duration and VOT, accompanied by other acoustic features, were also used in the categorization of Thai initial consonants [11].

Closure duration (CLS\_DUR), which is the duration from the point where the closure is formed to the point where it is released, and VOT are investigated and used in this study.

##### 3.2.2 Energy-based Measurements

Energy in different frequency regions together with other acoustic cues were successfully applied in classifying both Thai consonants and vowels [10]. Energy was also used for detecting Thai syllable segmentation [12] and vowel landmarks detection [13] in continuous speech. Suchato [3],[4] relies on energy-based measurements in classifying the English stop consonant place of articulation.

Here, speech energy is used to capture the presence of vocal fold vibration during the closure (voiced bar) as well as aspiration noise after the release. The energy in the 60-300 Hz frequency range of the speech signal in the closure interval ( $E_c$ ) is used to capture the former acoustic event while the time-averaged energy in the interval starting from the release to the point where the voicing of  $V_2$  starts ( $E_{rel}$ ) is used to capture the latter one.

For the measurement extraction, both CLS\_DUR and VOT for each token of  $V_1C_3V_2$  were obtained by manually marking the locations of the closure, the release, and the voicing onset of  $V_2$ . Once these important locations were marked,  $E_{rel}$  and  $E_c$  were extracted automatically by summing samples of the Discrete Fourier Transforms of the speech signal in appropriate frequency ranges. Speech signals were normalized so that the amplitude of the biggest speech samples was unity.

### 3.3 Experiments and Analyses

#### 3.3.1 Analysis of variance

In order to verify the validity of using the selected acoustic measurements in voicing classification, Analysis of Variance (ANOVA) was used to test whether the value distributions of the measurements obtained from the three types of stop consonants with the same place of articulation were significantly different. ANOVA tested whether the differences in means of the acoustic measurements extracted from the three types were significant.

#### 3.3.2 Classifier and Classification Experiments

The acoustic measurements were used to form classification feature vectors. Linear Discriminant Analysis (LDA) was deployed as our classification algorithm. Knowing the stop consonant's place of articulation, LDA classifies the stop consonant into either voiced stop, voiceless unaspirated stop, or voiceless aspirated stop by applying a linear transformation that maximizes separation among the three groups to the space spanned by the acoustic measurements to obtain a new space in which distance measures which have already taken into account variances of the acoustic measurements. Classification results are determined based on these distances. Multivariate normal density was assumed.

In the classification experiments of the non-CLP speech data, the Leave-One-Out Cross Validation (LOOCV) technique was used for estimating the classification accuracy. In this technique, a speech token is classified based on the models trained on the rest of speech data. For the CLP case, non-CLP speech data, from all ten speakers, was used to train the LDA model. The classification accuracy was evaluated on all the CLP tokens.

3.3.3 Subjective evaluation

Classification results of the CLP speech data obtained from automatic classification were compared against experts' judgment. Three experts, including a professional speech therapist, a speech scientist, and a graduate student with specialization in speech analysis, were asked to judge all CLP speech tokens whether each of them was either voiced, voiceless, or voiceless aspirated. The type of each stop consonant was determined based on the type that the majority of the three experts voted on.

4. RESULTS

The results obtained from ANOVA show that the four acoustic measurements are statistically different (at 1% significance level) among voiced, voiceless and voiceless aspirated stop consonants with the same place of articulation, except for  $E_b$  in the velar case. Value distributions of the four measurements for stop consonants with the three places of articulation are shown with box plots in Figure 2 to Figure 4. Table 2 to Table 4 lists the associated means and standard deviations.

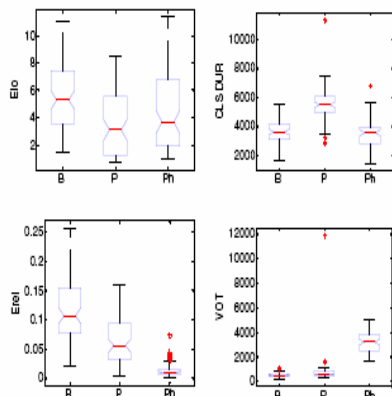


Figure 2. Distribution of measurement values for labial stops

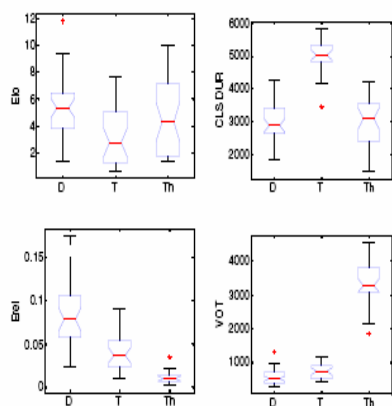


Figure 3. Distribution of measurement values for alveolar stops

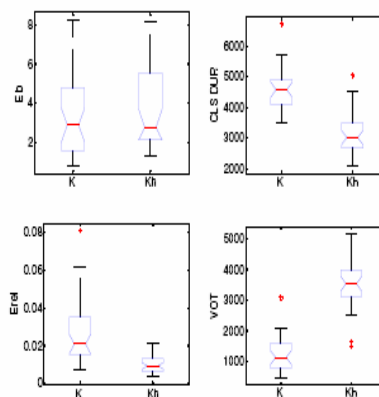


Figure 4. Distribution of measurement values for velar stops

Table 2. Means and standard deviations of the acoustic measurements extracted from labial stop consonants

Labial	CLS DUR		
	Voiced	Voiceless	Aspirated
Mean	3651.1	5492.5	3348.3
S.D.	763.8	1272.2	956.4
$E_b$			
Mean	5.6	3.4	4.4
S.D.	2.4	2.4	2.7
VOT			
Mean	494.2	791.1	3215.4
S.D.	162.9	1285.2	838
$E_{rel}$			
Mean	0.1194	0.0643	0.0134
S.D.	0.0580	0.0378	0.0118

Table 3. Means and standard deviations of the acoustic measurements extracted from alveolar stop consonants

Alveolar	CLS DUR		
	Voiced	Voiceless	Aspirated
Mean	2976.5	5009.1	3003.1
S.D.	566.6	509.2	694.8
$E_b$			
Mean	5.3	3.3	4.8
S.D.	2.3	2.2	3.1
VOT			
Mean	583.7	732.8	3385.7
S.D.	231.9	223.4	629.1
$E_{rel}$			
Mean	0.0851	0.0404	0.0117
S.D.	0.0380	0.0195	0.0069

Table 4. Means and standard deviations of the acoustic measurements extracted from velar stop consonants

Velar	CLS DUR	
	Voiceless	Aspirated
Mean	4559.5	3138.8
S.D.	667.7	650.9

		$E_b$	
Mean		3.3	3.7
S.D.		1.9	2.0
		VOT	
Mean		1205.9	3486.4
S.D.		557.4	781.3
		$E_{rel}$	
Mean		0.0257	0.0096
S.D.		0.0164	0.0045

Classification experiments, adopting LOOCV, were performed to evaluate how well the four measurements were distinguished among the three types of stops in the same place of articulation. Table 5 shows the confusion matrices obtained from the classification experiments.

Table 5. Confusion matrices for the three places of articulation

Labial	Classified as		
	Voiced	Voiceless	Aspirated
Voiced	65	15	0
Voiceless	12	67	1
Aspirated	0	1	79
Alveolar	Classified as		
	Voiced	Voiceless	Aspirated
Voiced	32	0	0
Voiceless	0	32	0
Aspirated	1	0	31
Velar	Classified as		
	Voiced	Voiceless	Aspirated
Voiced			
Voiceless		31	1
Aspirated		2	30

Classification accuracies are 88%, 99%, and 95% for the labial, alveolar and velar cases, respectively. The overall classification accuracy is 92%.

To evaluate how well the four measurements can perform in the voicing classification in the CLP cases, LDA whose parameters were trained upon the non-CLP speech data were used to classify the real CLP speech data. Results of the LDA classification experiments on CLP speech data were compared with the subjective evaluation of the experts' judgments. The three CLP patients were asked to say a balanced set of tokens. However, the type of each speech token was determined based on the aural perception of the three experts. Table 6 lists the number of each CLP speech token type.

Table 6. Number of CLP speech tokens

	Labial	Alveolar	Velar
Voiced	8	12	
Voiceless	64	48	33
Aspirated	10	12	15

Table 7 shows the confusion matrices obtained from the classification experiments.

Table 7. Confusion matrices of CLP stops for the three places

Labial	Classified as		
	Voiced	Voiceless	Aspirated
Voiced	8	0	0
Voiceless	6	42	6
Aspirated	0	0	10

Alveolar	Classified as		
	Voiced	Voiceless	Aspirated
Voiced	12	0	0
Voiceless	6	38	4
Aspirated	0	0	12
Velar	Classified as		
	Voiced	Voiceless	Aspirated
Voiced			
Voiceless		31	2
Aspirated		1	14

The classification accuracies in the CLP case are 83%, 86%, and 94% for labial, alveolar and velar stop consonants, respectively. The overall classification accuracy is 87%.

## 5. DISCUSSIONS

From the ANOVA results, the value distributions of CLS\_DUR, VOT,  $E_b$ , and  $E_{rel}$  have the same tendencies for every place of articulation. Regardless of the places, voiced stop consonants have the highest average  $E_b$ , which reflects the presence of their corresponding voiced bars. They also have the highest average  $E_{rel}$ , while the lowest belongs to the aspirated ones. If  $E_{rel}$  was not normalized with the corresponding VOT, aspirated stops would have the highest average due to the presence of aspiration noise. However, here, we adopt the normalized one and the trends are as shown in the box plot result. Also from the boxplots, we can see that voiceless unaspirated stop consonants have the longest closure duration. And, as expected, voiceless aspirated stop consonants have the longest VOT.

All of the acoustic measurements are shown to be significantly different among different stops with the same place of articulation except  $E_b$  in the velar case. This does not come as a surprise to us since the measurement was aimed at indicating the presence of voiced bars which have never occurred in velar stop consonants in the Thai sound system.  $E_b$  is still useful in other cases.

From the separation among stop consonants with different voicing characteristic observed in the boxplots, we can see that different acoustic measurements have different levels of discriminating ability. The acoustic measurement that demonstrates a clear separation between unaspirated stops (voiced and voiceless unaspirated stops) and aspirated stops is VOT. Specifically for the data set that we used here, VOT can separate aspirated stop consonants from the rest in most cases. Only a few data points in the velar case, as can be observed from the bottom right plot in Figure 4, possess VOT values that could be confusing among the aspirated and the unaspirated groups. This is understandable since the VOT values of velar stop consonants are generally longer than ones of the other two places of articulation. Therefore, some of the longer unaspirated velar tokens could be confused with some aspirated tokens.

From the boxplots, even though VOT does a reasonable job in separating unaspirated stop consonants from the aspirated ones, it is obvious that no single acoustic measurement can distinguish between the three types of voicing characteristic. This is why the four acoustic measurements are combined and used together in the classification experiments. The confusion matrices in Table 5 show that, when the LDA model is trained on non-CLP data and

the classification accuracy is evaluated on another set of non-CLP data, the overall classification accuracy obtained is 92%. This indicates the separating ability of the combination of our four acoustic measurements. The worse case is with labial stop consonants where the classification accuracy is still close to 90%. For the other two cases, the classification accuracy is quite high, especially in the alveolar case. Only one alveolar token out of the total of 96 tokens is misclassified.

However, we would like to find out how well the four acoustic measurements perform not only in the non-CLP cases but also with the real CLP speech data. From a set of  $V_1C_3V_2$  where voicing characteristics are balanced, the result of the experts' judgment shows that a number of stop consonants from these CLP patients are perceived as voiceless unaspirated even when the patients intend to utter other types of stop. Still, what we are really concerned with is whether the classification based on our four acoustic measurements yields the classification results identical with the experts.

The overall classification accuracy of the real CLP stop consonants is 87%, despite that the LDA model is trained upon non-CLP tokens. Compared to the classification results of the non-CLP cases, the accuracy obtained in the CLP case is 5% less than the non-CLP case. The fact that the classification accuracy percentage is in the high 80's demonstrates that the four acoustic measurements serve as good classification features even when no CLP speech data has been used in the training of the classifier, which is usually the case due to limited numbers of accessible CLP patients and complications in preparing a large body of CLP speech data.

## 6. CONCLUSION

It has been shown in this work that some combinations of acoustic measurements derived from acoustic-phonetic principle can be used to distinguish between voice, voiceless and voiceless aspirated stop consonants. The four acoustic measurements proposed for doing the classification task in this work have been shown to perform well and satisfactory classification accuracy percentages have been achieved, even in the case where real CLP speech data is not accessible in the training process. Although, the algorithm for automatic classification of stop consonant voicing characteristics based on acoustic-phonetic knowledge has not yet been completed in this work and further research is needed for automatically locating precise time-stamp for the stop closures as well as the releases, the findings should serve as an important foundation to the development of a fully functional computer-assisted speech training program for the cleft lip and palate patients in the future.

## 7. ACKNOWLEDGEMENTS

The authors would like to thank the native Thai speakers and the CLP patients who dedicate their time for the recordings, Nantana Pracharitpukdee of Chulalongkorn Hospital, Bangkok, who provided general knowledge of the cleft lip and palate as well as valuable suggestions, and who facilitated the recording of patients' speech data.

## 8. REFERENCES

- [1] Massaro, D.M., A computer-animated tutor for spoken and written language learning, in *Proc. of the 5th international conference on Multimodal interfaces*, Vancouver, British Columbia, Canada, 2003, pp.172-175.
- [2] Jiang, X., Wang, Y. and Zhang, F., Visual speech analysis and synthesis with application to Mandarin speech training, in *Proc. Of the ACM symposium on Virtual reality software and technology*, London, UK,1999, pp.111-115.
- [3] Suchato, A., Classification of Stop Consonant Place of Articulation: Combining Acoustic Attributes, in *Proc. From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, Cambridge, MA, June 2004, pp. 197-202.
- [4] Suchato, A., Classification of Stop Consonant Place of Articulation. *Journal of Acoustical Society of America*, Vol. 115, No.5, May 2004
- [5] Slifka, J., Acoustic cues to vowel-schwa sequences for high front vowels. *Journal of the Acoustical Society of America*, Vol.118, 2037, 2005
- [6] Ali, A.M.A., Van der Spiegel, J., and Mueller, P., Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants, *IEEE Trans. Speech Audio Processing*, vol. 9, no. 8, November 2001.
- [7] Luksaneeyanawin, S., A Three Dimensional Phonology: A Historical Implication, *Proceedings of the 3<sup>rd</sup> International Symposium on Language and Linguistics*, pp. 75-90, 1992.
- [8] Lunksaneeyanawin, S., Linguistics Research and Thai Speech Technology, *Proceeding of the 5<sup>th</sup> International Conference on Thai Studies*, School of Oriental and African Studies, University of London, pp. 1-29 1993.
- [9] Maneenoi, E., An Acoustic Study of Syllable Rhymes: A Basis for Thai Continuous Speech Recognition System, *Ph.D. Thesis*, Chulalongkorn University, Bangkok, 2003.
- [10] Tharnsakun, W., The Acoustic Analysis of Thai Stops, *Master thesis*, Department of Linguistics, Chulalongkorn University, Bangkok, 1988.
- [11] Thubthong, N., A Thai Speech Recognition System based on Phonemic Distinctive Features, *Master thesis*, Department of Computer Engineering, Chulalongkorn University, Bangkok, 1995.
- [12] Nutthacha, J., Somchai, J., Sudaporn, L., Visarut, A. and Chai, W., Thai Syllable Segmentation for Connected Speech Based on Energy, *IEEE Trans. Speech Audio Processing*, pp. 169-172, 1998.
- [13] Dareeyoah, P., Vowel Landmark Detection in Thai Continuous Speech, *Master thesis*, Department of Computer Engineering, Chulalongkorn University, Bangkok, 2007



## ประวัติผู้เขียนวิทยานิพนธ์

นางสาว บุศมาส พลกุล เกิดเมื่อวันที่ 29 ตุลาคม พ.ศ. 2523 ที่จังหวัด กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาบัณฑิต ในสาขาวิศวกรรมคอมพิวเตอร์ จากคณะ วิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2545 และเข้าศึกษาต่อหลักสูตรวิทยาศา สตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ (ภาคนอกเวลาราชการ) ภาควิชา วิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2549 ขณะศึกษาได้มีโอกาสไปเสนอผลงานเรื่อง Stop Consonant Voicing Classification for Computer-assisted Speech Training of Patients with Cleft Lips and Palates ในงานประชุม วิชาการ “International Convention for Rehabilitation Engineering & Assistive Technology (i-Create 2007)” ณ ประเทศสิงคโปร์ ปัจจุบันทำงานอยู่ที่ บริษัท รอยเตอร์ ซอฟต์แวร์ ไทยแลนด์ จำกัด