พารามิเตอร์ทางเสียงสำหรับการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย

นายวิทยา โรจน์กิตติเจริญ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

ACOUSTIC PARAMETERS FOR MANNER OF ARTICULATION CLASSIFICATION IN THAI CONTINUOUS SPEECH

Mr. Wittaya Rochkittichareon

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2011

หัวข้อวิทยานิพนธ์	พารามิเตอร์ทางเสียงสำหรับการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย
โดย	นายวิทยา โรจน์กิตติเจริญ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์.ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

……. ประธานกรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

…….. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร. อติวงศ์ สุชาโต)

……. อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ)

……. กรรมการภายนอกมหาวิทยาลัย
(ดร.ชัย วุฒิวิวัฒน์ชัย)

วิทยา โรจน์กิตติเจริญ : พารามิเตอร์ทางเสียงสำหรับการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย. (ACOUSTIC PARAMETERS FOR MANNER OF ARTICULATION CLASSIFICATION IN THAI CONTINUOUS SPEECH) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร. อติวงศ์ สุชาโต, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ.ดร.โปรดปราน บุณยพุกกณะ, 47 หน้า.

ในการพัฒนาระบบรู้จำเสียงแบบอื่นเช่น ระบบรู้จำเสียงแบบแลนมาร์ค จะต้องทำการหาตำแหน่งของแลนมาร์ค ของเสียงที่เราให้ความสนใจ เช่นตำแหน่งของเสียงพยัญชนะหรือตำแหน่งของเสียงสระ เป็นต้น เพื่อใช้เป็นข้อมูลขาเข้าในการรู้จำเสียงพูด ดังนั้นเป้าหมายงานวิทยานิพนธ์นี้จึง ได้เน้นไปที่การจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย เพื่อสามารถนำไปใช้ในการพัฒนาระบบรู้จำเสียงพูดแบบแลนแลนมาร์คได้

โดยที่งานวิทยานิพนธ์นี้ได้ทำการปรับปรุงชุดพารามิเตอร์ทางเสียงสำหรับเพื่อให้เหมาะสมกับภาษาไทย ซึ่งประกอบด้วยโดยได้ปรับให้มีการใช้ 1) จุดศูนย์ถ่วงของสเปกตรัม 2) อัตราการตัดศูนย์ในช่วงเวลา 3) อัตราส่วนพลังงานในช่วงความถี่ [0-400] Hz ต่อ พลังงานในช่วงความถี่ [400-6000] Hz เพิ่มเติม จากผลการทดลองจำแนกสมบัติทางสมบัติทางสวนสัทศาสตร์ แสดงให้เห็นว่ามีความผิดพลาดในการจำแนกสมบัติทางสมบัติทางสวนสัทศาสตร์ลดลง 28.09%, 11.0%, 2.41% สำหรับการจำแนกสมบัติทางสวนสัทศาสตร์ [คอนทินิวแอนท์], [ซิลลาบิค] และ [ไซเรนท์] ตามลำดับ เมื่อทำการเปรียบเทียบกับ ชุดพารามิเตอร์ทางเสียงที่ใช้ในการจำแนกสมบัติทางสวนสัทศาสตร์สำหรับเสียงภาษาอังกฤษ และเมื่อทำการตัดแบ่งเสียงเพื่อทำการหาตำแหน่งเสียงพยัญชนะ และ เสียงสระ พบว่าได้ความถูกต้องในการตัดแบ่ง 80.46% โดยมีความผิดพลาดในการตัดแบ่งลดลง 23.46% เมื่อเทียบกับระบบอ้างอิงที่ใช้การรู้จำเสียงพูดแบบอาศัยแบบจำลองฮิดเดนมาร์คอฟ ในการทดลองสุดท้ายพบว่าเมื่อทำการเทียบผลการรู้จำในระดับพยางค์ ในรูปแบบ พยัญชนะต้น-สระ-ตัวสะกด ระบบที่เสนอกับระบบอ้างอิงให้ความถูกต้องในระดับเดียวกัน

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2554 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

517 04571 21 : MAJOR COMPUTER ENGINEERING

KEYWORDS : ACOUSTIC PARAMETERS / SIGNAL PROCESSING / SPEECH RECOGNITION / SPEECH SEGMENTATION

WITTAYA ROCHKITTICHAREON : ACOUSTIC PARAMETERS FOR MANNER OF ARTICULATION CLASSIFICATION IN THAI CONTINUOUS SPEECH. ADVISOR : ASST. PROF.ATIWONG SUCHATO, Ph.D., CO-ADVISOR : ASST. PROF.PROADPRAN PUNYABUKKANA, Ph.D., 47 pp.

In landmark-based speech recognition system. We need to locate the landmark of speech such a consonant landmark or a vowel landmark. For using that kind of landmark as an input data to speech recognition system. This thesis focuses on finding broad manner class of Thai speech. For developing the landmark-based speech recognition system

This thesis is aimed at the improvement of the acoustic parameters for the Thai automatic speech recognition system. We proposed acoustic parameters that capture the characteristics of broad manner class of Thai speech. These acoustic parameters are: 1) spectral center of gravity 2) short time zero crossing rate to 3) the energy ratio E[0-400] to E[400-6000]. The results showed 28.09%, 11.0% and 2.41% error reductions for the continuant, the syllabic and the silence features, respectively, when compared to acoustic parameters used in English. The accuracy of 80.46% was obtained from the speech segmentation task and also introduced a 23.46% error reduction when compared to the baseline HMM-MFCC based broad class segmentation. We also found similar performance for word classification in the CVC context when compared to the baseline HMM-MFCC in word recognition tasks.

Department : Computer Engineering	Student's Signature .....
Field of Study : Computer Engineering	Advisor's Signature .....
Academic Year : 2011	Co-advisor's Signature .....

กิตติกรรมประกาศ

ในโอกาสนี้ข้าพเจ้าขอขอบคุณ ผศ.ดร. อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ผศ.ดร. โปรดปราน บุณยพุกกณะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ท่านทั้งสองได้ให้ความช่วยเหลือ ให้คำแนะนำและข้อคิดที่เป็นประโยชน์ อันเป็นส่วนสำคัญที่ทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณคณะกรรมการสอบ ศ.ดร.บุญเสริม กิจศิริกุล และ ดร.ชัย วุฒิวิวัฒน์ชัย ที่สละเวลามาดำเนินการสอบวิทยานิพนธ์ให้ข้าพเจ้า และให้คำแนะนำและข้อคิดต่างๆ ที่เป็นประโยชน์ในการทำวิทยานิพนธ์

สุดท้ายนี้ขอขอบคุณบิดา มารดา รวมถึง พี่ๆ เพื่อนๆ และ น้องๆ ในห้องปฏิบัติการ SLS-ATL ที่ได้ให้ความช่วยเหลือ จนกระทั่งวิทยานิพนธ์นี้สำเร็จได้ด้วยดี

สารบัญ

หน้า

สารบัญภาพ

สารบัญตาราง

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการพัฒนาระบบรู้จำเสียงพูด (Speech Recognition) มีความก้าวหน้าไปอย่างมาก ซึ่งทำให้มนุษย์สามารถทำการติดต่อสื่อสารกับเครื่องคอมพิวเตอร์โดยใช้เสียงได้ จะเห็นได้จากการนำเอาเทคโนโลยีการรู้จำเสียงพูดมาใช้ในระบบต่างๆเช่น การค้นหาข้อมูลผ่านทางอินเตอร์เน็ตบนโทรศัพท์เคลื่อนที่ โดยใช้เสียง ระบบให้บริการข้อมูลต่างๆ เป็นต้น

ระบบรู้จำเสียงพูดที่ได้รับการยอมรับในปัจจุบันคือ ระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิด-เดนมาร์คอฟโมเดล (Hidden Markov Model, HMM)[1] ซึ่งเป็นระบบรู้จำเสียงพูดที่ใช้ความรู้ทางด้านสถิติเข้ามาใช้ในการสร้างระบบรู้จำเสียงพูด และให้ผลการรู้จำเสียงพูดที่มีประสิทธิภาพดี อย่างไรก็ตาม การจะทำให้ระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟโมเดลมีประสิทธิภาพดี จำเป็นต้องใช้ข้อมูลจำนวนมากพอในการฝึกและใช้ทรัพยากรของระบบค่อนข้างมาก เพื่อที่จะทำให้ระบบรู้จำสามารถทำการรู้จำเสียงได้ครอบคลุมได้จำนวนมากๆ เพื่อทำการลดข้อจำกัดที่เกิดขึ้นจากระบบรู้จำเสียงพูดที่ใช้องค์ความรู้ทางด้านสถิติเป็นพื้นฐาน จึงได้มีเกิดการพัฒนาระบบรู้จำเสียงแบบอื่นเช่น ระบบรู้จำเสียงแบบแลนมาร์ค (Landmark Based Speech Recognition) [2] ซึ่งมีการประยุกต์ความรู้จากสวนสัทศาสตร์ (Acoustic-phonetics) มาใช้ร่วมกับองค์ความรู้ทางสถิติ เข้ามาในการรู้จำเสียงพูด โดยทำใช้ความรู้จากสวนสัทศาสตร์มาใช้ในการหาหน่วยเสียงสระและหน่วยเสียงพยัญชนะจากสัญญาณเสียง ซึ่งให้ผลการรู้จำเสียงพูดที่ได้ใกล้เคียงกับระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟโมเดล ทำให้เกิดแนวคิดที่จะทำงานวิจัยที่เสนอวิธีการหาตำแหน่งของหน่วยเสียงพยัญชนะ เพื่อสามารถที่จะนำไปใช้ในการพัฒนาระบบรู้จำเสียงพูดให้มีประสิทธิภาพที่ดีมากขึ้นด้วย

ในการพัฒนาระบบรู้จำเสียงพูดขึ้นมานั้น จะต้องประกอบด้วยกระบวนการต่างๆ หลายกระบวนการ เช่น ขั้นตอนการหาคุณลักษณะต่างๆ ของเสียงเพื่อใช้เป็นตัวแทนของสัญญาณเสียง ขั้นตอนการตัดแบ่งเสียง (Speech Segmentation) คือการทำการกำหนดขอบเขตของหน่วยเสียง ซึ่งอาจจะทำการขอบเขตของเสียงในระดับคำ หรือ ระดับหน่วยเสียงที่ประกอบหน่วยเสียงพยัญชนะ, หน่วยเสียงสระ ขั้นตอนการสร้างโมเดลการออกเสียง คือโมเดลที่ทำหน้าที่ในการบอกว่าคำ (Word) แต่ละคำมีการออกเสียงอย่างไร ขั้นตอนการสร้างโมเดลภาษา เพื่อนำไปใช้ในการ

บอกว่าคำพูดใดๆ จะ มีโอกาสตามด้วยคำพูดใดบ้าง หรือก็คือความน่าจะเป็นที่คำใดๆจะพูดต่อกันได้ นอกจากนี้ยังสามารถนำไปใช้ เพื่อที่จะบอกได้ว่าประโยคแต่ละประโยคแต่ละประโยคมีโอกาสเกิดขึ้นได้เท่าไหร่

จากขั้นตอนต่างๆในการพัฒนาระบบรู้จำเสียงพูดจะเห็นได้ว่าการตรวจหาตำแหน่งของพยัญชนะและสระ เป็นหนึ่งในขั้นตอนที่มีความสำคัญ เพราะถ้าเราสามารถทำการตรวจหาตำแหน่งของพยัญชนะและสระได้ถูกต้องแม่นยำ ก็จะทำให้เราสามารถพัฒนาระบบรู้จำที่มีประสิทธิภาพสูงขึ้นได้

ดังนั้นในงานวิจัยนี้จึงมีวัตถุประสงค์ที่ ทำการศึกษาและพัฒนาวิธีในการลักษณะการเปล่งเสียงในเสียงพูดภาษาไทย เพื่อให้มีความถูกต้องแม่นยำมากที่สุด เพื่อที่จะสามาราถนำไปใช้ในการพัฒนาระบบรู้จำเสียงพูดต่อไปได้

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและพัฒนาวิธีการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย โดยใช้วิธีทางสวนสัทศาสตร์
2. เพื่อใช้เป็นงานวิจัยพื้นฐานที่สามารถนำไปใช้พัฒนาระบบรู้จำเสียงพูดแบบแลนมาร์กสำหรับภาษาไทยต่อไปได้
3. เพื่อนำไปใช้ประยุกต์ใช้หรือพัฒนาระบบรู้จำเสียงพูดต่อเนื่องภาษาไทยแบบอื่น ให้มีประสิทธิภาพที่ดียิ่งขึ้น

ขอบเขตของการวิจัย

1. วิธีการจำแนกลักษณะการเปล่งเสียงในงานวิจัยนี้ เป็นการตรวจหาแลนมาร์กของเสียงพยัญชนะและเสียงสระในเสียงพูดต่อเนื่องภาษาไทย

2. ฐานข้อมูลเสียงที่ใช้ในวิทยานิพนธ์นี้คือ ฐานข้อมูลเสียงภาษาไทย LOTUS[3] โดยจะใช้ชุดหน่วยเสียงสมดุล (Phonetically Distributed Set: PD) ของฐานข้อมูลเสียงภาษาไทย LOTUS

3. การวัดผลของการจำแนกลักษณะการเปล่งเสียงของงานวิจัยนี้จะทำได้โดยทำการเปรียบเทียบผลการจำแนกลักษณะการเปล่งเสียงได้ เทียบกับระบบอ้างอิง(base line) ที่ใช้วิธีการตัดแบ่งเสียงออกเป็นประเภทของเสียง 5 ประเภท โดยใช้ Hidden Markov Model (HMM)

4. งานวิจัยนี้เป็นการจำแนกลักษณะการเปล่งเสียงจากเสียงพูดต่อเนื่องภาษาไทยเท่านั้น

5. งานวิจัยนี้ได้ทำการตั้งสมมุติฐานว่าเสียงพยัญชนะในกลุ่มเสียงกักที่เป็นเสียงพยัญชนะท้ายพยางค์หรือเสียงตัวสะกด (/p^/, /t^/, /k^/) มีลักษณะของเสียงเป็นเสียงเงียบ

6. งานวิจัยนี้ได้ทำการตั้งสมมุติฐานว่าเสียงพยัญชนะในกลุ่มเสียงกึ่งสระที่เป็นเป็นเสียงพยัญชนะท้ายพยางค์หรือเสียงตัวสะกด (/j^/, /w^/) มีลักษณะเสียงเป็นเสียงสระ

ประโยชน์ที่คาดว่าจะได้รับ

การจำแนกลักษณะการเปล่งเสียงที่นำเสนอนั้น สามารถนำไปใช้ในการพัฒนาระบบรู้จำเสียงพูดแบบแลนมาร์ก หรือนำประยุกต์ใช้กับระบบรู้จำเสียงแบบต่างๆได้ เพื่อให้ระบบรู้จำเสียงพูดมีประสิทธิภาพที่ดียิ่งขึ้น

วิธีดำเนินการวิจัย

1. ขั้นตอนการศึกษาเบื้องต้น

   1.1.ศึกษาข้อมูลเกี่ยวกับสารสนเทศสวนสัทศาสตร์ และการหาค่าพารามิเตอร์ทางเสียง

1.2.ศึกษางานวิจัยที่มีนำสารสนเทศสวนสัทศาสตร์ และงานวิจัยที่ทำการหาค่าพารามิเตอร์ทางเสียงเพื่อนำมาใช้ในการตรวจหาเสียงพยัญชนะ, เสียงสระ หรือการรู้จำเสียงพูด

1.3.วิเคราะห์สัญญาณเสียง เพื่อหาสารสนเทศสวนสัทศาสตร์และค่าพารามิเตอร์ทางเสียง ที่จะนำมาใช้มาใช้ในการจำแนกลักษณะการเปล่งเสียง

2. ขั้นตอนการทดลองและทดสอบประสิทธิภาพ

2.1.การนำความรู้ทางสถิติมาทำการวิเคราะห์ข้อมูล เพื่อทำการหาค่าพารามิเตอร์ทางเสียงที่เหมาะสม สำหรับใช้ในการฝึกซัพพอร์ตเวกเตอร์แมชชีน เพื่อใช้ในการจำแนกคุณสมบัติทางสวนสัทศาสตร์

2.2. ทำการการตรวจหาเสียงพยัญชนะแล้วทำการเปรียบเทียบผลที่ได้กับการวิธีการตัดแบ่งเสียงออกเป็นประเภทของเสียง 5 ประเภท โดยใช้ Hidden Markov Model (HMM)

3. ทำการวิเคราะห์ผลที่ได้และสรุปผลการทดลอง

ลำดับขั้นตอนในการเสนอผลการวิจัย

ในวิทยานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาในการเสนอผลการวิจัยออกเป็น 4 ส่วนคือ

ในบทที่ 2 จะกล่าวถึงทฤษฎีทางด้านเสียงภาษาไทย ทฤษฎีที่ใช้ในการจำแนกลักษณะการเปล่งเสียงในเสียงต่อเนื่องภาษาไทย รวมทั้ง วรรณกรรมที่เกี่ยวข้องกับการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย

ในบทที่ 3 ได้กล่าวถึง การหาค่าลักษณะสำคัญของเสียง และวิธีการที่ใช้ในการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย

ในบทที่ 4 ได้กล่าวถึง ฐานข้อมูลที่ใช้ในการวัดประสิทธิภาพในการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย วิธีและขั้นตอนในการทดสอบการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย การเสนอและวิเคราะห์ข้อมูลของผลการทดลอง

ในบทที่ 5 ได้กล่าวถึง การสรุปผลการทดลองและข้อเสนอแนะในการพัฒนางานวิจัยในการจำแนกลักษณะการเปล่งเสียงนี้ต่อไป

ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตอบรับให้ตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง "Broad Phonetic Class Segmentation Study for Thai Automatic Speech Recognition" โดย วิทยา โรจน์กิตติเจริญ อติวงศ์ สุชาโต และ โปรดปราน บุณยพุกกณะ ในงานประชุมวิชาการ “2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON 2012) (IEEE Catalog Number CFP1206E-ART, ISBN: 978-1-4673-2025-2)” ประเทศไทย ในระหว่างวันที่ 16-18 พฤษภาคม 2555

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เสียงในภาษาไทย

เสียงในภาษาไทย[4] ประกอบด้วยหน่วยเสียงจากตัวอักษรภาษาไทยมีทั้งหมด 44 หน่วยเสียงได้แก่หน่วยเสียงพยัญชนะ 21 หน่วยเสียง, หน่วยเสียงสระ 18 หน่วยเสียง และหน่วยเสียงวรรณยุกต์ 5 หน่วยเสียง

1. เสียงพยัญชนะ ในภาษาไทยมีหน่วยเสียงพยัญชนะทั้งหมด 21 หน่วยเสียง สำหรับงานวิจัยนี้เน้นความแตกต่างของหน่วยเสียงพยัญชนะจึงแบ่งหน่วยเสียงพยัญชนะออกเป็นกลุ่มเสียงพยัญชนะตามลักษณะของลมที่ผ่านช่องทางเดินของเสียง (vocal tract) ได้ดังนี้

    1. เสียงหยุด (Stop) สามารถแบ่งได้เป็น 2 กลุ่มย่อยคือ เสียงพยัญชนะระเบิด (Plosive Stop) และเสียงพยัญชนะกัก (Unreleased Stop) เสียงพยัญชนะระเบิดเกิดจากการที่มีการปิดกั้นลมไว้ภายในปาก แล้วจุดที่กักลมไว้เปิดออกมาทำให้เกิดลมพุ่งออกมาทันที โดยสียงพยัญชนะระเบิดแบ่งได้เป็น เสียงพยัญชนะระเบิดธนิต (Aspirated Plosive) ซึ่งจะมีกลุ่มลมพ่นออกมา หลังเปล่งเสียง และเสียงพยัญชนะระเบิดสิถิล (Unaspirated Plosive) ซึ่งจะไม่มีกลุ่มลมพ่นออกมา ส่วนเสียงพยัญชนะกักเกิดจากลมที่เปล่งออกถูกกักไว้ ณ ที่ใดที่หนึ่งในปาก และไม่ได้ถูกปล่อยออกมา ซึ่งเสียงพยัญชนะกักนี้จะเป็นเสียงตัวสะกดท้ายพยางค์

    2. เสียงนาสิก (Nasal) เป็นเสียงพยัญชนะที่เกิดจากมีการปิดกั้นลมไว้ในปาก และเกิดการลดระดับของเพดานอ่อนและช่องวีลิคเปิดทำให้ลมถูกส่งออกผ่านโพรงจมูก

    3. เสียงเสียดแทรก (Fricative) เป็นเสียงพยัญชนะที่เกิดจากการปิดกั้นลมที่ไม่สมบูรณ์ในช่องเสียงทำให้ลมที่ผ่านออกมาต้องผ่านช่องแคบเล็กๆ ณ ที่ใดที่หนึ่งในช่องปาก ทำให้ลมแทรกผ่านไปในลักษณะเสียดสี เกิดเป็นเสียงเหมือนเสียงรบกวน

4. เสียงกึ่งเสียดแทรก (Affricate) เป็นเสียงพยัญชนะที่เกิดจากมีการปิดกั้นลมไว้ภายในปาก เหมือนเสียงหยุดหรือเสียงระเบิด แต่แทนที่จะปล่อยลมออกมาทันที อวัยวะที่ใช้ในการออกเสียงจะเปิดออกอย่างช้าๆ ตามด้วยการเกิดเสียงเสียดแทรก และเพดานอ่อนยกขึ้นปิดช่องวีลิค ทำให้ลมออกทางปาก

5. เสียงข้างลิ้น (Lateral) เป็นเสียงพยัญชนะที่เกิดจากมีการปิดกั้นลมไว้ภายในปาก และมีจุดปิดกั้นอยู่ภายในปาก โดยใช้ลิ้นปิดบริเวณปุ่มเหงือกหรือเพดานแข็งส่วนกลางไว้ ทำให้ลมไหลผ่านออกมาท้างข้างลิ้น ซึ่งจะไหลออกมาข้างเดียวหรือสองข้างก็ได้

6. เสียงรัว (Trill) เป็นเสียงพยัญชนะที่เกิดจากปลายลิ้นกระดกขึ้นไปแตะปุ่มเหงือกอย่างเร็วและแตะหลายครั้งจนได้ยินเป็นเสียงรัว

7. เสียงพยัญชนะกึ่งสระ (Semi-Vowel) หรือ เสียงเปิด (Approximant) เป็นเสียงพยัญชนะที่เกิดขึ้น โดยการเปิดกว้างของช่องปาก ทำให้ลมผ่านออกมาได้โดยสะดวก โดยไม่มีการปิดกั้นของลม หรือไม่มีการบังคับให้ลมแทรกออกมาผ่านตามช่องแคบๆ ในลักษณะเสียดสี

นอกจากนี้เรายังสามารถแบ่งเสียงพยัญชนะตามตำแหน่งที่เกิดเสียงในพยางค์คือ เสียงที่สามารถเกิดในตำแหน่งต้นคำหรือต้นพยางค์ได้ทั้ง 21 หน่วยเสียง เกิดในตำแหน่งท้ายคำหรือท้ายพยางค์ได้ 9 หน่วยเสียง เกิดในตำแหน่งควบกล้ำได้ 3 หน่วยเสียง โดยที่

หน่วยเสียงพยัญชนะต้น มี 21 หน่วยเสียง ได้แก่ เสียงระเบิดมี 9 หน่วยเสียงได้แก่ /p/, /t/, /k/, /ph/, /th/, /kh/, /b/, /d/, /ʔ/ เสียงนาสิก มี 3 หน่วยเสียงได้แก่ /m/, /n/, /ŋ/, เสียงเสียดแทรกมี 3 หน่วยเสียงได้แก่ /f/, /s/, /h/ เสียงกึ่งเสียดแทรกมี 2 หน่วยเสียงได้แก่ /c/, /ch/, เสียงข้างลิ้นมี 1 หน่วยเสียงได้แก่ /l/, เสียงรัวมี 1 หน่วยเสียงได้แก่ /r/, เสียงพยัญชนะกึ่งสระมี 2 หน่วยเสียงได้แก่ /w/, /j/

หน่วยเสียงพยัญชนะท้าย มี 9 หน่วยเสียง ได้แก่ เสียงกักมี 4 หน่วยเสียงได้แก่ /p/, /t/, /k/, /ʔ/ เสียงนาสิกมี 3 หน่วยเสียงได้แก่/m/, /n/, /ŋ/ และเสียงกึ่งสระ มี 2 หน่วยเสียงได้แก่ /w/, /j/

หน่วยเสียงพยัญชนะควบกล้ำ (Consonant Cluster) ในภาษาไทย เสียงพยัญชนะที่เกิดได้ในตำแหน่งที่ 2 หรือตำแหน่งพยัญชนะควบกล้ำ มี 3 หน่วยเสียง ได้แก่ /r/, /l/, /w/ และเกิดได้เฉพาะตำแหน่งควบกล้ำต้นคำหรือต้นพยางค์เท่านั้น ส่วนเสียงพยัญชนะที่เกิดในตำแหน่งที่หนึ่งมี 6 หน่วยเสียงได้แก่ /p/, /t/, /k/, /ph/, /th/, /kh/, โดยที่เสียงระเบิด-เพดานอ่อน (Velar Plosive) /k/, /kh/ เท่านั้น ที่สามารถเกิดเสียงควบกล้ำกับหน่วยเสียง /w/ ส่วนหน่วยเสียง /p/,/t/, /ph/, /th/ เกิดควบได้เฉพาะกับ /r/, /l/ ดังนั้นเสียงควบกล้ำในภาษาไทย จึงมีทั้งหมด 12 หน่วยเสียงได้แก่ /pr/, /tr/, /kr/, /phr/, /thr/, /khr/, /pl/, /phl/, /kl/, /khl/, /kw/, /khw/ หน่วยเสียงพยัญชนะในภาษาไทยสามารถสรุปได้ดังตารางที่ 2.1

2. เสียงสระ เสียงสระเป็นเสียงที่ทำหน้าที่เป็นใจกลาง หรือแกน (Nucleus) ของพยางค์ ซึ่งเสียงพยัญชนะที่ไม่มีเสียงสระเป็นใจกลาง จะไม่สามารถแประกอบกันเป็นพยางค์ได้ นอกจากนี้เสียงสระเป็นเสียงที่ผ่านออกมาจากปากได้ โดยไม่มีการกักลมหรือการเสียดสีที่จุดใดจุดหนึ่งในปากขณะที่ทำการเปล่งเสียง อวัยวะที่สำคัญที่ทำให้เกิดเสียงสระต่างๆ คือ ลิ้น ริมฝีปาก

การแบ่งชนิดของสระ สามารถแบ่งได้ตามระดับความสูงต่ำของลิ้นภายในปาก, ความกว้างหรือชิดกันของปาก, ส่วนของลิ้นที่ใช้ในการออกเสียงคือใช้ลิ้นส่วนหน้า, กลางลิ้น หรือ หลังลิ้น ลักษณะของริมฝีปากกลมหรือไม่กลม, ความเกร็งหรือไม่เกร็งของลิ้น และความยาวของเสียงว่าเป็นเสียงสั้นหรือเสียงยาว

ในภาษาไทยมีทั้งสระเดี่ยว (Monophthong) , สระประสม (Diphthong) และสระเกิน (Vowel Letter) โดยมีรายละเอียดดังนี้

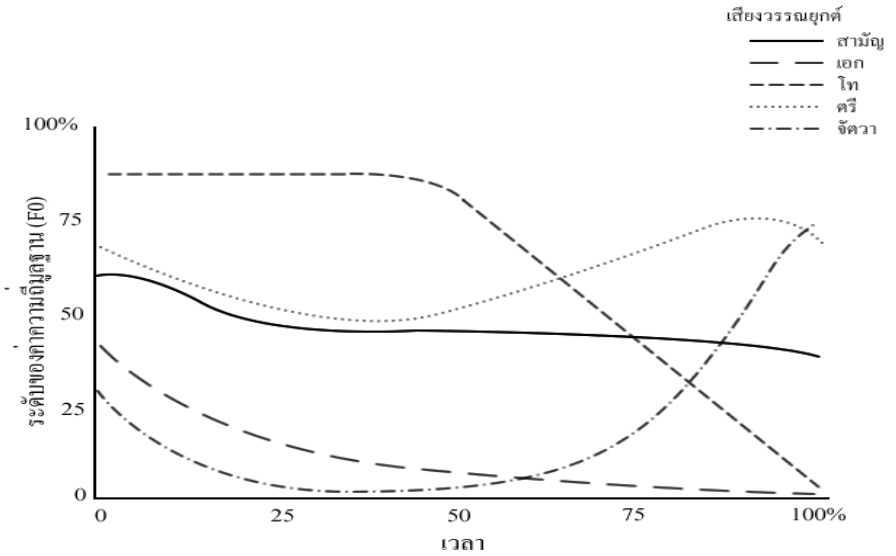
1. สระเดี่ยว(Monophthong) สระเดี่ยวในภาษาไทยมีทั้งสระเสียงสั้น (Short Vowel) จำนวน 9 หน่วยเสียง และสระเสียงยาว (long vowel) จำนวน 9 หน่วยเสียง รวม 18 หน่วยเสียงดังนี้ /i/, /iː/, /e/, /eː/, /æ/, /æː/, /ɨ/, /ɨː/, /ɜ/, /ɜː/, /a/, /aː/, /u/, /uː/, /o/, /oː/, /ɔ/, /ɔː/ • สระประสม (diphthong)ในภาษาไทยมีลักษณะสำคัญ คือ เป็นสระที่เน้นเสียงแรกเวลาออกเสียง ประกอบด้วยสระประสมเสียงสั้น 3 หน่วยเสียง และสระประสมเสียงยาว 3 หน่วยเสียง รวมทั้งสิ้น 6 หน่วยเสียงดังนี้ /ia/, /iːa/, /ɨa/, /ɨːa/, /ua/, /uːa/

2. สระเกิน (Vowel Letter) เป็นเสียงสระที่เกิดจากการรวมเสียงพยัญชนะที่เป็นเสียงพยัญชนะต้นพยางค์หรือเสียงพยัญชนะท้ายพยางค์เข้าด้วยกัน มีทั้งหมด 7 หน่วยเสียงดังนี้ /am/, /aj/, /aw/, /rv/, /rvv/, /lv/, /lvv/

หน่วยเสียงสระในภาษาไทยสามารถสรุปได้ดังตารางที่ 2.2, ตารางที่ 2.3 และตารางที่ 2.4

3. เสียงวรรณยุกต์ เสียงวรรณยุกต์นั้นถือได้ว่าเป็นหน่วยเสียงที่สำคัญ เพราะสามารถใช้แยกแยะความแตกต่างทางความหมายของคำในภาษาไทยได้ ในบางภาษาไม่จัดว่าเสียงวรรณยุกต์เป็นหน่วยเสียงในภาษา เช่นภาษาอังกฤษ โดยเสียงวรรณยุกต์นั้นคือเสียงสูงต่ำในภาษาที่เกิดจากการสั่นสะเทือนของเส้นเสียงในอัตราความถี่ที่ต่างกันไป ดังนั้นเสียงวรรณยุกต์จะปรากฏอยู่ในส่วนของเสียงสระ เพราะเสียงสระเป็นเสียงที่เกิดจากการสั่นของเส้นสียง แต่บางครั้งอาจมีเสียงวรรณยุกต์ปรากฏอยู่บ้างในส่วนของเสียงพยัญชนะที่เป็นพยัญชนะเสียงก้อง เพราะเสียงพยัญชนะไม่ก้องนั้นไม่ได้เกิดจากการสั่นของเส้นเสียง จึงไม่สามารถเกิดมีเสียงวรรณยุกต์ร่วมอยู่ได้

เสียงวรรณยุกต์ภาษาไทยสามารถแบ่งออกเป็น 5 ชนิดคือ เสียงวรรณยุกต์สามัญ (Mid Toneเสียงวรรณยุกต์เอก (Low Tone) เสียงวรรณยุกต์โท เสียงวรรณยุกต์ตรี (High Tone) และเสียงวรรณยุกต์จัตวา (Rising Tone) ซึ่งการเปลี่ยนแปลงความถี่ของเสียงในวรรณยุกต์ภาษาไทยสามารถแสดงได้ดังภาพที่ 2.1



ภาพที่ 2. 1 การเปลี่ยนแปลงความถี่ของเสียงในวรรณยุกต์ภาษาไทย (รูปจาก[6])

ตารางที่ 2.2 หน่วยเสียงพยัญชนะภาษาไทย (ปรับปรุงจาก[5])

1 (*) ปรากฏท้ายพยางค์ได้

2 (**) ปรากฏท้ายพยางค์เฉพาะในคำไทยทับศัพท์ภาษาอังกฤษ

3 (/…/) ปรากฏเฉพาะในคำไทยทับศัพท์ภาษาอังกฤษ

4 [/…/] ปรากฏในคำไทยทับศัพท์ภาษาอังกฤษ หรือคำไทยที่ยืมมาจากภาษาสันสกฤต

หน่วยเสียง[1]	หน่วยเสียงควบกล้ำ	ลักษณะของลม	การพ่นลม	ความก้อง	ฐานที่เกิด	รูปพยัญชนะ
/p/ (*)	/pr/, /pl/	กัก	ไม่พ่นลม	ไม่ก้อง	ริมฝีปาก	ป
/p^h/	/p^hr/,/p^hl/	กัก	พ่นลม	ไม่ก้อง	ริมฝีปาก	ผ พ ภ
/b/	(/br/), (/bl/)	กัก	ไม่พ่นลม	ก้อง	ริมฝีปาก	บ
/t/ (*)	/tr/	กัก	ไม่พ่นลม	ไม่ก้อง	ฟัน หรือ ปุ่มเหงือก	ฏ ต
/t^h/	[/t^hr/]	กัก	พ่นลม	ไม่ก้อง	ฟัน หรือ ปุ่มเหงือก	ฐ ฑ ฒ ถ ท ธ
/d/	(/dr/)	กัก	ไม่พ่นลม	ก้อง	ฟัน หรือ ปุ่มเหงือก	ฎ ด
/c/		กัก	ไม่พ่นลม	ไม่ก้อง	เพดานแข็ง	จ
/c^h/(**)		กัก	พ่นลม	ไม่ก้อง	เพดานแข็ง	ฉ ช ฌ
/k/ (*)	/kr/, /kl/, /kw/	กัก	ไม่พ่นลม	ไม่ก้อง	เพดานอ่อน	ก
/k^h/	/k^hr/, /k^hl/, /k^hw/	กัก	พ่นลม	ไม่ก้อง	เพดานอ่อน	ข ฃ ค ฅ ฆ
/ʔ/ (*)		กัก	ไม่พ่นลม	ไม่ก้อง	เส้นเสียง	อ
/m/ (*)		นาสิก		ก้อง	ริมฝีปาก	ม
/n/ (*)		นาสิก		ก้อง	ฟัน หรือ ปุ่มเหงือก	ณ น
/ŋ/ (*)		นาสิก		ก้อง	เพดานอ่อน	ง
/f/(**)	(/fr/), (/fl/)	เสียดแทรก		ไม่ก้อง	ริมฝีปาก	ฝ ฟ
/s/(**)		เสียดแทรก		ไม่ก้อง	ฟัน หรือ ปุ่มเหงือก	ซ ศ ษ ส
/h/		เสียดแทรก		ไม่ก้อง	เส้นเสียง	ห ฮ
/r/		รัว		ก้อง	ฟัน หรือ ปุ่มเหงือก	ร
/l/(**)		ข้างลิ้น		ก้อง	ฟัน หรือ ปุ่มเหงือก	ล ฬ
/w/ (*)		กึ่งสระ		ก้อง	ริมฝีปาก–เพดานอ่อน	ว
/j/ (*)		กึ่งสระ		ก้อง	เพดานแข็ง	ญ ย

[1]ใช้หน่วยเสียงตามสัทอักษรสากล (International Phonetic Alphabet – IPA)

ตารางที่ 2.3 หน่วยเสียงสระเดี่ยวภาษาไทย[5]

หน่วยเสียง[1]	ส่วนของลิ้นที่ใช้เปล่งเสียง	ความสูงของลิ้น	การห่อริมฝีปาก	ความยาวเสียง	รูปสระ
/i/	หน้า	ปิด	ไม่ห่อ	สั้น	อิ
/i̯/	หน้า	ปิด	ไม่ห่อ	ยาว	อี
/e/	หน้า	กึ่งปิด	ไม่ห่อ	สั้น	เอะ
/e̯/	หน้า	กึ่งปิด	ไม่ห่อ	ยาว	เอ
/æ/	หน้า	กึ่งเปิด	ไม่ห่อ	สั้น	แอะ
/æ̯/	หน้า	กึ่งเปิด	ไม่ห่อ	ยาว	แอ
/ɨ/	หลัง ค่อนมาทางกลาง	ปิด	ไม่ห่อ	สั้น	อึ
/ɨ̯/	หลัง ค่อนมาทางกลาง	ปิด	ไม่ห่อ	ยาว	อือ
/ɜ/	หลัง ค่อนมาทางกลาง	กึ่งปิด	ไม่ห่อ	สั้น	เออะ
/ɜ̯/	หลัง ค่อนมาทางกลาง	กึ่งปิด	ไม่ห่อ	ยาว	เออ
/a/	กลาง	เปิด	ไม่ห่อ	สั้น	อะ
/a̯/	กลาง	เปิด	ไม่ห่อ	ยาว	อา
/u/	หลัง	ปิด	ห่อ	สั้น	อุ
/u̯/	หลัง	ปิด	ห่อ	ยาว	อู
/o/	หลัง	กึ่งปิด	ห่อ	สั้น	โอะ
/o̯/	หลัง	กึ่งปิด	ห่อ	ยาว	โอ
/ɔ/	หลัง	กึ่งเปิด	ห่อ	สั้น	เอาะ
/ɔ̯/	หลัง	กึ่งเปิด	ห่อ	ยาว	ออ

[1]ใช้หน่วยเสียงตามสัทอักษรสากล (International Phonetic Alphabet – IPA)

ตารางที่ 2. 4 หน่วยเสียงสระประสมภาษาไทย [5]

หน่วยเสียง[1]	ส่วนประกอบ	ความยาวเสียง	รูปสระ
/ia/	/i/ + /a/	สั้น	เอียะ
/i̯a/	/i̯/ + /a/	ยาว	เอีย
/ɨa/	/ɨ/ + /a/	สั้น	เอือะ
/ɨ̯a/	/ɨ̯/ + /a/	ยาว	เอือ
/ua/	/u/ + /a/	สั้น	อัวะ
/u̯a/	/u̯/ + /a/	ยาว	อัว

[1]ใช้หน่วยเสียงตามสัทอักษรสากล (International Phonetic Alphabet – IPA)

ตารางที่ 2.5 หน่วยเสียงสระเกินภาษาไทย [5]

หน่วยเสียง[1]	ส่วนประกอบ	ความยาวเสียง	รูปสระ
/am/	/a/ + /m/	สั้น	อำ
/aj/	/a/ + /j/	สั้น	ไอ, ใอ
/aw/	/a/ + /w/	สั้น	เอา
/ri/, /r̥/	/r/ + /i/, /r/ + /̥/	สั้น	ฤ
/rí/, /ŕ̥/	/r/ + /í/, /r/ + /̥́/	ยาว	ฤา
/li/, /l̥/	/l/ + /i/, /l/ + /̥/	สั้น	ฦ
/lí/, /ĺ̥/	/l/ + /í/, /l/ + /̥́/	ยาว	ฦๅ

[1]ใช้หน่วยเสียงตามสัทอักษรสากล (International Phonetic Alphabet – IPA)

โครงสร้างพยางค์ในภาษาไทย

พยางค์ (syllable) หมายถึงหน่วยหนึ่งขององค์ประกอบเสียงที่ใช้ในการสื่อสารด้วยคำพูด โดยพยางค์ จะประกอบด้วย แกนพยางค์ (syllable nucleus) ซึ่งเป็นเสียงสระ และอาจจะมีเสียงพยัญชนะเกิดขึ้นด้วยทั้งในส่วนต้นพยางค์หรือท้ายพยางค์

ในภาษาไทย หน่วยเสียงพยัญชนะทุกหน่วยเสียงสามารถปรากฏในตำแหน่งต้นพยางค์ในฐานะพยัญชนะเดี่ยวได้ทั้งสิ้น นอกจากนี้ยังสามารถปรากฏในรูปพยัญชนะควบกล้ำได้เพียง 2 เสียง (CC) ซึ่งพยัญชนะควบกล้ำปรากฏได้เฉพาะในตำแหน่งต้นเท่านั้น การปรากฏของพยัญชนะควบกล้ำในภาษาไทย ไม่สามารถนำหน่วยเสียงพยัญชนะใดๆ ก็ได้ 2 หน่วยมาควบกล้ำกัน แต่จะต้องเป็นการควบกล้ำคู่ใดคู่หนึ่งใน 12 คู่ ดังต่อไปนี้ /pr/, /tr/, /kr/, /phr/, /thr/, /khr/, /pl/, /phl/, /kl/, /khl/, /kw/, /khw/ส่วนพยัญชนะที่สามารถปรากฏในตำแหน่งท้ายพยางค์ได้ มีเพียง 9 หน่วยเสียง คือ /p, t, k, ʔ, m, n, ŋ , w, j/

ดังนั้น โครงสร้างพยางค์ของภาษาไทย จึงพอสรุปได้เป็นดังนี้ (C) C V (V) (C) ซึ่งแบบโครงสร้างที่ปรากฏนี้ถือเป็นแบบรวม และมีโครงสร้างที่เป็นไปได้ทั้งหมด ดังตารางที่ 2.5

ตารางที่ 2.6 โครงสร้างพยางค์ในภาษาไทย

แบบที่	ส่วนประกอบ	ตัวอย่าง
1	CV	มา, หา
2	CVV	เพื่อ, เมื่อ
3	CVC	จาน, พาน
4	CVVC	เพื่อน, เรื่อง
5	CCV	ปลา
6	CCVV	เกลือ
7	CCVC	ควาย, กวาง
8	CCVVC	เปลือง, เครื่อง

สมบัติทางสวนสัทศาสตร์ (Phonetic Features)

Chomsky และ Halle [7] ได้ศึกษาและให้นิยามของสมบัติทางสวนสัทศาสตร์ (Phonetic Feature) ที่ใช้ในการวิเคราะห์และอธิบายคุณสมบัติของหน่วยเสียงออกเป็น 3 ประเภทคุณสมบัติของแหล่งกำหนดเสียง (Source feature), คุณสมบัติของวิธีการออกเสียง (Manner feature) และคุณสมบัติของอวัยวะที่เป็นฐานในการออกเสียง (Place feature) โดยกำหนดให้แต่ละ สมบัติทางสวนสัทศาสตร์ มีค่าเป็น + หรือ – เท่านั้น โดยที่ + แทนการมีคุณสมบัติ และ – แทนการไม่มีคุณสมบัติ

1. คุณสมบัติของแหล่งกำหนดเสียง (Source Feature) เป็นคุณสมบัติของแหล่งกำเนิดเสียง ในการออกเสียงพูดถ้าอากาศถูกดันออกมาจากปอดด้วยแรงดันที่มากพอทำให้เกิดการสั่นของเส้นเสียง ซึ่งเป็นผลให้สัญญาณเสียงที่เกิดขึ้นมีแหล่งกำเนิดเสียงที่มีลักษณะเป็นคาบ (Periodic signal) และแสดงได้ด้วยค่าคุณสมบัติความเป็นเสียงที่มีการสั่นของเส้นเสียง (voiced) แทนด้วยสัญลักษณ์ [+voiced] และ [-voiced] เมื่อสัญญาณเสียงที่เกิดจากแหล่งกำเนิดเสียงไม่เกิดการสั่นของเส้นเสียง

2. คุณสมบัติของวิธีการออกเสียง (Manner Feature) เป็นคุณสมบัติของวิธีการออกเสียง ซึ่งในการออกเสียงจะทำการพิจารณาจากลักษณะของช่องเสียง (Vocal tract)

ว่ามีการเปิด/ปิดอย่างไร รวมถึงมีการกักเสียงไว้มากหรือน้อยอย่างไร และการออกเสียงนั้นอากาศไหลผ่านบริเวณช่องปากหรือผ่านไปในช่องโพรงจมูกหรือไม่ ได้แก่

2.1 [+/- Sonorant] เป็นคุณสมบัติของเสียงที่ผ่านออกมาจากช่องเสียงได้โดยไม่มีการกักลมหรือโดนกักเพียงเล็กน้อย และไม่เกิดการเสียดสีของลมที่ออกมากับอวัยวะจุดใดจุดหนึ่งในปาก เพราะลักษณะการบังคับลมเป็นแบบเปิดกว้าง ซึ่งประกอบไปด้วยเสียง สระ (vowel) เสียงกึ่งสระ (semi-vowels) และเสียงนาสิก (nasals) คุณสมบัติความเป็น sonorant ด้วยค่า [+sonorant] ส่วนเสียงที่ไม่มีคุณสมบัติความเป็น sonorant ได้แก่ เสียงพยัญชนะกัก (stop consonants) และเสียงเสียดแทรก (Fricatives) แสดงได้ด้วยค่า [-sonorant]

2.2 [+/- Syllabic] เป็นคุณสมบัติของเสียงที่เปล่งออกมาได้ดังกว่า จึงทำหน้าที่เป็นใจกลางหรือแกนของพยางค์ ซึ่งเป็นลักษณะของเสียงสระ (vowel) คุณสมบัติความเป็น syllabic ด้วยค่า [+syllabic] ส่วนเสียงที่ไม่มีคุณสมบัติความเป็น syllabic ประกอบไปด้วยเสียง เสียงกึ่งสระ (semi-vowels) และเสียงนาสิก (nasals) แสดงได้ด้วยค่า [-syllabic]

2.3 [+/- Continuant] เป็นคุณสมบัติของเสียงที่เปล่งที่เปล่งออกมาได้อย่างต่อเนื่องโดยไม่เกิดการปิดกันของลมในช่องเสียง จนทำให้เสียงที่เกิดขึ้นไม่สามารถออกเสียงได้อย่างต่อเนื่อง ซึ่งเป็นคุณลักษณะที่ใช้สำคัญที่ใช้แยกเสียงเสียดแทรก (Fricatives) และ เสียงพยัญชนะกัก (stop consonants) โดยที่มีคุณสมบัติความเป็น continuant แสดงได้ด้วยค่า [+continuant] ถ้าเสียงนั้นที่ไม่มีคุณสมบัติความเป็น continuant แสดงได้ด้วย [-continuant]

3. คุณสมบัติของอวัยวะที่เป็นฐานในการออกเสียง (Place Feature) เป็นคุณสมบัติของอวัยวะที่เป็นฐานในการออกเสียง ว่าในการออกเสียงแต่ล่ะหน่วยเสียงมีอวัยวะใดเป็นฐานในการออกเสียง โดยที่ตำแหน่งที่มีการจำกัดสัญญาณเสียงในกรณีของเสียงพยัญชนะจะพิจารณาจากตำแหน่งที่มีใช้อวัยวะเช่นฟัน ลิ้น หรือริมฝีปากในการกั้นไม่ให้อากาศไหลผ่านช่องปากออกไปได้โดยตรง ในขณะที่กรณีของเสียงสระจะพิจารณาจากตำแหน่งของลิ้นในขณะที่อากาศเดินทางผ่านช่องปากออกไป

ตารางที่ 2.7 ตารางแสดงความสัมพันธ์ระหว่างประเภทของหน่วยเสียงเปรียบเทียบกับคุณสมบัติขิงวิธีการออกเสียง

Phonetic feature	[+continuant]	[-continuant]
[-sonorant]	Fricatives	Stop consonant
[+sonorant, -syllabic]	Nasal and Semi-vowels	-
[+sonorant, +syllabic]	Vowels	-

ทฤษฎีที่เกี่ยวกับการวิเคราะห์สัญญาณเสียง

1. การแปลงฟูริเยร์แบบช่วงเวลาสั้น (Short-Time Fourier Transform)

เนื่องจากในปัจจุบันได้มีการทำการวิเคราะห์สเปกตรัมของสัญญาณเสียง โดยใช้แนวคิดของแปลงฟูริเยร์ไม่อิสระทางเวลา (Time-Dependent Fourier Transform) หรือเรียกว่า การการแปลงฟูริเยร์แบบช่วงเวลาสั้น (Short-Time Fourier Transform) ซึ่งสามารถแสดงได้ในรูปของสมการคณิตศาสตร์ ดังสมการ ที่ (2.1)

$$X_{STFT}(e^{j\omega},\ n) = \sum_{n=-\infty}^{\infty} x[n-m]w[m]e^{-j\omega m} \qquad (2.1)$$

โดยที่ $x[n]$ คือ สัญญาณเสียง, $w[m]$ คือฟังก์ชันหน้าต่าง m, n คือค่าตัวแปรทางเวลาที่มีค่าไม่ต่อเนื่อง ค่าความถี่ ω คือค่าความถี่มีค่าต่อเนื่อง

จากสมการ (2.1) เมื่อทำการสุ่มตัวอย่าง $X_{STFT}(e^{j\omega},\ n)$ ที่ความถี่ที่มีระยะห่างเท่าๆ กัน N จำนวนจะได้

$$X_{STFT}(k,n) = \sum_{m=0}^{R-1} x[n-m]w[m]e^{-j\frac{2\pi}{N}km} \qquad (2.2)$$

โดยที่ $0 \leq k \leq N-1$ เมื่อ คือจำนวนจุดที่ใช้ในการคำนวณการแปลงฟูริเยร์สำหรับสัญญาณไม่เป็นคาบ (Discrete-Time Fourier Transform) และ $w[m]$ คือ ฟังก์ชันหน้าต่าง ซึ่ง $m \geq R$

การแสดงสเปกตรัมขนาดของ $X_{STFT}(k,n)$ จะเรียกว่าสเปกโตรแกรม(Spectrogram) โดยที่สามารถแบ่งได้เป็น 2 ประเภทคือ สเปกโตรแกรมแถบกว้าง (wideband) เกิดจากการเลือกฟังก์ชันหน้าต่าง ที่มีความยาวน้อย ซึ่งจะมีความสามารถแยกเวลาได้ดี ในทางตรงข้าม สเปกโตรแกรมแถบแคบ (narrowband) เกิดจากการเลือกฟังก์ชันหน้าต่างที่มีความยาวมาก จะมีความสามารถแยกความถี่ได้ดี

2. การหาค่าพารามิเตอร์ทางเสียง (Acoustic Parameter)

เป็นการศึกษาเกี่ยวกับการหาค่าพารามิเตอร์ทางเสียง โดยการวิเคราะห์ค่าลักษณะสำคัญของเสียงที่ได้เพื่อนำไปเป็นตัวแทนของสัญญาณเสียงเช่นการวิเคราะห์หาค่าลักษณะสำคัญของเสียงเสียดแทรกเพื่อนำไปใช้ในการระบุหาเสียงเสียดแทรกจากสัญญาณเสียงพูด โดยทั่วไปการวิเคราะห์หาค่าพารามิเตอร์ทางเสียง โดยสามารถแบ่งแนวทางในการพิจารณาได้เป็นการหาค่าบนโดเมนทางเวลา ซึ่งสามารถ วัดค่าพลังงาน วัดค่าอัตราการตัดศูนย์ เป็นต้น และการหาค่าบนโดเมนทางความถี่ซึ่งสามารถ วัดค่าต่างที่แสดงอยู่ในรูปแบบทางความถี่เช่น การค่าความถี่ฟอร์แมนท์ เป็นต้น โดยในงานวิจัยนี้ได้มีการใช้ค่าพารามิเตอร์ทางเสียงต่างดังนี้

1. การหาค่าพลังงาน (Energy)

ค่าค่าพารามิเตอร์ทางเสียงที่ใช้ในการวิเคราะห์เสียงพูดที่นิยมใช้กันอย่างแพร่หลายอย่างหนึ่ง คือ ค่าพลังงานของสัญญาณเสียงพูด เนื่องจากค่าพลังงานของสัญญาณเสียงพูดแสดงให้เห็นว่ามีสัญญาณเสียง เกิดขึ้น ณ เวลานั้นหรือไม่และเสียงชนิดต่างๆมีค่าพลังงานที่แตกต่างกัน อีกทั้งง่ายต่อการคำนวณ โดยใน การคำนวณค่าพลังงานจะทำการคำนวณทีละกรอบเสียงพูด เมื่อกำหนดให้ $E(m)$ คือค่าพลังงานของกรอบเสียงพูดที่จุด m และในแต่ละกรอบเสียงพูด จะทำประกอบด้วยสัญญาณเสียงพูดจำนวน n ตัวอย่าง ซึ่งสามารถทำการคำนวณหาค่าพลังงานของสัญญาณ $s(n)$ ใด ๆ ที่แปรตามเวลา ดังสมการ (2.3)

$$E = \sum_{n=-\infty}^{\infty} s[n]^2 \quad (2.3)$$

ในการคำนวณเราจะทำการพิจารณาสัญญาณเป็นช่วงกรอบเล็ก ๆ ขนาดประมาณ 10-30 มิลลิวินาที โดยที่ในงานวิจัยนี้จะแบ่งโดยใช้ขนาดกรอบประมาณ 10 มิลลิวินาที โดยการคำนวณพลังงานของเสียงในแต่ละกรอบ สามารถทำการคำนวณได้ตามสมการ (2.4)

$$E(m) = \sum_{n=0}^{N-1} [w(m)s(m-n)]^2 \quad (2.4)$$

โดยที่ $w(m)$ คือฟังก์ชันหน้าต่าง ที่ใช้กำหนดรูปร่างในการพิจารณาของสัญญาณเสียง $s(n)$ ในหนึ่งกรอบ

N คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ในกรอบของฟังก์ชันหน้าต่าง

2. การคำนวณค่าพลังงานในช่วงองค์ประกอบความถี่

ค่าพลังงานตามองค์ประกอบความถี่ สามารถคำนวณได้จากผลรวมค่าพลังงานในช่วงองค์ประกอบความถี่ที่ปรากฏบนสเปกโตรแกรม ที่อยู่ภายในกรอบพิจารณาช่วงเล็กๆ ดังสมการ (2.5)

$$E[f_a, f_b](n) = \sum_{f=f_a}^{f_b} E_f(n) \quad (2.5)$$

โดยที่ $[f_a, f_b](n)$ คือผลรวมค่าพลังงานของสัญญาณเสียงที่กรอบเสียงพูด n ในช่วงองค์ประกอบความถี่ f_a กับ f_b

$E_f(n)$ คือค่าพลังงานของสัญญาณเสียงที่กรอบเสียงพูดที่ n ที่องค์ประกอบความถี่ f

3. การคำนวณค่าอัตราส่วนพลังงาน

ค่าอัตราส่วนพลังงานที่ใช้ในงานวิจัยนี้ คือ อัตราส่วนของค่าพลังงานในช่วงองค์ประกอบความถี่ต่ำต่อค่าพลังงานในช่วงองค์ประกอบความถี่สูง สามารถแสดงได้ตามดังสมการ (2.6)

$$Ratio\ of\ E[f_a, f_b]\ to\ E[f_c, f_d]\ = \frac{E[f_a, f_b]}{E[f_c, f_d]} \tag{2.6}$$

โดยที่ $E[f_a, f_b](n)$ คือผลรวมค่าพลังงานของสัญญาณเสียงที่กรอบเสียงพูด n ในช่วงองค์ประกอบความถี่ f_a กับ f_b

$E[f_c, f_d](n)$ คือผลรวมค่าพลังงานของสัญญาณเสียงที่กรอบเสียงพูด n ในช่วงองค์ประกอบความถี่ f_c กับ f_d

4. การคำนวณค่าอัตราส่วนความเข้มสูงสุดในช่วงองค์ประกอบความถี่

ค่าอัตราส่วนความเข้มสูงสุดในช่วงองค์ประกอบความถี่ที่ใช้ในงานวิจัยนี้ คือ อัตราส่วนของค่าความเข้มสูงสุดในช่วงองค์ประกอบความถี่ต่ำที่ปรากฏบนสเปกโตรแกรม ต่อค่าความเข้มสูงสุดในช่วงองค์ประกอบความถี่สูงที่ปรากฏบนสเปกโตรแกรม สามารถแสดงได้ตามดังสมการ (2.7)

$$Ratio\ of\ max(A[f_a, f_b])\ to\ max(A[f_c, f_d]) = \frac{max(A[f_a, f_b])}{max(A[f_c, f_d])} \tag{2.7}$$

โดยที่ $max(A[f_a, f_b])$ คือค่าความเข้มสูงสุดของสัญญาณเสียงที่กรอบเสียงพูดบนสเปกโตรแกรม ในช่วงองค์ประกอบความถี่ f_a กับ f_b

$max(A[f_c, f_d])$ คือผลรวมค่าพลังงานของสัญญาณเสียงที่กรอบเสียงพูด บนสเปกโตรแกรม ในช่วงองค์ประกอบความถี่ f_c กับ f_d

5. ค่าความถี่ที่มีค่าความเข้มสูงสุดในช่วงองค์ประกอบความถี่

คือความถี่ที่มีค่าของความเข้มสูงสุดที่ปรากฏบนสเปกโตรแกรม ในช่วงองค์ประกอบความถี่ที่ทำการพิจารณา

6. ค่าพลังงานออนเซตและค่าพลังงานออฟเซต(Energy Onset and Energy Offset)

ในการคำนวณค่าพลังงานออนเซตและค่าพลังงานออฟเซต เป็นการคำนวณที่ใช้หาการจุดที่มีการเปลี่ยนแปลงของพลังงานที่เกิดการเพิ่มขึ้นหรือลดลงอย่างรวดเร็ว หาได้จากการหาผลต่างของพลังงานของกรอบสัญญาณเสียงที่อยู่ติดกัน และไม่มีการซ้อนทับของกรอบสัญญาณเสียงในแต่ละช่ององค์ประกอบความถี่ของสัญญาณเสียง แสดงได้ดังสมการ (2.8) ผลความต่างที่ได้มีหน่วยเป็นเดซิเบล (dB)

$$D_{i,k} = \begin{aligned} & 20log \sum_{m=-\infty}^{\infty} x_i(n+m)w(m) \\ & -20log \sum_{m=-\infty}^{\infty} x_i(n+m-k)w(m-k) \end{aligned} \quad (2.8)$$

โดยที่ x_i คือสัญญาณขาเข้าที่ช่องความถี่ i

k คือเวลาที่ต่างกันระหว่างกรอบสัญญาณ 2 กรอบสัญญาณ

$w(n)$ คือฟังชันก์หน้าต่างที่มีความยาวเท่ากับ k

โดยเมื่อทำการคำนวณผลต่างของพลังงานแต่ละกรอบที่อยู่ติดกันในแต่ละช่วงความถี่ได้แล้วเราสามารถทำการหาค่าพลังงานออนเซตและค่าพลังงานออฟเซต ได้จากสมการ (2.9) และ (2.10)

$$energy\ onset(n) = \frac{1}{N} \sum_{i:D_{i,k}>0} D_{i,k}(n) \quad (2.9)$$

$$energy\ onset(n) = \frac{1}{N} \sum_{i:D_{i,k}<0} D_{i,k}(n) \quad (2.10)$$

โดยที่ N คือจำนวนช่องความถี่ที่ใช้ในการคำนวณ

7. จุดศูนย์ถ่วงของสเปกตรัม(Spectral center of gravity)

เป็นค่าศูนย์ถ่วงของพลังงานสเปกตรัม ในแต่ละกรอบสัญญาณเสียง หาได้จากการดังสมการที่ (2.12)

$$SCG = \frac{\sum_{i=1}^{N} f(i)E_f(i)}{\sum_{i=1}^{N} E_f(i)} \quad (2.12)$$

โดยที่ $f(i)$ คือค่าความถีขององค์ประกอบความถี่ที่ i ของสเปกโตรแกรม

$E_f(i)$ คือลังงานของที่ความถี่ f ที่องค์ประกอบความถี่ที่ i ของสเปกโตรแกรม

8. อัตราการตัดศูนย์ในช่วงเวลาสั้น (Short time zero crossing rate)

คืออัตราการเปลี่ยนแปลงของสัญญาณจากค่าบวกไปเป็นค่าลบ หรือจากค่าลบไปเป็นค่าบวกในช่วงกรอบเวลาที่สนใจ ซึ่งสามารถใช้แยกเสียงก้อง และเสียงไม่ก้องออกจากกัน รวมถึงยังใช้ในการแยกเสียงกัก และ เสียงเสียดแทรกออกจากกันได้ด้วย

โดยเมื่อทำการนิยาม ฟังชันก์เครื่องหมาย(sign function)ตามสมการที่ (2.11) และฟังก์ชันหน้าต่าง $w(m)$ เราสามารถแสดงสมการที่ใช้ในการหาค่าอัตราการตัดศูนย์ Z_n ได้ดัง สมการที่ (2.12)

$$sgn[x(n)] = \begin{array}{l} 1 \;\; x(n) \geq 1 \\ -1 \; x(n) < 0 \end{array} \tag{2.11}$$

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[(m-1)]| w(n-m) \tag{2.12}$$

งานวิจัยที่เกี่ยวข้อง

ในงานวิจัยเรื่องการหาตำแหน่งของพยัญชนะ ได้มีนักวิจัยหลายกลุ่มทำการค้นคว้าและทดลองเพื่อหาวิธีการที่ใช้สำหรับการหาตำแหน่งของพยัญชนะ และนำไปประยุกต์ใช้ในระบบรู้จำเสียงพูด โดยงานวิจัยที่เกี่ยวข้องมีดังต่อไปนี้

Liu[8] ได้ทำการเสนอวิธีการหาแลนมาร์กของเสียงพยัญชนะโดยทำการกำหนดแลนมาร์กของเสียงพยัญชนะออกเป็น 3 ชนิด ได้แก่ g (glottis) เป็นเหตุการณ์เกิดการสั่นของเส้นเสียง, s (sonorant) เป็นเหตุการณ์ที่เกิดการเปิดหรือปิดของเพดานอ่อน, b (burst) เป็นเหตุการณ์ที่เกิดเสียงเสียดแทรกหรือเสียงระเบิด โดยใช้การหาจุดเปลี่ยนแปลงของพลังงาน จากการคำนวณค่าพลังงานจากสเปกโตรแกรม โดยแบ่งค่าพลังงานเป็นช่วงพลังงาน ต่าง ๆ ออกเป็น 6 ช่วง

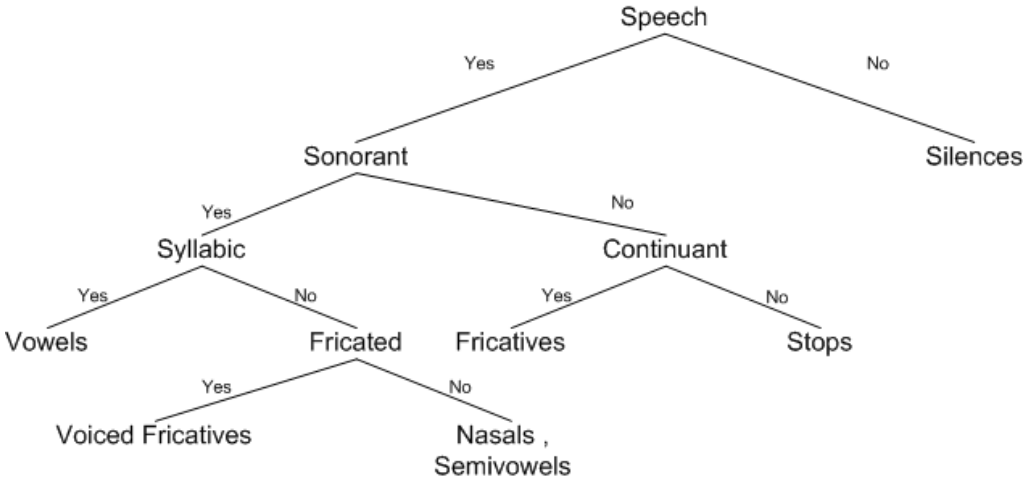
ประกอบด้วย ช่วงที่หนึ่ง [0.0-0.4] กิโลเฮิรตซ์ ช่วงที่สอง [0.8-1.5] กิโลเฮิรตซ์ ช่วงที่สาม [1.2-2.0] กิโลเฮิรตซ์ ช่วงที่สี่ [2.0-3.5] กิโลเฮิรตซ์ ช่วงที่ห้า[3.5-5.0] กิโลเฮิรตซ์ ช่วงที่หก [5.0-8.0] กิโลเฮิรตซ์ แล้วทำการหาจุดที่มีอัตราการเปลี่ยนแปลงของพลังงานมากที่สุดจากแต่ละช่วงพลังงาน เพื่อนำมาใช้ในการหาแลนมาร์กของเสียงพยัญชนะ โดยใช้ลักษณะเด่นของเสียงที่มีการเปลี่ยนแปลงพลังงานในช่วงความถี่ต่างๆ ที่แตกต่างกัน เช่น ตำแหน่งการที่เกิดการกักและปล่อยเสียงกัก ตำแหน่งการเริ่มหรือสิ้นสุดการสั่นของเส้นเสียง เป็นต้น โดยมีความถูกต้องในการหาตำแหน่ง 79% และความแม่นยำ 54% และ Chen[9] ได้ทำการพัฒนาเพิ่มเติมในส่วนค้นหาเสียงนาสิก เพื่อนำไปใช้กับระบบ LAFF (Lexical Access From Features)[10] โดยมีความถูกต้องของการจำแนกเสียงสระและเสียงนาสิก 83.6 %

ต่อมา Park[11] ได้ทำการเสนอแนวทางการหาแลนมาร์กของเสียงพยัญชนะโดยได้ทำการปรับปรุงงานของ Liu[7] โดยได้ทำการใช้ bigram model ของการเกิดขึ้นของคู่ลำดับแลนมาร์กของเสียงพยัญชนะ (g , b, s) มาใช้ ทำการสร้างกราฟของแลนมาร์กขึ้นมา แล้วทำการค้นหาคำตอบที่ดีที่สุดโดยใช้อัลกอริทึมวิเทอร์บี ในการหาคำตอบที่ดีที่สุดจากกราฟของแลนมาร์ก ซึ่งผลความถูกต้องที่ได้ใกล้เคียงกับผลการทดลองของ Liu[7] โดยมีความถูกต้องในการหาตำแหน่ง 76.8% และความแม่นยำ 62.1% โดยผลของความผิดพลาดที่เกิดขึ้นแบบเกิน (Insert Error) มีค่า 14.7% ลดลง 41.2% เมื่อเทียบกับอัลกอริทึมของ Liu[7] ซึ่งมีผลความผิดพลาดที่เกิดขึ้นแบบเกินมีค่า 25%

Lee [12] ได้ปรับปรุงอัลกอริทึมของ Liu[7] โดยทำการคำนวณค่าการเปลี่ยนแปลงของจุดศูนย์ถ่วงความถี่ของสเปกตรัม (Difference of Spectral Center of Gravity) และค่าการเปลี่ยนแปลงรากกำลังสองเฉลี่ยของพลังงาน(Difference of Rooted Mean Squared Energy) มาใช้เพื่อปรับปรุงกฎการตัดสินใจที่นำเสนอ โดย Liu[7] ซึ่งผลที่ได้จากการทดลองนี้ให้ผลความถูกต้องเพิ่มขึ้น 3% และลดความผิดพลาดที่เกิดขึ้นแบบเกิน(Insertion Error)ลง 42% เมื่อเทียบกับวิธีที่เสนอโดย Liu

Bitar และ Espy-Wilson[13] ได้ทำการเสนอค่าพารามิเตอร์ทางเสียง (acoustic parameter) เพื่อที่จะนำมาใช้แยกสมบัติทางสวนสัทศาสตร์ (Phonetic Features) ของเสียง กลุ่มและแบ่งการจำแนกเป็นระดับชั้นและคำนวณความน่าจะเป็นตามระดับชั้น ตามภาพที่ 2.2 ทำให้สามารถทำการแบ่งประเภทของหน่วยเสียงออกเป็นประเภทของหน่วยเสียง 5 ประเภทได้แก่ เสียงสระ, เสียงกึ่งสระกับเสียงนาสิก, เสียงกัก, เสียงเสียดแทรก และเสียงเงียบ แล้วจึงนำ acoustic

parameter ที่ได้มาใช้เป็นเวกเตอร์ข้อมูลเข้า (Input Vector) แทนการใช้ค่าสัมประสิทธิ์เมลฟรี-เคว็นซีเคปสตรัม (Mel-Frequency Cepstrum Coefficient) สำหรับใช้ในการสร้างระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟโมเดล ซึ่งผลที่ได้จากการทดลองบนฐานข้อมูลเสียง TIMIT พบว่า การใช้ Acoustic Parameter ให้ผลการรู้จำประเภทของเสียงที่จำแนกตามวิธีการออกเสียงได้ผลใกล้เคียงกันคือได้ความถูกต้อง 70.7 เมื่อใช้ acoustic parameter เป็นข้อมูลเข้า และ 66.75 เมื่อใช้ MFCC เป็นข้อมูลเข้าโดย นอกจากนี้ยังได้ทำการทดลองความขึ้นต่อผู้พูดโดยทำการฝึกโมเดลโดยใช้เสียงผู้ชายแล้วทำการทดสอบโมเดลโดยใช้เสียงผู้หญิง พบว่าระบบที่ได้ไม่มีความเปลี่ยนแปลงอย่างมีนัยยะสำคัญ หรือคือมีความคงทนต่อการเปลี่ยนเพศของผู้พูด

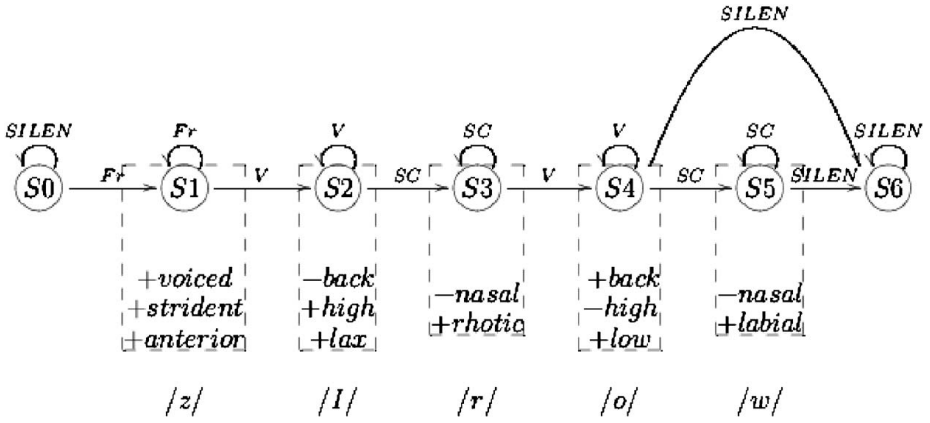


ภาพที่ 2.2 ต้นไม้ตัดสินใจของประเภทหน่วยเสียงกับคุณสมบัติวิธีการออกเสียง

Juneja และ Espy-Wilson [14, 15] ได้ทำการตัดแบ่งส่วนของเสียงออกเป็น 5 ประเภท ได้แก่ เสียงเงียบ, เสียงกัก, เสียงเสียดแทรก, เสียงกึ่งสระกับเสียงนาสิก และเสียงสระ ซึ่งทำการคำนวณหาค่าพารามิเตอร์ทางเสียง (acoustic parameter) ต่างๆ จากสัญญาณเสียง ทุกๆ 5 มิลลิวินาที แล้วทำการจำแนกเสียงในแต่ละกรอบเวลาที่ได้เป็นออกเป็นประเภทของหน่วยเสียง โดยใช้ซัพพอร์ตเวกเตอร์แมชชีนจำนวน 4 ตัวในการแบ่งประเภทของหน่วยเสียงออกเป็น 5 ประเภท แล้วจึงทำการรวมผลที่ได้จากแต่ละเฟรมที่เป็นหน่วยเสียงประเภทเดียวกันเข้าด้วยกัน โดยได้ผลความถูกต้องของการจำแนกประเภทของหน่วยเสียง 79.8% และความแม่นยำ 68.1 %

หลังจากนั้น Juneja และ Espy-Wilson ยังได้ทำการเสนอระบบรู้จำเสียงพูดแบบแลนมาร์ก[2] โดยได้ทำการเสนอโมเดลการออกเสียง (pronunciation model) โดยใช้สมบัติทางสวน

สัทศาสตร์ มาใช้ในการสร้างโมเดลการออกเสียงแทนการใช้หน่วยเสียง ตามภาพที่ 2.3 เมื่อนำโมเดลการออกเสียงที่ได้มาใช้ในการหาความน่าจะเป็นของลำดับ (sequence) การเกิดของแลนมาร์ก มาใช้เป็นเงื่อนข้อบังคับในการเกิดขึ้นของลำดับของแลนมาร์ก ทำให้การตัดแบ่งเสียงออกเป็น 5 ประเภท ได้แก่ เสียงเงียบ, เสียงกัก, เสียงเสียดแทรก, เสียงกึ่งสระกับเสียงนาสิก และเสียงสระ มีความแม่นยำของการตัดแบ่งประเภทของเสียงที่ดีขึ้นกว่าเดิม โดยผลที่ได้บนฐานข้อมูลเสียง TIDIGITS มีความแม่นยำ 85.2% โดยมีความแม่นยำสูงกว่าระบบที่ทำการตัดแบ่งโดยไม่ได้ใช้โมเดลการออกเสียงมาใช้อยู่ 10.9 % นอกจากนี้ผลที่ได้มีผลความแม่นยำดีกว่าผลการรู้จำที่ใช้แบบจำลองฮิดเดนมาร์คอฟโมเดลที่ใช้ MFCC เป็นตัวแทนสัญญาณเสียงอยู่ 0.9%



ภาพที่ 2.3 โมเดลการออกเสียงโดยใช้พื้นฐานสมบัติทางสวนสัทศาสตร์ ของคำว่า “zero” รูปจาก [3]

นอกจากนี้ยังมีงานวิจัยที่ทำการศึกษาเกี่ยวกับการจำแนกกลุ่มเสียงต่างๆ ทำให้สามารถศึกษาลักษณะของแต่ละหน่วยเสียงและหาพารามิเตอร์ทาง สำหรับกลุ่มเสียงต่าง ๆ ได้ตัวอย่างเช่นการจำแนกกลุ่มเสียงนาสิก ที่บอกลักษณะที่เหมาะสมในของกลุ่มเสียงนาสิกโดย Pruthi [16, 17] ต้องการจะจำแนกกลุ่มเสียงนาสิกและเสียงกึ่งสระออกจากโดยใช้ ค่าอัตราส่วนพลังงานบางองค์ประกอบความถี่ ค่าความหนาแน่นของความถี่ฟอร์แมนท์และวัดจุดสูงสุดของสเปกตรัมเพื่อค้นหาเสียงนาสิก ซึ่งใช้ตัวจำแนกซับพอร์ตเวกเตอร์แมชชีน เป็นเครื่องมือในการจำแนกลักษณะ นอกจากนี้ Abdelatty Ali[18] ยังได้ทำการนำค่าพารามิเตอร์ทางเสียงจำนวน 6 ชนิดมาใช้ร่วมกับระบบการประมวลเสียงทางโสตประสาท (auditory-based speech

processing) เพื่อทำการจำแนกเสียงกักออกเป็นหน่วยเสียงกัก /p/, /t/, /k/, /b/, /d/, /g/ และ /dx/ โดยได้ความแม่นยำรวม 86% ในการจำแนกเสียงกักออกเป็นหน่วยเสียงกัก

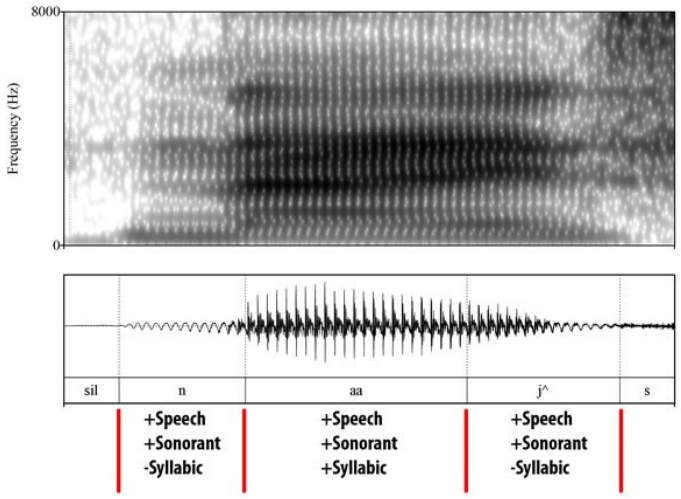
จากงานวิจัยต่างๆ ดังที่ได้กล่าวมาข้างต้นแสดงให้เห็นว่าการใช้สมบัติทางสวนสัทศาสตร์และการศึกษาหาค่าพารามิเตอร์ทางเสียงนั้นสามารถนำมาใช้ในการหาตำแหน่งของหน่วยเสียงตามลักษณะการออกเสียงและนำมาใช้แยกหน่วยเสียงต่างๆได้ ดังนั้นจึงเป็นเหตุผลหนึ่งที่ทำให้งานวิจัยนี้ได้นำสมบัติทางสวนสัทศาสตร์ และค่าพารามิเตอร์ทางเสียง มาใช้ในการหาตำแหน่งของหน่วยเสียง

บทที่ 3
วิธีการจำแนกลักษณะการเปล่งเสียง

บทนี้จะนำเสนอเกี่ยวกับการจำแนกลักษณะการเปล่งเสียง โดยจะเริ่มจาก นิยามของลักษณะการเปล่งเสียง การเลือกพารามิเตอร์ทางเสียงเพื่อใช้สำหรับตรวจหาเสียงพยัญชนะ และวิธีการตรวจหาเสียงพยัญชนะ

นิยามแลนด์มาร์ก

แลนด์มาร์ก (Landmark) คือตำแหน่งของสัญญาณเสียง ที่ใช้เป็นตัวแทนของแต่ละหน่วยเสียงในสัญญาณเสียงพูด โดยสามารถทำการแบ่งได้เป็น 2 ประเภทใหญ่ๆ คือ แลนมาร์กของพยัญชนะ และ แลนมาร์กของสระ โดยทั่วไป แลนมาร์กของสระนั้นจะทำการเลือกในตำแหน่งที่มีลักษณะทางสวนสัทศาสตร์ของสัญญาณเสียงเด่นชัดที่สุด ซึ่งจะเกิดในบริเวณที่มีค่าพลังงานหรือความเข้มของพลังงานมีค่าสูงที่สุดในหน่วยเสียงนั้น ในขณะที่ แลนมาร์กของเสียงพยัญชนะจะเกิดขึ้นในตำแหน่งที่เป็นรอยต่อระหว่างการกัก (closure) หริอปล่อยลม (release) หรือก็คือตำแหน่งที่มีการเปลี่ยนลักษณะทางสวนสัทศาสตร์จาก + ไปเป็น – เช่นเปลี่ยนจาก [+sonorant, +syllabic] ไปเป็น [+sonorant, -syllabic] ซึ่งเทียบได้กับการเปลี่ยนจากเสียงสระไปเป็นเสียงนาสิก หรือเสียงกึ่งสระ ซึ่งงานวิจัยนี้ได้ใช้ตำแหน่งที่เกิดการเปลี่ยนของลักษณะทางสวนสัทศาสตร์มาใช้เป็นแลนมาร์กโดยสามารถแสดงได้ ดังภาพที่ 3.1



ภาพที่ 3.1 แสดงสมบัติสวนสัทศาสตร์ของคำว่า “นาย”

โดยในงานวิจัยนี้ได้แบ่งประเภทของแลนด์มาร์กออกเป็น 5 ประเภทได้แก่ เสียงเงียบ (SIL), เสียงสระ(V), เสียงกึ่งสระและเสียงนาสิก(SC), เสียงกัก(ST), เสียงเสียดแทรก(FR)

ตารางที่ 3.1 ตารางแสดงประเภทของกับประเภทของแลนมาร์กของเสียงภาษาไทย

ประเภทของหน่วยเสียง	ประเภทของแลนมาร์กที่เกิดขึ้น	หน่วยเสียง
เสียงเงียบ (Silence)	SIL (silence)	/p^/, /t^/, /k^/, /sil/, /sp/
เสียงสระ(Vowel)	V (vowel)	/i/, /i:/, /e/, /e:/, /æ/, /æ:/, /ɨ/, /ɨ:/, /ɜ/, /ɜ:/, /a/, /a:/, /u/, /u:/, /o/, /o:/, /ɔ/, /ɔ:/,/ia/, /i:a/, /ɨa/, /ɨ:a/, /ua/, /u:a/, /w^/, /j^/
เสียงนาสิกและ(Nasal)	SC (Sonorant Consonant)	/m/, /n/, /ŋ/, m^/, /n^/, /ŋ^/
เสียงกึ่งสระ(Semivowel)	SC ((Sonorant Consonant)	/l/, /r/, /w/, /j/
เสียงกักไม่พ่นลม (Unaspirated Stop)	SIL (silence) ST (Release Burst)	/p/, /t/, /k/,/b/, /d/, /ʔ/
เสียงกักพ่นลม (Aspirated Stop)	SIL(silence) ST(Release Burst) FR (fricative)	/p^h/, /t^h/, /k^h/
เสียงกึ่งเสียดแทรก(Affricate)	FR (fricative)	/c/, /c^h/
เสียงเสียดแทรก(Fricative)	FR (fricative)	/f/, /s/, /h/ /c/,/s^/, /f^/,

(^) เป็นเสียงตัวสะกดที่ปรากฏท้ายพยางค์

การเลือกพารามิเตอร์ทางเสียง

งานวิจัยนี้จะเริ่มจากการทำการศึกษาการหาค่าพารามิเตอร์ทางเสียง (Acoustic Parameter) ที่ได้มีการนำเสนอโดยกลุ่มนักวิจัยต่างๆ [3, 13, 16, 17, 19] โดยจะทำการคำนวณหาค่าพารามิเตอร์ทางเสียงในแต่ละกรอบเวลา และได้ทำการปรับปรุงค่าพารามิเตอร์ให้เหมาะสมกับเสียงภาษาไทย โดยทำการศึกษาหาพารามิเตอร์ทางเสียงที่สามารถใช้ในการจำแนกสมบัติสวนสัทศาสตร์ และทำการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) ว่าพารามิเตอร์ทางเสียง สามารถทำการจำแนกสมบัติสวนสัทศาสตร์ ได้ดีมากน้อยเพียงใด แล้วจึงทำชุดการเลือกโดยค่าพารามิเตอร์ทางเสียงเพื่อนำมาใช้ในการฝึกสอนซัพพอร์ตเวกเตอร์แมชชีน เพื่อ

ใช้ในการตรวจการจำแนกสมบัติสวนสัทศาสตร์ได้ถูกต้องและแม่นยำ โดยค่าพารามิเตอร์ทางเสียงที่นำมาศึกษาแสดงได้ดังตารางที่ 3.2

ตารางที่ 3.2 ตารางแสดงพารามิเตอร์ทางเสียงที่นำมาศึกษาเพื่อใช้กับเสียงภาษาไทย

พารามิเตอร์ทางเสียง
1. E[0,F3-1000]
2. E[F3,FS/2]
3. E[100,400]
4. E[640,2800]
5. E[2000,3000]
6. E[F3-1000,FS/2]
7. Ratio of E[0,F3-1000] to [F3-1000,FS/2];
8. Ratio of spectral peak in [0, 400] to the spectral peak in [400, fs/2]
9. Energy peak in [0, 900]
10. Location in Hz of peak in [0, 900]
11. Ratio of E[0,358] to [358, 5373]
12. First formant estimation
13. Envelope avg
14. Envelope std
15. Energy Onset
16. Energy Offset
17. SCG
18. Probability of voicing
19. Short time zero crossing rate
20. E[500, 5000]
21. E[1000, 5000]
22. E[2000, 4000]
23. Hilbert envelope standard division;
24. E[1000,2000]
25. E[2000,5000]
26. E[5000,8000]

วิธีการจำแนกลักษณะการเปล่งเสียง

หลังจากที่ได้พารามิเตอร์ทางเสียงจากขั้นตอนการหาพารามิเตอร์ทางเสียง ตามตารางที่ 3.3 เพื่อเอาไปใช้ในการทำการฝึกซัพพอร์ตเวกเตอร์แมชชีน ทั้งหมด 4 ตัว สำหรับใช้ในการหาเพื่อสมบัติทางสวนสัทศาสตร์ 4 ประเภทได้แก่ [speech], [sonorant], [syllabic] และ [continuant] โดยได้ทำการแบ่งเสียงพูดออกเป็นกรอบเวลาขนาด 10 มิลลิวินาที และทำการเลื่อนกรอบเวลาไปทีละ 5 มิลลิวินาที โดยเมื่อใช้ทฤษฎีของเบส์ (Bayes theorem) และความรู้ก่อนหน้า (prior knowledge) เข้ามาร่วมกับสมบัติทางสวนสัทศาสตร์ มาใช้ในการแบ่งประเภทของเสียงตามลักษณะการออกเสียงตามที่เสนอโดย Juneja และ Espy-Wilson [14, 15] แสดงได้ดังภาพที่ 3.2 ทำการแบ่งจำแนกเป็นระดับชั้นและคำนวณความน่าจะเป็นตามระดับชั้น ทำให้เราสามารถหาความน่าจะเป็นภายหลังของกรอบเวลาที่เวลา t ว่าเป็นเสียงสระได้จากสมการ (3.1)

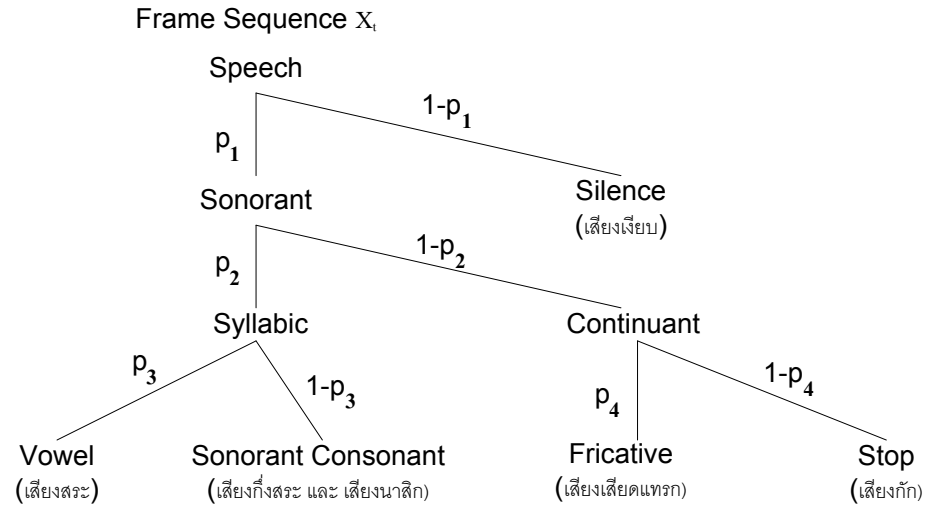
$$\begin{aligned} P(Vowel|x_t) &= P(speech, sonorant, syllabic|x_t) \\ &= P(speech|x_t)P(\ sonorant|speech, x_t) \\ &\qquad P(syllabic|speech, sonorant, x_t) \\ &= p_1 p_2 p_3 \end{aligned} \tag{3.1}$$

โดยเราสามารถทำการคำนวณหาค่าความน่าจะเป็นของเสียงประเภทต่างๆ ได้ในทำนองเดียวกันกับวิธีข้างต้น และเมื่อเรานำ สมมติฐานภายหลังมากสุด - เอ็มเอพี (Maximum A Posterior hypothesis - MAP) โดยทำการตั้งสมมติฐานตามสมการ (3.2) ทำให้เราสามารถทำการจำแนกกรอบเสียงในแต่ละกรอบออกเป็นเสียงประเภทต่างๆได้ เมื่อเราทำการทำการยุบรวมกรอบเสียงที่มีประเภทของเสียงชนิดเดียวกันเข้าด้วยกันจะทำให้เราสามารถทำการตรวจหาหน่วยเสียงประเภทต่างได้ๆ

$$\widehat{B_t} = \arg\max_B P(B|x_t) \tag{3.2}$$

โดยที่ $B \in \{Sonorant\ Consonant, Vowel, Stop, Fricative, Silent\}$

x_t เป็น เวกเตอร์ของพารามิเตอร์ทางเสียงที่เวลา t



ภาพที่ 3. 2 โครงสร้างลำดับชั้นของสมบัติทางสวนสัทศาสตร์ กับลักษณะการออกเสียง

ในงานวิจัยนี้ได้ทำการปรับปรุงวิธีการจำแนกลักษณะการเปล่งเสียง ให้มีความเหมาะสมกับเสียงพูดต่อเนื่องภาษาไทย เนื่องจากเมื่อเทียบเสียงภาษาไทยกับเสียงภาษอังกฤษจะพบว่ามีความแตกต่างกัน โดยความแตกต่างที่สำคัญ ได้แก่

1. เมื่อเสียงตัวสะกดเป็นเสียงกัก จะไม่เกิดการปล่อยลมที่กักเอาไว้ออกมา จึงไม่มีเสียงที่เป็นเสียงระเบิด (stop burst) เกิดขึ้นในส่วนเสียงตัวสะกดที่เป็นเสียงกัก ซึ่งในงานวิจัยนี้ได้กำหนดให้เสียงกักที่เป็นเสียงตัวสะกดเป็นเสียงเงียบ
2. ภาษาไทยเสียงสั้น และเสียงยาว เป็นลักษณะสำคัญที่ใช้แยกเสียงออกจากกันเป็นคนละเสียง ในขณะที่ภาษาอังกฤษ เสียงสั้น เสียงยาวไม่ใช่ลักษณะสำคัญในการแยกเสียง แต่จะใช้การเกร็งและไม่เกร็งของกล้ามเนื้อลิ้นด้วย (tense and lax)

ทำให้ในงานวิจัยนี้ได้ใช้ โครงสร้างลำดับชั้นของสมบัติทางสวนสัทศาสตร์ กับลักษณะการออกเสียง เพื่อใช้ในการตรวจหาเสียงพยัญชนะ โดยได้ทำการใช้โครงสร้างลำดับชั้นของสมบัติทางสวนสัทศาสตร์ ตามภาพที่ 3.2 มาใช้ในการตรวจหาเสียงพยัญชนะออกเป็น เสียงสระและเสียงกึ่งสระ, เสียงกักและเสียงเสียดแทรก, เสียงเงียบ หลังจากนั้นจึงทำการแยกเสียงสระและเสียงกึ่งสระออกเป็น เสียงสระ กับ เสียงกึ่งสระ กับ เสียงกักและเสียงเสียดแทรก ออกเป็น เสียงกัก กับ เสียงเสียดแทรก สามารถแสดงขั้นตอนการทำงานของการตรวจหาได้ดังภาพที่ 3.3

ตารางที่ 3.3 พารามิเตอร์ทางเสียงสำหรับภาษาไทยที่เสนอในงานวิจัยนี้ใช้ในการแบ่งประเภทของเสียงตามลักษณะการออกเสียง

สมบัติทางสวนสัทศาสตร์	พารามิเตอร์ทางเสียง
[speech]	1. Ratio of spectral peak in [0, 400] to the spectral peak in [400, FS/2] 2. E[0,F3-1000] 3. E[F3,FS/2] 4. Spectral center of gravity* 5. Short time zero crossing rate*
[sonorant]	1. E[0,F3-1000] 2. E[F3,FS/2] 3. E[100,400] 4. Ratio of E[0,F3-1000] to [F3-1000,FS/2]
[syllabic]	1. E[640,2800] 2. E[2000,3000] 3. Spectral peak frequency in [0Hz, 900Hz] 4. Ratio of E[0-400] to [400-6000]*
[continuant]	1. E[0, F3-1000] 2. E[F3-1000,FS/2] 3. Energy onset 4. Energy offset 5. Spectral center of gravity* 6. Short time zero crossing rate*

F3 ค่าเฉลี่ยของความถี่ฟอร์แมนท์ที่สามในส่วนที่เป็นเสียงก้อง (voiced)

FS ความถี่ของการสุ่มตัวอย่าง

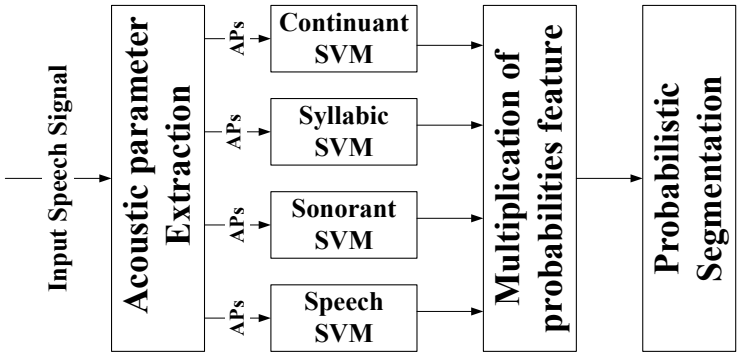
*พารามิเตอร์ทางเสียงสำหรับภาษาไทยที่เสนอเพิ่มเติมในงานวิจัยนี้

ตารางที่ 3.4 พารามิเตอร์ทางเสียงที่เสนอโดย Juneja[15] ใช้ในการแบ่งประเภทของเสียงตามลักษณะการออกเสียง

สมบัติทางสวนสัทศาสตร์	พารามิเตอร์ทางเสียง
[speech]	1. E[0,F3[1]-1000] 2. E[F3,FS/2] 3. Ratio of spectral peak in [0, 400] to the spectral peak in [400, FS/2] 4. Energy Onset 5. Energy Offset
[sonorant]	1. E[0,F3-1000] 2. E[F3,FS/2] 3. E[100,400] 4. Ratio of E[0,F3-1000] to [F3-1000,FS/2]
[syllabic]	1. E[640,2800] 2. E[2000, 3000] 3. Spectral peak frequency in [0Hz, 900Hz] 4. Location in Hz of peak in [0, 900]
[continuant]	1. E[0,F3-1000] 2. E[F3-1000,FS/2] 3. Energy Onset 4. Energy Offset

F3 ค่าเฉลี่ยของความถี่ฟอร์แมนท์ที่สามในส่วนที่เป็นเสียงก้อง (voiced)

FS ความถี่ของการสุ่มตัวอย่าง



ภาพที่ 3. 3 ขั้นตอนการจำแนกลักษณะการเปล่งเสียงที่ใช้ในงานวิจัยนี้

โดยในงานวิจัยนี้ได้ใช้ LIBSVM[20] ซึ่งสามารถทำการประมาณผลของการจำแนกข้อมูลออกมาเป็นความน่าจะเป็นได้ ทำให้สามารถทำการหาค่าความน่าจะเป็นของ แต่ละสมบัติทางสวนสัทศาสตร์ ในแต่ละกรอบเสียงพูดได้

บทที่ 4

การทดลองการจำแนกลักษณะการเปล่งเสียง

ฐานข้อมูลเสียงภาษาไทยโลตัส (LOTUS)

ฐานข้อมูลเสียงภาษาไทยโลตัส (LOTUS) [3] ที่นำมาใช้ในการทดลองการจำแนกลักษณะการเปล่งเสียงคือข้อมูลเสียงจากฐานข้อมูลเสียงโลตัส ซึ่งเป็นฐานข้อมูลเสียงพูดภาษาไทยขนาดใหญ่ที่มีคำศัพท์จำนวนมากแบบเสียงพูดต่อเนื่อง (Large Vocabulary Continuous Speech Recognition: LVCSR) จำนวน 5,000 คำ โดยฐานข้อมูลนี้จะประกอบด้วยข้อมูลเสียง ชุดหน่วยเสียงสมดุล (Phonetically Distributed Set) ใช้สำหรับการฝึกฝนแบบจำลองเสียง ซึ่งมีการกำกับหน่วยเสียง (Phoneme label) และนอกจากนี้ฐานข้อมูลยังประกอบด้วยชุดเสียงอีก 3 ชุดสำหรับฝึกฝนแบบจำลองเสียง และแบบจำลองภาษา ชุดสำหรับทดสอบเพื่อการพัฒนา และชุดสำหรับทดสอบเพื่อประเมินผล

ในการทดลองจะใช้ชุดหน่วยเสียงสมดุล (Phonetically Distributed Set: PD) ของฐานข้อมูลเสียงภาษาไทยโลตัส LOTUS[3] ที่บันทึกเสียงจากสภาพแวดล้อมแบบห้องเงียบ โดยประโยคในชุดหน่วยเสียงสมดุลจะครอบคลุมการเกิดของ “หน่วยเสียงคู่” (Bi-phone) ที่เกิดขึ้นในฐานข้อมูลข้อความภาษาไทยทั้งในพยางค์ ระหว่างพยางค์ และระหว่างคำ โดยไม่คำนึงถึงระดับเสียงวรรณยุกต์ (Tonal Level) และมาจากชุดประโยคที่ครอบคลุมคำศัพท์ภาษาไทยจำนวน 2,269 คำ ซึ่งประกอบไปด้วยเสียงผู้พูด 48 คนแบ่งเป็นเพศชายและหญิงในจำนวนเท่ากัน ประกอบด้วยไฟล์เสียงจำนวน 1680 ประโยค ในการฝึกฝนและทดสอบแบบจำลองทางเสียง โดยจะทำการแบ่งข้อมูล จากฐานข้อมูลชุด PD นี้ออกเป็น 2 ชุด คือ

ชุดที่ 1 ชุดข้อมูลที่ใช้ในการฝึก ประกอบด้วยไฟล์เสียงจำนวน 840 ประโยค

ซึ่งเป็นเสียงของผู้หญิง 420 ประโยค

ซึ่งเป็นเสียงของผู้ชาย 420 ประโยค

ชุดที่ 2 ชุดข้อมูลที่ใช้ในการทดสอบระบบ ประกอบด้วยไฟล์เสียงจำนวน 840 ประโยค โดยทำการเทียบกับประเภทของหน่วยเสียงที่ได้จากไฟล์กำกับหน่วยเสียง

ซึ่งเป็นเสียงของผู้หญิง 420 ประโยค

ซึ่งเป็นเสียงของผู้ชาย 420 ประโยค

โดยทำการแบ่งฐานข้อมูลที่ใช้ในงานวิจัยนี้ออกเป็นชุดๆ เพื่อแยกข้อมูลที่ใช้ในการฝึกฝนชุดข้อมูลที่ทดสอบ ออกจากกันเพื่อป้องกันไม่ให้เกิดการใช้ข้อมูลในการพัฒนาระบบการจำแนกลักษณะการเปล่งเสียงขึ้นกับข้อมูลที่ใช้ฝึกฝนนี้เท่านั้น

การทดลองและผลการทดลอง

การทดลองการจำแนกลักษณะการเปล่งเสียงในที่นี้ จะทดลองเพื่อวัดประสิทธิภาพของวิธีการจำแนกลักษณะการเปล่งเสียงที่เสนอไว้ในบทที่ 3 โดยจะประกอบไปด้วยการทดลองทั้งหมด 3 การทดลอง คือ การทดลองเพื่อวัดประสิทธิภาพการจำแนกสมบัติทางสวนสัทศาสตร์ การทดลองเพื่อเปรียบเทียบประสิทธิภาพการจำแนกลักษณะการเปล่งเสียง และการทดลองเพื่อเปรียบเทียบประสิทธิภาพการตรวจหาตำแหน่งของพยัญชนะและสระเมื่อทำการวัดผลเป็นพยางค์ ซึ่งการทดลองแต่ละส่วนมีรายละเอียดดังต่อไปนี้

ระบบอ้างอิงที่ใช้ในการเปรียบเทียบผลการทดลอง

ในงานวิจัยนี้ได้ทำเปรียบเทียบผลกับระบบอ้างอิงที่เป็นระบบรู้จำเสียงพูดแบบอาศัยแบบจำลองฮิดเดนมาร์คอฟ[1] โดยได้ทำการเชื่อมโยงหน่วยเสียงภาษาไทยออกเป็นหน่วยเสียงตามประเภทของวิธีการออกเสียง 5 ประเภทดังนี้ เสียงเงียบ(SIL), เสียงสระ(V), เสียงกึ่งสระและเสียงนาสิก(SC), เสียงกัก(ST), เสียงเสียดแทรก(FR) แสดงการเชื่อมโยงได้ดังตารางที่ 4.1 เพื่อใช้ในการรู้จำเสียงสำหรับ ระบบรู้จำเสียงพูดแบบอาศัยแบบจำลองฮิดเดนมาร์คอฟ โดยทำการตั้งค่าดังนี้

- คุณลักษณะทางเสียง (Acoustic Features) ใช้ 39 MFCCs
- สำหรับระบบอ้างอิง 1 หน่วยเสียง (Sound Unit) เป็นหน่วยเสียงตามประเภทของวิธีการออกเสียง 5 ประเภทดังนี้ เสียงเงียบ(SIL), เสียงสระ(V), เสียงกึ่งสระและเสียงนาสิก(SC), เสียงกัก(ST), เสียงเสียดแทรก(FR) ในขณะที่ระบบอ้างอิง 2 หน่วยมีเสียงเป็นหน่วยเสียงภาษาไทยจำนวน 64 หน่วยเสียงแล้วจึงทำการเชื่อมโยงกลับเป็นหน่วยเสียงตามประเภทของวิธีการออกเสียง 5 ประเภท

- แบบจำลองทางเสียง (Acoustic Model) แบบจำลองที่ใช้เป็นแบบจำลองเสียงพูดแบบที่ไม่ขึ้นกับบริบทรอบข้าง (Context-independent Phone Model) และประมาณค่าความน่าจะโดยใช้เกาส์เซียนมิกเจอร์โมเดล(Gaussian Mixture Model)
- แบบจำลองทางภาษา ใช้เสียงเดี่ยว (Mono-phone) ที่กำหนดรูปแบบแกรมมาเป็น /C/ /V/ /C/
- สร้างแบบจำลองโดยใช้ข้อมูลเสียงจาก LOTUS โดยชุดข้อมูลที่ใช้ในการฝึก ประกอบด้วยไฟล์เสียงจำนวน 840 ประโยค

ตารางที่ 4.1 การเชื่อมโยงหน่วยเสียงภาษาไทยออกเป็นหน่วยเสียงตามประเภทของวิธีการออกเสียง

ประเภทของหน่วยเสียง	หน่วยเสียง
เสียงเงียบ (Silence)	/p^/, /t^/, /k^/, /sil/, /sp/
เสียงสระ(Vowel)	/i/, /i˞/, /e/, /e˞/, /æ/, /æ˞/, /ɨ/, /ɨː/, /ɜ/, /ɜː/, /a/, /a˞/, /u/, /u˞/, /o/, /o˞/, /ɔ/, /ɔː/,/ia/, /i˞a/, /ɨa/, /ɨːa/, /ua/, /u˞a/, /w^/, /j^/
เสียงกึ่งสระและเสียงนาสิก	/m/, /n/, /ŋ/, m^/, /n^/, /ŋ^/ /l/, /r/, /w/, /j/
เสียงกัก	/p/, /t/, /k/,/b/, /d/, /ʔ/
เสียงเสียดแทรก(Fricative)	/p^h/, /t^h/, /k^h/ /c/, /c^h/ /f/, /s/, /h/ /c/,/s^/, /f^/,

ผลการทดลองการจำแนกสมบัติทางสวนสัทศาสตร์

1. **การทดลองการจำแนกสมบัติทางสวนสัทศาสตร์**

ทำโดยป้อนข้อมูลที่ได้จากกระบวนการสกัดลักษณะสำคัญโดยใช้ พารามิเตอร์ทางเสียงดังที่แสดงไว้ในตารางที่ 3.2 เข้าเป็นอินพุตของเอสวีเอ็มที่ได้จากการฝึกฝน ผลที่ได้คือค่าของฟังก์ชันตัดสินใจ ของแต่ละกรอบเวลาของสัญญาณเสียงซึ่งได้จากซัพพอร์ตเวกเตอร์แมชชีน

(SVM) โดยหากค่าที่ได้คือ +1 ก็หมายความว่าสัญญาณเสียงกรอบเวลานั้นมีสมบัติทางสวนสัทศาสตร์นั้น แต่หากค่าที่ได้เป็น -1 ก็หมายความว่าไม่มีสมบัติทางสวนสัทศาสตร์ดังกล่าว

การจำแนกสมบัติทางสวนสัทศาสตร์ ในแต่ละกรอบเวลาของสัญญาณเสียงจะใช้ เครื่องจำแนกสมบัติทางสวนสัทศาสตร์ ที่ใช้ซัพพอร์ตเวกเตอร์แมชชีน(SVM) ที่ใช้ฟังก์ชันเคอร์เนลแบบฟังก์ชันเรเดียลเบซิส (Radial Basis Function) ผลลัพธ์ที่ได้จากการจำแนกสมบัติทางสวนสัทศาสตร์ จะนำมาวัดประสิทธิภาพโดยเปรียบเทียบกับการจำแนกสมบัติทางสวนสัทศาสตร์ โดยใช้พารามิเตอร์ทางเสียงที่เสนอโดย Juneja [15] ซึ่งแสดงผลการวัดประสิทธิภาพและผลการทดลองได้ดังตารางที่ 4.2

ตารางที่ 4.2 ผลการทดลองการจำแนกสมบัติทางสวนสัทศาสตร์

สมบัติทางสวนสัทศาสตร์		ความแม่น (%)	
		ชุดพารามิเตอร์ทางเสียงที่เสนอ	ชุดพารามิเตอร์ทางเสียงอ้างอิง
[Silence]	ชุดฝึกฝน	96.14	96.00
	ชุดทดสอบ	95.55	95.44
[Sonorant]	ชุดฝึกฝน	88.87	88.87
	ชุดทดสอบ	88.63	88.63
[Syllabic]	ชุดฝึกฝน	84.76	82.52
	ชุดทดสอบ	84.47	82.55
[Continuant]	ชุดฝึกฝน	78.41	64.08
	ชุดทดสอบ	74.84	65.01

2. วิเคราะห์ผลการทดลอง

จากการทดลองวัดประสิทธิภาพการจำแนกสมบัติทางสวนสัทศาสตร์ โดยพิจารณาจากผลการแบ่งสมบัติทางสวนสัทศาสตร์ ในตารางที่ 4.1 พบว่า ชุดพารามิเตอร์ทางเสียงที่เสนอสามารถแบ่งแยกสมบัติทางสวนสัทศาสตร์ ได้เปอร์เซ็นต์ความแม่นโดยรวมอยู่ที่ 75 - 96% และเฉลี่ยที่ 86.46% ในขณะที่ชุดพารามิเตอร์ทางเสียงที่อ้างอิง สามารถแบ่งแยกสมบัติทางสวนสัทศาสตร์ได้เปอร์เซ็นต์ความแม่นยำโดยรวมอยู่ที่ 65 – 96% และเฉลี่ยที่ 82.89% ซึ่งชุด

พารามิเตอร์ทางเสียงที่เสนอมีความผิดพลาดในการจำแนกสมบัติทางสวนสัทศาสตร์ ลดลง 28.09%, 11.0% และ 2.41% สำหรับสมบัติทางสวนสัทศาสตร์ [continuant], [syllabic] และ [silence]

นอกจากนี้ยังพบว่าผลการจำแนกสมบัติทางสวนสัทศาสตร์ นี้มีความคงทนต่อการเปลี่ยนแปลงผู้พูด โดยเห็นได้จากผลที่ได้ จากชุดทดสอบ และชุดฝึกฝนมีความแม่นยำใกล้เคียงกัน

ผลการทดลองการจำแนกลักษณะการเปล่งเสียง

1. **การทดลองการจำแนกลักษณะการเปล่งเสียง**

การจำแนกลักษณะการเปล่งเสียงของงานวิจัยนี้จะทำการเปรียบเทียบผลการการจำแนกลักษณะการเปล่งเสียง ที่ได้ โดยใช้วิธีการตัดแบ่งเสียงออกเป็นประเภทของเสียงตามลักษณะการเปล่งเสียง 5 ประเภทที่ Juneja[15] เป็นผู้นำเสนอโดยใช้ชุดพารามิเตอร์ทางเสียงที่เสนอสำหรับเสียงภาษาไทย เทียบกับระบบอ้างอิง(base line) ที่ทำการจำแนกลักษณะการเปล่งเสียงแบบอาศัยเครื่องรู้จำเสียงพูดแบบ HMM ที่ใช้วิธีการตัดแบ่งเสียงออกเป็นประเภทของเสียง 5 เช่นเดียวกัน

การวัดผลของการหาตำแหน่งพยัญชนะของงานวิจัยนี้จะทำการเปรียบเทียบผลการจำแนกลักษณะการเปล่งเสียง เทียบกับระบบอ้างอิง(base line) นั้น สามารถทำการคำนวณความแม่น(accuracy) ได้จาก สมการที่ (4.1)

$$Accuracy = \frac{Total - TotalError}{Total} \times 100\% \quad (4.1)$$

โดยที่ $Total$ คือ จำนวนตำแหน่งทั้งหมดที่ระบุในไฟล์กำกับหน่วยเสียง $TotalError$ คือ ผลรวมของตำแหน่งที่เกิน และ ตำแหน่งที่ขาด

และวัดผลการลดลงของค่าความผิดพลาดของระบบเปรียบเทียบกับค่าความผิดพลาดของระบบอ้างอิง ตามสมการที่ (4.2)

$$\text{ความผิดพลาดที่ลดลง}(\%) = \frac{error(Baseline) - error(Proposed)}{error(Baseline)} \times 100\% \quad (4.2)$$

โดยที่ ความผิดพลาดที่ลดลง หมายถึง ร้อยละความผิดพลาดที่ลดลงเมื่อใช้วิธีที่เสนอ $error(Baseline)$ หมายถึงค่าความผิดพลาดของตำแหน่งพยัญชนะของชุดอ้างอิง เครื่องรู้จำเสียงพูดแบบ HMM

$error(Proposed)$ หมายถึงค่าความผิดพลาดของตำแหน่งพยัญชนะเมื่อใช้พารามิเตอร์ทางเสียงที่เสนอร่วมกับวิธีการตัดแบ่งเสียงที่เสนอโดย Juneja[15]

โดยผลการหาตำแหน่งของพยัญชนะแสดงได้ตามตารางที่ 4.3 และแสดงคอนฟิวชันเมตริกซ์ของการหาตำแหน่งเสียงพยัญชนะแต่ละประเภทที่ทำการตรวจหาได้ ซึ่งประกอบไปด้วยเสียง 5 ประเภทได้แก่ เสียงเงียบ(SIL), เสียงสระ (V), เสียงกึ่งสระและเสียงนาสิก (SC), เสียงเสียดแทรก (FR) และเสียงกัก (ST) รวมทั้งจำนวนตำแหน่งที่ตรวจหาได้เกิน(Insert)และตำแหน่งที่ขาด (Delete) แสดงได้ตามตารางที่ 4.4

ตารางที่ 4.3 ผลการจำแนกลักษณะการเปล่งเสียง

	ความถูกต้อง (%)	ความแม่นยำ(%)
การจำแนกลักษณะการเปล่งเสียงโดยใช้ชุดพารามิเตอร์ทางเสียงที่นำเสนอ	80.46	68.60
การจำแนกลักษณะการเปล่งเสียงโดยใช้ระบบอ้างอิง 1	74.47	69.08
การจำแนกลักษณะการเปล่งเสียงโดยใช้ระบบอ้างอิง 2	81.21	62.4

1. **วิเคราะห์ผลการทดลอง**

จากการทดลองวัดประสิทธิภาพการหาตำแหน่งเสียงพยัญชนะ ในตารางที่ 4.2 พบว่า ชุดพารามิเตอร์ทางเสียงที่เสนอสามารถทำการการหาตำแหน่งเสียงพยัญชนะ ได้โดยมีความถูกต้อง 80.46% และมีความแม่นยำ 68.60% ซึ่งชุดพารามิเตอร์ทางเสียงที่เสนอมีความผิดพลาดโดยรวมในการหาตำแหน่งเสียงพยัญชนะ ลดลง 23.46% เมื่อเทียบกับระบบอ้างอิง เมื่อทำการพิจารณาคอนฟิวชันเมตริกซ์ของการหาตำแหน่งเสียงพยัญชนะ พบว่า การใช้ การใช้พารามิเตอร์ทางเสียงที่เสนอสามารถทำการการหาตำแหน่งเสียงพยัญชนะได้ดีในพยัญชนะเสียงกึ่งสระและเสียงนาสิก

และเสียงสระ ในขณะที่ระบบอ้างอิงสามารถทำการหาตำแหน่งเสียงเสียดแทรกและเสียงกักได้แม่นยำมากกว่า

ตารางที่ 4.4 แสดงคอนฟิวชันเมตริกซ์ของการหาตำแหน่งเสียงพยัญชนะ

การหาตำแหน่งของพยัญชนะโดยใช้ชุดพารามิเตอร์ทางเสียงที่นำเสนอ								
	Total	SIL	V	SC	FR	ST	Delete	Correct (%)
SIL	8527	6892	2	33	78	64	1458	80.83
V	13270	16	12072	325	21	20	861	90.97
SC	11062	89	68	8099	72	228	2279	73.21
FR	2637	53	7	103	2025	222	227	76.79
ST	6797	45	17	276	168	4939	1352	72.66
Insert	-	824	405	3887	1006	964	-	-
การหาตำแหน่งของพยัญชนะโดยใช้ระบบอ้างอิง 1								
	Total	SIL	V	SC	FR	ST	Delete	Correct (%)
SIL	8527	6778	50	12	96	335	1256	79.49
V	13270	2	9479	2044	5	17	1723	71.43
SC	11062	154	1	6838	37	611	3421	61.82
FR	2637	28	21	70	2251	150	117	85.36
ST	6797	15	33	81	210	6150	308	90.48
Insert.	-	864	2039	1953	203	869	-	-
การหาตำแหน่งของพยัญชนะโดยใช้ระบบอ้างอิง 2								
	Total	SIL	V	SC	FR	ST	Delete	Correct (%)
SIL	8527	4884	143	48	589	1020	1843	57.28
V	13270	0	12711	269	6	9	275	95.79
SC	11062	5	249	8896	59	427	1426	80.42
FR	2637	4	10	27	2163	160	273	82.03
ST	6797	5	10	137	156	5693	796	83.76
Insert.	-	1775	2327	1945	571	1336	-	-

การตรวจหาตำแหน่งของพยัญชนะและสระเมื่อทำการวัดผลเป็นพยางค์

1. **การทดลองการหารู้จำเสียงในรูปแบบ พยางค์** /C//V//C/

ในการทดลองนี้ได้ทำการวัดผลความถูกต้องของกาดตัดแบ่งเสียงพยัญชนะและเสียงสระ โดยทำการวัด เป็นพยางค์เสียง ในรูปแบบพยางค์ที่ประกอบด้วย พยัญชนะต้น สระ และตัวสะกด /C//V//C/ โดยที่ไม่ได้ทำการระบุว่าเสียงพยัญชนะและเสียงสระเป็นเสียงใด แต่ทำการระบุเป็นชนิดของเสียง 5 ชนิด ได้แก่ เสียงสระ, เสียงกึ่งสระและเสียงนาสิก, เสียงเสียดแทรก, เสียงกัก และเสียงเงียบ เพื่อดูความถูกต้องของการรู้จำเสียงเมื่อทำการพิจารณาในรูปแบบพยางค์แล้วมีความถูกต้องแม่นยำแค่ใด โดยในการทดลองทำการวัดประสิทธิภาพการตรวจหาของระบบโดยใช้การคำนวณหาค่าแม่น, และวัดความสามารถในการสืบค้นคำโดยการวัดหาค่า ความเที่ยง(Precision) ความระลึก (Recall) และเอฟวันเมเซอร์ (F1 Measure)

นอกจากนี้ความถูกต้องและความผิดพลาดที่เกิดขึ้น จากตำแหน่งของพยัญชนะที่ตรวจหาได้นั้น สามารถนำมาวิเคราะห์ประสิทธิภาพของระบบการตรวจหาได้ดังนี้

1. เปอร์เซ็นต์ความแม่น (Accuracy)

$$Accuracy = \frac{Total - TotalError}{Total} \times 100\% \quad (4.3)$$

โดยที่ $Total$ คือ จำนวนตำแหน่งทั้งหมดที่ระบุในไฟล์กำกับหน่วยเสียง และ $TotalError$ คือ ผลรวมของตำแหน่งที่เกิน และ ตำแหน่งที่ขาด

2. ความเที่ยง (Precision)

$$Precision = \frac{T_p}{T_p + F_p} \quad (4.4)$$

3. ความระลึก (Recall)

$$Recall = \frac{T_p}{T_p + F_n} \quad (4.5)$$

4. เอฟวันเมเซอร์ (F1 Measure)

$$F1\ Measure = \frac{2xPrecisionxRecall}{Precision + Recall} \quad (4.5)$$

โดยที่ T_p คือ จำนวนตำแหน่งที่จำแนกถูกต้อง F_p คือ จำนวนตำแหน่งที่จำแนกผิด โดยตำแหน่งไม่อยู่ในกลุ่มแต่จำแนกว่าอยู่ในกลุ่ม F_n คือ จำนวนตำแหน่งที่จำแนกผิด โดยตำแหน่งอยู่ในกลุ่มแต่จำแนกว่าไม่อยู่ในกลุ่ม

โดยได้ผลการทดลองดังที่แสดงในตารางที่ 4.5

2. วิเคราะห์ผลการทดลอง

ในการทดสอบนี้ เมื่อทำการเปรียบเทียบค่าของ เปอร์เซ็นความแม่น เปอร์เซ็นความเที่ยง และเปอร์เซ็นความระลึก จากวิธีที่นำเสนอมีค่าใกล้เคียงกัน โดยที่ระบบที่เสนอมีค่าเปอร์เซ็นความถูกต้องความเที่ยง และเปอร์เซ็นความระลึก สูงกว่าเล็กน้อย เนื่องจาก ระบบอ้างอิงมีค่าความแม่นมากกว่า 2.51% เนื่องมากจากระบบอ้างอิงเกิดความผิดพลาดแบบเกินน้อยกว่า ในขณะที่ระบบที่เสนอนั้นมีความผิดพลาดแบบขาดน้อยกว่าทำให้มีค่าเปอร์เซ็นความระลึกสูงกว่า ระบบอ้างอิง 2.12% เมื่อพิจารณาประสิทธิภาพการสืบค้นด้วย เอฟวันเมเซอร์ (F1 Measure) พบว่าทั้งสองระบบที่เสนอและระบบอ้างให้ค่าเอฟวันเมเซอร์ (F1-measure) ใกล้เคียงกัน ที่ 0.76 และ 0.75 ตามลำดับ

ตารางที่ 4. 5 แสดงผลการรู้จำเสียงพูดเป็นพยางค์ /C//V//C/

	ระบบที่เสนอ	ระบบอ้างอิง1	ระบบอ้างอิง2
	ชุดทดสอบ	ชุดทดสอบ	ชุดทดสอบ
Correct (%)	81.16	81.26	84.92
Accuracy (%)	67.86	70.37	70.15
Precision (%)	71.64	71.45	73.99
Recall (%)	81.16	79.04	84.92
F1 Measure	0.76	0.75	0.79

บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

สรุปผลการวิจัย

การจำแนกลักษณะการเปล่งเสียงในงานวิจัยนี้ประกอบด้วยขั้นตอน หลัก ดังนี้ คือ ขั้นตอนการการจำแนกสมบัติทางสวนสัทศาสตร์ และขั้นตอนการรวมผลสมบัติทางสวนสัทศาสตร์ เพื่อทำการจำแนกลักษณะการเปล่งเสียง ในวิทยานิพนธ์นี้ได้ทำการนำเสนอ ชุดพารามิเตอร์ทางเสียงที่ใช่ในการจำแนกสมบัติทางสวนสัทศาสตร์ สำหรับเสียงพูดภาษาไทย เพื่อนำไปใช้กับระบบการตัดแบ่งเสียงพูดที่ใช้สมบัติทางสวนสัทศาสตร์[15] โดยได้ทำการเปรียบเทียบผลที่ได้ กับระบบรู้จำเสียงพูดแบบอาศัยแบบจำลองฮิดเดนมาร์คอฟ (HMM) โดยคาดหวังว่า ระบบที่นำเสนอจะสามารถทำการจำแนกลักษณะการเปล่งเสียง ได้ถูกต้องและแม่นยำ เพื่อสามารถนำไปใช้ในการพัฒนาระบบรู้จำเสียงพูดสำหรับภาษาไทยได้ต่อไปในอนาคต โดยสามารถสรุปผลการวิจัยได้ดังนี้

การตรวจหาตำแหน่งเสียงพยัญชนะ

ในการจำแนกลักษณะการเปล่งเสียงในงานวิทยานิพนธ์นี้เริ่ม จากค่าพารามิเตอร์ทางเสียงเพื่อใช้ในการจำแนกสมบัติทางสวนสัทศาสตร์ในแต่ละเฟรม โดยนำค่าพารามิเตอร์ทางเสียงที่ได้มาเป็นข้อมูลตัวแทนสัญญาณเสียงให้กับซัพพอร์ตเวกเตอร์แมชชีน (SVM) เพื่อใช้ในการหาสมบัติทางสวนสัทศาสตร์ ซึ่งเป็นการนำความรู้ทางทางสวนสัทศาสตร์มาใช้กับการรู้จำเสียงพูด หลังจากนั้นจึงนำผลการจำแนกสมบัติทางสวนสัทศาสตร์ มาใช้ในการการจำแนกลักษณะการเปล่งเสียงในเสียงพูดภาษาไทย วิธีที่ใช้มีลักษณะเด่นดังนี้

1. สามารถเลือกใช้สมบัติทางสวนสัทศาสตร์ ให้เหมาะสมกับประเภทของเสียงแต่ละประเภทได้ นอกจากนี้ยังสามารถปรับเพิ่มเติมค่าพารามิเตอร์ทางเสียง เพื่อใช้ในการหาสมบัติทางสวนสัทศาสตร์ ในภายหลังได้โดยง่าย เพื่อเพิ่มประสิทธิภาพและความแม่นยำให้มากขึ้น
2. วิธีที่ใช้นี้สามารถทำการฝึกฝนโดยใช้ชุดฐานข้อมูลเสียงที่มีขนาดเล็กได้ และผลที่ได้ไม่ขึ้นกับเสียงของผู้พูดที่ใช้ในการฝึกฝน

ในงานวิทยานิพนธ์นี้ ได้ทำการปรับปรุงค่าพารามิเตอร์ทางเสียง ให้เหมาะสมกับการใช้งานกับเสียงภาษาไทย โดยได้ปรับให้มีการใช้ 1) จุดศูนย์ถ่วงของสเปกตรัม(Spectral center of gravity) และ อัตราการตัดศูนย์ในช่วงเวลาสั้น (Short time zero crossing rate) แยกเสียงเงียบ และ นำมาใช้ในการหาสมบัติทางสมบัติทางสวนสัทศาสตร์ [continuant] เพื่อใช้ในการแบ่งแยก

เสียงเสียดแทรกและเสียงกัก 2) อัตราส่วนพลังงานในช่วงความถี่ [0-400] Hz ต่อ พลังงานในช่วงความถี่ [400-6000] Hz ในการหาสมบัติทางสมบัติทางสวนสัทศาสตร์ [syllabic] โดยผลการทดลองที่ได้แสดงให้เห็นว่ามีความผิดพลาดในการจำแนกสมบัติทางสมบัติทางสวนสัทศาสตร์ลดลง 28.09%, 11.0%, 2.41% สำหรับการจำแนกสมบัติทางสวนสัทศาสตร์ [continuant], [syllabic] และ [silent] ตามลำดับ เมื่อทำการเปรียบเทียบกับ พารามิเตอร์ทางเสียงที่ใช้ในการจำแนกสมบัติทางสวนสัทศาสตร์สำหรับเสียงภาษาอังกฤษ และเมื่อทำการตัดแบ่งเสียงเพื่อทำการหาตำแหน่งเสียงพยัญชนะ และ เสียงสระ พบว่าได้ความถูกต้องในการตัดแบ่ง 80.46% โดยมีความผิดพลาดในการตัดแบ่งลดลง 23.46% เมื่อเทียบกับระบบอ้างอิงที่ใช้การรู้จำเสียงพูดแบบอาศัยแบบจำลองฮิดเดนมาร์คอฟ ในการทดลองสุดท้ายพบว่าเมื่อทำการเทียบผลการรู้จำในระดับพยางค์ ในรูปแบบ พยัญชนะต้น-สระ-ตัวสะกด(/C_i/ /V/ /C_f/) ระบบที่เสนอกับระบบอ้างอิงให้ความถูกต้องในระดับเดียวกัน

โดยสรุปแล้ววิทยานิพนธ์นี้เสนอพารามิเตอร์ทางเสียงสำหรับภาษาไทย โดยนำไปใช้กับวิธีการตัดแบ่งเสียงพูดที่ใช้สมบัติทางสวนสัทศาสตร์[15] เพื่อทำการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย โดยระบบที่เสนอให้ความถูกต้องในระดับที่ดี ดังนั้นระบบที่นำเสนออาจนำไปประยุกต์ใช้ในการเพิ่มประสิทธิภาพการรู้จำเสียงพูด หรือนำไปใช้ในการตรวจจับการออกเสียงว่ามีการออกเสียงได้ถูกต้องหรือไม่

ข้อเสนอแนะ

แม้ว่าการจำแนกลักษณะการเปล่งเสียงที่ได้จะอยู่ในระดับที่ดี แต่ยังไม่สามารถเทียบกับการจำแนกโดยใช้คนได้ จึงเห็นว่างานวิจัยนี้ยังมีแนวทางในการพัฒนาได้อีกหลายทาง หรือนำไปประยุกต์ใช้ในงานอื่นๆ เช่น

1. ทำการเพิ่มเติมศึกษาพารามิเตอร์ทางเสียงชนิดใหม่ๆ เพิ่มเติมเพื่อให้สามารถจำแนกสมบัติทางสวนสัทศาสตร์ หรืออาจเพิ่มสมบัติทางสวนสัทศาสตร์ที่เกี่ยวข้องกับคุณสมบัติของอวัยวะที่เป็นฐานในการออกเสียง (place feature) เพิ่มเติมเข้าไปเพื่อให้ระบบสามารถระบุได้ว่า ตำแหน่งของเสียงพยัญชนะ และ เสียงสระ เป็นเสียงพยัญชนะ หรือเสียงสระ เสียงใด

2. นำไปประยุกต์ใช้ในการตรวจจับการออกเสียงว่ามีการออกเสียงได้ถูกต้องหรือไม่ตามหลักการออกเสียงหรือไม่โดยนำสมบัติทางสวนสัทศาสตร์ ที่จำแนกได้ว่าถูกต้องกับคำที่ทำการออกเสียงหรือไม่

รายการอ้างอิง

[1] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE 77 (1989) : 257-286.

[2] Kasuriyam, S. Sornlertlamvanich, V. Cotsomrong, P. Kanokphara, S. and Thatphithakkul, N. Thai Speech Corpus for Thai Speech Recognition. The Oriental COCOSDA (2003) : 54-61.

[3] A. Juneja and C. Espy-Wilson. A probabilistic framework for landmark detection based on phonetic features for automatic speech recognition. Journal of the Acoustical Society of America 123 (2008) : 1154-1168.

[4] กาญจนา นาคสกุล, ระบบเสียงภาษาไทย, พิมพ์ครั้งที่ 4, โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2541.

[5] บุญเสริม กิจศิริกุล และ ณัฐกร ทับทอง, การพัฒนาระบบรู้จำเสียงพูดภาษาไทย รายงานวิจัยฉบับสมบูรณ์ โครงการวิจัยร่วมภาครัฐและเอกชน ปีงบประมาณ 2546, ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย 2548.

[6] Pairote, L. Speech Segmentation for Thai Segment-Based Speech Recognition Using Acoustic-Phonetic Information. Master's Thesis Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, 2006.

[7] Chomsky, N. and Halle, N. The Sound Pattern of English: MIT Press, 1968.

[8] Liu , S. A. Landmark detection for distinctive feature-based speech recognition. Journal of the Acoustical Society of America 100 (1996) : 3417-3430.

[9] Chen, M. Y. Nasal detection module for a knowledge-based speech recognition system. ICSLP-2000 (2000) : 636-639.

[10] Stevens, K. N. Models of Phonetic Recognition II: An Approach to Feature based Recognition. in Proceedings of Montreal Symposion on Speech Recogntion, McGill University, (1986).

[11] Park , C. Y., Consonant Landmark Detection for Speech Recognition. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 2008.

[12] Lee , J. I. and Choi, J. Y. Detection of Obstruent Consonant Landmark for Knowledge Based Speech Recgonition System. The Journal of the Acoustical Society of America 123 (2008) : 2417-2421.

[13] Bitar , N. N. and Espy-Wilson , C. Knowledge-based parameters for HMM speech recognition. in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings (1996) : pp. 29-32.

[14] Juneja, A. and Espy-Wilson, C. Segmentation of Continuous Speech Using Acoustic-Phonetic Parameters and Statistical Learning in Proc. The Ninth International Conference on Neural Information Processing 2002 (ICONIP 2002) (2002) : 726-730.

[15] Juneja, A. and Espy-Wilson, C. Speech Segmentation Using Probabilistic Phonetic Feature Hierarchy and Support Vector Machines in Proceedings of the International Joint Conference on Neural Networks (2003) : 675-679.

[16] Pruthi, T. and Espy-Wilson, C. Automatic Classification of Nasals and Semivowels. in Proc. The Fifteenth International Congress of Phonetic Sciences (ICPhS 2003) (2003) : 3061-3064.

[17] Pruthi, T. and Espy-Wilson, C. Acoustic Parameters for Automatic Detection of Nasal Manner. Speech Communication 43 (2004) : 225-239.

[18] Abdelatty Ali, A. M.; Van Der Spiegel, J. and Mueller, P. Acoustic-phonetic features for the automatic classification of stop consonants. in IEEE Transactions on Speech and Audio Processing 9 (2001) : 833-841.

[19] Salomon, A.; Espy-Wilson, C. and Deshmukh, O. Detection of speech landmarks: Use of temporal information. Journal of the Acoustical Society of America 115 (2004) : 1296-1305

[20] Chang and C.C. and Lin, C.J. LIBSVM: a library for support vector machines 2001.

ประวัติผู้เขียนวิทยานิพนธ์

นายวิทยา โรจน์กิตติเจริญ เกิดเมื่อวันที่ 17 ตุลาคม พ.ศ. 2522 ที่จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นและตอนปลาย จากโรงเรียนนวมินทราชูทิศ กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาบัณฑิต ในสาขาวิชาวิศวกรรมคอมพิวเตอร์ไฟฟ้า จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2543