การตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทย

นางสาวเพียงจิต ดารีเย๊าะ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

VOWEL LANDMARK DETECTION IN THAI CONTINUOUS SPEECH

Miss Piengjit Dareeyoah

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 2006

Copyright of Chulalongkorn University

491389

หัวข้อวิทยานิพนธ์	การตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทย
โดย	นางสาวเพียงจิต ดารีเย๊าะ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาร่วม	อาจารย์ ดร. โปรดปราน บุณยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษา

(อาจารย์ ดร.อติวงศ์ สุชาโต)

..... อาจารย์ที่ปรึกษาร่วม

(อาจารย์ ดร. โปรดปราน บุณยพุกกณะ)

..... กรรมการ

(อาจารย์ ดร.พิษณุ คนองชัยยศ)

..... กรรมการ

(ดร.ชัย วุฒิวิวัฒน์ชัย)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เพียงจิต ดารีเย๊าะ : การตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทย. (VOWEL LANDMARK DETECTION IN THAI CONTINUOUS SPEECH) อ. ที่ปรึกษา : อ.ดร.อติวงศ์ สุชาโต, อ.ที่ปรึกษาร่วม : อ.ดร.โปรดปราน บุณยพุกกณะ, 84 หน้า.

ระบบรู้จำเสียงพูด ตั้งแต่อดีตจนถึงปัจจุบันได้ถูกพัฒนาขึ้นอย่างต่อเนื่อง เพื่อเพิ่มความสามารถในการรู้จำเสียงของระบบให้มีประสิทธิภาพมากที่สุด การพัฒนาระบบรู้จำเสียงพูดนั้นสามารถทำหลายวิธี ซึ่งหนึ่งในวิธีเหล่านั้นคือ การปรับปรุงกระบวนการในการรู้จำเสียงพูดให้มีความถูกต้องมากที่สุด หรือแม้แต่การเพิ่มตัวแปรบางชนิด เพื่อให้ระบบรู้จำเสียงพูดใช้เป็นความรู้เพิ่มเติมสำหรับการรู้จำเสียงพูดในแต่ละครั้ง

สระเป็นหน่วยเสียงแบบหนึ่งที่ทำหน้าที่เป็นแกนของพยางค์ และตำแหน่งของสระในประโยคคำพูดนั้น สามารถทำให้การตัดแบ่งเป็นหน่วยเสียงมีความถูกต้องมากขึ้นได้ นอกจากนี้ตำแหน่งของสระทั้งหมดสามารถบอกจำนวนพยางค์ที่เกิดขึ้นในประโยค ซึ่งใช้เป็นความรู้หนึ่งในการรู้จำเสียงพูดของระบบรู้จำเสียงพูดประเภทต่างๆ ทำให้ระบบรู้จำเสียงพูดเหล่านั้น สามารถรู้จำเสียงพูดได้ถูกต้องยิ่งขึ้น ดังนั้น วิทยานิพนธ์นี้นำเสนอวิธีการในการตรวจหาตำแหน่งสระ ด้วยลักษณะทางสวนสัทศาสตร์ ซึ่งประกอบไปด้วย ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ สำหรับใช้ในการหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงพูด และค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ สำหรับการคัดเลือกตำแหน่งที่จะเป็นสระด้วยระเบียบวิธีคอนเวกซ์ฮัลล์ การประเมินประสิทธิภาพของการตรวจหาสระนั้น ได้ทำการทดลองกับฐานข้อมูลเสียงโลตัส พบว่า ในเสียงพูดต่อเนื่องแบบบันทึกในห้องเงียบได้ความถูกต้องของการตรวจหาสระเท่ากับ 84.98 % และในเสียงพูดต่อเนื่องแบบบันทึกในสภาวะปกติ ได้ความถูกต้องของการตรวจหาสระเท่ากับ 85.33 % นอกจากนี้ ยังได้ทำการทดลองกับฐานข้อมูลเสียงชุดตัวเลข และได้ความถูกต้องของการตรวจหาสระเท่ากับ 95.80 % ในเสียงพูดตัวเลขแบบบันทึกผ่านทางไมโครโฟนในห้องเงียบ และในเสียงพูดตัวเลขแบบบันทึกผ่านทางโทรศัพท์ในห้องปกติ ได้ความถูกต้องของการตรวจหาสระเท่ากับ 84.22 %

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมคอมพิวเตอร์..... ลายมือชื่อนิสิต เพียงจิต ดารีเย๊าะ

สาขาวิชา.....วิศวกรรมคอมพิวเตอร์..... ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา.....2549..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

487 04088 21 : MAJOR COMPUTER ENGINEERING

KEY WORD: VOWEL LANDMARK / SYLLABLE DETECTION / SPEECH RECOGNITION

PIENGJIT DAREEYOAH : VOWEL LANDMARK DETECTION IN THAI CONTINUOUS SPEECH. THESIS ADVISOR : ATIWONG SUCHATO, Ph.D., THESIS COADVISOR : PROADPRAN PUNYABUKKANA, Ph.D., 84 pp.

From the past to present, speech recognition systems have been continuously developed in order for them to achieve possible maximal accuracy. Speech recognition systems can be done in many ways such as improving or adding some parameters into the recognition process.

Vowels are the nuclei of syllable. Their locations in speech utterances can help segmentation to obtain better performance. Furthermore, the number of vowels in a speech utterance can be used as additional speech recognition constraints. In this thesis, a method of vowel landmark detection based on two acoustic measurements is proposed. The first measurement is the maximal autocorrelation value of speech signal in the equivalent frequency range of 60 ถึง 320 Hz. This measurement is used for classifying speech frames into voiced or voiceless frames. Another one is the low frequency removed energy. The convex hull algorithm is used for picking the peak of low frequency removed energy profile to mark the location of a vowel landmark. The evaluation of this method was done on three corpora.

In Large Vocabulary Thai for Continuous Speech Recognition Corpus, it performs with 84.98% accuracy for clean speech data set and 85.33% accuracy for office environment speech data set. In Spoken Digit corpus, it performs with 95.80% accuracy for clean number data set and 84.22% of accuracy for telephone number data set. Also, it performed with 86.16% for TIMIT corpus.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department ..... Computer Engineering	Student's signature .....
Field of study .... Computer Engineering	Advisor's signature .....
Academic year ..... 2006.....	Co-advisor's signature .....

กิตติกรรมประกาศ

ในโอกาสนี้ข้าพเจ้าใคร่ขอขอบคุณ อาจารย์ ดร. อดิวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ อาจารย์ ดร. โปรดปราน บุณยพุกกณะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งท่านทั้งสองได้ช่วยเหลือ ให้คำแนะนำ ข้อคิดที่เป็นประโยชน์ อันเป็นส่วนสำคัญที่ทำให้ วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณ รศ. ดร. บุญเสริม กิจศิริกุล อ. ดร.พิษณุ คนองชัยยศ และ ดร.ชัย วุฒิวิวัฒน์ชัย ที่ให้ข้อแนะนำดีๆ และข้อคิดที่เป็นประโยชน์

นอกจากนี้ข้าพเจ้าขอขอบคุณ พี่ๆ เพื่อนๆ และน้องๆห้องปฏิบัติการ SLS ที่ได้ให้ความร่วมมือ สนับสนุน ช่วยเหลือ จนกระทั่งวิทยานิพนธ์นี้สำเร็จไปได้ด้วยดี

ท้ายที่สุดขอขอบคุณ บิดา มารดา พี่ชาย น้องสาว ที่ได้มอบกำลังใจ ความห่วงใย ช่วยเหลือแก่ข้าพเจ้าเสมอมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญภาพ ..... ญ
สารบัญตาราง 1
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์ของการวิจัย 2
ขอบเขตของการวิทยานิพนธ์ 2
ขั้นตอนการดำเนินการวิจัย 3
ประโยชน์ที่คาดว่าจะได้รับ 3
โครงสร้างวิทยานิพนธ์ 3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5
ทฤษฎีที่เกี่ยวข้องทางด้านภาษาศาสตร์ 5
1. อวัยวะที่ทำให้เกิดเสียง (The Organs of Speech) 5
2. ลักษณะที่แตกต่างกันของเสียง 6
3. เสียงพยัญชนะ 10
4. เสียงวรรณยุกต์ 12
เสียงสระ (Vowel) 13
1. ทฤษฎีในการสร้างเสียงสระ 16
2. แบบจำลองแหล่งกำเนิด-ตัวกรอง (Source-Filter Model) 17
3. ลักษณะที่แตกต่างกันระหว่างเสียงสระและเสียงพยัญชนะ 20
ทฤษฎีที่ใช้ในการหาลักษณะสำคัญของเสียง 26
1. ค่าพลังงาน (Energy) 26
2. ค่าอัตสหสัมพันธ์ (Autocorrelation) 27
วรรณกรรมที่เกี่ยวข้อง 28
1. การตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากกฎ 28
2. การตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากสถิติ 31

บทที่ 3 วิธีการตรวจหาตำแหน่งสระ 33
นิยามของตำแหน่งสระ 33
ที่มาของแนวคิดวิธีการตรวจหาตำแหน่งสระ 34
ภาพรวมวิธีการตรวจหาตำแหน่งของสระ 35
1. การตรวจหาความก้องหรือความไม่ก้องของเสียง 36
2. การคัดเลือกตำแหน่งของสระ 36
3. การให้คะแนน 37
การหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ของสัญญาณเสียง 37
การหาค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของสัญญาณเสียง 40
ระเบียบวิธีคอนเวกซ์ฮัลล์ 42
การให้คะแนนตำแหน่งที่ตรวจหาได้ 44
บทที่ 4 ฐานข้อมูลเสียงที่ใช้ในการตรวจหาสระ 46
ฐานข้อมูลเสียงภาษาไทย (Lotus) 46
1. ชุดประโยค 47
2. ผู้พูด 47
3. การบันทึกเสียง 47
4. หน่วยเสียงภาษาไทย ที่ปรากฏในฐานข้อมูล 48
ฐานข้อมูลเสียงตัวเลข (Spoken Digit Corpus) 52
1. ฐานข้อมูลตัวเลขที่บันทึกในห้องเงียบ 52
2. ฐานข้อมูลตัวเลขที่บันทึกผ่านทางโทรศัพท์ 52
บทที่ 5 ผลการตรวจหาสระ 54
ผลการทดสอบการหาความก้องหรือความไม่ก้องของเสียง 54
1. การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยเปรียบเทียบกับค่าขีดแบ่ง 54
2. การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยใช้การจำแนกกลุ่มด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น 55
3. การวิเคราะห์ผลการการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง 56
ผลการทดสอบการหาตำแหน่งของสระ 57
1. ผลการตรวจหาตำแหน่งสระ 58
2. การวิเคราะห์ผลการการตรวจหาตำแหน่งที่ได้ 59
ผลการทดสอบการให้คะแนนของตำแหน่งสระ 62
1. ผลการทดสอบ 62

2. การวิเคราะห์ความถูกต้องของการให้คะแนนตำแหน่งที่ตรวจหาได้ 64
วิธีการตรวจหาตำแหน่งสระของ Howitt 65
1. ผลการทดสอบ 66
2. การวิเคราะห์ผลการเปรียบเทียบผลลัพธ์ของการหาตำแหน่งสระระหว่างวิธีของ Howitt และ วิธีที่นำเสนอในวิทยานิพนธ์ 67
การวิเคราะห์เชิงภาษา 68
บทที่ 6 ข้อสรุปและข้อเสนอแนะ 73
ข้อสรุป 73
คุณประโยชน์ต่อวงวิชาการ 74
ข้อเสนอแนะ 75
รายการอ้างอิง 76
ภาคผนวก ก ตัวอย่างการนำระบบการตรวจหาตำแหน่งสระไปประยุกต์ใช้ 78
การตรวจหาจำนวนพยางค์ 78
ประวัติผู้เขียนวิทยานิพนธ์ 84

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 2.1 แผนภาพแสดงทางเดินของกระแสลมที่ออกจากปอด [4] 6
รูปที่ 2.2 ลักษณะคลื่นเสียงของเสียงก้อง 7
รูปที่ 2.3 ลักษณะคลื่นเสียงของเสียงไม่ก้อง 8
รูปที่ 2.4 คลื่นเสียงของสระอะ 9
รูปที่ 2.5 คลื่นเสียงของสระอา 9
รูปที่ 2.6 คลื่นเสียงของสระอี 16
รูปที่ 2.7 สัญญาณรายคาบในทางเวลา และทางความถี่ [6] 17
รูปที่ 2.8 รูปร่างของช่องทางเดินเสียงที่สัมพันธ์กับความถี่ฟอร์แมนท์ที่หนึ่งและสอง [6] 18
รูปที่ 2.9 แสดงแอมปลิจูดของสัญญาณเสียงที่ความถี่ต่างๆ ที่เปลี่ยนแปลงไปตามความถี่ฟอร์แมนท์ที่เกิดขึ้น [6] 18
รูปที่ 2.10 แสดงความสัมพันธ์ระหว่างเสียงสระแต่ละตัวกับความถี่ฟอร์แมนท์อันดับที่ 1 และ ความถี่ฟอร์แมนท์อันดับที่ 2 [7] 19
รูปที่ 2.11 ลักษณะของแผนภาพกล่อง 22
รูปที่ 2.12 แผนภาพกล่องแสดงความเป็นรายคาบของสัญญาณเสียงก้อง (V) และสัญญาณเสียงไม่ก้อง (U) 22
รูปที่ 2.13 แผนภาพกล่องแสดงอัตราการตัดศูนย์ของกรอบสัญญาณเสียงก้อง (V) และสัญญาณเสียงไม่ก้อง (U) 23
รูปที่ 2.14 อัตราส่วนของพลังงานที่ความถี่ต่ำต่อพลังงานที่ความถี่สูงของคำว่า หา [8] 24
รูปที่ 2.15 ความถี่ฟอร์แมนท์ที่หนึ่งและความถี่ฟอร์แมนท์ที่สองของคำว่า นาย 25
รูปที่ 2.16 แผนภาพกล่องของพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของเสียงสระ (Vowel) และเสียงก้องที่ไม่เสียงสระ (Non-Vowel) 25
รูปที่ 2.17 แผนภาพคลื่นเสียงและพลังงานของสัญญาณเสียงคำว่า "นา" 26
รูปที่ 2.18 ค่าอัตสหสัมพันธ์ ของสัญญาณเสียงคำว่า "มา" (maa0) ณ เวลาที่ 0.0375 วินาที (ก) สัญญาณเสียง (ข) ค่าอัตสหสัมพันธ์ 28
รูปที่ 2.19 แสดงแผนภาพของการแยกหน่วยเสียงออกเป็นสระ/ไม่ใช่สระของ Sirigos [15] 31
รูปที่ 3.1 แผนภาพการตรวจหาตำแหน่งสระ 37
รูปที่ 3.2 การหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ของสัญญาณเสียง 38
รูปที่ 3.3 ตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด เมื่อพิจารณาเฉพาะค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงไม่ก้อง /ส/ 39

รูปที่ 3.4 ตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด เมื่อพิจารณาเฉพาะค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงก้อง /อา/ 40
รูปที่ 3.5 การหาค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของสัญญาณเสียง 41
รูปที่ 3.6 พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ 42
รูปที่ 3.7 ตัวอย่างการหาแลนด์มาร์กโดยระเบียบวิธีคอนเวกซ์ฮัลล์ [3] 43
รูปที่ 3.8 การหาค่าตำแหน่งของแลนด์มาร์กด้วยระเบียบวิธีคอนเวกซ์ฮัลล์ 44
รูปที่ 3.9 การหาความชันของแต่ละตำแหน่งที่ถูกเลือกเป็นสระ 45
รูปที่ 5.1 ความสัมพันธ์ของความผิดพลาดที่เกิดขึ้นในขั้นตอนที่ 1 และขั้นตอนที่ 2 61
รูปที่ 5.2 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ในฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ 62
รูปที่ 5.3 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ของฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ 63
รูปที่ 5.4 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ของฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ 63
รูปที่ 5.5 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ของฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยโทรศัพท์ในห้องปกติ 63
รูปที่ 5.6 ความน่าจะเป็นที่ตำแหน่งที่ตรวจหาได้นั้นจะเป็นสระ เมื่อมีคะแนนในช่วงต่างๆ 64
รูปที่ 5.7 วิธีการตรวจหาสระของ Howitt 65
รูปที่ 5.8 ประสิทธิภาพในการตรวจหาสระที่ปรากฏในแต่ละภาษา 72
รูปที่ ก.1 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ 80
รูปที่ ก.2 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ 80
รูปที่ ก.3 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ 80
รูปที่ ก.4 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียงตัวเลข ที่บันทึกด้วยโทรศัพท์ในห้องปกติ 81
รูปที่ ก.5 จำนวนประโยคที่เกิดความผิดพลาดที่จำนวนต่างๆ เมื่อเทียบกับจำนวนประโยคทั้งหมดของแต่ละฐานข้อมูล 82

สารบัญตาราง

หน้า

ตารางที่ 2.1 หน่วยเสียงพยัญชนะในภาษาไทย 11
ตารางที่ 2.2 แสดงสัทอักษรสากล ของหน่วยเสียงพยัญชนะต้นในภาษาไทย 11
ตารางที่ 2.3 แสดงสัทอักษรสากล ของหน่วยเสียงพยัญชนะท้ายในภาษาไทย 12
ตารางที่ 2.4 หน่วยเสียงสระเดี่ยว 15
ตารางที่ 2.5 แสดงสัทอักษรสากล ของหน่วยเสียงสระในภาษาไทย 19
ตารางที่ 2.6 วรรณกรรมที่เกี่ยวข้องกับการตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากกฎ 30
ตารางที่ 2.7 สรุปวรรณกรรมที่เกี่ยวข้องกับการตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากสถิติ 32
ตารางที่ 4.1 จำนวนของผู้พูดจากสถานที่ต่างๆ ที่ใช้สำหรับการบันทึกเสียง 47
ตารางที่ 4.2 หน่วยเสียงพยัญชนะต้นเดี่ยว ของเสียงภาษาไทย 48
ตารางที่ 4.3 ตัวอย่างของการใช้หน่วยเสียงพยัญชนะต้นเดี่ยว 49
ตารางที่ 4.4 หน่วยเสียงพยัญชนะท้ายเดี่ยวในเสียงภาษาไทย 49
ตารางที่ 4.5 ตัวอย่างของการใช้หน่วยเสียงพยัญชนะท้ายเดี่ยว 50
ตารางที่ 4.6 ตัวอย่างของหน่วยเสียงสระเดี่ยวในภาษาไทย 50
ตารางที่ 4.7 ตัวอย่างของการใช้เสียงควบกล้ำ 51
ตารางที่ 4.8 ตัวอย่างของการใช้เสียงสระผสม 51
ตารางที่ 4.9 ตัวอย่างของข้อมูลเสียงตัวเลขและไฟล์กำกับเสียง 53
ตารางที่ 5.1 การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยเปรียบเทียบกับค่าขีดแบ่ง ของข้อมูลเสียงต่างๆ 55
ตารางที่ 5.2 การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยใช้การจำแนกกลุ่มด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นในฐานข้อมูลเสียง Lotus 56
ตารางที่ 5.3 ผลการตรวจหาตำแหน่งของสระในฐานข้อมูลเสียงต่างๆ 59
ตารางที่ 5.4 ประสิทธิภาพของระบบการตรวจหาสระ 59
ตารางที่ 5.5 การเปรียบเทียบความถูกต้องของการตรวจหาสระที่ได้จากวิธีอ้างอิงและวิธีที่นำเสนอ 60
ตารางที่ 5.6 ผลต่างของความผิดพลาดที่เกิดขึ้นจากวิธีที่นำเสนอและวิธีอ้างอิง ของแต่ละขั้นตอน 61
ตารางที่ 5.7 ผลการตรวจหาตำแหน่งสระด้วยวิธีของ Howittในฐานข้อมูลเสียงต่างๆ 66
ตารางที่ 5.8 ประสิทธิภาพของระบบการตรวจหาสระด้วยวิธีของ Howitt 66
ตารางที่ 5.9 การเปรียบเทียบผลการตรวจหาสระของวิธีการที่นำเสนอในงานวิจัย และวิธีการตรวจหาสระของ Howitt 67

ตารางที่ 5.10 สระที่ปรากฏในภาษาไทยเท่านั้น 69
ตารางที่ 5.11 สระที่ปรากฏในภาษาอังกฤษ 70
ตารางที่ 5.12 สระที่ปรากฏในภาษาไทยและภาษาอังกฤษ 70
ตารางที่ 5.13 ผลการตรวจหาสระที่ปรากฏในภาษาไทยเท่านั้น 71
ตารางที่ 5.14 ผลการตรวจหาสระที่ปรากฏในภาษาอังกฤษเท่านั้น 71
ตารางที่ 5.15 ผลการตรวจหาสระที่ปรากฏในภาษาไทยและภาษาอังกฤษเท่านั้น 71
ตารางที่ ก.1 ผลการตรวจหาจำนวนพยางค์ของแต่ละฐานข้อมูลเสียง 79

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ระบบการรู้จำเสียงพูด (Speech Recognition) ตั้งแต่อดีตจนถึงปัจจุบันได้ถูกพัฒนาขึ้นอย่างต่อเนื่อง เพื่อเพิ่มความสามารถในการรู้จำเสียงของระบบให้มีประสิทธิภาพมากที่สุด ระบบการรู้จำเสียงพูดในปัจจุบันได้มีการนำองค์ความรู้ทางด้านสถิติ เข้ามาประยุกต์ใช้ในกระบวนการรู้จำเสียง เนื่องจากองค์ความรู้ทางด้านสถิตินี้ มีโครงสร้างและองค์ประกอบทางด้านคณิตศาสตร์ที่สมบูรณ์ เช่น ระบบการรู้จำเสียงที่ใช้ข่ายงานประสาทเทียม (Artificial Neural Networks, ANN) และระบบการรู้จำเสียงที่ใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model, HMM) เป็นต้น ระบบการรู้จำเสียงประเภทนี้ จะให้ประสิทธิภาพที่ดีในระบบการรู้จำเสียงพูดที่มีคำศัพท์เป็นจำนวนมาก (Large-Vocabulary) หรือระบบการรู้จำเสียงที่ไม่ขึ้นกับผู้พูด (Speaker-Independent Speech Recognition) [1] อย่างไรก็ตาม ระบบการรู้จำเสียงที่มีพื้นฐานอยู่บนองค์ความรู้ทางด้านสถิตินี้ มีข้อจำกัดคือ ในขั้นตอนการเรียนรู้ข้อมูลที่ใช้ฝึก ระบบการรู้จำเสียงพูดประเภทนี้ จำเป็นต้องมีการทำงานหรือใช้การคำนวณของระบบค่อนข้างมาก เนื่องจากข้อมูลที่ใช้ในการฝึกจะต้องมีจำนวนมากพอ เพื่อให้ระบบรู้จำสามารถเรียนรู้ข้อมูลเสียงได้หลากหลายที่สุด เพื่อที่จะหลีกเลี่ยงข้อจำกัดที่เกิดขึ้นจากระบบรู้จำเสียงที่มีพื้นฐานอยู่บนองค์ความรู้ทางด้านสถิตินี้ จึงมีระบบการรู้จำเสียงอีกประเภทหนึ่งที่ใช้พื้นฐานการตัดสินใจจากกฎ (Knowledge-based) ซึ่งเป็นระบบที่เรียนรู้โครงสร้างและลักษณะที่เด่นชัด ที่สามารถวัดได้โดยตรงจากสัญญาณเสียง

เสียงพูดในทุกภาษานั้น จะต้องประกอบด้วยหน่วยเสียงพยัญชนะและหน่วยเสียงสระร่วมกัน โดยที่หน่วยเสียงทั้งสองประเภทนั้นมีความแตกต่างกันดังนี้ คือ หน่วยเสียงสระเป็นหน่วยเสียงแบบหนึ่งที่สามารถเปล่งได้ดังกว่าหน่วยเสียงพยัญชนะ หน่วยเสียงสระจึงทำหน้าที่เป็นใจกลางหรือแกน (Nucleus) ของพยางค์ ซึ่งหมายความว่ากลุ่มพยัญชนะใดที่ไม่มีสระเป็นใจกลาง จะประกอบกันขึ้นเป็นพยางค์ไม่ได้ ดังนั้น จำนวนของสระที่เกิดขึ้นจากการเปล่งเสียงพูดในแต่ละครั้ง จึงเท่ากับจำนวนของพยางค์ทั้งหมดที่เกิดขึ้นในการเปล่งเสียงพูดครั้งเดียวกัน ด้วยเหตุผลดังกล่าว จึงได้มีการนำเอาจำนวนของสระที่เกิดขึ้นจากการเปล่งเสียงพูดในแต่ละครั้งนั้น ไปประยุกต์ใช้ในงานต่างๆ ที่เกี่ยวข้องกับระบบรู้จำเสียงพูด เช่น การนับจำนวนพยางค์จากประโยคที่พูดนั้น เพื่อนำจำนวนพยางค์ที่ได้ ไปใช้เป็นตัวแปรหนึ่งที่ช่วยให้การรู้จำเสียงของระบบรู้จำเสียงแบบฮิดเดนมาร์คอฟ หรือระบบรู้จำเสียงแบบที่ใช้สมบัติลักษณะเฉพาะ (Distinctive Feature-Based) สามารถรู้จำเสียงพูดได้ถูกต้องมากยิ่งขึ้น เป็นต้น

การเพิ่มประสิทธิภาพในการรู้จำเสียงพูด ของระบบรู้จำเสียงพูดเพื่อให้ระบบรู้จำเสียงสามารถรู้จำเสียงได้อย่างถูกต้องนั้น นอกจากที่จะต้องเลือกลักษณะที่จะใช้เป็นตัวแทนของสัญญาณ

เสียงพูดให้เหมาะสมแล้ว แต่ละกระบวนการในการรู้จำเสียงพูดก็มีความสำคัญที่จะต้องทำงานได้อย่างมีประสิทธิภาพเช่นเดียวกัน กลไกในการรู้จำเสียงพูดประกอบด้วยกระบวนการต่างๆ หลายกระบวนการด้วยกัน เช่น การตัดแบ่งเป็นหน่วยเสียง (Phoneme Segmentation) คือการกำหนดขอบเขตของหน่วยเสียงเป็นพยัญชนะต้น สระ และพยัญชนะท้าย การหาลักษณะสำคัญของเสียง (Feature Extraction) คือการแปลงสัญญาณเสียงที่ถูกตัดแบ่งแล้ว ให้อยู่ในรูปสัมประสิทธิ์ของแกนความถี่ ซึ่งค่าดังกล่าวจะมีขนาดเล็กเหมาะสมที่จะเป็นตัวแทนของสัญญาณเสียงได้เป็นอย่างดี เป็นต้น ตำแหน่งของสระจะช่วยให้กระบวนการตัดแบ่งเป็นหน่วยเสียง ซึ่งเป็นกระบวนการหนึ่งในการรู้จำเสียงพูดมีความถูกต้องมากยิ่งขึ้น เนื่องจากตำแหน่งของสระจะอยู่ใจกลางของพยางค์เสมอ นั่นคือสระจะต้องตามหลังพยัญชนะต้น และอาจจะนำหน้าพยัญชนะท้าย ส่งผลให้การตัดแบ่งเป็นหน่วยเสียงในระบบรู้จำเสียงแบบแบ่งส่วน (Segment-Based Speech Recognition) มีความถูกต้องมากยิ่งขึ้น

ดังนั้น วิทยานิพนธ์นี้ จึงมีวัตถุประสงค์ที่จะศึกษา และพัฒนาวิธีการที่ใช้ในการตรวจหาตำแหน่งของสระในเสียงพูดต่อเนื่องภาษาไทย เพื่อให้มีความถูกต้องมากที่สุด

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาลักษณะร่วมทางสวนสัทศาสตร์ ที่ใช้ในการตรวจหาสระในเสียงพูด
2. เพื่อนำไปสร้างระบบที่ใช้ในการตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทยที่มีประสิทธิภาพ โดยใช้ลักษณะร่วมทางสัทศาสตร์เหล่านั้นร่วมกัน

ขอบเขตของการวิทยานิพนธ์

1. การตรวจหาสระในวิทยานิพนธ์นี้ เป็นการตรวจหาสระของเสียงพูดในภาษาไทย
2. ฐานข้อมูลเสียงที่ใช้ในวิทยานิพนธ์นี้คือ ฐานข้อมูลเสียงภาษาไทย Lotus [2] และฐานข้อมูลเสียงอื่นๆ ที่สามารถเตรียมได้
3. วิทยานิพนธ์นี้เป็นการตรวจหาตำแหน่งของสระจากเสียงพูดต่อเนื่องเท่านั้น ซึ่งจะไม่มีการระบุว่าตำแหน่งของสระที่หาได้นั้นเป็นสระอะไร
4. ประสิทธิภาพของระบบที่ใช้ในการตรวจหาสระของงานวิจัยนี้ จะทำการเปรียบเทียบกับระบบที่ใช้ในการตรวจหาสระของ Howitt [3] เนื่องจาก ประสิทธิภาพของระบบที่เป็นการตรวจหาสระของ Howitt นี้ ได้ผลดีเป็นที่น่ายอมรับ เมื่อเปรียบเทียบกับงานวิจัยที่เป็นการตรวจหาสระของนักวิจัยท่านอื่นๆ

ขั้นตอนการดำเนินการวิจัย

1. ศึกษาทฤษฎี และวิธีการที่ใช้ในการตรวจหาสระ จากงานวิจัยอื่นๆ
2. ศึกษาลักษณะร่วมทางสวนสัทศาสตร์ อันได้แก่ ค่าพลังงาน ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ ของกรอบสัญญาณเสียงที่เป็นเสียงก้อง และเสียงไม่ก้อง
3. ศึกษาวิธีการของการประมวลผลสัญญาณเสียงพูด (Speech signal processing)
4. ออกแบบขั้นตอนที่ใช้ในการตรวจหาสระ
5. บันทึกเสียง และทำการกำกับขอบเขตเสียง (Manual Transcription) เพื่อใช้เป็นฐานข้อมูลเสียง ที่นอกเหนือจากฐานข้อมูลเสียงของ Lotus
6. ทดสอบวิธีการที่นำเสนอ
7. วิเคราะห์ผลการทดลองจากฐานข้อมูลเสียงทั้งหมด ซึ่งประกอบด้วย ฐานข้อมูลเสียงที่เตรียมขึ้นเอง ฐานข้อมูลเสียง Lotus และฐานข้อมูลเสียง TIMIT
8. ปรับปรุงแก้ไข
9. สรุปผลที่ได้

ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงลักษณะร่วมทางสวนสัทศาสตร์ต่างๆ ที่มีประสิทธิภาพในการตรวจหาสระ
2. สามารถนำตำแหน่งของสระที่ตรวจหาได้นี้ ไปใช้ในการแบ่งแยกเสียงออกเป็นส่วนๆ
3. สามารถใช้ในการบอกจำนวนพยางค์ของประโยคนั้นๆ เพื่อเพิ่มความถูกต้องให้กับกระบวนการรู้จำเสียงพูดต่อไป

โครงสร้างวิทยานิพนธ์

ในวิทยานิพนธ์ฉบับนี้จะแบ่งเนื้อหาในการนำเสนอออกเป็น 4 ส่วน คือ ในบทที่ 2 จะกล่าวถึง ทฤษฎีทางด้านภาษาศาสตร์ ลักษณะของเสียงสระ ทฤษฎีที่ใช้ในการหาลักษณะสำคัญของเสียงสำหรับการตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทย และได้กล่าวถึง วรรณกรรมที่เกี่ยวข้องกับการตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทยของวิทยานิพนธ์นี้

ในบทที่ 3 จะได้กล่าวถึง นิยามของตำแหน่งสระ วิธีการในการตรวจหาตำแหน่งสระโดยรวม และวิธีการการหาค่าลักษณะที่สำคัญของเสียง ซึ่งประกอบด้วย ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ และค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์

ในบทที่ 4 จะได้กล่าวถึง ฐานข้อมูลที่ใช้ในการวัดประสิทธิภาพ ระบบการตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทยของวิทยานิพนธ์นี้

ในบทที่ 5 และบทที่ 6 จะได้กล่าวถึงวิธีการและขั้นตอนในการทดสอบการตรวจหาตำแหน่งของสระ ตัวอย่างการนำระบบการตรวจหาตำแหน่งสระไปประยุกต์ใช้ ตลอดจนการนำเสนอผลและวิเคราะห์ผลการทดสอบ และสรุปผลที่ได้จากวิทยานิพนธ์นี้ รวมทั้งให้ข้อเสนอแนะในการพัฒนางานวิจัยส่วนการตรวจหาสระในเสียงพูดนี้ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะได้แบ่งเนื้อหาออกเป็นสี่ส่วน ส่วนแรกกล่าวถึงทฤษฎีทางภาษาศาสตร์ที่เกี่ยวข้องกับการทำวิทยานิพนธ์นี้ ในส่วนที่สองกล่าวถึงเรื่องราวที่เกี่ยวข้องกับเสียงสระ ในส่วนที่สามกล่าวถึงทฤษฎีที่ใช้ในการวิเคราะห์และหาลักษณะสำคัญของเสียง ซึ่งประกอบด้วย ค่าพลังงานและค่าอัตสหสัมพันธ์ ซึ่งเป็นค่าที่ใช้สำหรับการวิเคราะห์สัญญาณเสียงที่ใช้ในวิทยานิพนธ์นี้ และในส่วนสุดท้ายจะได้กล่าวถึงวรรณกรรมที่เกี่ยวข้องกับการตรวจหาตำแหน่งสระของวิทยานิพนธ์นี้

ทฤษฎีที่เกี่ยวข้องทางด้านภาษาศาสตร์

1. อวัยวะที่ทำให้เกิดเสียง (The Organs of Speech)

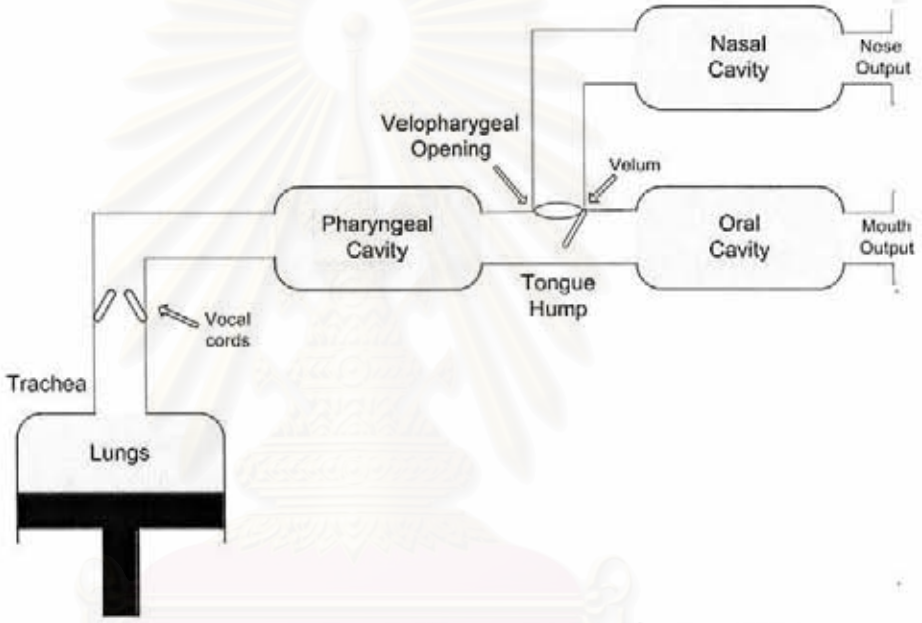
เสียงเกิดจากการสั่นสะเทือนของอากาศ ส่วนสำคัญที่เป็นบ่อเกิดของเสียงได้แก่ ลมหายใจหรือกระแสลม (Air Stream) ที่ผ่านเข้าออกจากปอด ปกติเสียงพูดจะเกิดขึ้นในขณะที่หายใจออก และในขณะที่ลมจากปอดไหลผ่านอวัยวะต่างๆ ไปนั้น ลมหรือกระแสลมจะถูกปิดกั้นหรือถูกบีบบังคับจากอวัยวะเหล่านั้นให้ไปในทิศทางต่างกัน จึงทำให้เกิดเสียงที่แตกต่างกันออกไป

อวัยวะที่ใช้ในการออกเสียงประกอบด้วย ส่วนสำคัญ 3 ส่วน ดังนี้

1. **ส่วนเริ่มต้น** (Initiation) หมายถึง ส่วนที่ทำให้เกิดการเคลื่อนไหวของลม หรือจุดเริ่มต้นของลม ได้แก่ ช่วงลำตัว ซึ่งอวัยวะที่สำคัญคือปอด ซึ่งทำหน้าที่ส่งลมออกจากปอด ให้ไหลผ่านไปตามอวัยวะต่างๆ
2. **ส่วนที่ทำให้เกิดเสียง** (Phonation) หมายถึง อวัยวะที่ใช้ในการเปล่งเสียง อันได้แก่ เส้นเสียง (Vocal Cords) ซึ่งเป็นส่วนที่สำคัญในการออกเสียง เพราะถ้าปราศจากเส้นเสียงจะไม่สามารถเปล่งเสียงออกมาไม่ได้
3. **ส่วนที่เปลี่ยนแปลงลักษณะเสียง** (Articulation) หมายถึง อวัยวะส่วนที่ทำให้การออกเสียงแตกต่างกัน ซึ่งได้แก่ ช่องคอ ช่องจมูก และช่องปาก ทิศทางของลมของลมทำให้เสียงที่เปล่งออกมานั้นแตกต่างกัน เช่น กระแสลมออกทางจมูกหรือทางปาก และเนื่องจากการเคลื่อนไหวของอวัยวะต่างๆ ในปากซึ่งทำให้ขนาดและรูปร่างของปากเปลี่ยนไป

แบบจำลองของกลไกในการเปล่งเสียง แสดงได้ดังรูปที่ 2.1 การเปล่งเสียงเริ่มต้นจากกล้ามเนื้อที่อยู่ในปอดผลักลมที่อยู่ภายในปอดให้ผ่านไปยังหลอดลม (Trachea) เพื่อที่จะเข้าสู่เส้นเสียง การเปรียบเทียบการไหลของลมหายใจที่ออกจากปอด อันเป็นกระบวนการในการออกเสียงกับการทดน้ำเข้านา สามารถช่วยให้การเข้าใจกลไกการเปล่งเสียงง่ายยิ่งขึ้น ดังนี้ ปอดซึ่งเป็น

จุดเริ่มต้นในการออกเสียงเปรียบได้กับแม่น้ำหรือแหล่งน้ำ กระแสลมที่ผ่านออกจากปอดไปตามหลอดลมเปรียบเสมือนน้ำที่ไหลไปตามท่อเพื่อเข้าสู่นา สิ่งกีดขวางแห่งแรกคือเส้นเสียงนั้นเปรียบได้กับประตูน้ำ ซึ่งอาจจะปิดสนิทไม่ให้น้ำไหลผ่าน หรือเปิดให้น้ำผ่านไปโดยไม่ปิดกั้นหรืออาจจะปิดไม่สนิท ยอมให้น้ำผ่านไปได้บ้างโดยใช้แรงดันมากๆ เมื่อกระแสลมจากปอดผ่านด่านแรกคือเส้นเสียง สามารถแยกไปได้ 2 ทิศทาง คือทางจมูกหรือทางปาก ถ้าออกทางจมูกกระแสลมจะผ่านวีลัม (Velum) ซึ่งเป็นช่องที่จะให้ลมผ่านเข้าไปยังช่องจมูก หรือโพรงจมูก ซึ่งเปรียบเสมือนด่านหรือประตูน้ำแห่งที่สอง ถ้าวีลัมปิด กระแสลมจากปอดจะไม่สามารถผ่านออกทางจมูกได้ จึงต้องออกทางปาก และในกรณีที่ลมออกทางปาก ก็ยังมีด่านหรือประตูน้ำอีกแห่งหนึ่ง คือริมฝีปาก



รูปที่ 2.1 แผนภาพแสดงทางเดินของกระแสลมที่ออกจากปอด [4]

2. ลักษณะที่แตกต่างกันของเสียง

เอกพล อนุสุเรนทร์ [5] รายงานว่าเสียงที่มนุษย์ใช้ในภาษาพูดนั้นจะมีลักษณะที่แตกต่างกันมากบ้างน้อยบ้าง อย่างไรก็ตามจะมีลักษณะสำคัญบางประการ ซึ่งปรากฏในเสียงพูดหลายเสียง ซึ่งอาจเรียกได้ว่าเป็นลักษณะร่วมของเสียงพูด ลักษณะร่วมของเสียงพูดมีหลายประเภท ดังนี้

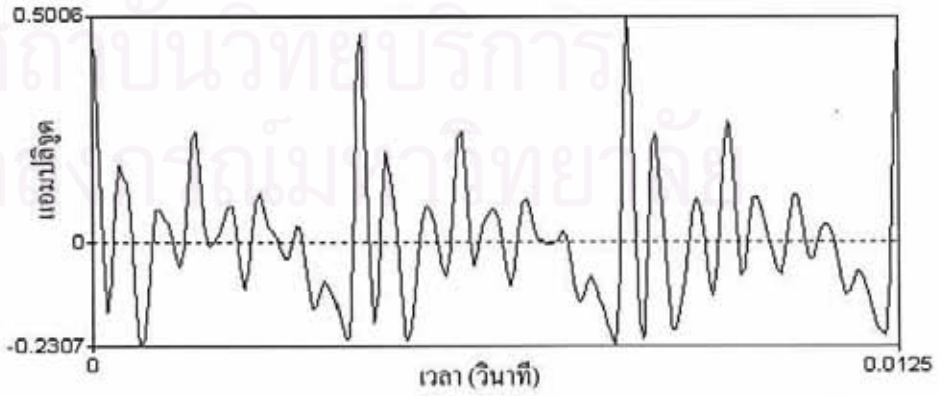
1. **ความก้องหรือไม่ก้องของเสียง** (Voice or Voiceless)

ความก้องหรือไม่ก้องของเสียง เกิดจากเส้นเสียงที่มีลักษณะเป็นแผ่นกล้ามเนื้อ 2 แผ่นวางขวางอยู่ตรงกล่องเสียงตรงปลายท่อหลอดลม ในเวลาที่ไม่พูดเส้นเสียงจะอยู่ห่างจากกันเปิดช่องระหว่างเส้นเสียงหรือช่องคอหอย (Glottis) ให้ลมหายใจผ่านเข้าออกได้โดยสะดวก เรียกเส้นเสียงในตำแหน่งนี้ว่า เส้นเสียงเปิด ในขณะที่ออกเสียงพูด

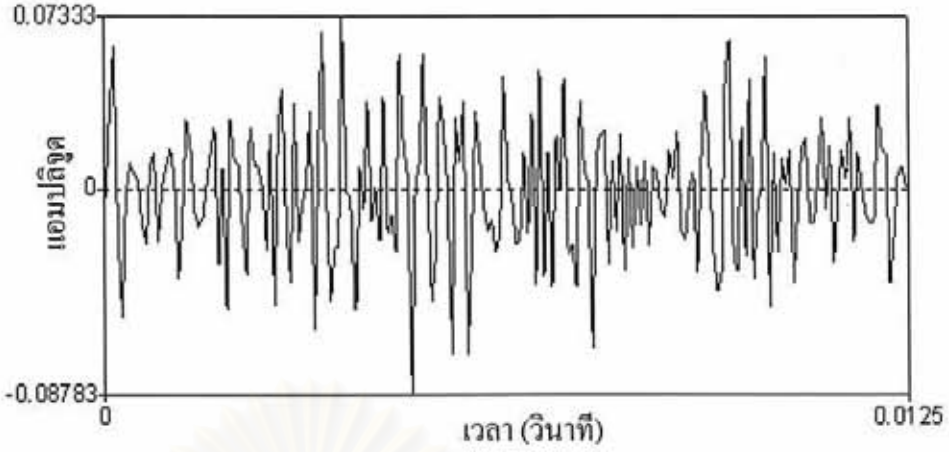
บางครั้งเส้นเสียงอาจเปิดกว้างกว่าเวลาหายใจตามปกติก็ได้ สิ่งที่เป็นตัวกำหนดว่าเสียงนั้นเป็นเสียงก้องหรือเสียงไม่ก้องก็คือ การสั่นของเส้นเสียง การสั่นของเส้นเสียงเกิดจากความตึงของกล้ามเนื้อที่เส้นเสียง ถ้ากล้ามเนื้อที่เส้นเสียงหย่อนจะทำให้เส้นเสียงเกิดการสั่น และถ้ากล้ามเนื้อที่เส้นเสียงตึงจะไม่สามารถทำให้เส้นเสียงนั้นสั่นได้ การสั่นของเส้นเสียงเริ่มต้นเมื่อเส้นเสียงถูกดึงให้ปิดเข้าหากัน แต่เมื่อกระแสลมจากปอดดันขึ้นมาอย่างรุนแรง ก็จะทำให้เส้นเสียงเปิดออกในลักษณะพลิ้วไหวและสั่น แรงดันของกระแสลมจากปอดที่ทำให้เส้นเสียงเปิดและสั่นนี้ ไม่ได้เกิดขึ้นครั้งเดียว เพราะเมื่อเส้นเสียงเปิดออกแล้ว ก็จะพยายามเข้ามาชิดกันดังเดิมและกระแสลมจากปอดก็จะดันให้เส้นเสียงเปิดออกอีก ถ้ายังไม่หยุดออกเสียง จะทำให้เส้นเสียงสั่นตลอดเวลา ลักษณะปิดๆ เปิดๆ นี้จะเป็นไปอย่างรวดเร็วติดต่อกันเหมือนการกระพริบตา

ในการออกเสียงพูดบางเสียงที่ทำให้เส้นเสียงเกิดการสั่น จะเรียกว่าเสียงก้อง (Voiced Sounds) เช่น การออกเสียงพยัญชนะ ม น ง หรือ การออกเสียงสระ และในทางตรงข้าม เสียงพูดหลายเสียงซึ่งเกิดขึ้นในลักษณะที่ไม่มีการสั่นของเส้นเสียง ซึ่งเรียกเสียงเหล่านี้ว่าเสียงไม่ก้อง (Voiceless Sounds) เช่นการออกเสียงพยัญชนะ พ ท ป ส ในภาษาไทย เป็นต้น

ลักษณะของเสียงก้องและเสียงไม่ก้องนี้สามารถสังเกตได้จากรูปคลื่นเสียง ถ้าเป็นเสียงก้อง คลื่นเสียงจะมีลักษณะเป็นคลื่นเสียงแบบรายคาบ ซึ่งจะเห็นรูปคลื่นที่มีรูปร่างคล้ายๆ กันเกิดซ้ำกัน ดังรูปที่ 2.2 ในกรณีของเสียงไม่ก้องหรือเสียงพยัญชนะส่วนมากในภาษาไทย คลื่นเสียงที่ได้จะมีรูปแบบไม่ซ้ำเดิม และมีรูปร่างไม่เป็นระเบียบ ดังรูปที่ 2.3



รูปที่ 2.2 ลักษณะคลื่นเสียงของเสียงก้อง



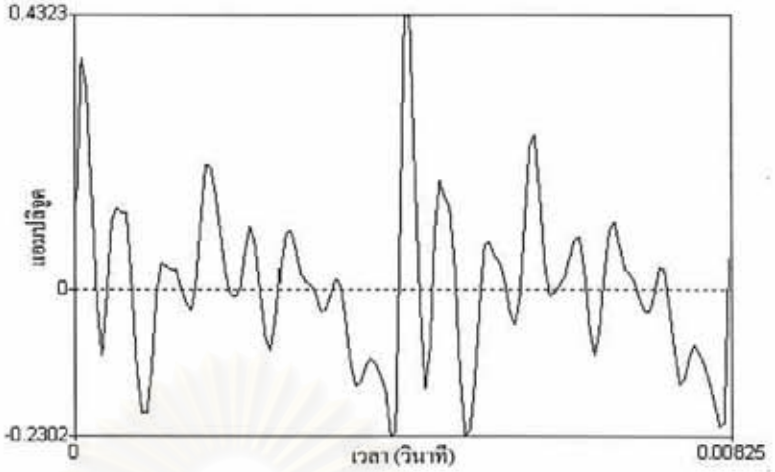
รูปที่ 2.3 ลักษณะคลื่นเสียงของเสียงไม่ก้อง

2. **ความยาวของเสียง** (Duration)

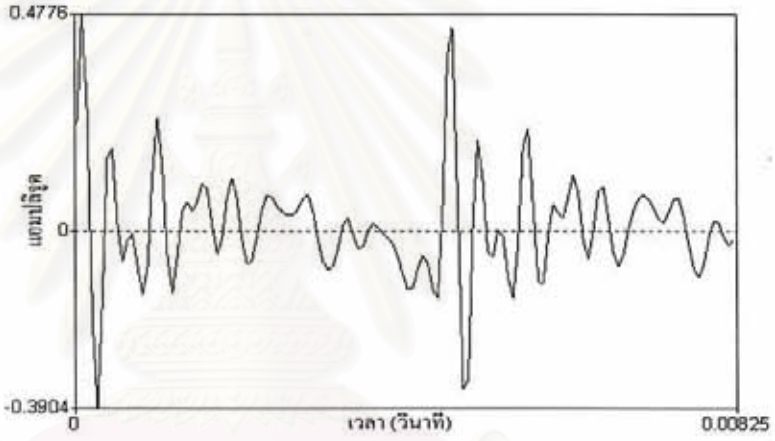
เสียงพูดบางเสียงอาจจะเปล่งออกมาติดต่อกันได้นาน เช่นเสียงสระ เสียงพยัญชนะนาสิก เสียงพยัญชนะเสียดแทรก เป็นต้น แต่บางเสียงไม่สามารถจะเปล่งได้นาน เช่น เสียงพยัญชนะระเบิดไม่ก้อง เป็นต้น ในภาษาไทยนั้นความสั้นยาวของเสียงจะแบ่งเป็น 2 ขนาดคือ สั้นและยาว ในภาษาไทยเสียงพูดซึ่งจะต้องพูดถึงความสั้นยาวก็มีเพียงเสียงสระเท่านั้น ความสั้นยาวของเสียงสระนั้นมีความสำคัญมากในภาษาไทย เพราะจะเป็นสิ่งที่แยกความหมายของคำต่างๆออกจากกันได้ เช่น คำว่า "ขุด" กับคำว่า "ขูด" จะมีความยาวในการออกเสียงแตกต่างกัน โดยคำว่า "ขูด" จะมีการลากเสียงให้ยาวกว่าคำว่า "ขุด" ทำให้คำสองคำนี้มีความหมายแตกต่างกัน

ถ้าดูจากคลื่นเสียง คลื่นเสียงที่มีความยาวมากจะมีคาบเวลาในการเกิดนานกว่าคลื่นเสียงที่มีความยาวน้อย ตัวอย่างที่เห็นได้ชัดคือ สระเสียงสั้นและสระเสียงยาว ซึ่งสระเสียงยาวจะมีคาบเวลาในการเกิดนานกว่าสระเสียงสั้น ดังแสดงในรูปที่ 2.4 และ รูปที่ 2.5

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.4 คลื่นเสียงของสระอะ



รูปที่ 2.5 คลื่นเสียงของสระอา

3. **ระดับเสียงสูงต่ำ** (Pitch)

เสียงจะมีระดับเสียงสูงต่ำอยู่ที่ความถี่เสียง ถ้าความถี่ต่ำเสียงที่ออกมาจะเป็นเสียงต่ำ ความถี่ของเสียงหมายถึง อัตราการสั่นสะเทือนของอุปกรณ์กำเนิดเสียง ในกรณีของเสียงพูด อวัยวะส่วนที่ทำให้เสียงมีระดับสูงหรือต่ำคือเส้นเสียง การที่เส้นเสียงเกิดการสั่นจะทำให้เกิดคลื่นเสียงแบบเป็นรายคาบ ซึ่งระดับเสียงสูงต่ำนี้ ขึ้นอยู่กับความถี่ของคลื่นเสียงที่เกิดจากการสั่นของเส้นเสียง ค่าความถี่ของเส้นเสียงนี้สามารถวิเคราะห์หาค่าออกมาได้โดยวิธีการทางคณิตศาสตร์ หรือการหาค่าอัตสหสัมพันธ์นั่นเอง ซึ่งผลของระยะห่างจากจุดเริ่มต้นของสัญญาณ ไปยังตำแหน่งที่มีค่าอัตสหสัมพันธ์สูงสุด จะเป็นค่าคาบเวลาพิทช์ และเมื่อนำค่าคาบเวลาพิทช์ที่ได้มาคำนวณกับความถี่ที่ใช้ในการสุ่มตัวอย่าง ผลที่ได้จะเป็นระดับเสียงสูงต่ำ หรือถ้าดูจากรูปคลื่นจะสังเกตได้จากคลื่น

เสียงแบบเป็นรายคาบที่เกิดจากซ้ำกันของเสียงสระ เสียงใดที่มีคาบเวลาของการครบรอบน้อยเสียงนั้นจะมีระดับเสียงสูง และตรงกันข้ามเสียงใดที่มีคาบเวลาของการครบรอบมากเสียงนั้นจะมีระดับต่ำ

4. **น้ำหนักของเสียง** (Stress)

การลงน้ำหนัก หมายถึง การออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นที่อยู่เคียงข้าง การลงน้ำหนักพยางค์จึงขึ้นอยู่กับกำลังแรงที่ใช้ในการเปล่งเสียงของพยางค์แต่ละพยางค์ ดังนั้น สามารถกล่าวได้ว่าการลงน้ำหนักคือการเพิ่มหรือการลดค่าแอมปลิจูดในคลื่นเสียง เพราะถ้าใช้แรงลมในการเปล่งเสียงมากจะทำให้แอมปลิจูดของคลื่นเสียงมีค่ามาก ถ้าลงน้ำหนักเบาหรือไม่ค่อยเน้นเสียงซึ่งใช้แรงลมน้อย จะทำให้คลื่นเสียงที่ได้มีค่าแอมปลิจูดไม่มาก

5. **ความดัง** (Loudness)

ความดัง คือ ค่าแอมปลิจูดของคลื่นเสียง ค่าแอมปลิจูดของคลื่นเสียงที่ได้ขึ้นอยู่กับแรงลมที่เปล่งออกมา ซึ่งจะขึ้นอยู่กับปริมาณของลมที่ผู้พูดเปล่งออกมาในช่วงหนึ่งๆ รวมกับลักษณะประจำตัวของเสียงและการลงน้ำหนักเสียง ดังนั้น พยางค์ที่ลงเสียงหนักจึงมีค่าความดังมากกว่าพยางค์ที่ลงเสียงเบา โดยทั่วไปคลื่นเสียงช่วงที่มีแอมปลิจูดมากจะมีความดังมาก และในทางตรงกันข้ามคลื่นเสียงช่วงที่มีแอมปลิจูดน้อยจะมีความดังน้อย และเนื่องจากเสียงมีการเปลี่ยนแปลงไปตามเวลาและการเน้นเสียง ดังนั้นค่าแอมปลิจูดของคลื่นเสียงจะมีค่าไม่เท่ากัน และจะเปลี่ยนแปลงไปตลอดเสียง

3. เสียงพยัญชนะ

เสียงพยัญชนะ เป็นเสียงที่เกิดจากลมซึ่งผ่านเส้นเสียง แล้วมาถูกดัดแปลงด้วยอวัยวะออกเสียงส่วนต่าง ๆ ในปาก ทำให้เกิดเสียงขึ้น เสียงพยัญชนะมีหลายประเภทมีลักษณะการออกเสียงแตกต่างกันหลายแบบ ความแตกต่างของเสียงจะมีความสำคัญยิ่งขึ้นเมื่อความแตกต่างนั้นทำให้ความหมายของคำในภาษาต่างกัน หรือเรียกได้ว่า เป็นความแตกต่างที่ทำให้เส้นเสียงสองเสียงทำหน้าที่เป็นสองหน่วยเสียงที่ต่างกันในภาษานั้น หน่วยเสียงพยัญชนะในภาษาไทยมี 21 หน่วยเสียง

หน่วยเสียงพยัญชนะในภาษาไทย สามารถปรากฏในตำแหน่งต่างๆได้ 4 ตำแหน่ง คือ

1. เกิดต้นคำ นำหน้าเสียงสระในพยางค์หนึ่งๆ เป็นหน่วยเสียงพยัญชนะต้น
2. เกิดนำเสียงพยัญชนะอื่นอีกเสียงหนึ่งในตำแหน่งต้นคำเรียกว่า พยัญชนะควบ
3. เกิดตามเสียงพยัญชนะอื่นในต้นคำ คือเกิดควบกับหน่วยเสียงพยัญชนะชนิดที่ 2
4. เกิดตามหลังสียงสระ เป็นเสียงพยัญชนะสะกด

ตารางที่ 2.1 หน่วยเสียงพยัญชนะในภาษาไทย

ลักษณะของเสียง		ริมฝีปาก	ปุ่มเหงือก	เพดานแข็ง	เพดานอ่อน	เส้นเสียง
พยัญชนะกัก	ไม่พ่นลม	ป *	ต ฏ *	จ	ก *	
	พ่นลม	ผ พ ภ	ฐ ท ฒ ถ ธ	ฉ ช ฌ	ข ค ฆ	
	ก้อง	บ	ฎ ด			
พยัญชนะไม่กัก	นาสิก	ม *	น ณ *		ง *	
	เสียดแทรก	ฝ ฟ	ซ ศ ษ ส			ห ฮ
	กระทบ		ร ฤ			
	ข้างลิ้น		ล ฬ			
	กึ่งสระ	ว *		ย *		

หมายเหตุ หน่วยเสียงพยัญชนะที่มีเครื่องหมาย * คือ หน่วยเสียงที่สามารถปรากฏที่ตำแหน่งท้ายพยางค์ได้

ตารางที่ 2.2 แสดงสัทอักษรสากล ของหน่วยเสียงพยัญชนะต้นในภาษาไทย และตารางที่ 2.3 แสดงสัทอักษรสากล ของหน่วยเสียงพยัญชนะท้ายในภาษาไทย

ตารางที่ 2.2 แสดงสัทอักษรสากล ของหน่วยเสียงพยัญชนะต้นในภาษาไทย

พยัญชนะต้นในภาษาไทย	สัทอักษรสากล	ตัวอักษร ASCII
ป	p	p
ต ฏ	t	t
จ จร	c	c
ก	k	k
อ	ʔ	#
พ ภ ผ	p^h	ph
ท ธ ฒ ฑ ถ ฐ	t^h	th
ช ฌ ฉ	c^h	ch
ข ค ฆ	k^h	kh
บ	b	b
ด ฎ ฑ	d	d
ม หม	m	m
น ณ หน	n	n

พยัญชนะต้นในภาษาไทย	สัทอักษรสากล	ตัวอักษร ASCII
ง หง	ŋ	ng
ฝ ฟ	f	f
ส ศ ษ ซ ทร สร ศร	s	s
ห ฮ	h	h
ร หร ฤ	r	r
ล ฬ หล ฦ	l	l
ว หว	w	w
ย ญ หย หญ อย	j	j

ตารางที่ 2.3 แสดงสัทอักษรสากล ของหน่วยเสียงพยัญชนะท้ายในภาษาไทย

พยัญชนะท้ายในภาษาไทย	สัทอักษรสากล	ตัวอักษร ASCII
แม่กก (ก ข ค ฆ)	k*	k
แม่กด (จ ช ซ ฎ ฏ ฑ ฒ ด ต ถ ท ธ ศ ษ ส)	t*	t
แม่กบ (บ ป ภ พ ฟ)	p*	p
แม่กน (น ณ ญ ร ล ฬ)	n	n
แม่กง (ง)	ŋ	ng
แม่กม (ม)	m	m
แม่เกอย (ย)	j	j
แม่เกอว (ว)	w	w

4. เสียงวรรณยุกต์

เสียงวรรณยุกต์ หมายถึง ระดับเสียงสูงต่ำในคำ ซึ่งทำให้ความหมายของคำแตกต่างกัน เสียงวรรณยุกต์ เป็นส่วนที่สำคัญของภาษาบางภาษา นอกเหนือไปจากเสียงพยัญชนะและเสียงสระ ไม่ว่าจะเป็นภาษาใดก็ตาม เสียงพูดแต่ละเสียงนั้นจะมีระดับเสียงสูงต่ำแตกต่างกัน เช่น ผู้หญิง และเด็กมีพื้นฐานระดับเสียงค่อนข้างสูงกว่าผู้ชาย และผู้พูดเดียวกันก็อาจใช้ระดับเสียงต่างกันได้

ระดับเสียงพื้นฐานจะแบ่งออกได้เป็น 3 ระดับ คือ ระดับเสียงสูง ระดับเสียงกลาง และระดับเสียงต่ำ และในแต่ละระดับเสียงยังแบ่งออกได้อีก 3 เสียง ตามลักษณะการลงเสียงท้ายคำ หรือหางเสียงของแต่ละคำ ดังนี้

1. หางเสียงเสมอกัน หรือที่เรียกว่าเสียงคงระดับ เช่น หางเสียงของคำว่า กา, กะ
2. หางเสียงลากสูงขึ้น เช่น หางเสียงของคำว่า ก๊า, ก๋า
3. หางเสียงทอดลง เช่น หางเสียงของคำว่า ก้า, ก่า

เสียงสระ (Vowel)

หน่วยเสียงที่สำคัญในภาษาทุกภาษาคือหน่วยเสียงสระ เสียงสระเป็นเสียงที่ถูกเปล่งออกมาทางช่องปากหรือช่องจมูก โดยไม่มีอวัยวะส่วนใดในปากมาเป็นอุปสรรคปิดกั้นทางลมไว้ สามารถทำให้เกิดเป็นเสียงสระเสียงใดเสียงหนึ่งขึ้นได้ ในเสียงพูดปกติ เสียงสระนั้นเกิดจากลมซึ่งผ่านเส้นเสียงในตำแหน่งปิดเกือบสนิท และลมที่ต้องดันตัวออกมาทำให้เส้นเสียงสั่นเกิดเป็นเสียงดังขึ้น เรียกว่า เสียงก้อง เสียงที่เรียกว่าเป็นเสียงสระจะต้องถูกเปล่งออกมาทางปากโดยที่ไม่มีอวัยวะส่วนใดในปากมาปิดกั้นทางลมไว้ แต่อวัยวะในปากนั้น อาจจะอยู่ในรูปแบบท่าทางและตำแหน่งต่างๆ ที่ทำให้โพรงปากมีลักษณะต่างกันได้หลายแบบ ลมที่ผ่านออกมาจึงเกิดเป็นเสียงต่างๆ กัน การห่อริมฝีปากหรือไม่ห่อริมฝีปาก การยกลิ้นส่วนหนึ่งส่วนใด การยกลิ้นสูงต่ำต่างๆ เป็นส่วนสำคัญที่ทำให้เกิดเสียงสระต่างๆ ขึ้น

การแบ่งชนิดของสระ สามารถแบ่งตามความสูงต่ำของลิ้นภายในปาก และรูปร่างของปากในขณะที่เปล่งเสียง รวมทั้งปริมาณของลมที่ผ่านออกไปทางจมูก ซึ่งคุณสมบัติทั้งหมดนี้ เมื่อรวมกัน จะสามารถแยกแยะเสียงสระได้อย่างมากมาย โดยรายละเอียดของแต่ละลักษณะ ซึ่งทำให้เสียงสระทั้งหลายแตกต่างกัน และใช้ในการบ่งชี้ถึงสระใด ๆ นั้น มีดังต่อไปนี้

1. **ส่วนของลิ้นที่ใช้ในการเปล่งเสียง** (Place of Articulation)

   จากการศึกษาภาพถ่ายที่ได้จากการฉายรังสีเอ็กซ์ของช่องปากมนุษย์ ในขณะที่ออกเสียงสระต่าง ๆ นั้น ทำให้เราทราบว่า มีลิ้นหลายส่วนด้วยกันที่ใช้ในการออกเสียงสระ ไม่ว่าจะเป็นลิ้นส่วนหน้า ลิ้นส่วนกลาง หรือลิ้นส่วนหลัง โดยลิ้นส่วนนั้น ๆ จะยกขึ้นใกล้เพดานปาก ในขณะที่ออกเสียงสระหนึ่ง ๆ ซึ่งลิ้นแต่ละส่วนที่ยกขึ้นนั้นจะทำให้เกิดเสียงสระที่แตกต่างกัน จึงสามารถที่จะแบ่งประเภทของเสียงสระตามส่วนของลิ้นที่ยกขึ้นได้ดังนี้ ถ้าลิ้นส่วนหน้าถูกยกขึ้น ให้จุดสูงสุดอยู่ใกล้เพดานแข็ง เราก็จะเรียกเสียงสระนั้นว่า เสียงสระส่วนเพดานแข็ง หรือสระหน้า (Font Vowel) เช่น สระ อี เอ แอ เป็นต้น แต่ถ้าการออกเสียงสระใดใช้ลิ้นส่วนหลัง โดยทำการยกลิ้นส่วนหลังขึ้นให้จุดสูงสุดอยู่ใกล้เพดานอ่อน จะเรียกเสียงสระนั้นว่าเป็น เสียงสระส่วนเพดานอ่อน หรือสระหลัง (Back Vowel) เช่น สระ อู โอ ออ เป็นต้น ส่วนถ้าในการออกเสียงสระใด ลิ้นส่วนกลางถูกยกขึ้นไปยังส่วนกลางของเพดานปาก จะเรียกเสียงสระนั้นว่า สระกลาง (Central Vowel) เช่น สระ อือ เออ อา เป็นต้น

2. **ระยะห่างระหว่างลิ้นและเพดานปากหรือความสูงของลิ้น** (Degree of Structure)

ในการเปล่งเสียงสระนั้น ระยะห่างระหว่างลิ้นและเพดานปาก ซึ่งวัดได้จากการที่ลิ้นเคลื่อนที่ลงในแนวดิ่งนั้น จะเป็นลักษณะที่สำคัญอย่างหนึ่งในการแบ่งชนิดของเสียงสระ กล่าวคือ ระยะห่างระหว่างลิ้นกับเพดานปากในตำแหน่งซึ่ง ลิ้นได้ทำการยกตัวขึ้นสูงที่สุดในการเปล่งเสียงสระหนึ่ง ๆ จะเป็นตัวกำหนดว่าเสียงสระนั้น ๆ จะเป็นสระเปิดหรือสระปิด ถ้าหากตำแหน่งของลิ้นอยู่ห่างจากเพดานปากมาก หรือลิ้นอยู่ในระดับต่ำ ทำให้ช่องโพรงในปากเปิดกว้าง ลมก็จะสามารถผ่านออกมาได้ในปริมาณมาก เสียงสระที่ได้จะเป็นเสียงสระเปิด (Open Vowel) เช่น สระอา ในทางตรงกันข้าม ถ้าหากตำแหน่งของลิ้นอยู่ใกล้กับเพดานปากมาก หรือลิ้นอยู่ในระดับสูง ช่องโพรงในปากก็แคบ ทำให้ลมผ่านออกมาได้น้อย จะเรียกสระนั้นว่าเป็นสระปิด (Close Vowel) เช่น สระ อี อือ อู เป็นต้น ซึ่งถ้าระยะห่างระหว่างลิ้นและเพดานปาก หรือความสูงของลิ้นอยู่ในระหว่างสระปิดและสระเปิด ซึ่งจะมีได้หลายระดับ เช่น เสียงสระที่เปิดกว้างกว่าสระปิดเล็กน้อย จะเรียกสระนั้นว่าเป็นสระกลางปิดหรือกึ่งปิด (Close – mid / Half – close Vowel) เช่น สระ เอ เออ โอ เป็นต้น แต่ถ้าเปิดกว้างขึ้นอีก จะเรียกว่าเป็นสระกลางเปิดหรือกึ่งเปิด (Open – mid / Half – open Vowel) เช่น สระ เอ ออ เป็นต้น ทั้งนี้ขึ้นอยู่กับระดับความสูงของลิ้นในปากนั่นเอง

3. **การห่อริมฝีปาก** (Labialization)

การห่อริมฝีปาก หมายถึง การที่ริมฝีปากทั้งสองเคลื่อนไหวโดยยื่นตัวไปข้างหน้าแล้วห่อกลมมากน้อยเพียงใด ถ้าริมฝีปากยื่นออกไปข้างหน้าแล้วห่อกลมมาก เสียงสระที่ได้จะเป็นสระกลม (Round Vowel) เช่น สระ อู โอ ออ เป็นต้น และในทางตรงกันข้าม ถ้าริมฝีปากทั้งสองฉีกออกหรือไม่ห่อกลมในขณะเปล่งเสียง สระที่ได้จะเป็นสระไม่กลม (Unrounded Vowel) เช่น สระ อี แอ อา อือ เออ เป็นต้น

4. **ลักษณะนาสิก** (Nasalization)

เป็นลักษณะในการออกเสียงสระ ที่ทำให้เกิดเสียงสระขึ้นจมูกหรือสระนาสิก (Nasal Vowel) ขึ้น ซึ่งจะทำให้แตกต่างจากสระโอษฐชะ (Oral Vowel) กล่าวคือ ในการเปล่งเสียงสระโอษฐชะนั้น เพดานอ่อนจะยกขึ้นปิดโพรงจมูก อากาศจึงไม่สามารถผ่านออกไปทางจมูกได้ แต่กลับปล่อยออกมาทางปากทั้งหมด ส่วนในการเปล่งเสียงสระนาสิกนั้น เพดานอ่อนจะลดต่ำลง และปล่อยให้อากาศผ่านออกทางช่องจมูกในเวลาเดียวกัน ส่งผลให้ลมส่วนหนึ่งผ่านออกทางช่องจมูกและลมอีกส่วนหนึ่งผ่านออกทางช่องปากไปพร้อม ๆ กันนั่นเอง สำหรับในภาษาไทยนั้น ตามปกติแล้ว ไม่มีการออกเสียงสระนาสิก แต่ในบางครั้ง เมื่อมีการเปล่งเสียงสระโอษฐชะร่วมกับพยัญชนะนาสิก เสียงสระก็อาจจะมีลักษณะนาสิกปนมาด้วย เช่น ในคำว่า นั้น เป็นต้น แต่นั่นก็

เป็นเพราะอิทธิพลของการเปล่งเสียงพยัญชนะนาสิกที่อยู่ใกล้เคียง ไม่ใช่เพราะลักษณะนาสิกที่แท้จริงของเสียงสระ

5. **ความยาวในการออกเสียง** (Duration)

ความสั้นยาวของเสียงสระนั้นมีความสำคัญมากในภาษาไทย เพราะจะเป็นสิ่งที่แยกความหมายของคำต่าง ๆ ออกจากกันได้ เช่น คำว่า ขุด กับคำว่า ขูด จะมีความยาวในการเปล่งเสียงแตกต่างกัน โดยคำว่า ขูด จะมีการลากเสียงสระให้ยาวกว่าคำว่า ขุด ทำให้คำสองคำนี้มีความหมายแตกต่างกัน ในภาษาไทยนั้น ความสั้นยาวของสระ จะแบ่งประเภทของสระออกเป็น 2 ประเภท คือ สระเสียงสั้น (รัสสระ) ได้แก่ สระ อิ อึ อุ เอะ โอะ แอะ อะ เอาะ และสระเสียงยาว (ฑีฆสระ) ได้แก่ สระ อี อือ อู เอ เออ โอ แอ อา และ ออ

หน่วยเสียงสระในภาษาไทยมีทั้งหมด 21 หน่วยเสียง ประกอบด้วย หน่วยเสียงสระเดี่ยวเสียงสั้น 9 หน่วยเสียง หน่วยเสียงสระเดี่ยวเสียงยาว 9 หน่วยเสียง และหน่วยเสียงสระประสม 3 หน่วยเสียง

หน่วยเสียงสระเดี่ยวเสียงสั้น 9 หน่วยเสียงประกอบด้วย หน่วยเสียงสระ อิ เอะ แอะ อึ อะ อุ โอะ เอาะ และหน่วยเสียงสระเดี่ยวเสียงยาว 9 หน่วยเสียงประกอบด้วย หน่วยเสียงสระ อี เอ แอ อือ เออ อา อู โอ และ ออ โดยแต่ละหน่วยเสียงของสระเสียงเดี่ยว มีลักษณะเป็นดังตารางที่ 2.4

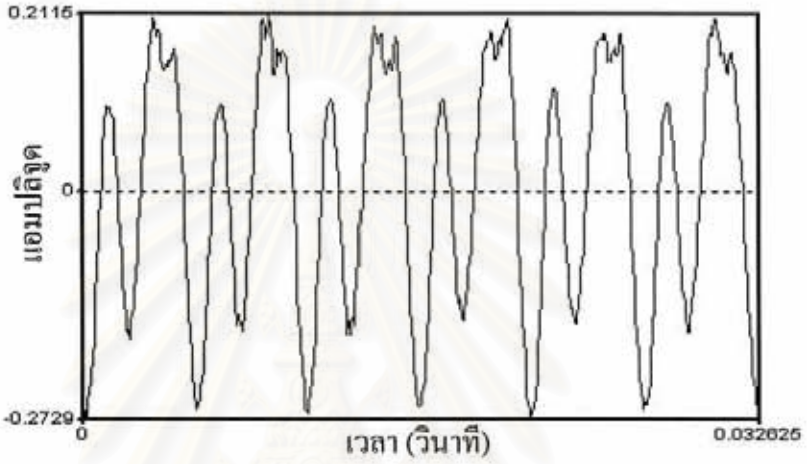
ตารางที่ 2.4 หน่วยเสียงสระเดี่ยว

ลักษณะริมฝีปาก	ริมฝีปากรี		ริมฝีปากห่อ
ลักษณะของลิ้น	หน้า	กลาง	หลัง
สูง	อิ อี	อึ อือ	อุ อู
กลาง	เอะ เอ	เออะ เออ	โอะ โอ
ต่ำ	แอะ แอ	อะ อา	เอาะ ออ

หน่วยเสียงสระประสม 3 หน่วยเสียงประกอบด้วย เสียงย่อยหน่วยละ 2 เสียงเป็นสระเสียงสั้นและสระเสียงยาวคือ สระ เอียะ เอีย เอือะ เอือ อัวะ และ อัว

เสียงสระเป็นเสียงก้องที่ดังกระจายไปได้ไกลกว่าเสียงพยัญชนะ เสียงสระจึงเป็นแกนของพยางค์ซึ่งทำให้เสียงอื่นๆ ในพยางค์นั้นสามารถได้ยินได้ หน่วยเสียงสระในภาษาไทยสามารถเกิดร่วมกับวรรณยุกต์ได้ทุกหน่วย แต่เกิดตามหลังหน่วยเสียงพยัญชนะต้นและนำหน้าหน่วยเสียงพยัญชนะท้าย

เนื่องจากเสียงสระเป็นเสียงก้อง ซึ่งเกิดจากการสั่นของเส้นเสียง ผ่านอวัยวะกล่องเกลาเสียงในช่องทางเดินเสียงและออกไปทางปาก โดยไม่มีอวัยวะอื่นใดมาขวางทางออกของเสียง เมื่อพิจารณารูปคลื่นของเสียงสระจะพบว่าเป็นรูปคลื่นที่เป็นรายคาบ โดยมีคลื่นเสียงที่มีความถี่หลักมูลเท่ากับอัตราที่เส้นเสียงสั่น และเนื่องจากเกิดการกำธรจากช่องทางเดินเสียงคือช่องปากและลำคอส่งผลให้มีค่าความถี่ฮาร์มอนิกส์อื่นๆ ผ่านออกมาได้มากน้อยต่างๆ กัน ปนเข้าไปกับคลื่นเสียงจากเส้นเสียง จึงเกิดเป็นคลื่นเสียงที่ซับซ้อน ดังรูปที่ 2.6



รูปที่ 2.6 คลื่นเสียงของสระอี

1. ทฤษฎีในการสร้างเสียงสระ

ลักษณะหลายๆอย่างที่เกิดขึ้นในเสียงสระ และสามารถที่จะบ่งบอกได้ว่าเสียงสระเหล่านั้นเกิดขึ้นได้อย่างไร โดยที่ลักษณะที่เกิดขึ้นกับเสียงสระเหล่านั้น ประกอบด้วย

1. สัญญาณเสียงที่เป็นสระนั้น ต้องเป็นสัญญาณแบบรายคาบ
2. เนื่องจาก สระเสียงเดียวกันสามารถที่จะออกเสียงได้ในระดับเสียงสูงต่ำที่ต่างกัน (Pitch) และค่าของระดับเสียงสูงต่ำที่เหมือนกันก็สามารถสร้างสระที่แตกต่างกันได้ ด้วยเหตุนี้ ค่าของระดับเสียงสูงต่ำจึงไม่สามารถบอกความแตกต่างของสระเสียงต่างๆ ได้ เช่น การออกเสียงอา อ่า อ้า อ๊า อ๋า จะเห็นว่าเป็นเสียงสระเดียวกันคือสระอา แต่ออกเสียงในระดับที่มีเสียงสูงต่ำต่างกัน
3. อวัยวะที่ทำให้เกิดเสียงต่างๆ มีผลทำให้เสียงสระแตกต่างกันดังนี้ ระดับของลิ้น คือ อยู่สูง กลาง หรือ ต่ำ (High, Mid or Low) ส่วนของลิ้นที่ใช้ในการออกเสียง คือ ใช้ลิ้นส่วนหน้า (Front) ลิ้นส่วนกลาง (Central) หรือลิ้นส่วนหลัง (Back) ลักษณะของริมฝีปาก คือ กลม หรือ ไม่กลม (Rounded or Unrounded) และเสียงสระที่ต่างกันนี้ เมื่อ

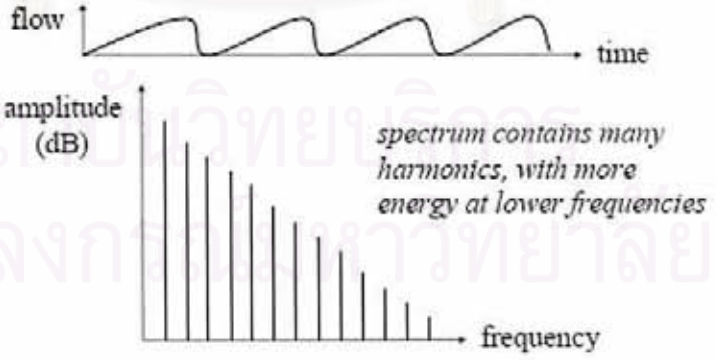
พิจารณาสัญญาณเสียงในทางความถี่จะพบว่า สระที่ต่างกันจะมีแอมปลิจูดของความถี่ต่างๆ ในสเปกตรัม (Spectrum) ที่ต่างกัน

4. ความเกร็งและไม่เกร็งของกล้ามเนื้อลิ้น ก็มีผลทำให้เสียงสระแตกต่างกันดังนี้ ถ้ากล้ามเนื้อลิ้นและกล้ามเนื้อปากเกร็งในขณะเปล่งเสียงสระ เสียงสระนั้นจัดเป็นสระเกร็ง (Tense Vowel) แต่ถ้ากล้ามเนื้อลิ้นและกล้ามเนื้อปากไม่เกร็ง ก็จัดเป็นสระคลาย (Lax Vowel) ส่วนมากสระเกร็งมักจะเป็นสระเสียงยาว ส่วนสระคลายมักจะเป็นสระเสียงสั้น เช่น สระอะ สระอิ สระอุ จะเป็นเสียงสระคลาย และสระอา สระอี สระอู จะเป็นเสียงสระเกร็ง เป็นต้น

2. แบบจำลองแหล่งกำเนิด-ตัวกรอง (Source-Filter Model)

การอธิบายถึงกลไกในการสร้างเสียงพูดจำนวนมากมาย ให้สามารถเข้าใจได้ง่ายขึ้นนั้น มักจะนิยมใช้แบบจำลองที่เรียกว่า แบบจำลองแหล่งกำเนิด-ตัวกรอง ซึ่งแบบจำลองนี้จะพิจารณาว่าส่วนประกอบสองส่วนของแหล่งกำเนิดเสียง (Source) และ ตัวกรองเสียง (Filter) ซึ่งเป็นส่วนที่ทำให้เสียงแตกต่างกันตามรูปร่างของช่องทางเดินเสียง ว่าเป็นส่วนที่แยกออกจากกันและไม่เกี่ยวข้องกัน ทำให้เราสามารถที่จะวัดและบอกปริมาณของสิ่งที่มาจากแหล่งกำเนิดเสียงแยกจากตัวกรองได้

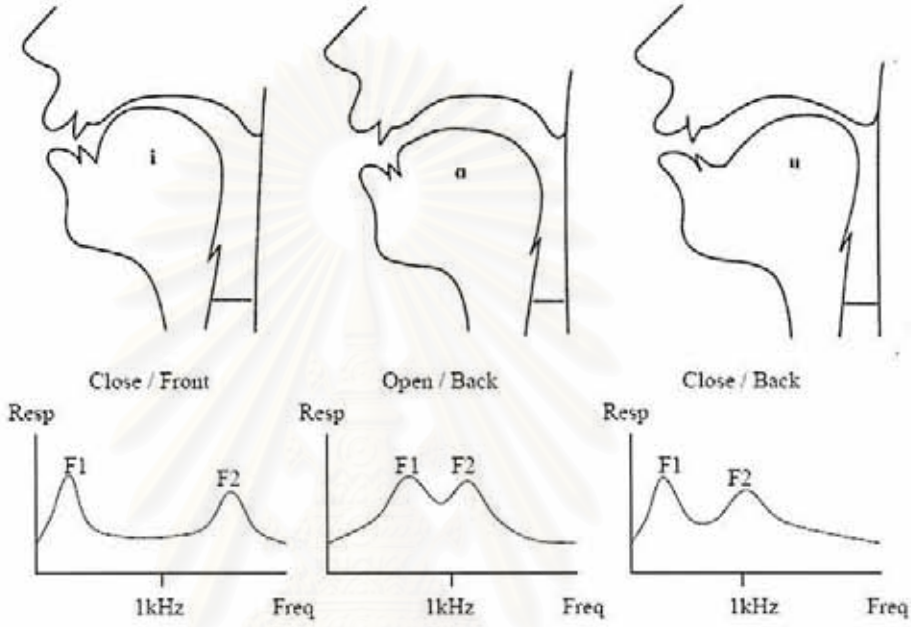
ในการสร้างเสียงสระ ต้นกำเนิดของเสียงสระคืออากาศที่ไหลผ่านเส้นเสียงและทำให้เส้นเสียงเกิดการสั่น ก่อให้เกิดสัญญาณที่มีลักษณะเป็นรายคาบ เมื่อพิจารณาสัญญาณรายคาบนี้ในทางความถี่ จะพบว่าประกอบด้วยความถี่ในช่วงต่างๆ มากมาย และจะมีพลังงานมากในช่วงความถี่ต่ำ ดังรูป 2.7 และความถี่มูลฐานที่เกิดขึ้นจากเสียงสระของผู้ชาย โดยทั่วไปจะอยู่ในช่วง 100-200 เฮิรตซ์ และในผู้หญิงจะอยู่ในช่วง 150-300 เฮิรตซ์



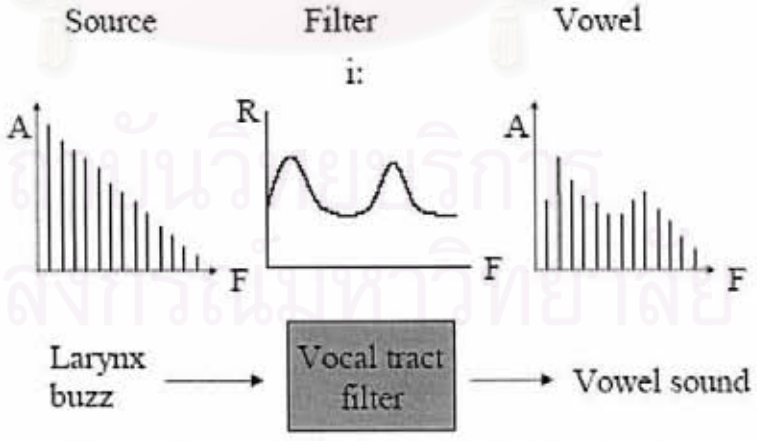
รูปที่ 2.7 สัญญาณรายคาบในทางเวลา และทางความถี่ [6]

ตัวกรองเสียงในการสร้างเสียงสระ คือ รูปร่างของช่องทางเดินเสียงทั้งหมด ที่อยู่ระหว่างเส้นเสียงและริมฝีปาก ผลตอบสนองเชิงความถี่ของตัวกรองนี้ ทำให้เกิดยอดที่สูงเด่นขึ้นมาใน

บริเวณความถี่เรโซแนนท์ (Resonant Peak) ที่เรียกว่า ความถี่ฟอร์แมนท์ (Formant Frequency) การศึกษาถึงความถี่ฟอร์แมนท์ในเสียงสระที่ต่างกัน พบว่าความสัมพันธ์ระหว่างความถี่ของฟอร์แมนท์ที่หนึ่งและสอง (F1, F2) ที่ต่างกันจะทำให้รูปร่างของช่องทางเดินเสียงเปลี่ยนไป ดังรูป 2.8 และแอมปลิจูดของสัญญาณเสียงที่ความถี่ต่างๆ ก็จะเปลี่ยนแปลงไปตามความถี่ฟอร์แมทน์ที่เกิดขึ้นด้วย ดังรูปที่ 2.9

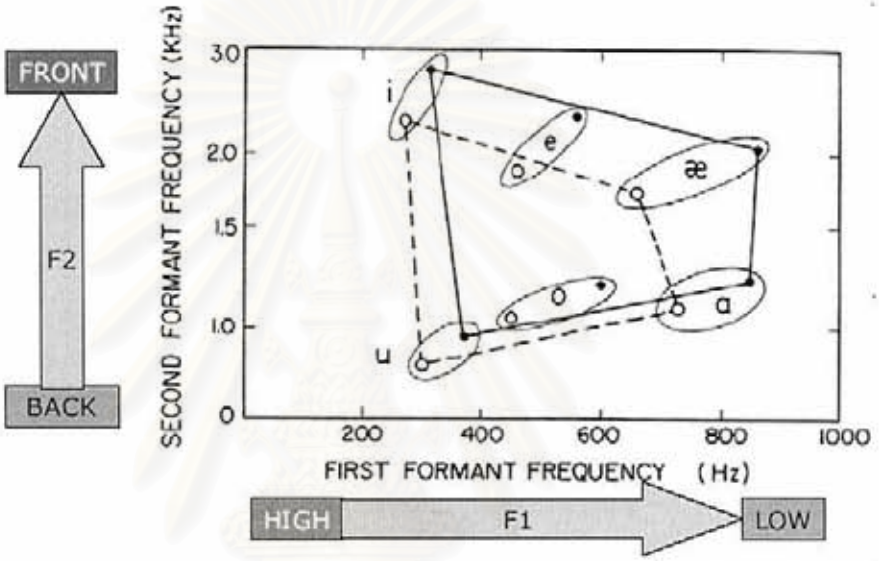


รูปที่ 2.8 รูปร่างของช่องทางเดินเสียงที่สัมพันธ์กับความถี่ฟอร์แมนท์ที่หนึ่งและสอง [6]



รูปที่ 2.9 แสดงแอมปลิจูดของสัญญาณเสียงที่ความถี่ต่างๆ ที่เปลี่ยนแปลงไปตามความถี่ฟอร์แมนท์ที่เกิดขึ้น [6]

ดังที่ได้กล่าวมาแล้วว่า เสียงสระแต่ละตัวจะมีลักษณะเฉพาะตัวของความถี่เสียงที่ต่างกัน ถ้านำสระต่างๆ มาวิเคราะห์หาองค์ประกอบทางความถี่ดู จะพบว่าสระต่างๆ จะมีรูปแบบของความถี่ฟอร์แมนท์ที่ต่างกัน เนื่องจากรูปร่างของช่องทางเดินเสียงนี้จะเป็นตัวกำหนดความถี่ที่จะเกิดเรโซแนนท์ และรูปร่างของช่องทางเดินเสียงนี้จะเปลี่ยนแปลงไปตามแต่ละสระ นั่นคือ จะทำให้ค่าความถี่ฟอร์แมนท์ของสระต่างๆ มีค่าต่างกัน รูปที่ 2.10 แสดงความสัมพันธ์ระหว่างเสียงสระแต่ละเสียงกับความถี่ฟอร์แมนท์อันดับที่ 1 และ ความถี่ฟอร์แมนท์อันดับที่ 2 และความสัมพันธ์ระหว่างเสียงสระแต่ละเสียงกับส่วนของลิ้นที่ใช้ในการออกเสียงสระนั้นๆ



รูปที่ 2.10 แสดงความสัมพันธ์ระหว่างเสียงสระแต่ละตัวกับความถี่ฟอร์แมนท์อันดับที่ 1 และความถี่ฟอร์แมนท์อันดับที่ 2 [7]

ตารางที่ 2.5 แสดงสัทอักษรสากล (International Phonetic Alphabet) ของหน่วยเสียงสระในภาษาไทย

ตารางที่ 2.5 แสดงสัทอักษรสากล ของหน่วยเสียงสระในภาษาไทย

สระในภาษาไทย	สัทอักษรสากล	ตัวอักษร ASCII
อิ	i`	i
อี	i:	ii
เอะ	e`	e
เอ	e:	ee
แอะ	ɛ	x

สระในภาษาไทย	สัทอักษรสากล	ตัวอักษร ASCII
แอ	ɛː	xx
อึ	ɯ`	v
อือ	ɯː	vv
เออะ	ə`	q
เออ	əː	qq
อะ	a`	a
อา	aː	aa
อุ	u`	u
อู	uː	uu
โอะ	o`	o
โอ	oː	oo
เอาะ	ɔ`	@
ออ	ɔː	@@
เอียะ	i`a	ia
เอีย	iaː	iia
เอือะ	ɯ`a	va
เอือ	ɯːa	vva
อัวะ	ua	ua
อัว	uaː	uua
อำ	am	am
ไอ ใอ	aj	aj
เอา	aw	aw

3. ลักษณะที่แตกต่างกันระหว่างเสียงสระและเสียงพยัญชนะ

โดยทั่วไป เมื่อพิจารณาถึงความแตกต่างระหว่างหน่วยเสียงทั้งสองประเภทนั้น จะสังเกตได้ว่า เสียงสระมักจะมีลักษณะที่สำคัญคือ เป็นเสียงก้องทุกหน่วยเสียง และเป็นการเปล่งเสียงที่ลมสามารถผ่านออกมาได้โดยสะดวก การเปล่งเสียงสระนั้นสามารถออกเสียงได้ยาวนานและชัดเจน โดยลักษณะของเสียงสระนั้น จะขึ้นอยู่กับ ลิ้น ริมฝีปาก ปริมาณของลมที่ผ่านออกมา และระยะเวลาในการเปล่งเสียง ในขณะที่เสียงพยัญชนะนั้น มีลักษณะที่สำคัญคือ เป็นได้ทั้งเสียงก้องและเสียงไม่

ก้อง และเป็นการเปล่งเสียงที่ลมผ่านออกมาได้ไม่สะดวก การเปล่งเสียงพยัญชนะสามารถออกเสียงได้เพียงสั้นๆ และไม่ชัดเจน โดยลักษณะของเสียงพยัญชนะนั้น ขึ้นอยู่กับอวัยวะกล่อมเกลาเสียงเป็นปัจจัยสำคัญ

และเมื่อพิจารณาถึงความแตกต่างที่เกิดขึ้นระหว่างสัญญาณของหน่วยเสียงทั้งสองประเภทนั้น ลักษณะที่แตกต่างกันซึ่งสามารถเห็นได้ชัดเจนจากสัญญาณเสียง ประกอบด้วยลักษณะดังต่อไปนี้ คือ

1. ภาวะเป็นรายคาบ (Periodicity)

ภาวะเป็นรายคาบของสัญญาณ คือ ลักษณะของสัญญาณที่ซ้ำกันทุกๆ คาบเวลาของสัญญาณนั้น ซึ่งภาวะเป็นรายคาบของสัญญาณนั้นจะสัมพันธ์กับหน่วยเสียงสองประเภทคือ เสียงก้องและเสียงไม่ก้อง ดังนี้

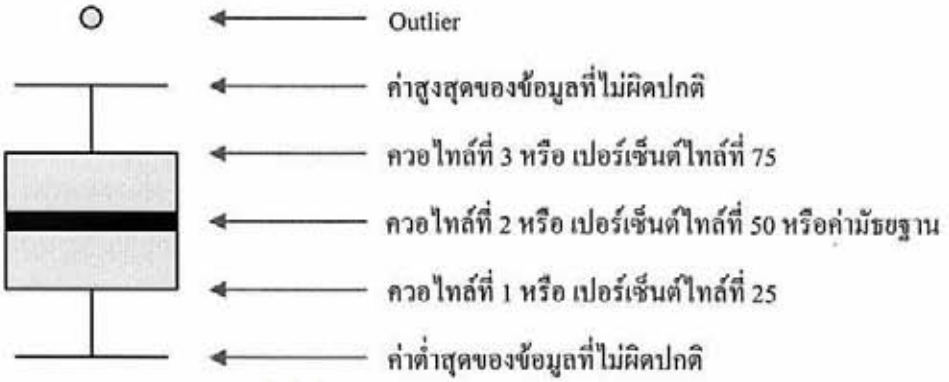
เสียงก้อง ประกอบด้วย หน่วยเสียงสระ หน่วยเสียงพยัญชนะกึ่งสระ หน่วยเสียงพยัญชนะนาสิก หน่วยเสียงพยัญชนะข้างลิ้น หน่วยเสียงพยัญชนะกระทบ และหน่วยเสียงพยัญชนะกักแบบก้อง หน่วยเสียงเหล่านี้จะมีภาวะความเป็นรายคาบสูง กล่าวคือ เมื่อสังเกตลักษณะของรูปคลื่นทางเวลา จะเห็นว่ารูปคลื่นทั้งหมดนั้นมีรูปร่างคล้ายๆ กันเกิดซ้ำกัน

ในทางตรงกันข้าม เสียงไม่ก้อง ประกอบด้วย เสียงพยัญนะกักแบบไม่ก้อง และเสียงพยัญชนะเสียดแทรก หน่วยเสียงเหล่านี้จะมีภาวะความเป็นรายคาบต่ำ กล่าวคือ เมื่อสังเกตลักษณะของรูปคลื่นทางเวลา จะเห็นว่ารูปคลื่นทั้งหมดนั้นมีรูปร่างไม่เป็นระเบียบ

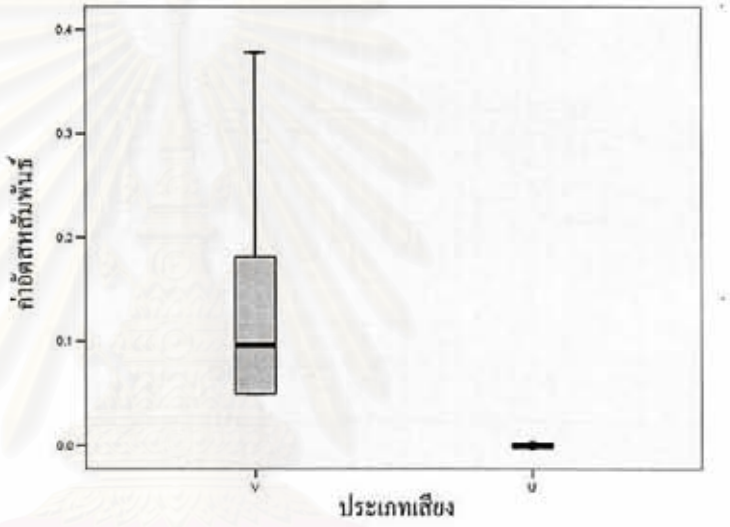
เครื่องมือที่ใช้สำหรับวัดความเป็นรายคาบของสัญญาณ คือ ค่าอัตสหสัมพันธ์ รายละเอียดของวิธีการหาค่าอัตสหสัมพันธ์นั้น จะได้กล่าวต่อไป

แผนภาพกล่อง (Boxplot) เป็นเทคนิคที่ให้รายละเอียดเกี่ยวกับการแจกแจงของข้อมูลค่าสถิติ ได้แก่ ค่ามัธยฐาน เปอร์เซ็นต์ไทล์ที่ 25 และ 75 และให้ค่าข้อมูลที่ผิดปกตินั่นคือ ค่าที่สูงมากหรือต่ำมาก (Outlier) จากค่ากลาง ลักษณะของแผนภาพกล่องเป็นดังรูปที่ 2.11

รูปที่ 2.12 แสดงแผนภาพกล่องของภาวะความเป็นรายคาบของกรอบสัญญาณเสียงก้องและกรอบสัญญาณเสียงไม่ก้อง



รูปที่ 2.11 ลักษณะของแผนภาพกล่อง



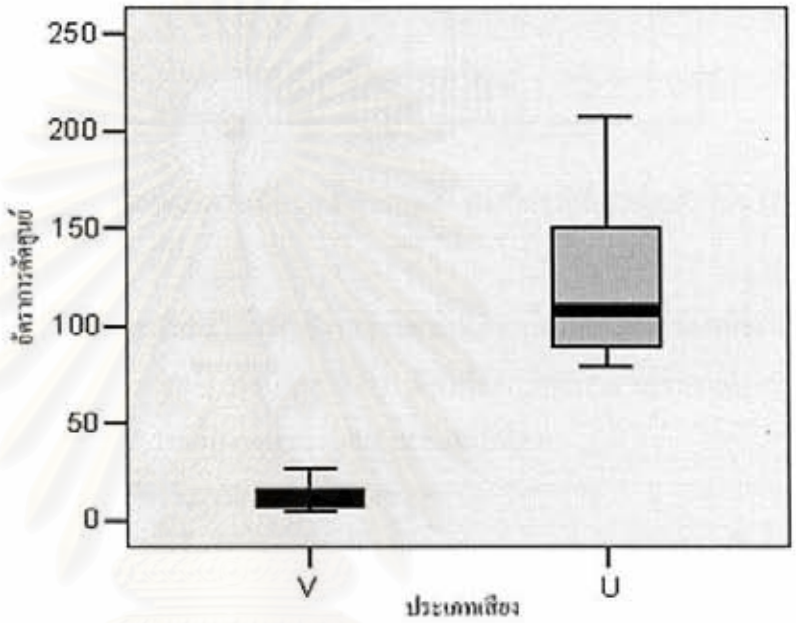
รูปที่ 2.12 แผนภาพกล่องแสดงความเป็นรายคาบของสัญญาณเสียงก้อง (V) และสัญญาณเสียงไม่ก้อง (U)

จากรูปที่ 2.12 จะเห็นได้ว่า ค่าเฉลี่ยภาวะเป็นรายคาบของเสียงไม่ก้อง (U) มีค่าน้อยกว่า ค่าเฉลี่ยภาวะเป็นรายคาบของเสียงก้อง (V)

2. อัตราการตัดศูนย์ (Zero Crossing Rate)

การเกิดจุดตัดศูนย์จะเกิดขึ้นจากการที่สัญญาณมีการตัดกับแกนเวลา นั่นคือ ค่าของสัญญาณจะมีการเปลี่ยนสัญลักษณ์ทางคณิตศาสตร์ (Algebraic Sign) นั่นเอง อัตราการตัดศูนย์ คือ จำนวนจุดตัดศูนย์ในหนึ่งช่วงเวลา และเป็นเครื่องมืออย่างง่ายที่ใช้ในการวิเคราะห์หาความแตกต่างระหว่างสัญญาณเสียง หรืออธิบายการเปลี่ยนแปลงข้อมูลของสัญญาณ โดยทั่วไป อัตราการตัดศูนย์นั้นถูกนำมาใช้ในการตัดสินว่าสัญญาณเสียงนั้นเป็นสัญญาณเสียงก้องหรือสัญญาณเสียงไม่ก้อง เนื่องจากเสียงก้องส่วนใหญ่จะมีความถี่ต่ำ ในขณะที่เสียงไม่ก้องจะมีความสูง และค่าอัตราการตัดศูนย์นี้ก็สัมพันธ์

โดยตรงกับความถี่ของสัญญาณ ดังนั้นจึงสรุปได้ว่า สัญญาณเสียงที่มีค่าอัตราการตัดศูนย์มากจะเป็นเสียงไม่ก้อง และสัญญาณเสียงที่มีค่าอัตราการตัดศูนย์น้อยจะเป็นเสียงก้อง อย่างไรก็ตาม การกำหนดขนาดของค่าอัตราการตัดศูนย์ที่แน่นอนเพื่อจำแนกประเภทของเสียงนั้น จะต้องอาศัยผลจากการทดลองเป็นหลัก รูปที่ 2.13 แสดงแผนภาพกล่องของอัตราการตัดศูนย์ของกรอบสัญญาณเสียงก้อง (V) และกรอบสัญญาณเสียงไม่ก้อง (U)



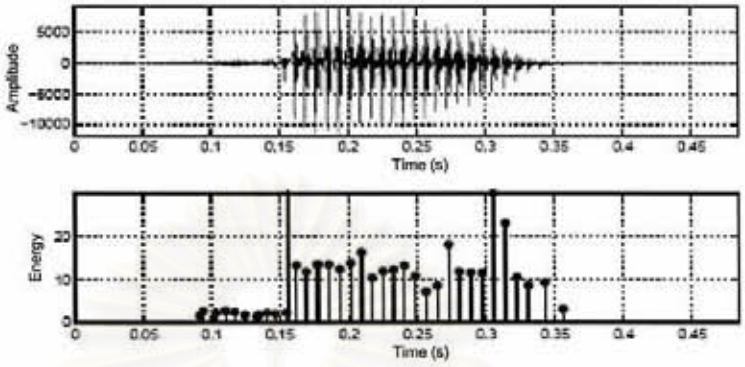
รูปที่ 2.13 แผนภาพกล่องแสดงอัตราการตัดศูนย์ของกรอบสัญญาณเสียงก้อง (V) และสัญญาณเสียงไม่ก้อง (U)

จากรูปที่ 2.13 จะเห็นได้ว่า ค่าเฉลี่ยอัตราการตัดศูนย์ของเสียงไม่ก้อง (U) มีค่ามากกว่า ค่าเฉลี่ยอัตราการตัดศูนย์ของเสียงไม่ก้อง (V)

3. อัตราส่วนของพลังงานที่ความถี่ต่ำและพลังงานที่ความถี่สูง

ค่าพลังงานนั้นจะถูกนำมาใช้ในการวิเคราะห์สัญญาณเสียงโดยทั่วไป เนื่องจากสามารถคำนวณได้ง่าย ในทางทฤษฎีพลังงานที่ความถี่ต่ำของสัญญาณเสียงก้องจะมีค่ามากกว่าพลังงานที่ความถี่ต่ำของสัญญาณเสียงไม่ก้อง และในทางตรงกันข้าม พลังงานที่ความถี่สูงของสัญญาณเสียงก้อง จะมีค่าน้อยกว่าพลังงานที่ความถี่สูงของสัญญาณเสียงไม่ก้อง ดังนั้น อัตราส่วนของพลังงานที่ความถี่ต่ำต่อพลังงานที่ความถี่สูงของสัญญาณเสียงก้อง จึงมีค่ามากกว่าอัตราส่วนของพลังงานที่ความถี่ต่ำต่อพลังงานที่

ความถี่สูงของสัญญาณเสียงไม่ก้อง ซึ่งอัตราส่วนพลังงานสามารถแสดงได้ดังรูปที่ 2.14

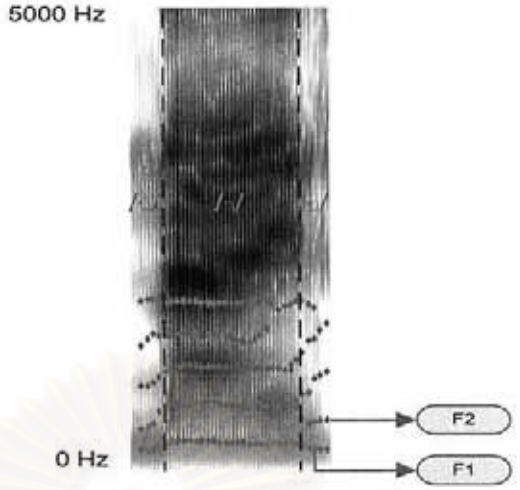


รูปที่ 2.14 อัตราส่วนของพลังงานที่ความถี่ต่ำต่อพลังงานที่ความถี่สูงของคำว่า หา [8]

จากรูปที่ 2.14 จะสังเกตได้ว่า อัตราส่วนของพลังงานที่ความถี่ต่ำต่อพลังงานที่ความถี่สูงของพยัญชนะ /ห/ (ก่อนแถบสีดำ) มีค่าน้อยกว่าอัตราส่วนของพลังงานที่ความถี่ต่ำต่อพลังงานที่ความถี่สูงของสระ /อา/ (หลังแถบสีดำ)

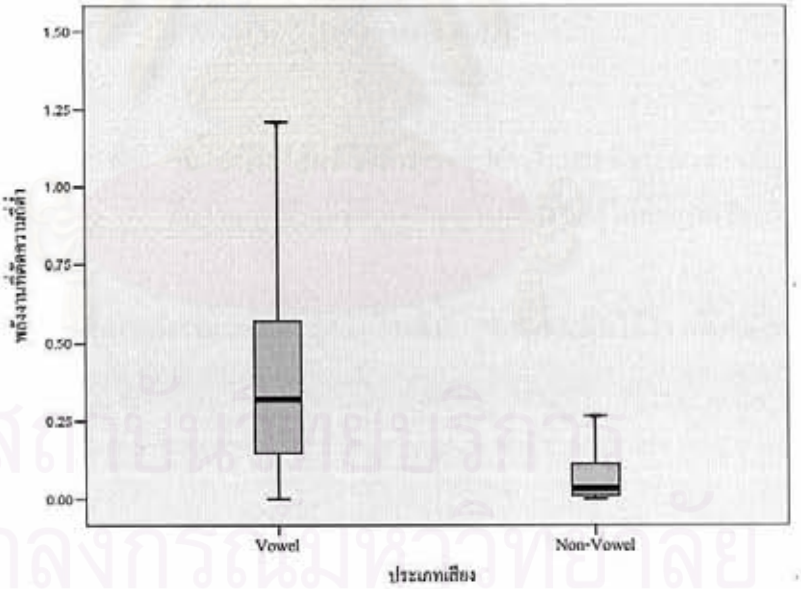
4. พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์

เนื่องจากสัญญาณเสียงก้องมักประกอบด้วยส่วนประกอบของความถี่ต่ำ จึงทำให้ค่าพลังงานที่ความถี่ต่ำของสัญญาณเสียงก้องนั้นมีค่าสูง แต่เมื่อพิจารณาพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ พบว่า เสียงสระ ซึ่งเป็นเสียงก้องประเภทหนึ่ง จะมีค่าพลังงานที่สูงกว่าเสียงพยัญชนะที่เป็นประเภทเสียงก้องเดียวกัน รูปที่ 2.15 แสดงรูปร่างของความถี่ฟอร์แมนท์ของคำว่า นาย ซึ่งสังเกตได้ว่า โครงสร้างความถี่ฟอร์แมนท์ที่หนึ่งและความถี่ฟอร์แมนท์ที่สองของหน่วยเสียงพยัญชนะ /น/ จะเริ่มเปลี่ยนแปลงสู่โครงสร้างความถี่ฟอร์แมนท์ที่หนึ่งและสองที่มีความถี่ที่เพิ่มสูงขึ้นของหน่วยเสียงสระ /อา/ และเปลี่ยนแปลงสู่โครงสร้างความถี่ฟอร์แมนท์ที่หนึ่งและสองซึ่งมีความถี่ต่ำลงอีกครั้งของหน่วยเสียงพยัญชนะ /ย/ ด้วยเหตุนี้ จึงทำให้พลังงานของสัญญาณเสียงสระซึ่งมีส่วนประกอบทางความถี่ต่ำที่สูงกว่า แม้จะมีการตัดส่วนประกอบที่ความถี่ต่ำออกไป ก็ยังคงมีค่าสูงกว่าเสียงพยัญชนะที่เป็นประเภทเสียงก้องเดียวกันซึ่งมีส่วนประกอบทางความถี่ต่ำที่ต่ำกว่าเสียงสระ



รูปที่ 2.15 ความถี่ฟอร์แมนท์ที่หนึ่งและความถี่ฟอร์แมนท์ที่สองของคำว่า นาย

แผนภาพกล่องแสดงค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของกรอบสัญญาณเสียงสระ (V) และกรอบสัญญาณเสียงก้องที่ไม่ใช่เสียงสระ (NV) แสดงได้ดังรูปที่ 2.16



รูปที่ 2.16 แผนภาพกล่องของพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของเสียงสระ (Vowel) และเสียงก้องที่ไม่เสียงสระ (Non-Vowel)

ทฤษฎีที่ใช้ในการหาลักษณะสำคัญของเสียง

ลักษณะสำคัญของเสียงคือ ค่าที่สกัดจากสัญญาณเสียง เพื่อใช้เป็นตัวแทนของเสียงนั้นๆ ลักษณะสำคัญที่ดี จะมีขนาดเล็กและเก็บสารสนเทศที่เพียงพอในการบ่งชี้ และแยกแยะเสียงนั้นๆ ออกจากเสียงอื่นได้

1. ค่าพลังงาน (Energy)

พลังงานของสัญญาณเป็นลักษณะสำคัญอย่างหนึ่ง ที่มักจะถูกใช้ในการวิเคราะห์สัญญาณทั่วๆ ไป โดยพลังงานของสัญญาณ $s[n]$ ใดๆ ที่แปรตามเวลาสามารถนิยามได้ว่า

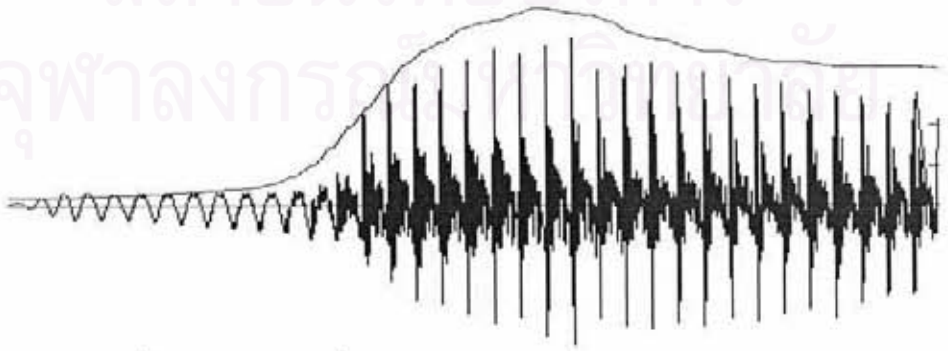
$$E = \sum_{n=-\infty}^{\infty} s^2[n] \tag{2.1}$$

สำหรับสัญญาณเสียง ซึ่งเป็นสัญญาณที่แปรเปลี่ยนอยู่ตลอดเวลา ไม่มีเสถียรภาพตามเวลา เราจะต้องแบ่งสัญญาณออกมาพิจารณาเป็นช่วงเล็กๆ ตามแกนเวลาหรือเรียกว่าแบ่งออกเป็น กรอบ เช่น กรอบละประมาณ 10-30 มิลลิวินาที ดังนั้น สามารถหาพลังงานของเสียงในแต่ละกรอบ ได้เป็น

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

โดยที่ $w(m)$ คือ วินโดว์ฟังก์ชันที่ใช้กำหนดรูปร่างในการพิจารณาของสัญญาณเสียง $s(n)$ ในหนึ่งกรอบ และ N คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ในกรอบหรือภายในกรอบของฟังก์ชันหน้าต่าง

ในรูปที่ 2.17 แสดงพลังงานของสัญญาณเสียงคำว่า "คา" จะเห็นได้ว่า เส้นของพลังงานจะเป็นซองครอบคลื่นเสียง



รูปที่ 2.17 แผนภาพคลื่นเสียงและพลังงานของสัญญาณเสียงคำว่า "นา"

2. ค่าอัตสหสัมพันธ์ (Autocorrelation)

อัตสหสัมพันธ์เป็นเครื่องมือที่ใช้สำหรับวัดความเป็นรายคาบของสัญญาณ หรือลักษณะของสัญญาณที่ซ้ำกันทุกๆ คาบเวลาของสัญญาณนั้น สำหรับสัญญาณเสียงนั้น สัญญาณเสียงซึ่งมีค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่มูลฐาน (Pitch) มากและมีลักษณะเป็นรายคาบ จะเป็นลักษณะของสัญญาณเสียงก้อง และในทางกลับกันสัญญาณเสียงซึ่งมีค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ความถี่มูลฐาน น้อยและไม่เป็นรายคาบ เป็นลักษณะของสัญญาณเสียงไม่ก้อง

ค่าอัตสหสัมพันธ์ สามารถหาได้จากสมการที่ 2.3

$$r_x(\eta;m)=\frac{1}{N}\sum_{n=-\infty}^{\infty}s(n)w(m-n)s(n+|\eta|)w(m-n+|\eta|) \qquad (2.3)$$

เมื่อ $r_x(\eta;m)$ แทนค่าอัตสหสัมพันธ์ที่ตำแหน่ง m เมื่อหน่วงเวลาไปเป็นระยะ η

$s(n)$ แทนสัญญาณเสียงที่ตำแหน่ง n

$w(n)$ แทนฟังก์ชันหน้าต่าง

N แทนความกว้างของฟังก์ชันหน้าต่าง

ในสัญญาณที่เป็นสัญญาณคาบที่มีคาบเท่ากับ P จะทำให้ความสัมพันธ์ตามสมการข้างต้นมีค่ามากเมื่อ $\eta=0,\pm P,\pm 2P,...$

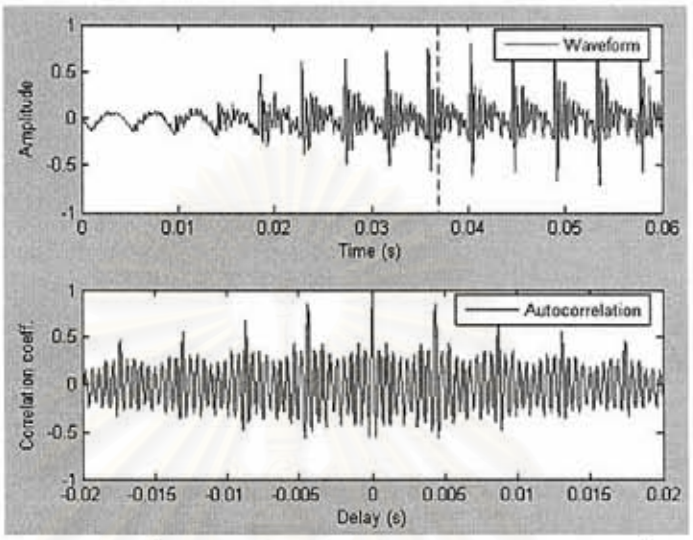
สำหรับเสียงพูดที่เป็นเสียงก้อง ค่าอัตสหสัมพันธ์จะมีค่ามากทุกๆ ตำแหน่งที่เป็นจำนวนเท่าของคาบพิทช์ ดังรูปที่ 2.18

รูปที่ 2.18 ก. แสดงสัญญาณเสียงคำว่า "กา" รูปที่ 2.18 ข. แสดงค่าอัตสหสัมพันธ์ของสัญญาณเสียง ณ จุดตัวอย่างที่ 2055 ค่าอัตสหสัมพันธ์จะมีลักษณะซ้ำเป็นคาบ และระยะห่างระหว่างค่าสูงสุดจะเท่ากัน ซึ่งเท่ากับคาบของสัญญาณเสียง

เพื่อลดความผิดพลาดที่เกิดจากความไม่ต่อเนื่องของข้อมูล นิยมกำจัดสัญญาณที่มีขนาดแอมปลิจูดต่ำซึ่งคาดว่าจะเป็นสัญญาณรบกวนออก ด้วยการขลิบกลาง (Central Clipping) สัญญาณและนำสัญญาณที่ผ่านการขลิบมาคำนวณหาค่าอัตสหสัมพันธ์ โดยวิธีการขลิบกลางมีฟังก์ชันดังสมการต่อไปนี้ [9]

$$C\{s(n)\}=\begin{cases}s(n)-C^{+}, & s(n)>C^{+}\\ 0, & C^{-}\le s(n)\le C^{+}\\ s(n)-C^{-}, & s(n)<C^{-}\end{cases} \qquad (2.4)$$

โดยทั่วไป ค่าของ C+, C- จะมีค่าประมาณ 30% ของค่าสูงสุดของค่าสัมบูรณ์ของสัญญาณ และเมื่อนำสัญญาณมาทำการขลิบกลางก่อนที่จะนำมาหาค่าอัตสหสัมพันธ์ จะทำให้ค่าสูงสุดที่ได้เด่นชัดขึ้น สามารถหาจุดสูงสุดของค่าอัตสหสัมพันธ์ได้แม่นยำขึ้น



รูปที่ 2.18 ค่าอัตสหสัมพันธ์ ของสัญญาณเสียงคำว่า "มา" (maa0) ณ เวลาที่ 0.0375 วินาที (ก) สัญญาณเสียง (ข) ค่าอัตสหสัมพันธ์

วรรณกรรมที่เกี่ยวข้อง

วิธีการในการตรวจหาสระตั้งแต่อดีตจนถึงปัจจุบันมีมากมาย ซึ่งแต่ละวิธีนั้นก็มีการใช้คุณลักษณะทางเสียงและวิธีการที่แตกต่างกัน อย่างไรก็ตาม สามารถที่จะแบ่งประเภทของการตรวจหาสระ ตามวิธีการที่ใช้ในการตรวจหาสระได้เป็น 2 ประเภท คือ

1. การตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากกฎ (Rule-Based Method)
2. การตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากสถิติ (Statistical-Based Method)

โดยที่แต่ละประเภทนั้น มีรายละเอียดและวิธีการที่ใช้ในการตรวจหาสระ ดังต่อไปนี้

1. การตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากกฎ

วิธีการในการตรวจหาสระ ที่จัดได้ว่าอยู่ในประเภทนี้คือ ในการตรวจหาสระนั้นจะมีหลักการหรือวิธีการเป็นของตัวเอง โดยที่หลักการหรือวิธีการของตัวเองนี้ จะขึ้นอยู่กับตัวแปรเสริม (Parameter) ที่จะใช้ในการตรวจหาสระ ซึ่งตัวแปร เหล่านั้นจะมีหลักการของตัวมันเองในการที่จะระบุว่า ตำแหน่งใดของสัญญาณเสียงเป็นสระ ที่แตกต่างกัน

Pellegrino และ Andre-obrecht [10] ใช้อัตราส่วนของพลังงานในย่านความถี่ต่ำเทียบกับพลังงานรวมทั้งหมดและความยาวของหน่วยเสียงที่หาโดย "Forward-backward Divergence Algorithm" ในการตรวจหาสระ พลังงานของแต่ละกรอบนี้ได้มาจากการทำเมลสเกลฟิลเตอร์แบงค์ (Mel-scale Filter Bank) กับสัญญาณเสียงที่ผ่านการกรองความถี่ออกเป็น 2 ย่านความถี่คือ ย่านความถี่ต่ำ (300 เฮิรตซ์ ถึง 1,000 เฮิรตซ์) และย่านความถี่สูง (1,000 เฮิรตซ์ ถึง 3200 เฮิรตซ์) เพื่อนำผลลัพธ์ที่ได้ไปคำนวณหาพลังงานทั้งในย่านความถี่ต่ำ และย่านความถี่สูง และใช้อัตราส่วนของพลังงานในย่านความถี่ต่ำเทียบกับพลังงานรวมทั้งหมด (ทั้งความถี่ต่ำและความถี่สูง) ทั้งอัตราส่วนของพลังงานและความยาวของหน่วยเสียง จะนำไปเปรียบเทียบกับค่าขีดแบ่ง (Threshold) เพื่อที่จะดูว่าตำแหน่งนั้นของสัญญาณเสียงเหมาะสมที่จะเป็นสระหรือไม่ ฐานข้อมูลเสียงที่ใช้ในการทดสอบครั้งนี้คือ OGI_TS [11] ผลการตรวจหาสระได้ความถูกต้อง 75% ในงานวิจัยนี้ได้มีการเลือกใช้ ความยาวของหน่วยเสียงเข้ามาช่วยในการตัดสินใจ ทำให้หน่วยเสียงที่มีความยาวสั้นๆ เช่น เสียงระเบิด ไม่มีผลกระทบต่อขั้นตอนในการตรวจหาตำแหน่งของสระ อย่างไรก็ตาม ความผิดพลาดที่เกิดขึ้นโดยส่วนใหญ่ เกิดจากตำแหน่งของ เสียงเสียดแทรก และตำแหน่งของเสียงก้องอื่นๆ เช่น เสียงนาสิก หรือ เสียงกึ่งสระ เป็นต้น

Pfau และ Ruske [12] ใช้ยอด (Peak) ของความดังที่ดัดแปร (Modified Loudness) ในการตรวจหาสระ โดยที่ความดังที่ดัดแปรนี้สามารถคำนวณได้จาก ผลต่างของฟังก์ชันความดัง (Loudness Function) ที่ความถี่ต่ำและความถี่สูง ผลต่างของฟังก์ชันความดังในงานวิจัยนี้คำนวณมาจากผลรวมของความดัง จากแต่ละช่วงความถี่ในสเกลบาร์ก (Bark Scale) โดยที่ค่าของความดังที่ความถี่ต่ำเกิดจากผลรวมของความดังที่ ช่วงความถี่ที่ 3 ถึง 15 และค่าความดังที่ความถี่สูงเกิดจากผลรวมของความดังที่ ช่วงความถี่ที่ 20 ถึง 22 และใช้ขั้นตอนวิธีการหายอด (Peak Picking Algorithm) ในการตรวจหายอดจากผลต่างของความดังที่ดัดแปรที่ความถี่ต่ำและความถี่สูงเพื่อนำไปเทียบกับค่าขีดแบ่ง ที่กำหนดขึ้นจากชุดข้อมูลฝึก (Training Data) นอกจากนี้ ยังได้ใช้อัตราการตัดศูนย์ (Zero Crossing Rate, ZCR) เป็นตัวแปรเสริมอีกตัวหนึ่งเพื่อป้องกันเสียงเสียดแทรก ฐานข้อมูลเสียงที่ใช้ในการทดสอบครั้งนี้คือ German Verbmobil ประกอบด้วย ชุดข้อมูลฝึก 50% และข้อมูลทดสอบ (Test Data) 50% หรือ ชุดข้อมูลฝึก 694 ประโยค (สระ17,040 ตัว) และ ชุดข้อมูลทดสอบ 693 ประโยค (สระ 18,539 ตัว) และได้ความถูกต้อง 77%

Zhimin Xie ตรวจหาสระโดยการใช้ขั้นตอนวิธีการคอนเวกซ์ฮัลล์ (Conven Hull) ของ Mermelstein [13] ซึ่งลักษณะทางเสียงที่ใช้ในงานวิจัยนี้คือ ความเป็นรายคาบ (Periodic) ของแต่ละกรอบสัญญาณ และพลังงานที่ความถี่ต่ำ ฐานข้อมูลเสียงที่ใช้ในการทดสอบครั้งนี้คือ TIMIT ซึ่งเป็น

เสียงอ่านที่บันทึกในสภาวะเงียบที่ประกอบด้วย ชุดข้อมูลฝึกจำนวน 253 ประโยค และชุดข้อมูลทดสอบจำนวน 152 ประโยค ได้ความถูกต้อง 86%

Howitt ได้ปรับปรุงขั้นตอนวิธีการ (Algorithm) ของ Mermelstein ใหม่เพื่อที่จะหายอดของความเข้ม (Intensity Peak) และความเข้มที่ตก (Intensity Dip) ในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์ โดยที่นำค่าความเข้มที่ตกนี้ไปเปรียบเทียบกับค่าขีดแบ่ง ซึ่งมีขนาด 2 เดซิเบล ถ้าค่าความเข้มที่ตกมีขนาดต่ำกว่าค่าขีดแบ่ง ที่จุดนั้นจะถือว่าเป็นขอบ (Boundary) และยอดที่ได้จะถือว่าเป็นตำแหน่งของสระ ถ้าอัตราการตัดศูนย์และความยาวของหน่วยเสียงมีค่ามากกว่าค่าขีดแบ่ง ฐานข้อมูลเสียงที่ใช้ในการทดสอบครั้งนี้คือ TIMIT ซึ่งเป็นเสียงอ่านที่บันทึกในสภาวะเงียบประกอบด้วย ชุดข้อมูลฝึก 619 ประโยค (7,585 สระ) และ ชุดข้อมูลทดสอบ 373 ประโยค (4,404 สระ) และได้ผลรวมของความผิดพลาดเท่ากับ 15% งานวิจัยนี้ ใช้ความเข้มที่ตกในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์ เป็นลักษณะทางเสียงที่ใช้ในการตรวจหาสระ ซึ่งความถี่ฟอแมนท์ของสระจะตกในช่วงนี้ ทำให้ตำแหน่งของการตรวจหาสระจากค่าดังกล่าวที่ได้ ไม่เป็นตำแหน่งของเสียงเสียดแทรก แต่ความผิดพลาดที่เกิดขึ้น เกิดจากตำแหน่งของเสียงก้องอื่นๆ เช่น เสียงกึ่งสระ หรือ เสียงนาสิก เป็นต้น เมื่อพิจารณางานวิจัยนี้กับงานวิจัยของ Zhimin Xie พบว่า งานวิจัยทั้งสองมีการทดสอบประสิทธิภาพของการตรวจหาสระโดยใช้ฐานข้อมูลเสียงเดียวกัน คือ ฐานข้อมูลเสียง TIMIT แต่งานวิจัยนี้ใช้ข้อมูลเสียงในการทดสอบมากกว่าเป็นครึ่งหนึ่งเมื่อเปรียบเทียบกับงานวิจัยของ Zhimin Xie ซึ่งได้ความถูกต้องใกล้เคียงกัน กล่าวคือ ในงานวิจัยของ Xie ได้ความถูกต้อง 86% และงานวิจัยนี้ ได้ความถูกต้อง 85% ด้วยเหตุผลดังกล่าว งานวิจัยนี้ จึงดีกว่างานวิจัยของ Xie

รายละเอียดโดยสรุปของแต่ละกระบวนวิธี ของการตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากกฎ สามารถแสดงได้ดังตารางที่ 2.6

ตารางที่ 2.6 วรรณกรรมที่เกี่ยวข้องกับการตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากกฎ

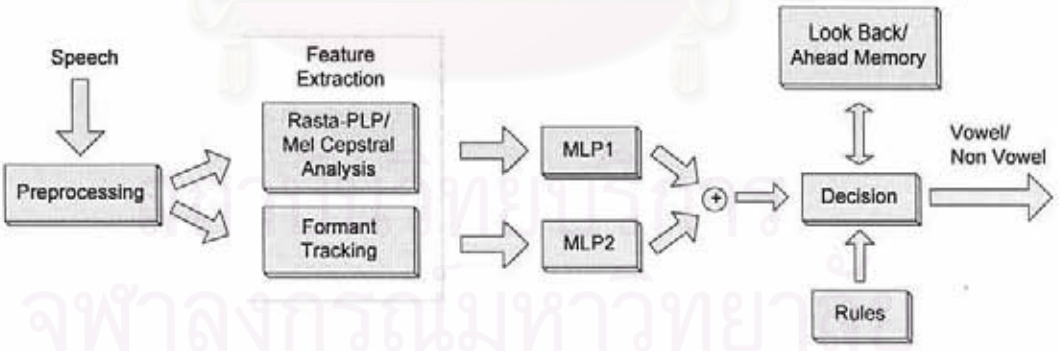
	ลักษณะทางเสียง	การตรวจหาสระ	ฐานข้อมูลเสียง	ความถูกต้อง
Pellegrino	1. อัตราส่วนของพลังงานในย่านความถี่ต่ำเทียบกับพลังงานรวมทั้งหมด 2. ความยาวของหน่วยเสียง	ค่าขีดแบ่ง	OGI_TS	88%
Pfau	1. ยอดของความดังดัดแปร 2. อัตราการตัดศูนย์	ขั้นตอนวิธีการหายอด	German Verbmobil	77%

	ลักษณะทางเสียง	การตรวจหาสระ	ฐานข้อมูลเสียง	ความถูกต้อง
Xie	1. ความเป็นรายคาบ 2. พลังงานที่ความถี่ต่ำ	ขั้นตอนวิธีการคอนเวกซ์ฮัลล์	TIMIT	86%
Howitt	1. พลังงานในช่วงความถี่ 300-900 เฮิรตซ์	ขั้นตอนวิธีการคอนเวกซ์ฮัลล์	TIMIT	85%

2. การตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากสถิติ

วิธีการในการตรวจหาสระ ที่จัดได้ว่าอยู่ในประเภทนี้คือ ในการตรวจหาสระ จะใช้วิธีการที่เกี่ยวข้องกับการเรียนรู้ของเครื่อง (Machine Learning) โดยการให้คอมพิวเตอร์ใช้การเรียนรู้ ซึ่งใช้สถิติเข้ามาช่วยในการสร้างแบบจำลองของเสียง มาเป็นวิธีการที่จะใช้ในการตรวจหาสระ

Sirigos, et al. [14] จำแนกกรอบของสัญญาณเสียงออกเป็น กรอบที่เป็นสระและกรอบที่ไม่ใช่สระด้วยข่ายงานประสาทเทียมที่มีเพอร์เซปตรอนหลายชั้น (Multi Layer Perceptron) ขนาด 4 ชั้น จำนวน 2 ตัวร่วมกับความยาวของหน่วยเสียง สำหรับการตรวจหาสระ โดยที่เพอร์เซปตรอนหลายชั้นตัวที่ 1 ใช้ ราสต้า พีแอลพี (RASTA PLP) เป็นเวกเตอร์ข้อมูลเข้า (Input Vector) สำหรับการเรียนรู้และ เพอร์เซปตรอนหลายชั้นตัวที่ 2 ใช้ ความถี่ฟอร์แมนท์และแบนด์วิดท์ของฟอร์แมนท์ (Formant Bandwidth) เป็นเวกเตอร์ข้อมูลเข้าสำหรับการเรียนรู้ นอกจากนี้ความยาวของสระได้ถูกนำมาใช้เป็นตัวแปรเสริมอีกตัวหนึ่ง เพื่อให้ผลการตรวจหาสระถูกต้องมากขึ้น ดังรูปที่ 2.19 ฐานข้อมูลเสียงที่ใช้ในการทดสอบครั้งนี้คือ TIMIT ที่เป็นแบบเสียงอ่าน



รูปที่ 2.19 แสดงแผนภาพของการแยกหน่วยเสียงออกเป็นสระ/ไม่ใช่สระของ Sirigos [15]

Wu, et al. [15] ใช้ราสต้า พีแอลพี และมอดดูเลชั่นสเปกโตรแกรม (Modulation Spectrogram) เป็นเวกเตอร์ข้อมูลเข้าสำหรับการเรียนรู้ให้กับ ข่ายงานประสาทเทียมที่มีเพอร์เซปตรอนหลายชั้นที่มีขนาด 3 ชั้น ซึ่งได้ทำการทดลองกับฐานข้อมูลเสียงที่เป็นของ OGI Number95

รายละเอียดโดยสรุปของแต่ละกระบวนวิธี ของการตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากสถิติ สามารถแสดงได้ดังตารางที่ 2.7

ตารางที่ 2.7 สรุปวรรณกรรมที่เกี่ยวข้องกับการตรวจหาสระโดยใช้พื้นฐานการตัดสินใจจากสถิติ

	ลักษณะทางเสียง	การตรวจหาสระ	ฐานข้อมูลเสียง	ความถูกต้อง
Sirigos	1. ราสต้า พีแอลพี 2. ฟอร์แมนท์	ข่ายงานประสาทเทียม (เพอร์เซปตรอนหลายชั้นที่มีขนาด 4 ชั้น)	TIMIT	77.8%
Wu, et al	1. ราสต้า พีแอลพี 2. มอดดูเลชั่นสเปกโตรแกรม	ข่ายงานประสาทเทียม (เพอร์เซปตรอนหลายชั้นที่มีขนาด 3 ชั้น)	OGI Number95	79%

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีการตรวจหาตำแหน่งสระ

ในบทที่ 2 ได้กล่าวถึง ลักษณะที่แตกต่างกันระหว่างเสียงก้องและเสียงไม่ก้อง ซึ่งลักษณะสำคัญที่ใช้ในการบ่งบอกว่า กรอบสัญญาณเสียงนั้นเป็นเสียงก้องหรือเสียงไม่ก้อง คือ ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ และลักษณะสำคัญของเสียงที่ใช้ในการแยกสระออกจากเสียงก้องประเภทอื่น ซึ่งประกอบด้วย ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ การตรวจหาตำแหน่งสระในวิทยานิพนธ์นี้ จะใช้ลักษณะสำคัญดังกล่าวร่วมกัน โดยเริ่มต้นจากการหาความก้องหรือไม่ก้องของเสียงด้วยค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ และจากกรอบสัญญาณเสียงก้องที่ได้ นำไปตรวจหาตำแหน่งที่เหมาะสมที่จะเป็นตำแหน่งของสระ โดยใช้ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์

ด้วยรายละเอียดที่กล่าวมาข้างต้น บทนี้จะได้นำเสนอ นิยามของตำแน่งสระ รายละเอียดของขั้นตอนและวิธีการที่ใช้สำหรับการตรวจหาตำแหน่งสระในเสียงพูดภาษาไทย ดังนี้

นิยามของตำแหน่งสระ

แลนด์มาร์ก (Landmark) คือ ตำแหน่งของสัญญาณเสียง ณ เวลาต่างๆ ที่ถูกเลือกขึ้นมา เพื่อใช้เป็นตัวแทนของแต่ละหน่วยเสียงในสัญญาณเสียงพูดนั้น โดยที่หน่วยเสียงแต่ละหน่วยเสียงนั้นสามารถมีแลนด์มาร์กได้พียงตำแหน่งเดียวเท่านั้น

Stevens [17] ได้ให้คำนิยามของแลนด์มาร์กสระไว้ว่า แลนด์มาร์กของสระคือ ตำแหน่งที่มีค่าแอมปลิจูดในช่วงความถี่ฟอร์แมนท์ที่หนึ่งมากที่สุด ในสัญญาณที่เป็นหน่วยเสียงสระนั้น เช่นเดียวกับ Choi [18] ที่กล่าวว่า แลนด์มาร์กของสระคือ ตำแหน่งทางเวลาของสัญญาณเสียงในหน่วยเสียงสระ ที่มีค่าแอมปลิจูดของความถี่ฟอร์แมนท์ที่หนึ่งมากที่สุด

เมื่อพิจารณาสัญญาณเสียงพูด พบว่า ความแตกต่างระหว่างหน่วยเสียงต่างๆ ที่สังเกตได้เด่นชัดที่สุด คือ ความแตกต่างระหว่างหน่วยเสียงสระและหน่วยเสียงพยัญชนะ ดังนั้น ประเภทของแลนด์มาร์กที่เกิดขึ้นในสัญญาณเสียง จึงสามารถแบ่งออกเป็น 2 ประเภทใหญ่ๆ คือ

1. แลนด์มาร์กของสระ
2. แลนด์มาร์กของพยัญชนะ

โดยทั่วไป ตำแหน่งของสัญญาณเสียงที่เวลาต่างๆ ที่ถูกเลือกมาเป็นแลนด์มาร์กของหน่วยเสียงนั้น มักจะถูกเลือกจากบริเวณที่มีลักษณะทางสวนสัทศาสตร์ของสัญญาณเสียงที่เด่นชัดที่สุด ที่ปรากฏขึ้นในหน่วยเสียงนั้น เช่น แลนด์มาร์กของสระ จะเกิดขึ้นในบริเวณที่มีค่าของพลังงานที่ความถี่ต่ำของหน่วยเสียงสระนั้นมากที่สุด หรือ แลนด์มาร์กของพยัญชนะ มักจะเกิดขึ้นในบริเวณที่

เป็นรอยต่อระหว่างการกักลม (Closure) และ การปล่อยลม (Release) ที่เกิดขึ้นในขณะที่เปล่งเสียงพยัญชนะนั้น เป็นต้น

ดังนั้น แลนด์มาร์กหรือตำแหน่งของสระในวิทยานิพนธ์นี้ คือ ตำแหน่งทางเวลาของสัญญาณเสียง ที่ถูกเลือกขึ้นมาจากลักษณะทางสวนสัทศาสตร์ ซึ่งเป็นค่าที่สูงที่สุดของพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของหน่วยเสียงสระนั้น และตำแหน่งที่ถูกเลือกขึ้นมานั้น จะต้องอยู่ในช่วงเวลาที่เป็นขอบเขตของหน่วยเสียงสระในไฟล์กำกับเสียง (Transcription) จึงจะเป็นแลนด์มาร์กที่ถูกต้อง โดยที่ขอบเขตช่วงเวลาของหน่วยเสียงต่างๆ ที่ระบุในไฟล์กำกับเสียงนั้น เกิดขึ้นจากการกำกับเสียงโดยผู้เชี่ยวชาญทางภาษาศาสตร์

ที่มาของแนวคิดวิธีการตรวจหาตำแหน่งสระ

การได้มาของแนวคิดนี้ เป็นการปรับปรุงระบบการตรวจหาสระให้มีประสิทธิภาพดีขึ้น โดยการหาสาเหตุของความผิดพลาดที่เกิดขึ้นจากงานวิจัยในอดีต รวมทั้งวิธีการของ Howitt เพื่อที่จะนำมาปรับปรุงวิธีการในการตรวจหาสระให้ดีขึ้นต่อไป ซึ่งวิธีการตรวจหาสระของ Howitt นั้นสามารถสรุปได้ ดังนี้

Howitt ได้ใช้พลังงานในช่วงความถี่ 300 ถึง 900 เฮิรตซ์ และความยาวเสียงเป็นลักษณะทางเสียงที่ใช้ในการตรวจหาสระ เหตุผลที่เลือกลักษณะทางเสียงดังกล่าวนั้น เนื่องจากความผิดพลาดที่เกิดขึ้นจากงานวิจัยของ Mermelstein มาจากการที่ใช้ค่าพลังงานในช่วงความถี่ที่ไม่เหมาะสมคือในช่วง 500 ถึง 4000 เฮิรตซ์ ดังนั้น Howitt จึงได้นำเสนอการตรวจหาตำแหน่งสระจากพลังงานที่ความถี่ต่ำ และได้เลือกพลังงานในช่วงความถี่ฟอร์แมนท์ที่หนึ่ง เนื่องจากเขาได้ทำการทดลองแล้วพบว่า ตำแหน่งที่สามารถเป็นตัวแทนของหน่วยเสียงสระที่ดีที่สุดนั้นคือ ยอดสูงเด่นของพลังงานในช่วงความถี่ฟอร์แมนท์ที่หนึ่ง หรือในช่วงความถี่ 300 ถึง 900 เฮิรตซ์ ซึ่งผลการทดลองที่ได้นี้ตรงกับคำนิยามของตำแหน่งสระที่ Stevens ได้กล่าวไว้ว่า แลนด์มาร์กของสระคือ ตำแหน่งที่มีค่าแอมปลิจูดในช่วงความถี่ฟอร์แมนท์ที่หนึ่งมากที่สุด โดยใช้การหายอดสูงเด่นจากระเบียบวิธีคอนเวกซ์ฮัลล์ เป็นวิธีการตรวจหาสระ ซึ่งเมื่อวิเคราะห์ถึงสาเหตุของความผิดพลาดที่เกิดขึ้นนั้น พบว่าวิธีการตรวจหาสระของ Howitt นี้ใช้ค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์เป็นลักษณะทางสวนสัทศาสตร์ในการคัดเลือกตำแหน่งสระ ซึ่งเมื่อพิจารณาค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์ของกรอบสัญญาณเสียงสระ และ ค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์ของกรอบสัญญาณเสียงนาสิก เสียงกึ่งสระ เสียงกระทบ เสียงกักแบบก้อง และเสียงข้างลิ้น ซึ่งเป็นส่วนหนึ่งของเสียงก้องแล้ว พบว่ามีค่าไม่ต่างกันมาก จึงส่งผลให้การคัดเลือกตำแหน่งสระนั้น มีหน่วยเสียงก้องประเภทอื่นที่ไม่ใช่หน่วยเสียงสระปนเข้ามาเป็นจำนวนมาก ทำให้เกิดความผิดพลาดขึ้น และด้วยสาเหตุของความผิดพลาดที่เกิดขึ้นจากวิธีการของ Howitt ทำให้

ในการตรวจหาสระของวิทยานิพนธ์นี้ เลือกลักษณะทางสวนสัทศาสตร์ที่จะนำมาใช้ในการตรวจหาสระ เพื่อให้ประสิทธิภาพของระบบการตรวจหาสระดีขึ้น โดยรายละเอียดนั้นสามารถสรุปได้ดังนี้

ลักษณะทางเสียงที่ใช้ประกอบด้วย ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ และพลังงานในช่วงความถี่ที่มากกว่า 300 เฮิรตซ์ โดยเหตุผลที่เลือกลักษณะทางเสียงดังกล่าวนั้น เนื่องจากด้วยความผิดพลาดที่เกิดขึ้นจากงานวิจัยในการตรวจหาตำแหน่งของสระในอดีตที่ผ่านๆมารวมถึงของ Howitt พบว่า ความผิดพลาดโดยส่วนใหญ่เกิดขึ้นจากตำแหน่งที่เกิน ซึ่งเป็นตำแหน่งของเสียงพยัญชนะที่ประกอบด้วย เสียงเสียดแทรก เสียงนาสิก หรือเสียงกึ่งสระปะปนเข้ามาในตำแหน่งที่ถูกเลือกว่าเป็นตำแหน่งสระ ดังนั้น เพื่อหลีกเลี่ยงผลกระทบที่เกิดจากตำแหน่งเกินของพยัญชนะดังกล่าว ระบบการตรวจหาสระด้วยวิธีนี้ จึงได้ออกแบบขั้นตอนวิธีการในการตรวจหาสระออกเป็น 2 ขั้นตอน คือการตรวจหาความก้องหรือความไม่ก้องของเสียง เพื่อหลีกเลี่ยงการเกิดเสียงเสียดแทรกที่เข้ามาปะปนกับตำแหน่งที่เป็นสระ เพราะเนื่องจากว่า เสียงเสียดแทรกเป็นเสียงประเภทหนึ่งในจำนวนของเสียงไม่ก้องทั้งหมด ด้วยการใช้ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ ในการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง และในขั้นตอนถัดมาคือ การคัดเลือกตำแหน่งของสระจากกรอบสัญญาณเสียงก้อง ในขั้นตอนนี้ได้มีการใช้พลังงานของกรอบสัญญาณเสียงที่มีความถี่มากกว่า 300 เฮิรตซ์เป็นลักษณะทางเสียง เพื่อหลีกเลี่ยงปัญหาที่เกิดจากเสียงก้องประเภทอื่นที่ไม่ใช่เสียงสระ ซึ่งมักจะปนเข้ามาเป็นตำแหน่งของสระซึ่งประกอบด้วย เสียงนาสิก เสียงกึ่งสระ เนื่องจากได้พบว่า พลังงานของกรอบสัญญาณเสียงที่มีความถี่มากกว่า 300 เฮิรตซ์ ทำให้เสียงก้องประเภทอื่นที่ไม่ใช่เสียงสระ มีพลังงานน้อยกว่าเสียงสระ ทำให้ตำแหน่งของเสียงก้องประเภทอื่นที่ไม่ใช่สระลดน้อยลง แม้ว่าพลังงานในช่วงนี้จะขัดแย้งกับช่วงพลังงานที่ Howitt นำเสนอ เนื่องจากมีพลังงานในช่วงความถี่สูงเข้ามาด้วย แต่เมื่อพิจารณาเฉพาะกรอบสัญญาณเสียงก้อง พบว่ากรอบสัญญาณเสียงก้องเหล่านั้นไม่มีส่วนประกอบของพลังงานที่ความถี่สูง ดังนั้น ค่าพลังงานที่ได้ในกรอบสัญญาณเสียงในช่วงความถี่ที่มากกว่า 300 เฮิรตซ์ จึงมีเฉพาะพลังงานในช่วงความถี่ต่ำ ซึ่งเป็นลักษณะหนึ่งของสัญญาณเสียงสระ

ภาพรวมวิธีการตรวจหาตำแหน่งของสระ

การตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทยในวิทยานิพนธ์นี้ ได้ออกแบบให้ใช้พื้นฐานการตัดสินใจจากกฎเป็นวิธีการที่ใช้ในการตรวจหาสระ เนื่องจากวิธีการนี้จะเป็นการเรียนรู้ลักษณะที่เด่นชัดที่สามารถวัดได้โดยตรงจากสัญญาณเสียง โดยไม่ต้องมีการเรียนรู้ข้อมูลฝึก และสามารถนำไปใช้กับฐานข้อมูลเสียงที่หลากหลายมากขึ้น และด้วยความผิดพลาดที่เกิดขึ้นจากงานวิจัยในการตรวจหาตำแหน่งของสระในอดีตที่ผ่านๆมา ซึ่งพบว่า ความผิดพลาดโดยส่วนใหญ่จะเกิดขึ้นจากการที่มีตำแหน่งของเสียงพยัญชนะซึ่งประกอบด้วย เสียงเสียดแทรก เสียงนาสิก หรือเสียงกึ่งสระ

ปะปนเข้ามาในตำแหน่งที่ถูกเลือกว่าเป็นตำแหน่งสระ ดังนั้น เพื่อหลีกเลี่ยงผลกระทบที่เกิดจากตำแหน่งของพยัญชนะดังกล่าว วิทยานิพนธ์นี้ จึงได้ออกแบบขั้นตอนวิธีการในการตรวจหาสระออกเป็น 2 ขั้นตอน คือการตรวจหาความก้องหรือความไม่ก้องของเสียง (Voicing Detection) เพื่อหลีกเลี่ยงการเกิดเสียงเสียดแทรกที่เข้ามาปะปนกับตำแหน่งที่ถูกเลือกว่าเป็นตำแหน่งของสระ เพราะเนื่องจากว่า เสียงเสียดแทรกเป็นเสียงประเภทหนึ่งในจำนวนของเสียงไม่ก้องทั้งหมด และในขั้นตอนถัดมาคือ การคัดเลือกตำแหน่งของสระ (Landmark Picking) ในขั้นตอนนี้ได้มีการใช้พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์เป็นลักษณะทางเสียง เพื่อหลีกเลี่ยงปัญหาที่เกิดจากเสียงก้องประเภทอื่นที่ไม่ใช่เสียงสระ ซึ่งประกอบด้วย เสียงกระทบ เสียงข้างลิ้น เสียงนาสิก เสียงกึ่งสระ หรือเสียงกักแบบก้อง ที่จะเข้ามาเป็นตำแหน่งที่ถูกเลือกว่าเป็นตำแหน่งของสระ เนื่องจากได้พบว่า ที่ความถี่ที่มากกว่า 300 เฮิรตซ์ เสียงกระทบ เสียงข้างลิ้น เสียงกึ่งสระ และ และเสียงนาสิกจะมีพลังงานน้อยกว่าเสียงสระ

ค่าของลักษณะทางเสียงต่างๆเหล่านี้ สามารถคำนวณได้จากสัญญาณเสียงซึ่งอยู่ในโดเมนเวลา (Time Domain) รายละเอียดของกระบวนการที่ใช้ในการตรวจหาสระในที่นี้แบ่งเป็นขั้นตอนย่อยๆ ดังนี้

1. การตรวจหาความก้องหรือความไม่ก้องของเสียง

ส่วนแรกของการตรวจหาสระเริ่มจาก การนำสัญญาณเสียงที่อยู่ในโดเมนเวลาไปตัดแบ่งออกเป็นกรอบ จากนั้นนำแต่ละกรอบของสัญญาณเสียงไปตรวจหาความก้องหรือไม่ก้องโดยวิธีการหาค่าอัตสหสัมพันธ์ ซึ่งสัญญาณเสียงซึ่งมีค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์มาก เป็นลักษณะของสัญญาณเสียงก้อง และในทางกลับกันสัญญาณเสียงซึ่งมีค่าอัตสหสัมพันธ์ที่มากที่สุดความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์น้อย เป็นลักษณะของสัญญาณเสียงไม่ก้อง ดังนั้น ค่านี้จะนำมาใช้ในการแบ่งกรอบสัญญาณเสียงออกเป็นเสียงก้อง และเสียงไม่ก้อง ในภาษาไทย เสียงประเภทเสียงก้อง จะประกอบด้วย เสียงกักแบบก้อง เสียงกระทบ เสียงข้างลิ้น เสียงสระ เสียงกึ่งสระ และเสียงนาสิก เสียงประเภทเสียงไม่ก้อง จะประกอบด้วย เสียงกักแบบไม่ก้อง และ เสียงเสียดแทรก นอกจากนี้ ทุกๆกรอบจะมีการคำนวณค่าของพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ เพื่อที่จะนำค่าพลังงานที่ได้นั้นไปตรวจหาตำแหน่งที่เหมาะสมจะเป็นตำแหน่งของสระ

2. การคัดเลือกตำแหน่งของสระ

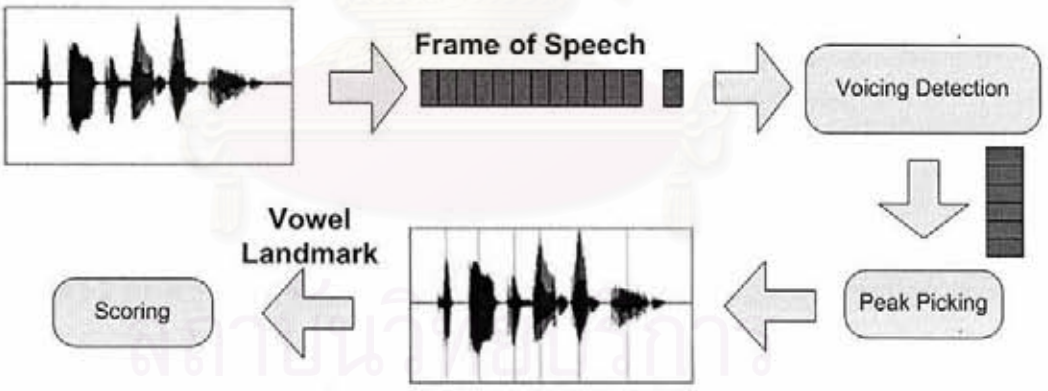
กรอบเสียงที่เป็นเสียงก้องเท่านั้นที่ถูกนำมาใช้ในขั้นตอนนี้ ซึ่งประกอบด้วย กรอบของเสียงกักแบบก้อง เสียงกระทบ เสียงข้างลิ้น เสียงสระ เสียงกึ่งสระ และเสียงนาสิก และเมื่อพิจารณาพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของเสียงสระ และ

เสียงก้องประเภทอื่น พบว่า พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของเสียงสระนั้น มีค่ามากกว่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของเสียงก้องประเภทอื่น ส่งผลให้ยอดพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของเสียงสระสูงเด่นกว่ายอดพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของเสียงก้องประเภทอื่น ด้วยเหตุนี้การตรวจหาตำแหน่งของสระในขั้นตอนนี้ จึงทำการคัดเลือกยอดที่สูงเด่น ด้วยขั้นตอนวิธีการของคอนเวกซ์ฮัลล์ (Convex Hull Algorithm) จากพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของกรอบสัญญาณเสียงก้องซึ่งคำนวณได้จากขั้นตอนแรก ซึ่งยอดที่หาได้จะกำหนดให้เป็นตำแหน่งของสระ

3. การให้คะแนน

ตำแหน่งของสระที่ได้จะมีการให้คะแนน ซึ่งช่วงของคะแนนนี้ จะมีค่าตั้งแต่ 0-1 ตำแหน่งของสระที่มีความถูกต้องมาก เมื่อเปรียบเทียบกับขอบเขตของหน่วยเสียงสระนั้นในไฟล์กำกับเสียง จะมีคะแนนมาก และในทางกลับกัน ตำแหน่งของสระที่มีความถูกต้องน้อยกว่าเมื่อเปรียบเทียบกับขอบเขตของหน่วยเสียงสระนั้นในไฟล์กำกับเสียง จะต้องมีคะแนนน้อย ซึ่งคะแนนที่ได้นี้จะต้องมีการตรวจสอบความสมเหตุสมผล เพื่อตรวจสอบว่าคะแนนที่ตั้งขึ้นมาเหมาะสมหรือไม่

ขั้นตอนที่ใช้ในการตรวจหาสระ สามารถแสดงได้ดังรูปที่ 3.1

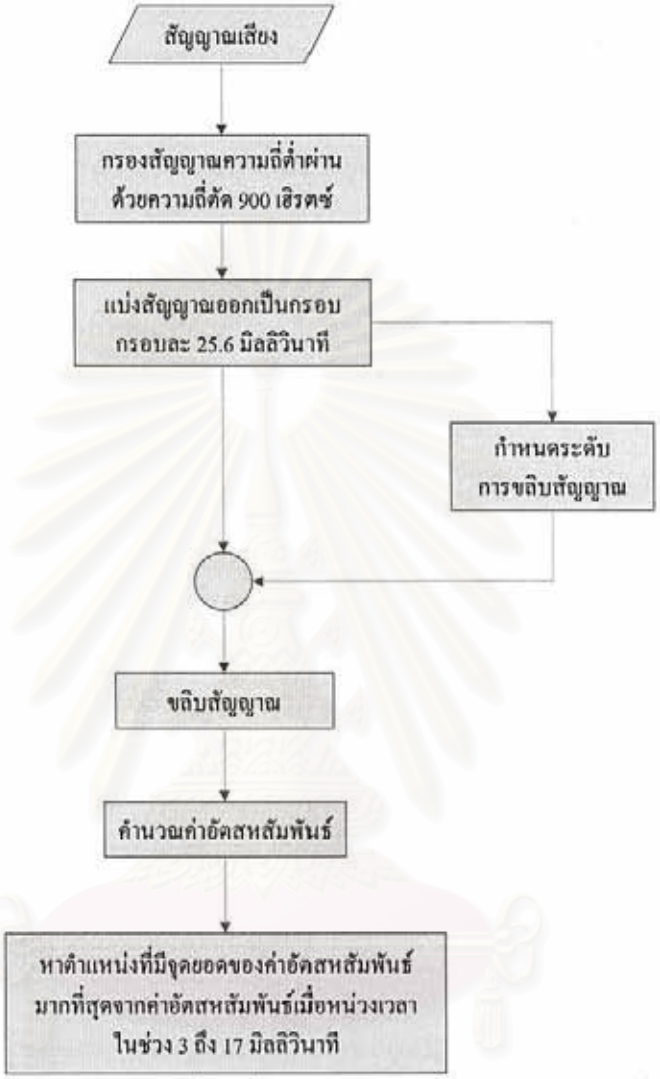


รูปที่ 3.1 แผนภาพการตรวจหาตำแหน่งสระ

การหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ของสัญญาณเสียง

ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ คือค่าที่สามารถบอกความเป็นรายคาบของสัญญาณเสียงภายในกรอบสัญญาณเสียงนั้นๆ โดยการหาค่าอัตสหสัมพันธ์นี้จะทำ

ที่ทุกๆ ส่วนของสัญญาณเสียง โดยการนำสัญญาณเสียงนั้นมาแบ่งเป็นกรอบของสัญญาณ การหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ มีวิธีการและขั้นตอนดังรูปที่ 3.2



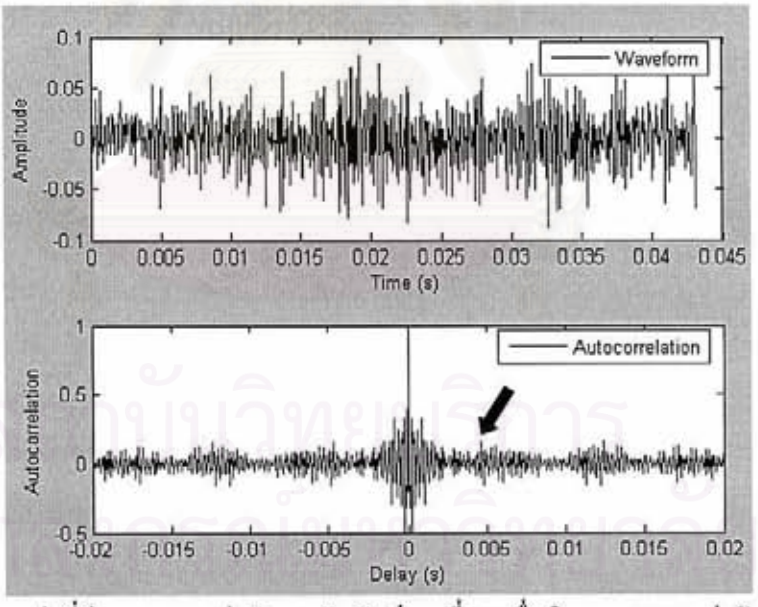
รูปที่ 3.2 การหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ของสัญญาณเสียง

รายละเอียดในแต่ละขั้นตอนของการหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ มีดังต่อไปนี้

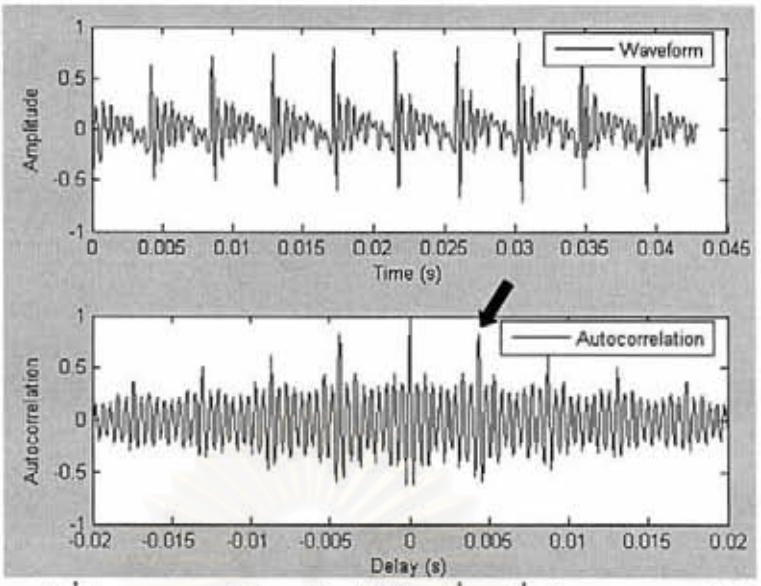
1. นำสัญญาณเสียงที่ต้องการจะหาค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ มากรองสัญญาณแบบความถี่ต่ำผ่านด้วยความถี่ตัด (Cut-off Frequency) 900 เฮิรตซ์ ด้วยตัวกรองดิจิตอลแบบบัตเตอร์เวิร์ธ (Butterworth)

2. แบ่งสัญญาณเสียงที่ผ่านการกรองเป็นกรอบเสียงพูด โดยให้แต่ละกรอบมีช่วงซ้อนทับกัน 3 ใน 4 เฟรม เพื่อลดความผิดพลาดที่เกิดจากความไม่ต่อเนื่องของข้อมูล แต่ละกรอบมีขนาด 25.6 มิลลิวินาที หรือ 410 ตัวอย่าง สำหรับสัญญาณเสียงที่มีอัตราสุ่ม (Sampling Rate) 16 กิโลเฮิรตซ์
3. กำหนดระดับของการขลิบกลาง (Clipping Level) ที่มีขนาดเท่ากับ 30 เปอร์เซ็นต์ของค่าสูงสุดของค่าสัมบูรณ์ของสัญญาณเสียงในกรอบสัญญาณนั้น เพื่อกำจัดสัญญาณที่มีขนาดแอมปลิจูดต่ำซึ่งคาดว่าจะเป็นสัญญาณรบกวนออก
4. นำแต่ละตัวอย่างที่อยู่ในกรอบของสัญญาณเสียงนั้น มาทำการขลิบกลาง (Central Clipping) ด้วยระดับของการขลิบกลางที่กำหนดไว้
5. คำนวณค่าอัตสหสัมพันธ์ของทุกๆ ตัวอย่างในกรอบสัญญาณเสียงที่ผ่านการขลิบกลางด้วยสมการที่ 2.3
6. หาตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด จากค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ หรือ 3 ถึง 17 มิลลิวินาที ซึ่งค่าที่ได้นี้กำหนดให้เป็นค่าอัตสหสัมพันธ์สูงสุดของกรอบสัญญาณนั้น

ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงก้องและกรอบสัญญาณเสียงไม่ก้อง มีลักษณะดังรูปที่ 3.3 และ รูปที่ 3.4



รูปที่ 3.3 ตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด เมื่อพิจารณาเฉพาะค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงไม่ก้อง /ส/



รูปที่ 3.4 ตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด เมื่อพิจารณาเฉพาะค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงก้อง /อา/

จากรูปที่ 3.3 และ รูปที่ 3.4 จะเห็นได้ว่า ตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด (ลูกศรสีดำ) เมื่อพิจารณาเฉพาะค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ ของกรอบสัญญาณเสียงไม่ก้อง จะมีค่าน้อยกว่า ตำแหน่งที่มีจุดยอดของค่าอัตสหสัมพันธ์มากที่สุด เมื่อพิจารณาเฉพาะค่าอัตสหสัมพันธ์ในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ ของกรอบสัญญาณเสียงก้อง

การหาค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของสัญญาณเสียง

ลักษณะทางสวนสัทศาสตร์ ที่นอกเหนือจากการใช้ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ในการหาความก้องหรือความไม่ก้องของกรอบสัญญาณเสียงแล้ว พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของกรอบสัญญาณเสียง จะนำมาใช้ในการคัดเลือกตำแหน่งแลนด์มาร์กของสระ ซึ่งการหาค่าพลังงงานที่ตัดความถี่ต่ำของสัญญาณเสียงนั้น มีวิธีการและขั้นตอนดังรูปที่ 3.5

รูปที่ 3.5 การหาค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของสัญญาณเสียง

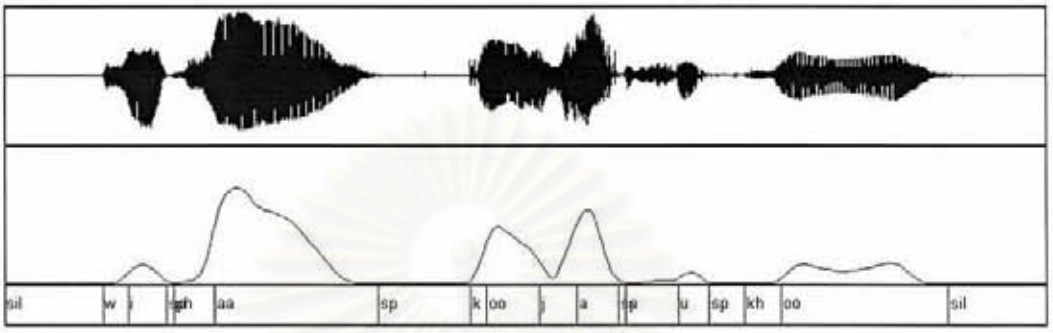
รายละเอียดในแต่ละขั้นตอนของการหาค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ มีดังต่อไปนี้

1. นำสัญญาณเสียงที่ต้องการจะหาค่าพลังงาน มากรองสัญญาณแบบความถี่สูงผ่านด้วยความถี่ตัด 300 เฮิรตซ์ ด้วยตัวกรองดิจิตอลแบบบัตเตอร์เวิร์ธ เนื่องจากที่ความถี่นี้พบว่า หน่วยเสียง เสียงกระทบ เสียงข้างลิ้น เสียงกึ่งสระ และเสียงนาสิก มีพลังงานน้อยกว่าเสียงสระ
2. แบ่งสัญญาณเสียงที่ผ่านการกรองเป็นกรอบเสียงพูด โดยให้แต่ล่ะกรอบมีช่วงซ้อนทับกัน 3 ใน 4 เฟรม แต่ละกรอบมีขนาด 30 มิลลิวินาที หรือ 480 ตัวอย่าง สำหรับสัญญาณเสียงที่มีอัตราสุ่ม 16 กิโลเฮิรตซ์
3. คำนวณค่าพลังงานของแต่ละกรอบสัญญาณเสียงพูด ตามสมการต่อไปนี้

$$E(m) = \sum_{n=0}^{N-1} [w(m)s(m-n)]^2 \qquad (3.1)$$

$w(m)$ คือ วินโดว์ฟังก์ชัน ที่ใช้กำหนดรูปร่างในการพิจารณาของสัญญาณเสียง $s(n)$ ในหนึ่งกรอบ และ N คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ในกรอบหรือภายในกรอบของฟังก์ชันหน้าต่าง

ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของสัญญาณเสียง มีลักษณะดังรูปที่ 3.6

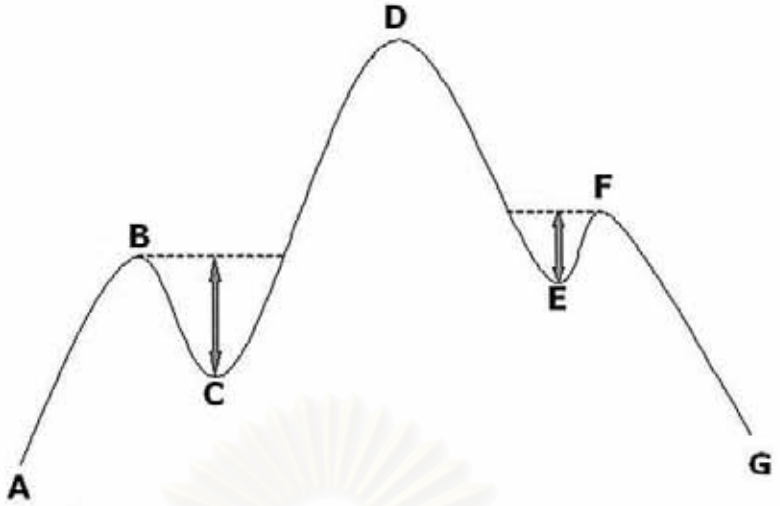


รูปที่ 3.6 พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์

ระเบียบวิธีคอนเวกซ์ฮัลล์

ระเบียบวิธีคอนเวกซ์ฮัลล์ เป็นการคัดเลือกตำแหน่งของสระที่แตกต่างจากการคัดเลือกตำแหน่งของสระด้วยวิธีทั่วๆไป ดังนี้ การคัดเลือกตำแหน่งของสระด้วยวิธีทั่วๆไปนั้น จะทำการคัดเลือกโดยค้นหาตำแหน่งที่เหมาะสมจะเป็นแลนด์มาร์กแบบตรงไปตรงมา กล่าวคือ ตำแหน่งของสระนั้นจะคัดเลือกจากค่าที่สูงที่สุดจากลักษณะทางสวนสัทศาสตร์ที่กำหนดขึ้น เช่น ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ดังนั้น ทุกตำแหน่งที่เป็นยอดสูง (Peak) ที่ปรากฏขึ้นทั้งหมดในรูปร่างของพลังงาน จะกำหนดให้เป็นแลนด์มาร์ก ในขณะที่ระเบียบวิธีคอนเวกซ์ฮัลล์นั้นเป็นการคัดเลือกตำแหน่งของสระ จากส่วนแบ่งที่เกิดขึ้นจากการค้นหาตำแหน่งที่มีค่าลักษณะทางสวนสัทศาสตร์ต่ำที่สุด (Dip) ก่อนในขั้นต้น และทำการคัดเลือกตำแหน่งที่จะเป็นแลนด์มาร์กจากค่าลักษณะทางสวนสัทศาสตร์ที่สูงที่สุดในส่วนแบ่งนั้น ดังรูปที่ 3.7 ซึ่งเป็นตัวอย่างการหาแลนด์มาร์กโดยระเบียบวิธีคอนเวกซ์ฮัลล์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

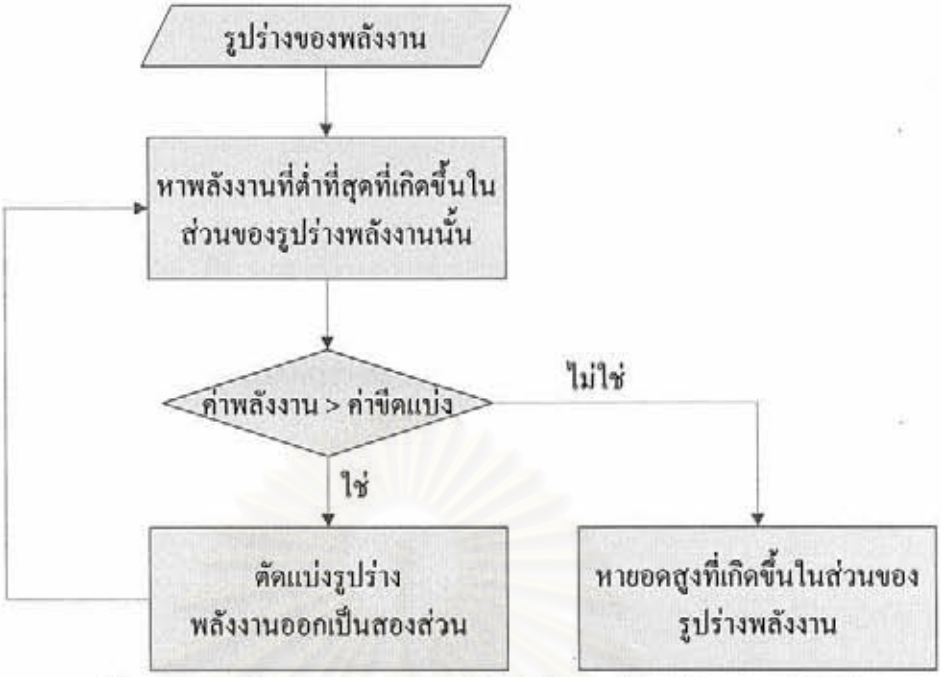


รูปที่ 3.7 ตัวอย่างการหาแลนด์มาร์กโดยระเบียบวิธีคอนเวกซ์ฮัลล์ [3]

จากรูปที่ 3.7 เป็นการหาแลนด์มาร์กจากรูปร่างของพลังงาน (A-G) การหาแลนด์มาร์กเริ่มต้นจาก การหาตำแหน่งที่มีค่าพลังงานต่ำๆ ทั้งหมดที่เกิดขึ้นตลอดรูปร่างของพลังงานจากซ้ายไปขวาโดยไม่รวมจุดเริ่มต้นและจุดปลาย ซึ่งตำแหน่งที่มีค่าพลังงานต่ำๆ ที่เกิดขึ้นในรูปที่ 3.7 นั้นคือ เส้นปะ B-C และเส้นปะ E-F

จากตำแหน่งที่มีค่าพลังงานต่ำๆทั้งหมดที่ได้นั้น จะถูกนำมาเปรียบเทียบกัน เพื่อหาตำแหน่งที่มีค่าพลังงานต่ำที่สุด จะได้ตำแหน่ง B-C และจะนำค่าพลังงานที่ตำแหน่ง B-C นี้มาเปรียบเทียบกับค่าขีดแบ่ง เพื่อที่จะดูว่าเหมาะสมที่จะตัดแบ่งรูปร่างของพลังงานออกเป็นสองส่วนหรือไม่ ถ้าค่าพลังงานที่ต่ำตำแหน่ง B-C นั้นมีค่ามากกว่าค่าขีดแบ่ง จะทำการตัดแบ่งพลังงานออกเป็นสองส่วน คือ ส่วนของ A-C และส่วนของ C-G โดยที่รูปร่างพลังงานในส่วนของ C-G นั้นจะถูกนำกลับไปเริ่มต้นการค้นหาตั้งแต่แรกอีกครั้ง จนกระทั่งไม่พบค่าพลังงานที่ต่ำที่สุด ดังที่เกิดขึ้นกับรูปร่างของพลังงานในส่วนของ A-C หรือค่าพลังงานที่ต่ำที่สุดที่ตำแหน่งนั้นมีค่าน้อยกว่าค่าขีดแบ่ง จะไม่มีการแบ่งตัดแบ่งรูปร่างของพลังงาน แต่จะนำส่วนของพลังงานนั้นไปหาตำแหน่งที่มีค่าพลังงานมากที่สุดที่เกิดขึ้นในส่วนแบ่งนั้นแทน และจะกำหนดให้ตำแหน่งนั้นเป็นตำแหน่งของแลนด์มาร์ก การหาค่าตำแหน่งของแลนด์มาร์กนั้น มีวิธีการและขั้นตอนดังรูปที่ 3.8

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.8 การหาค่าตำแหน่งของแลนด์มาร์กด้วยระเบียบวิธีคอนเวกซ์ฮัลล์

การให้คะแนนตำแหน่งที่ตรวจหาได้

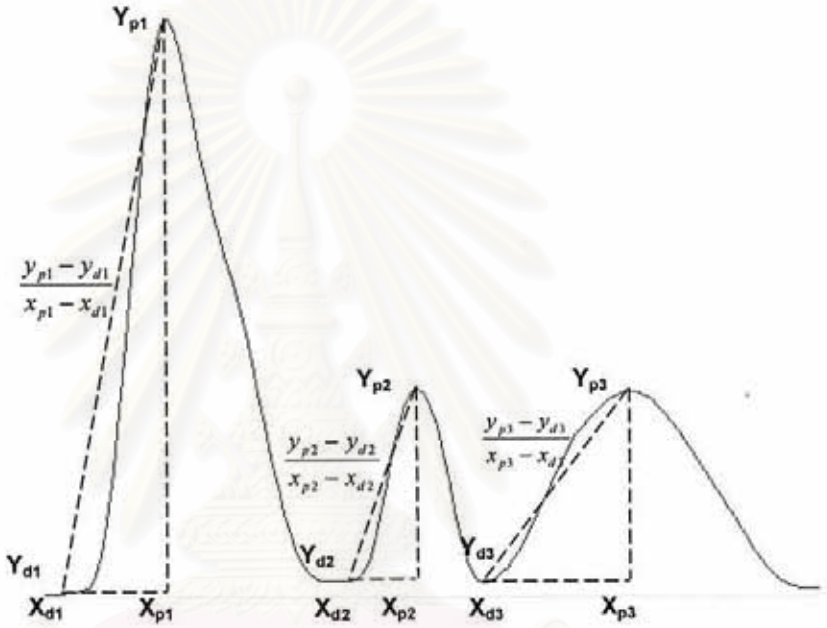
วิธีการในการให้คะแนนของตำแหน่งที่ถูกเลือกว่าเป็นสระนั้น เริ่มจาก การจำแนกกลุ่ม (Classify) ด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น (Linear Discriminant Analysis, LDA) ออกเป็นกลุ่มของตำแหน่งที่เป็นสระ และกลุ่มของตำแหน่งที่ไม่เป็นสระ โดยที่เวกเตอร์ลักษณะสำคัญ (Feature Vector) ของชุดข้อมูลฝึก (Training Set) และชุดข้อมูลทดสอบ (Test Set) แต่ละตัวอย่างนั้น ประกอบด้วย 2 ลักษณะ คือ

1. ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของกรอบสัญญาณเสียงที่ถูกเลือกว่าเป็นตำแหน่งของสระ
2. ค่าความชันของพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของกรอบสัญญาณเสียงที่ถูกเลือกว่าเป็นตำแหน่งของสระเดียวกัน ตัวอย่างการหาความชันสามารถแสดงได้ดังรูปที่ 3.9

คะแนนของตำแหน่งที่ถูกเลือกว่าเป็นสระนั้น มีค่าเท่ากับความน่าจะเป็นที่ตำแหน่งสระที่ตรวจหาได้นั้น (x) จะถูกจำแนกให้อยู่ในกลุ่มของตำแหน่งที่เป็นสระ ซึ่งสามารถคำนวณได้จากสมการที่ 3.2

$$P(g=\omega_i \mid x)=\frac{e^{-\frac{d_i}{2}}}{\sum_{k=1}^{K} e^{-\frac{d_k}{2}}} \qquad (3.2)$$

โดยที่ d_i คือ ระยะทางระหว่างตัวอย่างข้อมูลทดสอบ x กับ จุดศูนย์กลางของกลุ่มชุดข้อมูลฝึกที่เป็นสระ (ω_i) d_k คือ ระยะทางระหว่างตัวอย่างข้อมูลทดสอบ x กับ จุดศูนย์กลางของกลุ่มชุดข้อมูลฝึกที่เป็นไม่เป็นสระ และ K คือ จำนวนกลุ่มที่ต้องการจำแนก



รูปที่ 3.9 การหาความชันของแต่ละตำแหน่งที่ถูกเลือกเป็นสระ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4
ฐานข้อมูลเสียงที่ใช้ในการตรวจหาสระ

ในบทนี้จะได้กล่าวถึงฐานข้อมูลที่ใช้สำหรับหาประสิทธิภาพของระบบการตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทยของวิทยานิพนธ์นี้ ซึ่งประกอบด้วยฐานข้อมูล 2 ชนิดคือ

1. ฐานข้อมูลเสียงภาษาไทย
2. ฐานข้อมูลเสียงตัวเลข

รายละเอียดของแต่ละฐานข้อมูล เป็นดังต่อไปนี้

ฐานข้อมูลเสียงภาษาไทย (Lotus)

ฐานข้อมูลเสียง Lotus เป็นฐานข้อมูลเสียงพูดภาษาไทยขนาดใหญ่ ที่เหมาะสมสำหรับใช้ในการวิจัยและพัฒนาระบบที่เกี่ยวข้องกับการรู้จำเสียงพูดต่อเนื่องสำหรับภาษาไทย โดยใช้ลักษณะการพูดแบบอ่านในการบันทึกเสียง ซึ่งครอบคลุมคำศัพท์ภาษาไทย จำนวนไม่น้อยกว่า 5,000 คำ ที่คัดเลือกมาจากฐานข้อมูลบทความข่าวหรือบทความทั่วๆ ไป

ฐานข้อมูลเสียงนี้ สามารถแบ่งออกเป็น 2 ประเภท ตามชนิดของไมโครโฟนที่ใช้ในการบันทึกเสียงคือ ประเภทแรกเป็นไมโครโฟนที่มีคุณภาพสูง และประเภทที่สองเป็นไมโครโฟนที่มีคุณภาพปานกลาง และทำการบันทึกเสียงใน 2 สภาพแวดล้อมที่แตกต่างกันคือ

1. สภาพแวดล้อมแบบห้องเงียบ (Clean Speech)
2. สภาพแวดล้อมแบบสำนักงาน (Office Speech)

ผู้พูดในการบันทึกเสียงนั้น ประกอบด้วยผู้พูดทั้งที่เป็นเพศชายและเพศหญิงในจำนวนที่เท่ากัน โดยเก็บข้อมูลสียงผ่านเทปแม่เหล็กที่สามารถเก็บข้อมูลได้ 2 ถึง 5 จิกะไบต์ (Digital Audio Tape) ก่อนที่จะทำการแปลงเป็นไฟล์อิเล็กทรอนิกส์ ซึ่งจะนำใช้ในการวิจัยและพัฒนาระบบเสียงพูดต่อไป

ภายในฐานข้อมูลนี้ ประกอบด้วยชุดหน่วยเสียงสมดุล (Phonetically Balanced Set) ซึ่งใช้สำหรับการฝึกฝนแบบจำลองเสียง (Acoustic Model) และมีการกำกับหน่วยเสียงอัตโนมัติ (Automatic Phoneme Labeler) และนอกจากนี้ ฐานข้อมูลนี้ยังประกอบไปด้วย ชุดข้อมูลเสียงอีก 3 ชุด สำหรับฝึกฝน แบบจำลองเสียง และแบบจำลองภาษา (Language Model) ชุดสำหรับทดสอบเพื่อการพัฒนา และชุดสำหรับทดสอบเพื่อการประเมินผล

1. ชุดประโยค

ประโยคที่ใช้ในการอัดเสียงในฐานข้อมูลนี้ จะถูกตัดเครื่องหมายหรือสัญลักษณ์พิเศษออกทั้งหมด โดยการออกแบบชุดข้อมูลที่ใช้สำหรับบันทึกเสียงนั้น แบ่งได้เป็นชุดประโยค 2 ชุดใหญ่ ๆ ดังนี้

1. ชุดหน่วยเสียงสมดุล (Phonetically Distributed Set, PD)
2. ชุดประโยคที่ครอบคลุมคำศัพท์ภาษาไทยจำนวน 5,000 คำ

ในการหาประสิทธิภาพของการตรวจหาสระในเสียงพูดต่อเนื่องภาษาไทยของวิทยานิพนธ์นี้ จะใช้เพียงชุดหน่วยเสียงสมดุลเท่านั้น ซึ่งประกอบด้วย จำนวนประโยคทั้งหมด 801 ประโยค ซึ่งมาจากจำนวนคำศัพท์ทั้งหมด 2,269 คำ ในจำนวนประโยคทั้งหมดนั้นคิดเป็น 7,847 พยางค์

2. ผู้พูด

จำนวนผู้พูดทั้งหมดที่บันทึกเสียง ประกอบด้วย ผู้พูดจากมหาวิทยาลัยสงขลานครินทร์ จำนวน 100 คน ประกอบด้วยเพศชายจำนวน 50 คน และเพศหญิงจำนวน 50 คน จากมหาวิทยาลัยเทคโนโลยีมหานคร จำนวน 100 คน ประกอบด้วยเพศชายจำนวน 50 คน และเพศหญิงจำนวน 50 คน และจากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ จำนวน 48 คน ประกอบด้วยเพศชายจำนวน 24 คน และเพศหญิงจำนวน 24 คน รวมผู้พูดทั้งสิ้น 248 คน ซึ่งเป็นเพศชายจำนวน 124 คน และเพศหญิงจำนวน 124 คน ดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 จำนวนของผู้พูดจากสถานที่ต่างๆ ที่ใช้สำหรับการบันทึกเสียง

สถานที่บันทึกเสียง	ชาย	หญิง	รวม
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ	24	24	48
มหาวิทยาลัยสงขลานครินทร์	50	50	100
มหาวิทยาลัยเทคโนโลยีมหานคร	50	50	100

3. การบันทึกเสียง

1. สถานที่สำหรับใช้ในการบันทึกเสียง

การบันทึกเสียงในสภาพแวดล้อมแบบห้องเงียบ สถานที่ซึ่งใช้สำหรับการบันทึกเสียงประเภทนี้ คือห้องเงียบ และสำหรับการบันทึกเสียงที่มีสภาพแวดล้อมแบบสำนักงาน สถานที่ซึ่งใช้สำหรับการบันทึกเสียงประเภทนี้ คือสำนักงานทั่วไป

2. ไมโครโฟน

การบันทึกเสียงในสภาพแวดล้อมแบบห้องเงียบ ไมโครโฟนซึ่งใช้ในการบันทึกเสียงประเภทนี้ คือไมโครโฟนระดับคุณภาพสูง และการบันทึกเสียงที่มีสภาพแวดล้อมแบบสำนักงาน ไมโครโฟนซึ่งใช้สำหรับการบันทึกเสียงประเภทนี้ คือ ไมโครโฟนระดับคุณภาพปานกลาง

4. หน่วยเสียงภาษาไทย ที่ปรากฏในฐานข้อมูล

1. เสียงพยัญชนะต้นเดี่ยว (Initial Consonant)

หน่วยเสียงพยัญชนะต้นเดี่ยวในเสียงภาษาไทย สามารถแสดงได้ดังตารางที่ 4.2 และตัวอย่างของการใช้หน่วยเสียงพยัญชนะต้นเดี่ยว แสดงได้ดังตารางที่ 4.3

ตารางที่ 4.2 หน่วยเสียงพยัญชนะต้นเดี่ยว ของเสียงภาษาไทย

	Bilabial	Labio-Dental	Alveolar	Post-Alveolar	Palatal	Velar	Glottal
Plosive	p ph b		t th d			k kh	z
Nasal	m		n			ng	
Fricative		f	s				h
Affricate				c ch			
Trill			l				
Approximant					j	w	
Lateral Approximant			l				

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.3 ตัวอย่างของการใช้หน่วยเสียงพยัญชนะต้นเดี่ยว

พยัญชนะต้น			
เดี่ยว	ตัวอย่าง	เดี่ยว	ตัวอย่าง
p	ปาก	m	ไม่
t	เต้น, ฤฦี	n	นาน, เณร
z	อ่าน	ng	เงิน
c	จะ	l	เล่น, กีฬา
k	ก่อน	r	รอ, ฤทัย
ph	พบ, ภัย, ผ่าน	f	ฝน, ฟัน
th	ทิ้ง, ธง, เถ่า, ฐาน, มณโฑ	s	สาย, ศิลา, รักษา, ซ่อน
ch	ชอบ, เฌอ	h	โหน, เฮฮา
kh	คน, เขิน, ฆ่า	w	ว่า
b	บอก	j	ย้อน, หญิง
d	ด้าน, ฌฎา		

2. เสียงพยัญชนะท้ายเดี่ยว (Final Consonant)

หน่วยเสียงพยัญชนะท้ายเดี่ยวในเสียงภาษาไทย สามารถแสดงได้ดังตารางที่ 4.4 และตัวอย่างของการใช้หน่วยเสียงพยัญชนะท้ายเดี่ยว แสดงได้ดังตารางที่ 4.5

ตารางที่ 4.4 หน่วยเสียงพยัญชนะท้ายเดี่ยวในเสียงภาษาไทย

	Bilabial	Labio-Dental	Alveolar	Post-Alveolar	Palatal	Velar	Glottal
Plosive	p^		t^			k^	
Nasal	m^		n^			ng^	
Fricative		f^	s^				
Affricate				ch^			
Trill			l				
Approximant					j^	w^	
Lateral Approximant			l^				

ตารางที่ 4.5 ตัวอย่างของการใช้หน่วยเสียงพยัญชนะท้ายเดี่ยว

พยัญชนะท้าย			
เสียงภาษาไทย		เสียงทับศัพท์	
เดี่ยว	ตัวอย่าง	เดี่ยว	ตัวอย่าง
p^	พบ	f^	กราฟ
t^	เกร็ด	l^	แอล
k^	ปาก	s^	เอส
n^	หาร	ch^	คลัช
m^	ลม		
ng^	ฟาง		
j^	ขาย		
w^	กาว		

3. เสียงสระเดี่ยว (Vowel)

ตัวอย่างของหน่วยเสียงสระเดี่ยวในภาษาไทย สามารถแสดงได้ดังตารางที่ 4.6

ตารางที่ 4.6 ตัวอย่างของหน่วยเสียงสระเดี่ยวในภาษาไทย

สระ			
เดี่ยว	ตัวอย่าง	เดี่ยว	ตัวอย่าง
a	อะ	ee	เอ
aa	อา	x	แอะ
i	อิ	xx	แอ
ii	อี	o	โอะ
v	อึ	oo	โอ
vv	อือ	@	เอาะ
u	อุ	@@	ออ
uu	อู	q	เออะ
e	เอะ	qq	เออ

4. หน่วยเสียงประสม (Cluster Consonant)

- เสียงควบกล้ำ (Cluster Consonant) ได้แก่ pr phr tr kr khr pl phl kl khl kw และ khw ตัวอย่างของการใช้เสียงควบกล้ำ แสดงได้ดังตารางที่ 4.7

ตารางที่ 4.7 ตัวอย่างของการใช้เสียงควบกล้ำ

พยัญชนะต้น			
เสียงภาษาไทย		เสียงทับศัพท์	
ผสม	ตัวอย่าง	ผสม	ตัวอย่าง
pr	ประสาน	br	เบรน
phr	พราน	bl	บลู
tr	เตรียม	fr	ฟราย
kr	กราบ	fl	เฟลม
khr	คร่า	dr	ดรากอน
pl	ปลา		
phl	พลาด		
thr	จันทรา		
kl	เกลอ		
khl	เคลื่อน		
kw	กวาง		
khw	ขวา		

- เสียงสระผสม (Dipthong) ได้แก่ ia iia va vva ua และ uua ตัวอย่างของการใช้เสียงสระผสม แสดงได้ดังตารางที่ 4.8

ตารางที่ 4.8 ตัวอย่างของการใช้เสียงสระผสม

สระผสม	ตัวอย่าง	สระผสม	ตัวอย่าง
ia	เอียะ	vaa	เอือ
iaa	เอียะ	ua	อัวะ
va	เอือะ	uaa	อัว

ฐานข้อมูลเสียงตัวเลข (Spoken Digit Corpus)

ฐานข้อมูลเสียงพูดประเภทนี้ เป็นฐานข้อมูลเสียงพูดภาษาไทยเช่นเดียวกันกับฐานข้อมูลเสียง Lotus แต่มีความแตกต่างกัน กล่าวคือ ฐานข้อมูลเสียง Lotus เป็นฐานข้อมูลเสียงพูดภาษาไทยขนาดใหญ่ ซึ่งคำพูดต่างๆ ที่ใช้ในการบันทึก มาจากบทความข่าวและบทความทั่วไป ซึ่งบทความต่างๆ เหล่านั้นประกอบด้วยคำศัพท์ต่าง ๆ มากมาย รวมถึงคำพูดที่แทนตัวเลข ในขณะที่ฐานข้อมูลเสียงตัวเลขนี้ เป็นฐานข้อมูลที่ประกอบด้วยคำพูดที่แทนตัวเลขเท่านั้น ฐานข้อมูลเสียงตัวเลขที่ใช้ในวิทยานิพนธ์นี้แบ่งออกเป็น 2 ประเภท คือ

1. ฐานข้อมูลตัวเลขที่บันทึกในห้องเงียบ (Clean Number)
2. ฐานข้อมูลตัวเลขที่บันทึกผ่านทางโทรศัพท์ (Telephone Number)

รายละเอียดของฐานข้อมูลเสียงตัวเลข ในแต่ละประเภท มีดังต่อไปนี้

1. ฐานข้อมูลตัวเลขที่บันทึกในห้องเงียบ

สถานที่ซึ่งใช้สำหรับการบันทึกฐานข้อมูลเสียงตัวเลขนี้ คือ ห้องเงียบ โดยที่ในการบันทึกเสียงนั้น จะทำการพูดชุดตัวเลขผ่านไมโครโฟนที่ใช้สำหรับการบันทึกเสียง ซึ่งเสียงพูดตัวเลขที่บันทึกได้นี้จะไม่ถูกรบกวนด้วยสัญญาณรบกวนจากสภาพแวดล้อมภายนอก

1. ชุดประโยค

   ชุดประโยคตัวเลขที่ใช้ในการบันทึก เป็นชุดประโยคตัวเลขที่กำหนดขึ้นเอง ซึ่งประกอบด้วย ตัวเลขตั้งแต่ 0 ถึง 9 จำนวน 290 ประโยค หรือ 2,620 พยางค์ เช่น ชุดประโยคตัวเลข "ศูนย์ศูนย์ ศูนย์หนึ่ง ศูนย์สอง ศูนย์สาม ศูนย์สี่ ศูนย์ห้า" เป็นต้น

2. ผู้พูด

   จำนวนผู้พูดทั้งหมดที่ทำการบันทึกเสียง ประกอบด้วย ผู้พูดที่เป็นเพศชายจำนวน 3 คน และเพศหญิงจำนวน 5 คน รวมผู้พูดทั้งสิ้น 8 คน

2. ฐานข้อมูลตัวเลขที่บันทึกผ่านทางโทรศัพท์

เป็นฐานข้อมูลเสียงตัวเลขเช่นเดียวกับฐานข้อมูลตัวเลขที่บันทึกในห้องเงียบ แตกต่างกันที่ในการบันทึกเสียงนั้น จะทำการพูดชุดตัวเลขผ่านทางโทรศัพท์ แทนการพูดผ่านทางไมโครโฟน ซึ่งมีผลทำให้คุณภาพเสียงที่ได้ต่ำลง เนื่องจากข้อมูลเสียงที่พูดผ่านทางโทรศัพท์นี้ จะถูกรบกวนด้วยสัญญาณรบกวนที่ความถี่ต่างๆจากสายโทรศัพท์ รวมไปถึง ข้อจำกัดที่เกิดขึ้นจากช่วงความถี่ของข้อมูลเสียงที่สามารถผ่านได้ในสายโทรศัพท์ ทำให้ลักษณะของสัญญาณเสียงที่บันทึกได้ต่ำลง

1. ชุดประโยค

ชุดประโยคตัวเลขที่ใช้ในการบันทึก เป็นชุดประโยคตัวเลขที่กำหนดขึ้นเองเช่นเดียวกัน ซึ่งประกอบด้วย ตัวเลขตั้งแต่ 0 ถึง 9 จำนวน 200 ประโยค หรือ 1,172 พยางค์ โดยที่ชุดตัวเลขแต่ละประโยคนั้นมีเสียงพูดตัวเลข 8 ถึง 12 พยางค์

2. ผู้พูด

จำนวนผู้พูดทั้งหมดที่บันทึกเสียง ประกอบด้วย ผู้พูดที่เป็นเพศชายจำนวน 5 คน และเพศหญิงจำนวน 5 คน รวมผู้พูดทั้งสิ้น 10 คน

ตัวอย่างของข้อมูลเสียงตัวเลขและไฟล์กำกับประโยค แสดงได้ดังตารางที่ 4.9

ตารางที่ 4.9 ตัวอย่างของข้อมูลเสียงตัวเลขและไฟล์กำกับเสียง

ข้อมูลเสียงตัวเลข (Number)	ไฟล์กำกับเสียง (Transcription)	ข้อมูลเสียงตัวเลข (Number)	ไฟล์กำกับเสียง (Transcription)
ศูนย์	s oo n^	ห้า	h aa
หนึ่ง	n v ng^	หก	h o k^
สอง	s o ng^	เจ็ด	c e t^
สาม	s aa m^	แปด	p xx t^
สี่	s ii	เก้า	k a w^

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

ผลการตรวจหาสระ

ในบทนี้ ได้นำเสนอผลการทดลองและการวิเคราะห์ผลการทดลองในแต่ละขั้นตอนของการตรวจหาสระด้วยวิธีการที่นำเสนอ ผลการทดลองการตรวจหาสระด้วยวิธีการของ Howitt และการเปรียบเทียบความถูกต้องของการตรวจหาสระระหว่างวิธีที่นำเสนอกับวิธีของ Howitt ซึ่งรายละเอียดนั้น จะได้กล่าวต่อไป

ผลการทดสอบการหาความก้องหรือความไม่ก้องของเสียง

ขั้นตอนแรกของการตรวจหาสระ คือ การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง โดยใช้ค่าอัตสหสัมพันธ์ ซึ่งพบว่ากรอบสัญญาณที่เป็นสัญญาณเสียงก้องนั้นจะมีค่าอัตสหสัมพันธ์ที่สูงที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ มากกว่ากรอบสัญญาณที่เป็นเสียงไม่ก้อง

ในการทดสอบการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงในวิทยานิพนธ์นี้ ได้ทำการทดสอบ 2 วิธี คือ

1. การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยเปรียบเทียบกับค่าขีดแบ่ง
2. การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยใช้การจำแนกกลุ่มด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น

รายละเอียดในการทดสอบและผลการทดสอบที่ได้ของแต่ละวิธี แสดงได้ดังต่อไปนี้

1. การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยเปรียบเทียบกับค่าขีดแบ่ง

การทดสอบการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงประเภทนี้ จะนำค่าอัตสหสัมพันธ์ที่สูงที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงแต่ละกรอบที่ได้มาเปรียบเทียบกับค่าขีดแบ่งที่ได้กำหนดขึ้น เพื่อระบุว่ากรอบสัญญาณเสียงนั้นเป็นกรอบสัญญาณเสียงก้องหรือกรอบสัญญาณเสียงไม่ก้อง ดังนี้ กรอบสัญญาณเสียงก้องคือกรอบที่มีค่าอัตสหสัมพันธ์ที่สูงที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์มากกว่าค่าขีดแบ่ง และในทางตรงกันข้ามกรอบสัญญาณเสียงไม่ก้องคือกรอบที่มีค่าอัตสหสัมพันธ์ที่สูงที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์น้อยกว่าค่าขีดแบ่ง

ตารางที่ 5.1 แสดงผลการทดสอบการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงในฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ ฐานข้อมูลเสียง Lotus ที่บันทึกในห้อง

ปกติ ฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ ฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยโทรศัพท์ในห้องปกติ และฐานข้อมูลเสียง TIMIT ตามลำดับ

ตารางที่ 5.1 การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยเปรียบเทียบกับค่าขีดแบ่ง ของข้อมูลเสียงต่างๆ

ฐานข้อมูล	ชนิดของกรอบสัญญาณอ้างอิง	จำนวนกรอบสัญญาณเสียงทั้งหมด	ชนิดที่แบ่งแยกได้	
			เสียงก้อง (%)	เสียงไม่ก้อง (%)
Lotus (Clean Speech)	เสียงก้อง	253,229	79.89	20.11
	เสียงไม่ก้อง	209,126	17.44	82.56
Lotus (Office Speech)	เสียงก้อง	262,329	84.84	15.16
	เสียงไม่ก้อง	208,088	23.76	76.24
Spoken Digit (Clean Number)	เสียงก้อง	31,752	83.85	16.15
	เสียงไม่ก้อง	30,601	17.77	82.23
Spoken Digit (Telephone Number)	เสียงก้อง	12,249	76.73	23.27
	เสียงไม่ก้อง	19,773	19.32	80.68
TIMIT	เสียงก้อง	189,734	77.93	22.07
	เสียงไม่ก้อง	115,492	19.91	80.09

2. การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยใช้การจำแนกกลุ่มด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น

เพื่อป้องกันความผิดพลาดที่เกิดขึ้นจากการหาความก้องหรือไม่ก้องของสัญญาณเสียง โดยเปรียบเทียบกับค่าขีดแบ่ง จึงได้ทำการทดสอบการหาความก้องหรือไม่ก้องของสัญญาณเสียงโดยใช้การจำแนกกลุ่มออกเป็นกลุ่มของสัญญาณเสียงก้องและกลุ่มของสัญญาณเสียงไม่ก้อง ด้วยระเบียบวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น ซึ่งเวกเตอร์ลักษณะสำคัญของชุดข้อมูลฝึกสำหรับใช้ในการจำแนกนั้น คือ ค่าอัตสหสัมพันธ์ที่สูงที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงแต่ละกรอบ โดยหวังว่าคะแนนที่ได้จากการจำแนกกรอบสัญญาณเสียงออกเป็นเสียงก้องหรือเสียงไม่ก้องนั้น จะช่วยเพิ่มความมั่นใจในการระบุว่ากรอบสัญญาณเสียงที่จำแนกไปตามกลุ่มนั้นมีความถูกต้องเพียงใด

การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงในประเภทนี้ ได้ทำการทดสอบกับฐานข้อมูลเสียง Lotus เท่านั้น เนื่องจากฐานข้อมูลเสียง Lotus มีจำนวนประโยคเสียงพูดเพียงพอสำหรับการทำชุดข้อมูลฝึกและข้อมูลทดสอบ ตารางที่ 5.2 แสดงผลการจำแนกกรอบสัญญาณเสียงออกเป็นกรอบสัญญาณเสียงก้องและกรอบสัญญาณเสียงไม่ก้องในฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ (Clean) และฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ (Office) ตามลำดับ

ตารางที่ 5.2 การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยใช้การจำแนกกลุ่มด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นในฐานข้อมูลเสียง Lotus

ฐานข้อมูล	ชนิดของกรอบสัญญาณอ้างอิง	จำนวนกรอบสัญญาณเสียงทั้งหมด	ชนิดที่แบ่งแยกได้	
			เสียงก้อง (%)	เสียงไม่ก้อง (%)
Lotus (Clean Speech)	เสียงก้อง	76,264	77.78	22.22
	เสียงไม่ก้อง	64,017	16.5	83.50
Lotus (Office Speech)	เสียงก้อง	71,644	74.56	25.44
	เสียงไม่ก้อง	58,081	16.17	83.83

3. การวิเคราะห์ผลการการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง

เมื่อพิจารณาผลการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงในตารางที่ 5.1 และตารางที่ 5.2 พบว่า ในการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงในฐานข้อมูลแต่ละประเภทด้วยการใช้ค่าอัตสหสัมพันธ์ที่มากที่สุดในช่วงความถี่ 60 เฮิรตซ์ ถึง 320 เฮิรตซ์ของกรอบสัญญาณเสียงแต่ละกรอบเป็นลักษณะทางสวนสัทศาสตร์นั้น ได้ความถูกต้องโดยเฉลี่ยเท่ากับ 80.50% และ 79.92% ตามลำดับ

เมื่อพิจารณาผลการการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง ในตารางที่ 5.1 พบว่าความถูกต้องที่ได้จากฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ มีความถูกต้องมากที่สุด และในฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยโทรศัพท์ในห้องปกติได้ความถูกต้องน้อยที่สุด ที่เป็นเช่นนี้เนื่องจาก สัญญาณเสียงที่ได้จากการบันทึกด้วยโทรศัพท์ในห้องปกตินั้น จะมีสัญญาณรบกวนทั้งที่ความถี่สูงและความถี่ต่ำปนอยู่มาก จึงทำให้ค่าอัตสหสัมพันธ์ที่ได้นั้นผิดเพี้ยนไป ส่งผลให้การตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงมีความผิดพลาดขึ้น

ผลการทดสอบการหาตำแหน่งของสระ

ในขั้นตอนนี้ จะเป็นการเลือกตำแหน่งที่เหมาะสมจะเป็นตำแหน่งสระ โดยการคัดเลือกยอดที่สูงเด่นจากพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของกรอบสัญญาณเสียงก้องที่ได้จากขั้นตอนแรก ด้วยระเบียบวิธีคอนเวกซ์ฮัลล์ ซึ่งยอดสูงเด่นที่ได้นี้ถูกกำหนดให้เป็นตำแหน่งของสระ

ตำแหน่งที่ตรวจหาได้นั้น อาจจะประกอบด้วยตำแหน่งที่เป็นของหน่วยเสียงสระจริง หรือตำแหน่งของหน่วยเสียงอื่นที่ไม่ใช่หน่วยเสียงสระ อย่างไรก็ตาม สามารถจำแนกตำแหน่งที่ตรวจหาได้ออกเป็น 3 ประเภทคือ

1. ตำแหน่งที่ถูกต้อง (Correct)

   ตำแหน่งที่ถูกต้อง หมายถึง เมื่อทำการเปรียบเทียบตำแหน่งที่ตรวจหาได้ กับขอบเขตทางเวลาของหน่วยเสียงที่ระบุในไฟล์กำกับเสียง แล้วพบว่าตำแหน่งที่ตรวจหาได้นั้นอยู่ในขอบเขตของช่วงเวลาที่เป็นหน่วยเสียงสระ จะถือว่าตำแหน่งที่ตรวจหาได้นั้นเป็นตำแหน่งที่ถูกต้อง

2. ตำแหน่งเกิน (Insertion)

   เมื่อทำการเปรียบเทียบตำแหน่งที่ตรวจหาได้ กับขอบเขตทางเวลาของหน่วยเสียงที่ระบุในไฟล์กำกับเสียง แล้วพบว่าตำแหน่งที่ตรวจหาได้นั้นอยู่ในขอบเขตของช่วงเวลาที่เป็นหน่วยเสียงอื่นที่ไม่ใช่หน่วยเสียงสระ จะถือว่าตำแหน่งที่ตรวจหาได้นั้นเป็นตำแหน่งที่ไม่ถูกต้อง และเรียกตำแหน่งที่ตรวจหาได้นั้นว่า ตำแหน่งเกิน

3. ตำแหน่งขาด (Deletion)

   เมื่อพิจารณาขอบเขตทางเวลาของหน่วยเสียงสระที่ระบุในไฟล์กำกับเสียง แล้วพบว่าไม่มีตำแหน่งที่ตรวจหาได้ปรากฏขึ้น จะถือว่าสระที่ตำแหน่งนี้เกิดความผิดพลาดขึ้น และเรียกตำแหน่งนี้ว่า ตำแหน่งขาด

จำนวนความถูกต้องและความผิดพลาดที่เกิดขึ้น จากแต่ละตำแหน่งที่ตรวจหาได้นั้น สามารถนำมาวิเคราะห์ประสิทธิภาพของระบบการตรวจหาสระได้ ซึ่งในวิทยานิพนธ์นี้ได้รายงานออกมาในค่าต่างๆ ดังนี้ คือ

1. เปอร์เซ็นต์ความถูกต้อง (Accuracy) สามารถคำนวณได้จากสมการที่ 5.1

$$\%accuracy = \left(\frac{Total - TotalError}{Total} \right) \times 100 \qquad (5.1)$$

2. เปอร์เซ็นต์ความแม่นยำ (Precision) สามารถคำนวณได้จากสมการที่ 5.2

$$\%precision = \left(\frac{Correct}{Correct + Ins.Error}\right) \times 100 \qquad (5.2)$$

3. เปอร์เซ็นต์ความครอบคลุม (Recall) สามารถคำนวณได้จากสมการที่ 5.3

$$\%recall = \left(\frac{Correct}{Correct + Del.Error}\right) \times 100 \qquad (5.3)$$

โดยที่ *Ins.Error* คือ จำนวนตำแหน่งที่เกิน

Del.Error คือ จำนวนตำแหน่งที่ขาด

Total คือ จำนวนสระทั้งหมดที่ระบุในไฟล์กำกับเสียง

TotalError คือ ผลรวมของจำนวนตำแหน่งที่เกินและตำแหน่งที่ขาด

รายละเอียดของผลการตรวจหาตำแหน่งสระ แสดงได้ดังต่อไปนี้

1. ผลการตรวจหาตำแหน่งสระ

การตรวจหาตำแหน่งสระในวิธีนี้ ใช้การคัดเลือกยอดที่สูงเด่นจากค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ของกรอบสัญญาณเสียงก้องที่ได้มาจากขั้นตอนแรก คือการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงโดยการเปรียบเทียบกับค่าขีดแบ่ง ตารางที่ 5.3 และตารางที่ 5.4 แสดงผลการตรวจหาตำแหน่งของสระและประสิทธิภาพของระบบการตรวจหาสระในฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ หรือ Lotus (Clean) ฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ หรือ Lotus (Office) ฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ หรือ Clean Number ฐานข้อมูลเสียงตัวเลข ที่บันทึกด้วยโทรศัพท์ในห้องปกติ หรือ Telephone Number และฐานข้อมูลเสียง TIMIT ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.3 ผลการตรวจหาตำแหน่งของสระในฐานข้อมูลเสียงต่างๆ

ฐานข้อมูลเสียง	สระทั้งหมด (ตัว)	ตำแหน่งที่ถูกต้อง (ตำแหน่ง)	ตำแหน่งที่ขาด (ตำแหน่ง)	ตำแหน่งที่เกิน (ตำแหน่ง)	ความผิดพลาดทั้งหมด (ตำแหน่ง)
Lotus (Clean Speech)	21,514	19,960	1,554	1,680	3,234
Lotus (Office Speech)	23,935	22,184	1,751	1,760	3,511
Spoken Digit (Clean Number)	2,620	2,606	14	96	110
Spoken Digit (Telephone Number)	1,172	1,068	104	81	185
TIMIT	23,547	21,229	2,318	942	3,260

ตารางที่ 5.4 ประสิทธิภาพของระบบการตรวจหาสระ

ฐานข้อมูลเสียง	ความถูกต้อง (%)	ความแม่นยำ (%)	ความครอบคลุม (%)
Lotus (Clean Speech)	84.97	92.34	92.78
Lotus (Office Speech)	85.33	92.65	92.68
Spoken Digit (Clean Number)	95.80	96.47	99.47
Spoken Digit (Telephone Number)	84.21	92.95	91.13
TIMIT	86.16	95.75	90.16

2. การวิเคราะห์ผลการการตรวจหาตำแหน่งที่ได้

เมื่อพิจารณาความถูกต้อง ความแม่นยำ และความครอบคลุมที่เกิดขึ้นในฐานข้อมูลเสียงแต่ละประเภทซึ่งทำการตรวจหาสระด้วยวิธีการที่นำเสนอ พบว่า ความถูกต้องโดยเฉลี่ยของการตรวจหาตำแหน่งสระที่ได้จากฐานข้อมูลแต่ละประเภทนั้นมีค่าเท่ากับ 87.29% และเนื่องจากเสียงพูดในฐานข้อมูลเสียงตัวเลขนั้นมีจำนวนประโยค น้อยกว่าเสียงพูดในฐานข้อมูลเสียง Lotus มาก และที่ตำแหน่งพยัญชนะต้นของเสียงพูดตัวเลขนั้น จะเป็นเสียงเสียดแทรกโดยส่วนใหญ่ ในขณะที่ฐานข้อมูลเสียง Lotus นั้น ที่ตำแหน่งพยัญชนะต้นและพยัญชนะท้าย จะประกอบด้วย

หน่วยเสียงที่เป็นเสียงก้องจำนวนมากมาย ทำให้การตรวจหาตำแหน่งสระ ย่อมมีความผิดพลาดมากกว่า ด้วยเหตุนี้ประสิทธิภาพในการตรวจหาสระในฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบจึงดีที่สุด แต่ประสิทธิภาพในการตรวจหาสระในฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยโทรศัพท์ในห้องปกติที่ได้นั้น ไม่สูงนัก แม้จะมีจำนวนประโยคน้อยก็ตาม เนื่องจากเสียงพูดในฐานข้อมูลประเภทนี้ถูกรบกวนด้วยสัญญาณรบกวนมาก จึงส่งผลให้เกิดความผิดพลาดขึ้นในทุกๆ ขั้นตอนของการตรวจหาสระ นอกจากนี้ความผิดพลาดที่เกิดขึ้นจากขั้นตอนที่ 1 ย่อมส่งผลให้ประสิทธิภาพในการตรวจหาสระด้วยวิธีที่นำเสนอนี้ลดลงด้วยเช่นกัน

การทดลองเพื่อหาความผิดพลาดที่เกิดขึ้นจากแต่ละขั้นตอนในการตรวจหาสระนั้น มีลักษณะคล้ายกับ การทดลองเพื่อทดสอบประสิทธิภาพของวิธีการตรวจหาตำแหน่งสระที่นำเสนอ แต่มีความแตกต่างกันที่ขั้นตอนของการตรวจหาความก้องและไม่ก้องของสัญญาณเสียง ซึ่งสามารถแบ่งออกได้เป็น 2 กรณี คือ กรณีที่ 1 คือ กรณีที่ทราบความก้องของเสียงจากการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง และกรณีที่ 2 คือ กรณีที่ทราบความก้องของกรอบสัญญาณเสียงแน่นอน

ตารางที่ 5.5 การเปรียบเทียบความถูกต้องของการตรวจหาสระที่ได้จากวิธีอ้างอิงและวิธีที่นำเสนอ

ฐานข้อมูลเสียง	วิธีตรวจหาสระ	ขั้นตอนที่ 1		ขั้นตอนที่ 2	
		ความถูกต้อง (%)	ความผิดพลาด (%)	ความถูกต้อง (%)	ความผิดพลาด (%)
Lotus (Clean Speech)	กรณีที่ 1	81.23	18.77	84.97	15.03
	กรณีที่ 2	100	0	85.91	14.09
Lotus (Office Speech)	กรณีที่ 1	80.54	19.46	85.33	14.67
	กรณีที่ 2	100	0	86.06	13.94
Spoken Digit (Clean Number)	กรณีที่ 1	83.06	16.94	95.8	4.2
	กรณีที่ 2	100	0	95.98	4.02
Spoken Digit (Telephone Number)	กรณีที่ 1	78.71	21.29	84.21	15.79
	กรณีที่ 2	100	0	88.08	11.92
TIMIT	กรณีที่ 1	79.01	20.99	86.16	13.84
	กรณีที่ 2	100	0	87.59	12.41

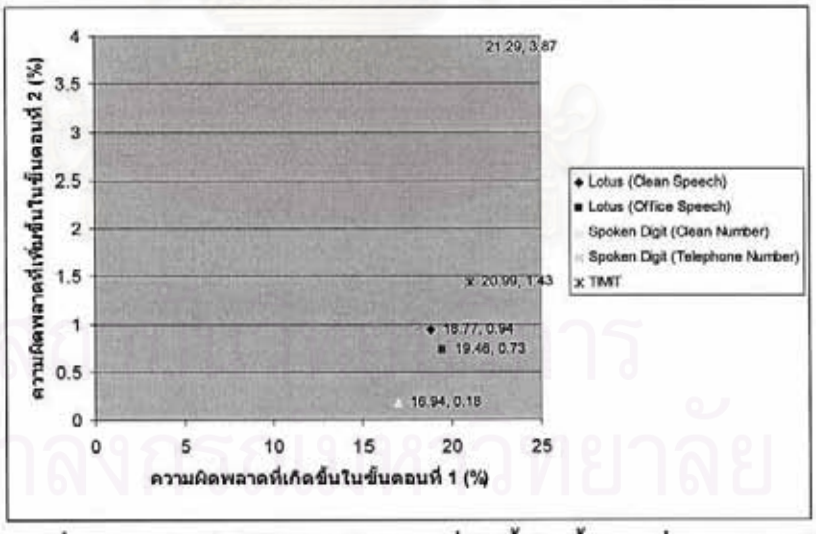
ขั้นตอนที่ 1 และขั้นตอนที่ 2 ในตารางที่ 5.5 และตารางที่ 5.6 คือ การตรวจหาความก้องและไม่ก้องของกรอบสัญญาณเสียง และการคัดเลือกตำแหน่งที่เหมาะสมจะเป็นสระตามลำดับ จากตารางที่ 5.5 สามารถหาผลต่างของความผิดพลาดที่เกิดขึ้นจากวิธีที่นำเสนอและวิธีอ้างอิงของแต่ละขั้นตอนได้ ซึ่งแสดงได้ดังตารางที่ 5.6

ตารางที่ 5.6 ผลต่างของความผิดพลาดที่เกิดขึ้นจากวิธีที่นำเสนอและวิธีอ้างอิง ของแต่ละขั้นตอน

ฐานข้อมูลเสียง	ผลต่างของความผิดพลาดที่เกิดขึ้นจากวิธีที่นำเสนอและวิธีอ้างอิง (%)	
	ขั้นตอนที่ 1	ขั้นตอนที่ 2
Lotus (Clean Speech)	18.77	0.94
Lotus (Office Speech)	19.46	0.73
Spoken Digit (Clean Number)	16.94	0.18
Spoken Digit (Telephone Number)	21.29	3.87
TIMIT	20.99	1.43

ผลต่างของความผิดพลาดที่เกิดขึ้นจากการทดลองในกรณีที่ 1 และกรณีที่ 2 จากแต่ละขั้นตอนของฐานข้อมูลแต่ละประเภทที่แสดงในตารางที่ 5.6 นั้น แสดงให้เห็นถึงความผิดพลาดที่เกิดขึ้นจากขั้นตอนที่ 1 มีผลต่อความผิดพลาดที่เกิดขึ้นในขั้นตอนที่ 2 อย่างไร ดังแสดงในรูปที่ 5.1

รูปที่ 5.1 ความสัมพันธ์ของความผิดพลาดที่เกิดขึ้นในขั้นตอนที่ 1 และขั้นตอนที่ 2



จากรูปที่ 5.1 จะเห็นได้ว่าความผิดพลาดที่เกิดขึ้นในขั้นตอนที่ 1 ส่งผลให้เกิดความผิดพลาดเพิ่มขึ้นในขั้นตอนที่ 2 ของแต่ละฐานข้อมูลมีขนาดต่างๆ กันซึ่งโดยเฉลี่ยพบว่า เมื่อเกิดความผิดพลาดขึ้นในขั้นตอนที่ 1 16.49% จะส่งผลให้ความผิดพลาดในขั้นตอนที่ 2 เพิ่มขึ้น 1.06% ดังนั้น จึงสามารถสรุปได้ว่า ในการทดสอบการตรวจหาตำแหน่งของสระด้วยวิธีการที่นำเสนอนี้

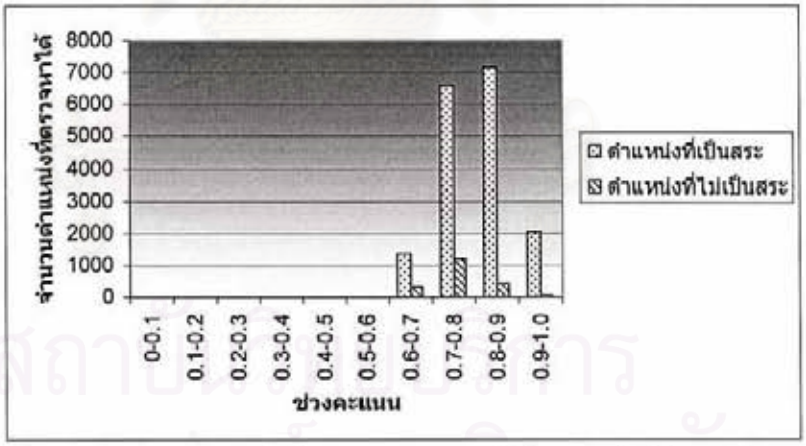
ซึ่งได้ความถูกต้องของตำแหน่งที่ตรวจหาได้โดยเฉลี่ยเท่ากับ 87.29% หรือเกิดความผิดพลาดขึ้น 12.71% นั้น เป็นความผิดพลาดที่มาจากขั้นตอนที่ 1 คือการตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียง 1.43% และจากขั้นตอนที่ 2 คือการคัดเลือกตำแหน่งที่เหมาะสมจะเป็นสระ 11.28%

ผลการทดสอบการให้คะแนนของตำแหน่งสระ

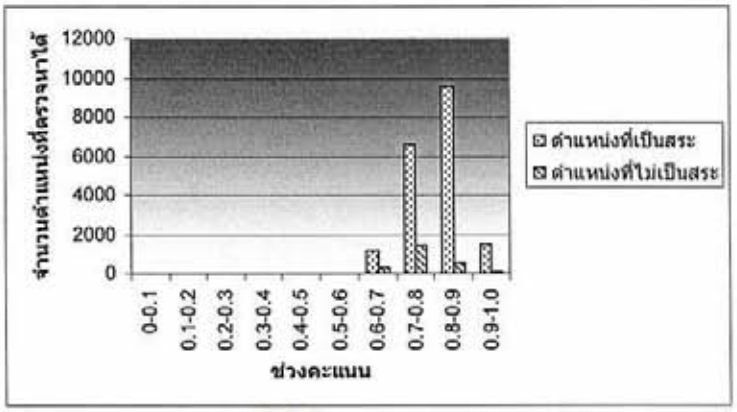
การให้คะแนนตำแหน่งที่ตรวจหาได้นั้น จะทำให้ตำแหน่งนั้นมีความมั่นใจมากขึ้นว่าเหมาะสมที่จะเป็นตำแหน่งของสระจริงหรือไม่ โดยดูจากคะแนนที่คำนวณได้ของแต่ละตำแหน่ง ซึ่งช่วงของคะแนนมีค่าตั้งแต่ 0-1 ถ้าคะแนนของตำแหน่งที่หาได้นั้นมีค่ามาก กล่าวคือ มีค่าใกล้เคียง 1 หมายความว่า ตำแหน่งนั้นมีโอกาสที่จะเป็นตำแหน่งของสระมาก และในทางตรงกันข้าม ถ้าตำแหน่งที่ตรวจหาได้นั้นมีคะแนนน้อย ตำแหน่งนั้นจะมีโอกาสที่จะเป็นตำแหน่งของสระน้อย

1. ผลการทดสอบ

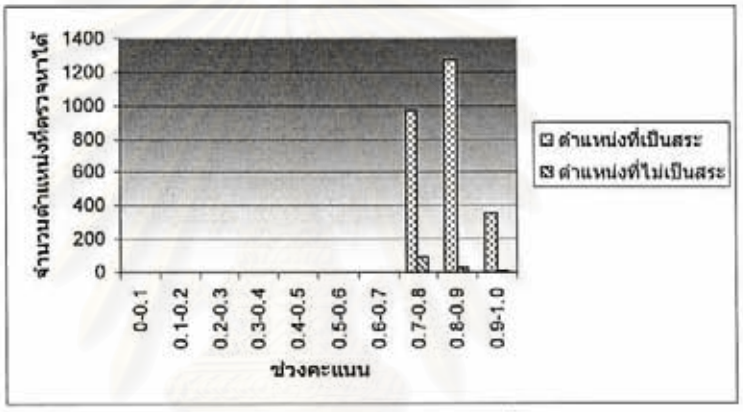
ผลการทดสอบการให้คะแนนของตำแหน่งที่ตรวจหาได้ในฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ ฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ ฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ และฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยโทรศัพท์ในห้องปกติ แสดงด้วยฮิสโทรแกรมดังรูปที่ 5.2 ถึงรูปที่ 5.5 ตามลำดับ



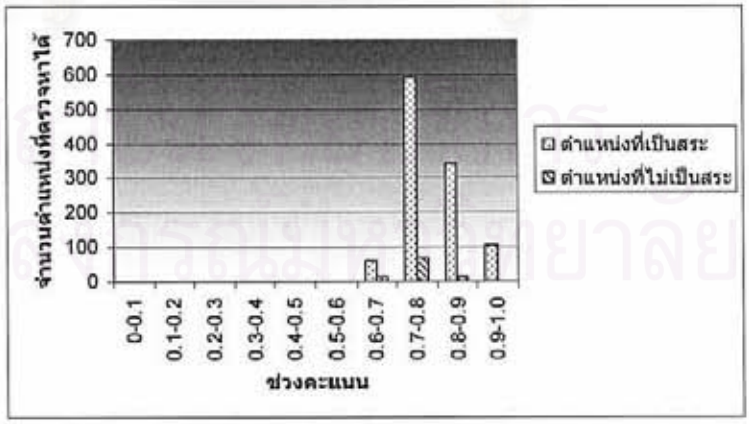
รูปที่ 5.2 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ในฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ



รูปที่ 5.3 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ของฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ



รูปที่ 5.4 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ของฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ



รูปที่ 5.5 ฮิสโทรแกรมแสดงผลการให้คะแนนของตำแหน่งที่ตรวจหาได้ของฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไทรศัพท์ในห้องปกติ

2. การวิเคราะห์ความถูกต้องของการให้คะแนนตำแหน่งที่ตรวจหาได้

การวิเคราะห์ความถูกต้องในการให้คะแนนของตำแหน่งที่ตรวจหาได้นั้น จะพิจารณาถึงค่าความน่าจะเป็น เมื่อตำแหน่งที่ตรวจหาได้นั้นมีคะแนนในช่วงต่างๆ ซึ่งแสดงได้ดังรูปที่ 5.6

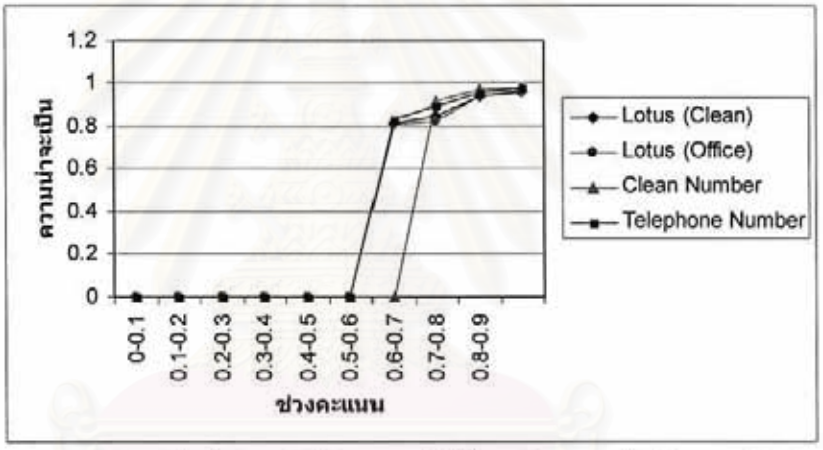
ค่าความน่าจะเป็นของตำแหน่งที่ตรวจหาได้ เมื่อมีคะแนนในช่วงต่างๆ สามารถคำนวณได้จากสมการที่ 5.4

$$P = \frac{N_{Vo}}{N_{Vo} + N_{NVo}} \tag{5.4}$$

โดยที่ P คือ ค่าความน่าจะเป็น

N_{Vo} คือ จำนวนตำแหน่งที่เป็นสระ

N_{NVo} คือ จำนวนตำแหน่งที่ไม่เป็นสระ

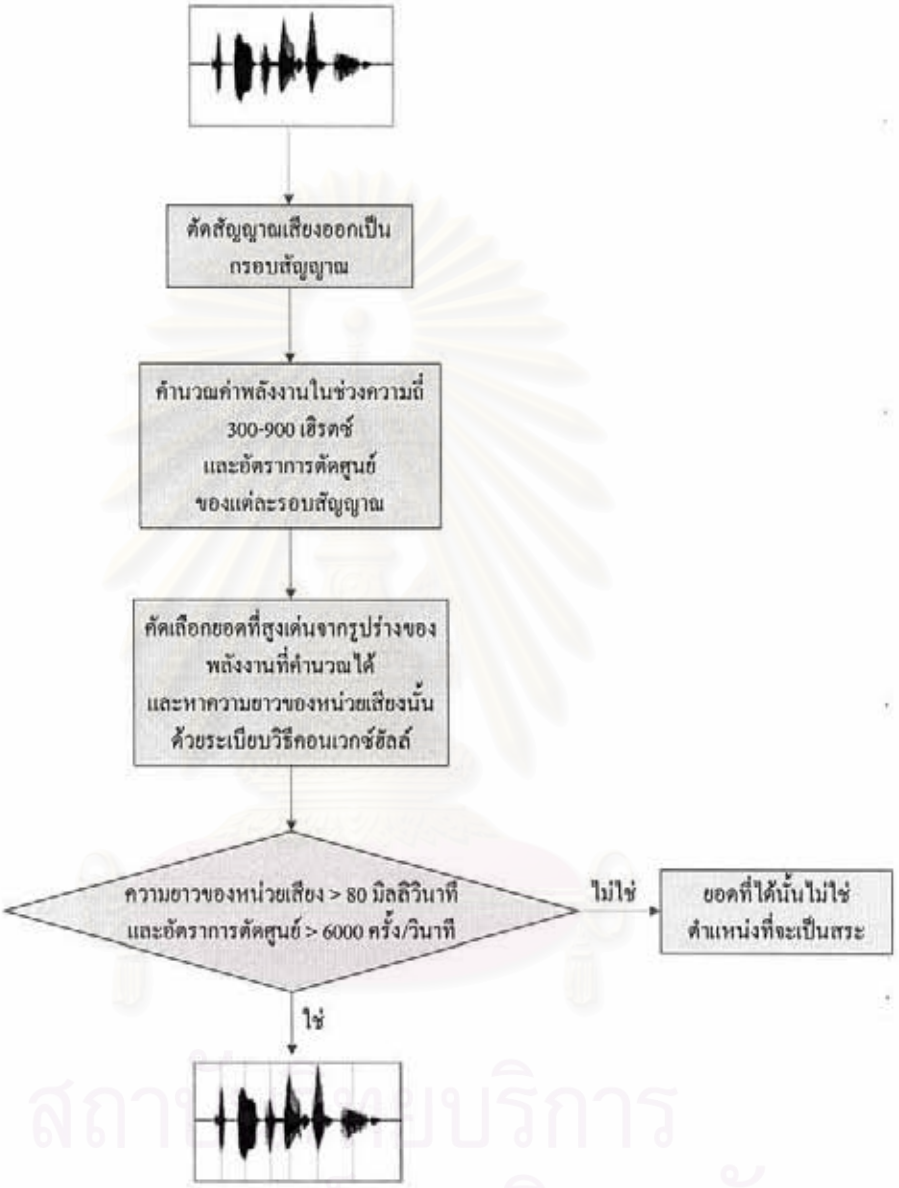


รูปที่ 5.6 ความน่าจะเป็นที่ตำแหน่งที่ตรวจหาได้นั้นจะเป็นสระ เมื่อมีคะแนนในช่วงต่างๆ

การให้คะแนนตำแหน่งสระที่ดีนั้น หมายความว่า ตำแหน่งที่เป็นหน่วยเสียงสระจะต้องมีคะแนนมาก และในทางตรงกันข้าม ตำแหน่งที่ไม่ใช่หน่วยเสียงสระจะต้องมีคะแนนน้อย การทดสอบการให้คะแนนตำแหน่งที่ตรวจหาได้ ซึ่งถูกกำหนดให้เป็นตำแหน่งของหน่วยเสียงสระแสดงด้วยกราฟเส้นในรูปที่ 5.5 ซึ่งอธิบายได้ว่า ตำแหน่งที่ได้คะแนนอยู่ในช่วงคะแนนที่สูงๆ นั้นจะมีความน่าจะเป็นที่ตำแหน่งนั้นจะเป็นตำแหน่งของหน่วยเสียงสระมาก และในทางตรงกันข้าม หากตำแหน่งที่ได้คะแนนอยู่ในช่วงคะแนนที่น้อยลง จะทำให้ความน่าจะเป็นที่ตำแหน่งนั้นจะเป็นตำแหน่งของหน่วยเสียงสระมีค่าน้อยตามไปด้วย

วิธีการตรวจหาตำแหน่งสระของ Howitt

การตรวจหาตำแหน่งสระของ Howitt นั้น มีขั้นตอนและวิธีการ ซึ่งสามารถแสดงได้ดังรูปที่ 5.7



รูปที่ 5.7 วิธีการตรวจหาสระของ Howitt

จากรูปที่ 5.7 นั้นวิธีการตรวจหาสระของ Howitt เริ่มต้นจากการนำสัญญาณเสียงมาตัดออกเป็นกรอบสัญญาณย่อยๆ เพื่อที่จะนำไปคำนวณหาค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์และค่าอัตราการตัดศูนย์ของแต่ละกรอบสัญญาณ ในขั้นตอนต่อมาคือการคัดเลือกยอดที่

สูงเด่นจากรูปร่างของพลังงานที่คำนวณได้และหาความยาวของหน่วยเสียงนั้น ด้วยระเบียบวิธีคอนเวกซ์ฮัลล์ โดยตำแหน่งที่ได้นั้นจะกำหนดให้เป็นตำแหน่งของสระเมื่อความยาวของหน่วยเสียงและอัตราการตัดสูนย์มากกว่าค่าขีดแบ่งที่กำหนด

1. ผลการทดสอบ

การทดสอบประสิทธิภาพของระบบการตรวจหาสระด้วยวิธีการของ Howitt นั้นได้ทำการทดลองกับฐานข้อมูลเสียงประเภทเดียวกับ การทดสอบประสิทธิภาพของระบบการตรวจหาสระด้วยวิธีการที่นำเสนอ ผลการตรวจหาตำแหน่งของสระและประสิทธิภาพของระบบการตรวจหาสระของ Howitt ในฐานข้อมูลเสียงต่างๆ นั้น แสดงได้ดังตารางที่ 5.7 และ 5.8 ตามลำดับ

ตารางที่ 5.7 ผลการตรวจหาตำแหน่งสระด้วยวิธีของ Howittในฐานข้อมูลเสียงต่างๆ

ฐานข้อมูลเสียง	สระทั้งหมด (ตัว)	ตำแหน่งที่ถูกต้อง (ตำแหน่ง)	ตำแหน่งที่ขาด (ตำแหน่ง)	ตำแหน่งที่เกิน (ตำแหน่ง)	ความผิดพลาดทั้งหมด (ตำแหน่ง)
Lotus (Clean Speech)	21,514	18,388	3,126	2,456	5,582
Lotus (Office Speech)	23,935	19,405	4,530	1,702	6,232
Spoken Digit (Clean Number)	2,620	24,73	147	168	315
Spoken Digit (Telephone Number)	1,172	1,044	128	92	220
TIMIT	23,547	20,180	3,367	2,013	5,380

ตารางที่ 5.8 ประสิทธิภาพของระบบการตรวจหาสระด้วยวิธีของ Howitt

ฐานข้อมูลเสียง	ความถูกต้อง (%)	ความแม่นยำ (%)	ความครอบคลุม (%)
Lotus (Clean Speech)	74.05	88.21	85.47
Lotus (Office Speech)	73.97	91.94	81.07
Spoken Digit(Clean Number)	87.98	93.64	94.39
Spoken Digit(Telephone Number)	81.22	91.90	89.08
TIMIT	77.15	90.93	85.70

2. การวิเคราะห์ผลการเปรียบเทียบผลลัพธ์ของการหาตำแหน่งสระระหว่างวิธีของ Howitt และ วิธีที่นำเสนอในวิทยานิพนธ์

ประสิทธิภาพของวิธีการในการตรวจหาตำแหน่งสระที่นำเสนอนั้น ได้นำมาเปรียบเทียบกับประสิทธิภาพของวิธีการตรวจหาสระของ Howitt เพื่อดูว่าระบบการตรวจหาสระที่นำเสนอนั้นมีประสิทธิภาพหรือความถูกต้องมากน้อยแค่ไหน การเปรียบเทียบผลลัพธ์ของการหาตำแหน่งสระระหว่างวิธีของ Howitt และ วิธีที่นำ แสดงได้ดังตารางที่ 5.9

ตารางที่ 5.9 การเปรียบเทียบผลการตรวจหาสระของวิธีการที่นำเสนอในงานวิจัย และวิธีการตรวจหาสระของ Howitt

	วิธีการที่นำเสนอ			วิธีของ Howitt		
ฐานข้อมูล	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Lotus (Clean)	84.98%	92.28%	92.24%	74.38%	81.50%	91.96%
Lotus (Office)	85.33%	92.68%	92.65%	73.96%	81.07%	91.94%
Clean Number	99.47%	95.80%	96.45%	87.98%	94.39%	93.64%
Telephone Number	84.22%	91.13%	92.95%	81.23%	89.08%	91.90%
TIMIT	86.16%	90.16%	95.75%	76.27%	86.59%	89.35%

จากตารางที่ 5.9 สามารถวิเคราะห์การเปรียบเทียบผลการตรวจหาสระระหว่างวิธีการตรวจหาสระที่นำเสนอในวิทยานิพนธ์ และวิธีการตรวจหาสระของ Howitt โดยแยกการวิเคราะห์ระหว่างฐานข้อมูลเสียงพูดภาษาไทย ซึ่งประกอบด้วย ฐานข้อมูล Lotus และ ฐานข้อมูลตัวเลข กับฐานข้อมูลเสียงภาษาอังกฤษ ซึ่งประกอบด้วย ฐานข้อมูล TIMIT ดังต่อไปนี้

ในการทดสอบกับฐานข้อมูลเสียงพูดภาษาไทย การเปรียบเทียบค่าของเปอร์เซ็นต์ความถูกต้อง เปอร์เซ็นต์ความแม่นยำ และเปอร์เซ็นต์ความครอบคลุมที่เกิดขึ้นจากฐานข้อมูลแต่ละประเภท ระหว่างวิธีการตรวจหาสระที่นำเสนอและวิธีการตรวจหาสระของ Howitt พบว่าเปอร์เซ็นต์ความถูกต้อง เปอร์เซ็นต์ความแม่นยำ และเปอร์เซ็นต์ความครอบคลุมที่เกิดขึ้นจากวิธีการตรวจหาสระที่นำเสนอ มีค่ามากกว่าค่าของเปอร์เซ็นต์ความถูกต้อง เปอร์เซ็นต์ความแม่นยำ และเปอร์เซ็นต์ความครอบคลุมที่เกิดขึ้นจากวิธีการตรวจหาสระของ Howitt เนื่องจากวิธีการตรวจหาสระของ Howitt นั้นใช้ค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์เป็นลักษณะทางสวนสัทศาสตร์ในการคัดเลือกตำแหน่งที่เหมาะสมจะเป็นสระ ซึ่งเมื่อพิจารณาค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์ของกรอบสัญญาณเสียงสระ และ ค่าพลังงานในช่วงความถี่ 300 เฮิรตซ์ ถึง 900 เฮิรตซ์ของกรอบสัญญาณเสียงนาสิก เสียงกึ่งสระ เสียงกระทบ เสียงกักแบบก้อง และเสียงข้างลิ้น ซึ่งเป็นส่วนหนึ่งของเสียงก้องแล้ว พบว่ามีค่าไม่ต่างกันมาก จึงส่งผลให้การคัดเลือก

ตำแหน่งที่เหมาะสมจะเป็นสระนั้น มีหน่วยเสียงก้องประเภทอื่นที่ไม่ใช่หน่วยเสียงสระปนเข้ามาเป็นจำนวนมาก ในขณะที่วิธีการตรวจหาสระที่นำเสนอนั้น ใช้พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ เป็นลักษณะทางสวนสัทศาสตร์ในการตรวจหาตำแหน่งที่เหมาะสมจะเป็นสระ ซึ่งค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของกรอบสัญญาณเสียงสระ และ ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของกรอบสัญญาณเสียงก้องประเภทอื่น มีความแตกต่างกันอย่างชัดเจน ทำให้ตำแหน่งของหน่วยเสียงอื่นที่ไม่ใช่หน่วยเสียงสระปนเข้ามาน้อย ทำให้ประสิทธิภาพในการตรวจหาสระด้วยวิธีการที่นำเสนอนั้น ดีว่าประสิทธิภาพในการตรวจหาสระของ Howitt

ในการทดสอบกับฐานข้อมูล TIMIT ซึ่งเป็นเสียงพูดภาษาอังกฤษ ผลการทดสอบที่ได้ยังคงพบว่า ประสิทธิภาพในการตรวจหาสระด้วยวิธีการที่นำเสนอนั้น ดีกว่าประสิทธิภาพในการตรวจหาสระของ Howitt เนื่องจาก ขั้นตอนของวิธีการตรวจหาสระที่นำเสนอนั้น ก่อนที่จะทำการตรวจหาตำแหน่งที่เหมาะสมจะเป็นสระ จะต้องนำสัญญาณนั้นมาตรวจหาความก้องหรือไม่ก้องของกรอบสัญญาณเสียงก่อน ทำให้ตำแหน่งที่ถูกเลือกขึ้นมานั้นมีโอกาสที่จะเป็นเสียงไม่ก้องน้อยกว่า เมื่อเปรียบเทียบกับวิธีการตรวจหาตำแหน่งสระของ Howitt

การวิเคราะห์เชิงภาษา

สระแบ่งออกเป็น 2 ชนิด คือ สระเดี่ยว ซึ่งเป็นสระที่ลิ้นไม่เปลี่ยนตำแหน่ง หรือลิ้นไม่เลื่อนขึ้นลงในขณะออกเสียง และสระประสม เป็นสระที่ตำแหน่งของลิ้นเปลี่ยนไป คือ เลื่อนจากเสียงหนึ่งไปยังอีกเสียงหนึ่ง นอกจากนี้ สระประสมยังแบ่งออกได้เป็น 2 ชนิด คือ สระประสมชนิดเสียงเลื่อน และสระประสมแท้

สระประสมชนิดเสียงเลื่อน เป็นสระที่เลื่อนจากสระเสียงแรกไปยังเสียงกึ่งสระ หรือเลื่อนไปยังสระอีกเสียงหนึ่งที่ตามมา เช่น ในภาษาอังกฤษ เสียง [ij] ในคำ beat เลื่อนจากเสียงสระ [i] ไปยังเสียงกึ่งสระ [j] เสียง [eɪ] ในคำ late เลื่อนจากเสียงสระ [e] ไปยังเสียงสระ [ː] เป็นต้น

สระประสมแท้ เป็นสระที่เกิดขึ้นจาก สระ 2 เสียงประสมกัน และได้ยินเป็นสระเสียงใหม่เสียงเดียว เช่น สระ เอีย เอือ ในภาษาไทย และสระ เสียง [aɪ] ในคำว่า like เสียง [aʊ] ในคำว่า house เป็นต้น

สระที่ปรากฏในภาษาอังกฤษบางสระนั้น คล้ายกับสระที่ปรากฏในภาษาไทย แต่ในทางภาษาศาสตร์จะถือว่า สระหล่านั้นแตกต่างกับสระที่ปรากฏในภาษาไทย ดังนี้

สระ [ij] ในภาษาอังกฤษไม่เหมือนสระอี [iː] ในภาษาไทย ตรงที่สระ [iː] ในภาษาไทยเป็นสระเดี่ยว ส่วนสระ [ij] ในภาษาอังกฤษเป็นสระประสมประเภทเสียงเลื่อน

สระ [eɪ] ในภาษาอังกฤษไม่เหมือนสระเอ [e:] ในภาษาไทย ตรงที่สระ [e:] ในภาษาไทยเป็นสระเดี่ยว ส่วนสระ [eɪ] ในภาษาอังกฤษเป็นสระประสมประเภทเสียงเลื่อน

สระ [uw] ในภาษาอังกฤษไม่เหมือนสระอู [u:] ในภาษาไทย ตรงที่สระ [u:] ในภาษาไทยเป็นสระเดี่ยว ส่วนสระ [uw] ในภาษาอังกฤษเป็นสระประสมประเภทเสียงเลื่อน

สระ [əʊ] ในภาษาอังกฤษไม่เหมือนสระโอ [o:] ในภาษาไทย ตรงที่สระ [o:] ในภาษาไทยเป็นสระเดี่ยว ส่วนสระ [əʊ] ในภาษาอังกฤษเป็นสระประสมประเภทเสียงเลื่อน

เพื่อที่จะดูว่า ระบบการตรวจหาสระที่นำเสนอนี้เหมาะสมสำหรับการตรวจหาสระในภาษาใดภาษาหนึ่งหรือไม่นั้น ในการทดลองนี้จึงได้ทำการทดสอบการตรวจหาสระ โดยแยกวิธีการในการตรวจหาสระ ออกเป็น 3 วิธีคือ

1. การตรวจหาสระที่ปรากฏในภาษาไทยเท่านั้น ซึ่งประกอบด้วยสระต่างๆ ซึ่งสามารถแสดงได้ดังตารางต่อไปนี้

ตารางที่ 5.10 สระที่ปรากฏในภาษาไทยเท่านั้น

สระ	IPA	ตัวอย่าง	สระ	IPA	ตัวอย่าง
อี	i:	จีบ	โอ	o:	โกง
อึ	ɯ	อึด	ออ	ɔ:	รอ
อือ	ɯ:	อืด	เออะ	ə	เลอะ
อู	u:	ขูด	เออ	ə:	เธอ
เอ	e:	เอน	เอีย	ia:	เตียง
แอ	ɛ	แม่	เอือ	ɯ:a	เดือน
โอะ	o	ลง	อัว	ua:	พวก

2. การตรวจหาสระที่ปรากฏในภาษาอังกฤษเท่านั้น ซึ่งประกอบด้วยสระต่างๆ ซึ่งสามารถแสดงได้ดังตารางต่อไปนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.11 สระที่ปรากฏในภาษาอังกฤษ

สระ	IPA	ตัวอย่าง	สระ	IPA	ตัวอย่าง
ij	i:	beat	ɔɪ	ɔ:	boy
eɪ	e:	late	əʊ	o:	boat
aɪ	aɪ	like	ʊw	u:	boot
ə	ə	be ในคำว่า believe	ɪə	ɪə	hear
aʊ	a:	house	ʊə	ʊə	poor
ɛə	ɜə	bear	ɒ	ɒ	not
ɔə	ɔə	pour			

3. การตรวจหาสระที่ปรากฏทั้งในภาษาไทยและภาษาอังกฤษ ซึ่งประกอบด้วย สระต่างๆ ดังต่อไปนี้

ตารางที่ 5.12 สระที่ปรากฏในภาษาไทยและภาษาอังกฤษ

สระ			ตัวอย่าง		สระ			ตัวอย่าง	
ไทย	อังกฤษ	IPA	ไทย	อังกฤษ	ไทย	อังกฤษ	IPA	ไทย	อังกฤษ
อะ	a	a	จะ	nothing	เอะ	e	e	เตะ	bet
อา	a:	a:	มา	farm	แอะ	ɛ	ɛ	แกะ	bat
อิ	I	ɪ	จิบ	sit	เอาะ	ɔ	ɔ	เกาะ	lost
อุ	ʊ	ʊ	ดุ	book					

ผลการทดลองการตรวจหาสระ โดยแยกสระที่ปรากฏในภาษาต่างๆ สามารถแสดงได้ดังตารางที่ 5.13 ถึงตารางที่ 5.15

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.13 ผลการตรวจหาสระที่ปรากฏในภาษาไทยเท่านั้น

ฐานข้อมูล	สระที่ปรากฏในภาษาไทย			
	Total	Correct	Delete	Insert
LOTUS (CLEAN)	11,459	11,049	410	585
LOTUS (OFFICE)	11,861	11,382	479	587
Spoken Digit (Clean)	1,784	1,775	9	22
Spoken Digit (Telephone)	651	602	49	1
รวม	25,755	24,808	947	1,195
คิดเป็นร้อยละ		**96.32**	**3.68**	**4.64**

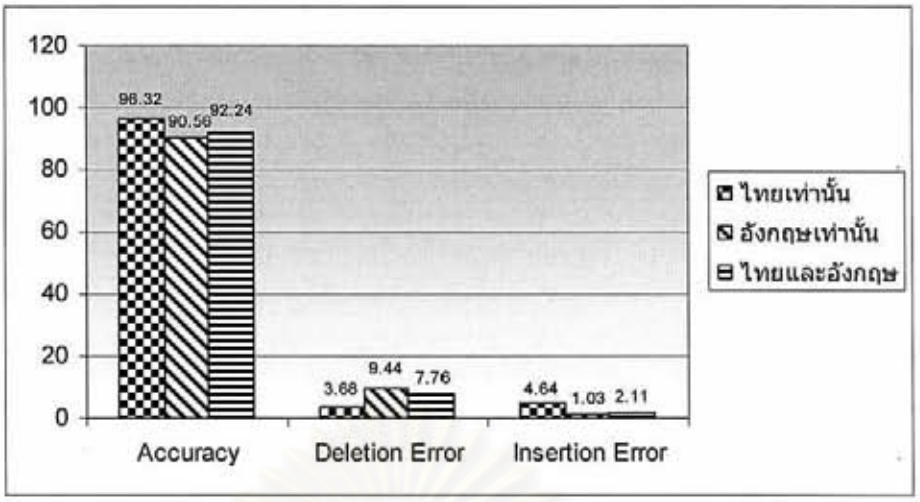
ตารางที่ 5.14 ผลการตรวจหาสระที่ปรากฏในภาษาอังกฤษเท่านั้น

ฐานข้อมูล	สระที่ปรากฏในภาษาไทย			
	Total	Correct	Delete	Insert
TIMIT	9,017	8,166	851	93
คิดเป็นร้อยละ		90.56	9.44	1.03

ตารางที่ 5.15 ผลการตรวจหาสระที่ปรากฏในภาษาไทยและภาษาอังกฤษเท่านั้น

ฐานข้อมูล	สระที่ปรากฏในภาษาไทย			
	Total	Correct	Delete	Insert
LOTUS(CLEAN)	10,083	9,318	765	214
LOTUS(OFFICE)	10,400	9,633	767	220
Spoken Digit(Clean)	826	823	3	41
Spoken Digit(Telephone)	507	482	25	7
TIMIT	5,499	4,940	559	94
รวม	27,315	25,196	2,119	576
คิดเป็นร้อยละ		92.24	7.76	2.11

การเปรียบเทียบร้อยละของความถูกต้อง ร้อยละของตำแหน่งเกิน และร้อยละของตำแหน่งขาดในสระที่ปรากฏในภาษาไทยเท่านั้น สระที่ปรากฏในภาษาอังกฤษเท่านั้น และสระที่ปรากฏทั้งในภาษาไทยและภาษาอังกฤษ แสดงได้ดังรูปที่ 5.8



รูปที่ 5.8 ประสิทธิภาพในการตรวจหาสระที่ปรากฏในแต่ละภาษา

รูปที่ 5.8 แสดงให้เห็นถึงประสิทธิภาพในการตรวจหาสระที่ปรากฏในภาษาไทยเท่านั้น สระที่ปรากฏในภาษาอังกฤษเท่านั้น และสระที่ปรากฏทั้งในภาษาไทยและภาษาอังกฤษ โดยประสิทธิภาพที่ได้นั้น จำแนกออกเป็นจำนวนสระที่ถูกต้อง จำนวนสระที่เกิน และจำนวนสระที่ขาด ดังนี้

1. ในจำนวนสระที่ถูกต้องนั้น สระที่ปรากฏในภาษาไทยเท่านั้นมีความถูกต้องมากที่สุด และสระที่ปรากฏในภาษาอังกฤษเท่านั้น มีความถูกต้องน้อยที่สุด
2. ในจำนวนสระที่ขาดนั้น สระที่ปรากฏในภาษาอังกฤษเท่านั้นที่มีจำนวนมากที่สุด และสระที่ปรากฏในภาษาไทยเท่านั้นที่มีจำนวนน้อยที่สุด
3. ในจำนวนสระที่เกินนั้น สระที่ปรากฏในภาษาไทยเท่านั้นที่มีจำนวนมากที่สุด และสระที่ปรากฏในภาษาไทยเท่านั้นที่มีจำนวนน้อยที่สุด

ผลการทดลองแสดงให้เห็นว่า ความถูกต้องในการตรวจหาสระที่ปรากฏในภาษาไทยเท่านั้นมีมากที่สุดและสระที่ปรากฏในภาษาอังกฤษเท่านั้นมีน้อยที่สุด แต่เมื่อพิจารณาจำนวนตำแหน่งเกินที่เกิดขึ้น พบว่าสระที่ปรากฏในภาษาไทยเท่านั้นมีจำนวนตำแหน่งเกินมากที่สุด และสระที่ปรากฏในภาษาอังกฤษเท่านั้นมีจำนวนตำแหน่งเกินน้อยที่สุด

อย่างไรก็ตาม เมื่อพิจารณาเฉพาะความถูกต้องที่เกิดขึ้นในการตรวจหาสระแต่ละประเภทนั้น จะเห็นได้ว่า ความถูกต้องที่เกิดขึ้นนั้นไม่แตกต่างกันมาก ซึ่งเมื่อนำจำนวนตำแหน่งสระที่ถูกต้อง จำนวนตำแหน่งที่เกิน และจำนวนตำแหน่งที่ขาดมาเฉลี่ยด้วยกันแล้ว จะทำให้ความถูกต้องที่เกิดขึ้นโดยรวมทั้งหมดมีขนาดใกล้เคียงกัน ดังนั้น จึงสามารถสรุปได้ว่า สามารถนำระบบการตรวจหาสระที่นำเสนอนี้ไปใช้ในการตรวจหาสระที่ปรากฏในภาษาไทย หรือสระที่ปรากฏในภาษาอังกฤษ

บทที่ 6
ข้อสรุปและข้อเสนอแนะ

ข้อสรุป

งานวิจัยทางด้านการตรวจหาสระในเสียงพูดของมนุษย์ ที่นำเสนอในวิทยานิพนธ์ฉบับนี้ เป็นวิธีการตรวจหาตำแหน่งสระในเสียงพูดต่อเนื่องภาษาไทย โดยการใช้ลักษณะทางสวนสัทศาสตร์ อันประกอบด้วย ค่าอัตสหสัมพันธ์มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์ และค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ ซึ่งลักษณะทางสวนสัทศาสตร์ดังกล่าวนั้นสามารถคำนวณได้ง่าย โดยไม่จำเป็นต้องแปลงสัญญาณเสียงพูดให้อยู่ในโดเมนความถี่ ทำให้ประหยัดเวลาในการคำนวณของระบบมากขึ้น

วิธีการในการตรวจหาสระที่นำเสนอในวิทยานิพนธ์นี้ ประกอบด้วยขั้นตอนการทำงาน 3 ขั้นตอน ดังนี้

1. การหาความก้องและความไม่ก้องของกรอบสัญญาณเสียงพูด

   ในขั้นตอนนี้จะทำการวิเคราะห์ว่า กรอบสัญญาณเสียงพูดนั้นเป็นเสียงก้องหรือหรือเสียงไม่ก้อง ซึ่งลักษณะทางสวนสัทศาสตร์ที่ใช้ในการหาความก้องหรือไม่ก้องของสัญญาณเสียงนั้น คือ ค่าอัตสหสัมพันธ์มากที่สุดในช่วงความถี่ 60 ถึง 320 เฮิรตซ์

2. การคัดเลือกตำแหน่งที่เหมาะสมจะเป็นสระ

   ในขั้นตอนนี้เป็นการคัดเลือกตำแหน่งที่เหมาะสมจะเป็นสระ ด้วยระเบียบวิธีคอนเวกซ์ฮัลล์ จากลักษณะทางสวนสัทศาสตร์ คือ ค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ โดยการคัดเลือกนั้นจะพิจารณาให้ยอดที่สูงเด่นจากรูปร่างพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ เป็นตำแหน่งของสระ

3. การให้คะแนน

   คะแนนของแต่ละตำแหน่งที่ถูกเลือกให้เป็นสระนั้น มาจากการใช้พลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์ของกรอบสัญญาณนั้น ร่วมกับความชันของพลังงานซึ่งวัดระหว่างกรอบสัญญาณที่มีค่าพลังงานของสัญญาณเสียงที่มากกว่า 300 เฮิรตซ์แล้วต่ำที่สุด กับกรอบที่ถูกเลือกให้เป็นสระ เพื่อใช้เป็นข้อมูลฝึกสำหรับการจำแนกตำแหน่งที่ได้ ให้เป็นตำแหน่งของสระหรือตำแหน่งที่ไม่ใช่สระ ด้วยวิธีการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น

ประสิทธิภาพของวิธีการในการตรวจหาตำแหน่งสระที่นำเสนอนี้ ได้ถูกเปรียบเทียบกับประสิทธิภาพในการตรวจหาตำแหน่งสระของ Howitt โดยทำการทดสอบกับฐานข้อมูล 3 ประเภท คือ ฐานข้อมูลเสียง Lotus ฐานข้อมูลเสียงตัวเลข และฐานข้อมูลเสียง TIMIT ซึ่งพบว่าวิธีการตรวจหาสระที่นำเสนอนั้น ให้ความถูกต้องมากกว่าวิธีการตรวจหาสระด้วยวิธีของ Howitt ผลการ

เปรียบเทียบการตรวจหาสระของวิธีการที่นำเสนอในวิทยานิพนธ์นี้ และวิธีการตรวจหาสระของ Howitt นั้นได้แสดงไว้ดังตารางที่ 5.9

นอกจากนี้ ยังได้มีการทดสอบว่าระบบการตรวจหาสระที่นำเสนอนี้นั้น เหมาะสมสำหรับการตรวจหาสระเฉพาะในภาษาไทยหรือไม่ ซึ่งจากผลการทดลองสามารถสรุปได้ว่า สามารถนำระบบการตรวจหาสระที่นำเสนอนี้ไปใช้ในการตรวจหาสระ ได้ทั้งในภาษาไทยและภาษาอังกฤษ

คุณประโยชน์ต่อวงวิชาการ

ระบบการตรวจหาสระที่นำเสนอนี้นั้น สามารถนำไปประยุกต์ใช้ในงานต่างๆ ที่เกี่ยวข้องกับการรู้จำเสียงพูด เพื่อให้ระบบการรู้จำเสียงพูดมีประสิทธิภาพในการรู้จำเสียงได้เป็นอย่างดีที่สุด โดยที่ระบบการตรวจหาสระนี้จะทำให้ได้ตำแหน่งของสระ ซึ่งจะนำไปประยุกต์ใช้ในงานต่างๆ ดังนี้

1. การตรวจหาพยางค์

   ด้วยโครงสร้างของพยางค์นั้น จะพบว่าหนึ่งพยางค์จะต้องประกอบด้วยสระจำนวนหนึ่งตัวเสมอ ดังนั้น ตำแหน่งของสระนี้จึงสามารถแทนจำนวนพยางค์ได้เพียงพยางค์เดียวเท่านั้น ด้วยเหตุนี้ การนำระบบการตรวจหาสระไปประยุกต์ใช้ในการตรวจหาพยางค์นั้น ทำให้ทราบว่า เสียงพูดประโยคนั้นประกอบด้วยพยางค์จำนวนกี่พยางค์ ซึ่งจำนวนพยางค์ที่ได้ดังกล่าวนั้น สามารถนำไปช่วยในการรู้จำเสียงพูดในระบบรู้จำเสียงพูดประเภทต่างๆ เช่น ระบบรู้จำเสียงแบบที่ใช้คุณสมบัติลักษณะเฉพาะและการตัดแบ่งหน่วยเสียงในระบบรู้จำเสียงแบบแบ่งส่วน เพื่อให้การรู้จำเสียงพูดมีความถูกต้องมากขึ้นได้ โดยนำจำนวนพยางค์ดังกล่าวนั้นไปใช้ในการระบุว่าประโยคที่ต้องการจะรู้จำเสียงพูดนั้นมีจำนวนกี่พยางค์ ระบบรู้จำเสียงแบบที่ใช้คุณสมบัติลักษณะเฉพาะ (Distinctive Feature-Based) และการตัดแบ่งหน่วยเสียงในระบบรู้จำเสียงแบบแบ่งส่วน (Segment-Based Speech Recognition)

2. การประมาณอัตราเร็วของเสียงพูด

   ดังที่ได้กล่าวไปแล้วว่า โครงสร้างของพยางค์นั้นประกอบด้วยสระจำนวนหนึ่งตัวเสมอ ดังนั้น จึงสามารถใช้ตำแหน่งของสระที่ได้นั้นแทนพยางค์หนึ่งพยางค์ได้ อัตราเร็วของสียงพูดคือ จำนวนพยางค์ในหนึ่งหน่วยเวลา ซึ่งโดยทั่วไปมักจะกำหนดให้หนึ่งหน่วยเวลาคือ 1 นาที ด้วยเหตุนี้ การนำระบบการตรวจหาสระไปประยุกต์ใช้ในการประมาณอัตราเร็วของเสียงพูดนั้น จะทำให้ทราบว่า ประโยคเสียงพูดนั้นมีอัตราเร็วในการพูดมากหรือน้อย โดยที่การประยุกต์นี้จะวัดจำนวนของตำแหน่งสระที่ได้ต่อหนึ่งหน่วยเวลา

ข้อเสนอแนะ

การปรับปรุงประสิทธิภาพของการตรวจหาสระ โดยเฉพาะอย่างยิ่งการปรับปรุงความถูกต้องของการตรวจหาตำแหน่งสระนั้นสามารถทำได้หลายวิธี ดังนี้

1. เนื่องจากความผิดพลาดที่เกิดขึ้นนั้น ส่วนใหญ่มาจากขั้นตอนการคัดเลือกตำแหน่งที่เหมาะสมจะเป็นสระ ซึ่งตำแหน่งที่ตรวจหาได้นั้นเป็นตำแหน่งของเสียงก้องประเภทอื่นที่ไม่ใช่เสียงสระ ดังนั้นการเลือกลักษณะทางสวนสัทศาสตร์ที่เหมาะสมที่จะเป็นตัวแทนของหน่วยเสียงสระเท่านั้น โดยไม่เหมาะสมที่จะเป็นตัวแทนของหน่วยเสียงก้องประเภทอื่น จึงเป็นสิ่งสำคัญที่จะช่วยให้ความถูกต้องของตำแหน่งสระที่ตรวจหาได้นั้นมีมากขึ้น
2. การนำวิธีการทางด้านสถิติ เข้ามาช่วยในการหาตำแหน่งที่เหมาะสมจะเป็นสระ แทนการคัดเลือกตำแหน่งของสระจากยอดที่สูงเด่น อาจทำให้ความถูกต้องของการจำแนกนั้นดีขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] Xie, Z. An acoustic feature based vowel detector. Master's Thesis, Department of Electrical Engineering, The University of Chicago. 2004.

[2] Kasuriya, S., Sornlertlamvanich, V., Cotsomrong, P., Kanokphara, S., and Thatphithakkul, N.. Thai Speech Corpus for Speech Recognition. Proceeding of the COCOSDA'03, 2003.

[3] Howitt, A. W.. Automatic Syllable Detection for Vowel Landmarks. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. 2000.

[4] Markowitz, J. A.. Using Speech Recognition. Prentice-Hall Inc., 1995.

[5] Anusurain, E.. Synthesis of tones and vowels in Thai open syllables using microphonemes. Master's Thesis, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University. 1998.

[6] Rosen, S., and Howell, P.. Signals and Systems for Speech and Hearing. Academic Press Inc., 1990.

[7] Stevens, K. N.. Acoustic Phonetics. Cambridge: MIT Press, 1999.

[8] Prasanna, S., Gangashetty, S., and Yegnanarayana, B.. Significance of vowel onset point for speech analysis. In Proceedings of Signal Processing and Communication (2001).

[9] Deller, Jr., Proakis, J., and Hansen, J.. Discrete-Time Processing of Speech Signals. New Jersey: Pentice-Hall Inc., 1993.

[10] Pellegrino, F., and Andre-Obrecht, R.. From Vocalic Detection to Automatic Emergence of Vowel System. In Proceedings of ICASSP. (1997).

[11] Lander,T. L., Cole, R. A., Ohsika, B., and Noel, M. The OGI 22 Language Telephone Speech Corpus. Eurospeech. (1995): 817-820.

[12] Pfau , T., and Ruske, G.. Estimating the Speaking Rate using Vowel Detection. In Prodeccings of ICASSP. (1998): 945-948.

[13] Mermelstein, P. Automatic Segmentation of Speech into Syllabic Units. Journal of the Acoustic Society of America. (1975): 880-883.

[14] Sirigos, J., Darsinos, V., Fakotakis, N., and Kokkiankis, G.. Vowel-Non Vowel Decision using Neural Networks and Rules. IEEE International Conference on Electronics, Circuits, and Systems. (1997): 510-513.

[15] Wu, S,. Shire, M., Greenberg , S., and Morgan, N.. Integrating syllable boundary information into speech recognition. In Proceedings of ICASSP. (1997): 987-990.

[16] Choi, J. Y., Chuang, E., Gow, D., Kwong, K., Shattuck, S., Stevens, K., and Zhang, Y.. Labeling a speech database with landmarks and features. The Journal of the Acoustical Society of America. (1997): 3163.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

ตัวอย่างการนำระบบการตรวจหาตำแหน่งสระไปประยุกต์ใช้

ตำแหน่งของสระที่ตรวจหาได้นั้น นอกจากที่จะนำมาช่วยในการตัดแบ่งเป็นหน่วยเสียงเพื่อให้ผลลัพธ์ที่ได้มีความถูกต้องมากขึ้นแล้ว ยังสามารถนำตำแหน่งของสระที่ตรวจหาได้นั้นมาประยุกต์ใช้ในงานอื่นๆ ได้เช่นกัน วิทยานิพนธ์นี้ได้นำระบบการตรวจหาตำแหน่งสระไปประยุกต์ใช้ใน การตรวจหาจำนวนพยางค์ (Syllable Detection) ซึ่งรายละเอียดในการตรวจหาพยางค์นั้น มีดังต่อไปนี้

การตรวจหาจำนวนพยางค์

พยางค์ คือ เสียงพูดที่เปล่งออกมาพร้อมกันทั้งเสียงสระ เสียงพยัญชนะและเสียงวรรณยุกต์ ซึ่งอาจจะมีความหมายหรือไม่มีความหมายก็ได้ พยางค์ในภาษาไทยสามารถเขียนในรูปของโครงสร้างได้ดังนี้

$$S = C(C)V(:)(C)$$

โดยที่ S = พยางค์

C = พยัญชนะ

V = สระ

: = สระเสียงยาว

จากรูปโครงสร้างของพยางค์นั้น สังเกตว่าเสียงสระที่เกิดขึ้นในพยางค์แต่ละพยางค์นั้นจะมีได้เพียงสระเดียวเท่านั้น ดังนั้น ตำแหน่งของสระแต่ละตำแหน่งที่ตรวจหาได้ จึงสามารถแทนจำนวนพยางค์ได้เพียงพยางค์เดียวเท่านั้น

การนำระบบการตรวจหาสระมาประยุกต์ใช้ในการตรวจหาจำนวนพยางค์ของวิทยานิพนธ์นี้ ได้ทำการทดลองกับฐานข้อมูลเสียงทั้ง 2 ชนิด คือ ฐานข้อมูลเสียงโลตัส และฐานข้อมูลเสียงตัวเลข ซึ่งรายละเอียดของฐานข้อมูลแต่ละชนิดนั้น ได้กล่าวไปแล้วในบทที่ 4

ขั้นตอนในการตรวจหาจำนวนพยางค์ มีดังต่อไปนี้

1. นำสัญญาณเสียงแต่ละประโยคนั้น มาทำการตรวจหาตำแหน่งสระด้วยวิธีการที่นำเสนอ
2. หาความถูกต้องของจำนวนพยางค์ที่ได้จากแต่ละประโยค โดยการเปรียบเทียบตำแหน่งที่ตรวจหาได้กับจำนวนของหน่วยเสียงสระที่ระบุในไฟล์กำกับเสียง หาก

ตำแหน่งของสระที่ได้ทั้งหมด น้อยกว่าจำนวนของหน่วยเสียงสระที่ระบุในไฟล์กำกับเสียง จะถือว่าประโยคนั้นมีจำนวนพยางค์ที่ขาดเกิดขึ้น และถ้าตำแหน่งของสระที่ได้ทั้งหมด มากกว่าจำนวนของหน่วยเสียงสระที่ระบุในไฟล์กำกับเสียง จะถือว่าประโยคนั้นมีจำนวนพยางค์ที่เกินเกิดขึ้น

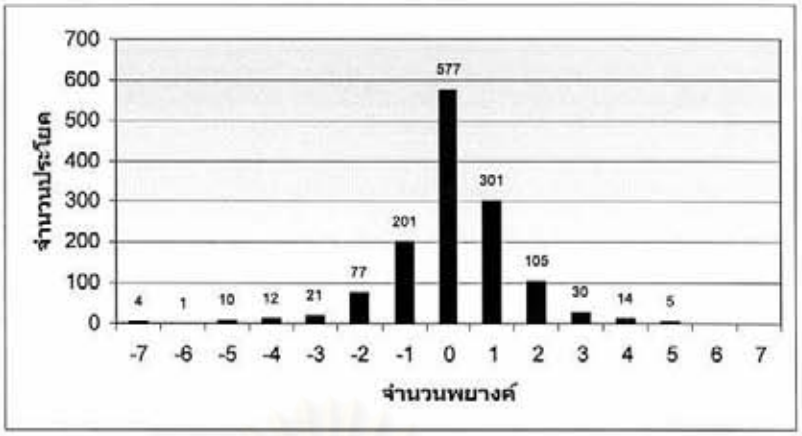
ผลการทดลองที่ได้ แสดงได้ดังตารางที่ ก.1

ตารางที่ ก.1 ผลการตรวจหาจำนวนพยางค์ของแต่ละฐานข้อมูลเสียง

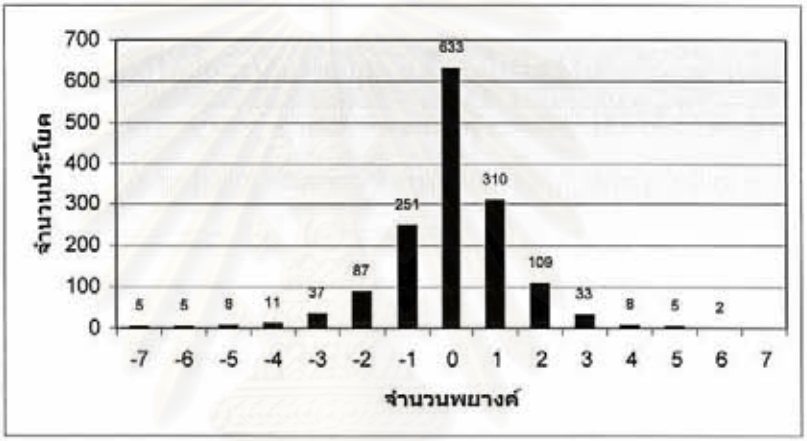
		จำนวนประโยคในฐานข้อมูล			
		Lotus (Clean)	Lotus (Office)	Clean Number	Telephone Number
จำนวนพยางค์ที่ขาด	7	4	5	0	1
	6	1	5	0	0
	5	10	8	0	0
	4	12	11	0	1
	3	21	37	0	3
	2	77	87	0	8
	1	201	251	1	17
ไม่ขาดและไม่เกิน (0)		577	633	218	85
จำนวนพยางค์ที่เกิน	1	301	310	57	18
	2	105	109	7	6
	3	30	33	4	0
	4	14	8	0	0
	5	5	5	0	0
	6	0	2	0	0
	7	0	0	0	0
รวมทั้งหมด		1,358	1,504	2,87	139

รูปที่ ก.1 ถึงรูปที่ ก.4 แสดงฮิสโทรแกรมของจำนวนประโยคที่สัมพันธ์กับจำนวนพยางค์ที่ขาด จำนวนพยางค์ที่เกิน และจำนวนพยางค์ที่ไม่ขาดและไม่เกินของฐานข้อมูลเสียงแต่ละประเภท

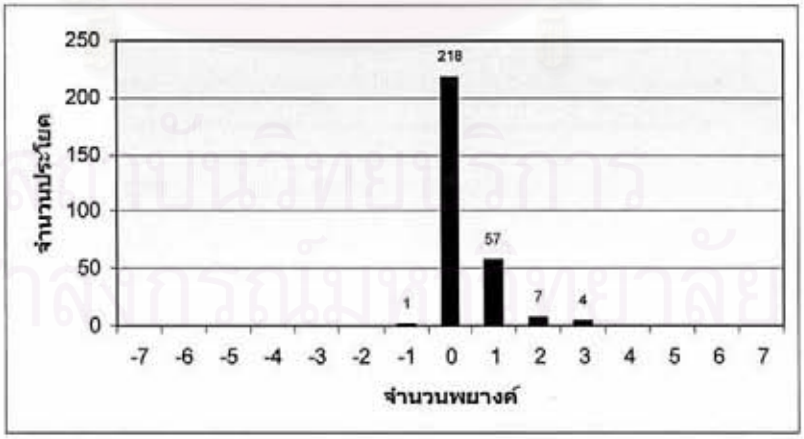
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



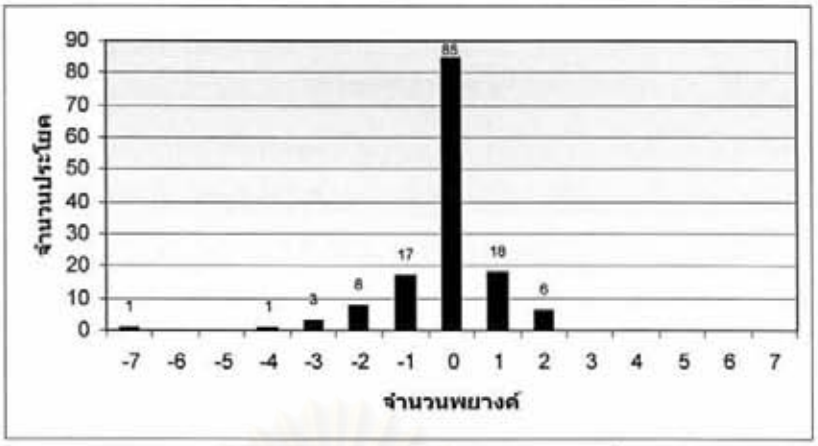
รูปที่ ก.1 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ



รูปที่ ก.2 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ

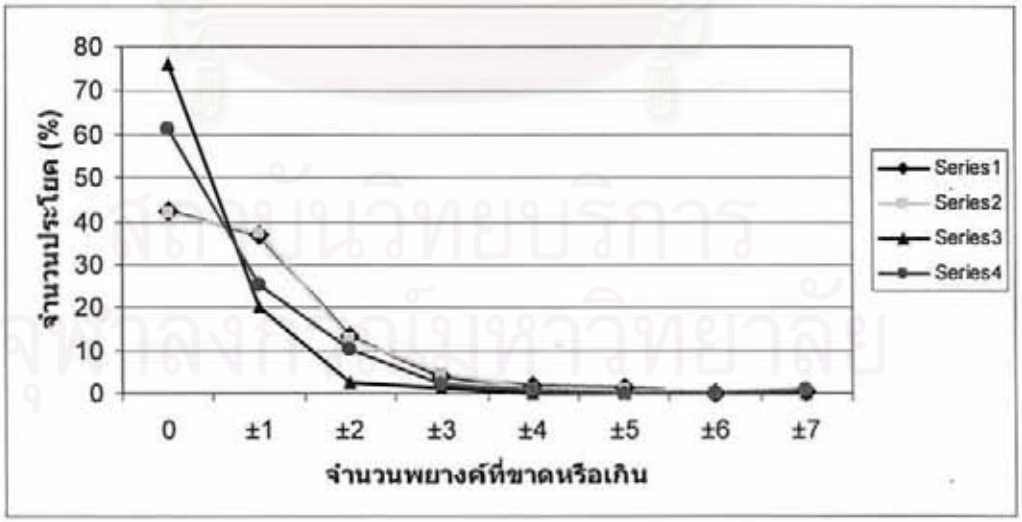


รูปที่ ก.3 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบ



รูปที่ ก.4 ผลการตรวจหาจำนวนพยางค์ของฐานข้อมูลเสียงตัวเลข ที่บันทึกด้วยโทรศัพท์ในห้องปกติ

จากฮิสโทรแกรมในรูปที่ ก.1 ถึงรูปที่ ก.4 ความผิดพลาดของการตรวจหาพยางค์เท่ากับ 0 หมายถึง จำนวนพยางค์ที่ได้นั้นตรงกับจำนวนสระที่ระบุในไฟล์กำกับเสียง ความผิดพลาดของการตรวจหาพยางค์เท่ากับ 1, 2, 3, 4, 5, 6, และ 7 หมายถึงจำนวนพยางค์ที่ได้นั้นมากกว่าจำนวนสระที่ระบุในไฟล์กำกับเสียง 1 พยางค์ 2 พยางค์ 3 พยางค์ 4 พยางค์ 5 พยางค์ 6 พยางค์ และ 7 พยางค์ ตามลำดับ และความผิดพลาดของการตรวจหาพยางค์เท่ากับ -1, -2, -3, -4, -5, -6, และ -7 หมายถึงจำนวนพยางค์ที่ได้นั้นน้อยกว่าจำนวนสระที่ระบุในไฟล์กำกับเสียง 1 พยางค์ 2 พยางค์ 3 พยางค์ 4 พยางค์ 5 พยางค์ 6 พยางค์ และ 7 พยางค์ ตามลำดับ การวิเคราะห์จำนวนประโยค ในการนับจำนวนพยางค์ที่มีความผิดพลาดของจำนวนพยางค์ที่ขาดหรือเกินจำนวนต่างๆ เมื่อเทียบกับจำนวนประโยคทั้งหมดของแต่ละฐานข้อมูลนั้น แสดงได้ดังรูปที่ ก.5

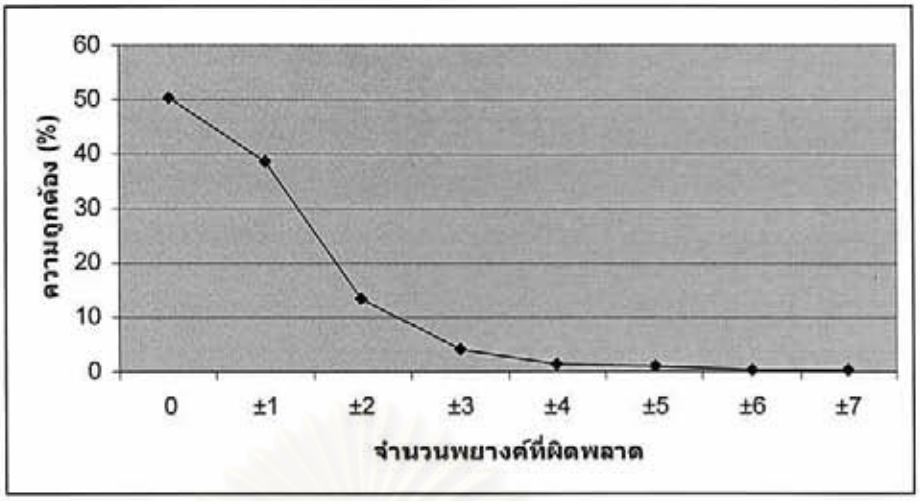


รูปที่ ก.5 จำนวนประโยคที่เกิดความผิดพลาดที่จำนวนต่างๆ เมื่อเทียบกับจำนวนประโยคทั้งหมดของแต่ละฐานข้อมูล

จากการพิจารณากราฟเส้นที่เกิดขึ้นในรูปที่ ก.5 พบว่าประกอบด้วยเส้นกราฟ 4 เส้น ซึ่งแต่ละเส้นนั้นแสดงถึง จำนวนประโยคที่เกิดความผิดพลาดขนาดต่างๆ เมื่อเทียบกับจำนวนประโยคทั้งหมดของฐานข้อมูลแต่ละประเภท ในการตรวจหาจำนวนพยางค์ของแต่ละประโยคโดยทั่วไปนั้น คาดหวังให้ความผิดพลาดของการตรวจหาจำนวนพยางค์เท่ากับ 0 หรือไม่มีจำนวนพยางค์ที่ขาดหรือเกินจากจำนวนสระที่ถูกต้องมีจำนวนมากๆ และให้ความผิดพลาดของการตรวจหาจำนวนพยางค์เท่ากับ ±1 ±2 ±3 ±4 ±5 ±6 และ ±7 หรือมีจำนวนพยางค์ที่ขาดหรือเกินมาจากจำนวนสระที่ถูกต้องจำนวน 1 พยางค์ 2 พยางค์ ไปเรื่อยๆไม่เกิดขึ้นเลยหรือยอมให้เกิดขึ้นได้ในจำนวนน้อยๆ และจากผลการทดลองของข้อมูลทั้ง 4 ประเภทนั้น ได้จำนวนประโยคที่มีจำนวนพยางค์ที่ขาดหรือเกินเท่ากับ 0 มีจำนวนมากที่สุด และลดลงเรื่อยๆ เมื่อความผิดพลาดของการตรวจหาจำนวนพยางค์เพิ่มขึ้นเรื่อยๆ แม้ผลลัพธ์ของจำนวนประโยคที่มีจำนวนความผิดพลาดมากกว่า 0 พยางค์นั้นจะไม่เป็น 0 ดังที่คาดหวังไว้ก็ตาม แต่จำนวนประโยคที่ได้นั้นก็มีจำนวนน้อยลงเรื่อยๆจนใกล้เคียง 0% โดยเฉพาะเมื่อจำนวนความผิดพลาดของพยางค์ที่ขาดหรือเกินมากๆ

นอกจากนี้ เมื่อพิจารณาจำนวนประโยคที่ความผิดพลาดจำนวนต่างๆ นั้น พบว่า ที่ความผิดพลาดเท่ากับ 0 จำนวนประโยคในฐานข้อมูลเสียงตัวเลขที่บันทึกด้วยไมโครโฟนในห้องเงียบและฐานข้อมูลเสียงตัวเลข ที่บันทึกด้วยโทรศัพท์ในห้องปกตินั้น มีจำนวนประโยคที่เกิดขึ้นมากกว่าจำนวนประโยคที่ได้จากฐานข้อมูลเสียง Lotus ที่บันทึกในห้องเงียบ และฐานข้อมูลเสียง Lotus ที่บันทึกในห้องปกติ ที่เป็นเช่นนี้เนื่องจากเสียงพูดในฐานข้อมูลเสียงตัวเลขนั้นมีความหลากหลายของคำพูด น้อยกว่าเสียงพูดในฐานข้อมูลเสียง Lotus มาก และที่ตำแหน่งพยัญชนะต้นของเสียงพูดตัวเลขนั้น จะเป็นเสียงเสียดแทรกโดยส่วนใหญ่ ในขณะที่ฐานข้อมูลเสียง Lotus นั้น ที่ตำแหน่งพยัญชนะต้นและพยัญชนะท้าย จะประกอบด้วยหน่วยเสียงที่เป็นเสียงก้องจำนวนมากมาย ทำให้การตรวจหาตำแหน่งสระ ย่อมมีความผิดพลาดมากกว่า

จำนวนพยางค์ของแต่ละประโยคที่ได้นี้ สามารถนำไปใช้เป็นความรู้หนึ่งที่ช่วยให้การรู้จำเสียงพูดของระบบรู้จำเสียงแบบต่างๆ มีความถูกต้องมากขึ้น ซึ่งจากผลการหาตรวจหาพยางค์ในฐานข้อมูลเสียงแต่ละประเภทที่แสดงในรูปที่ ก.5 นั้น สามารถหาความผิดพลาดที่เกิดขึ้นจากจำนวนพยางค์ที่ขาดหรือเกินที่จำนวนต่างๆ ได้ดังแสดงในรูปที่ ก.6



รูปที่ ก.6 ความถูกต้องของการนำระบบการตรวจหาสระไปประยุกต์ใช้ในการตรวจหาจำนวนพยางค์

จากรูปที่ ก.6 นี้ จะเห็นได้ว่า ความถูกต้องของการนำระบบการตรวจหาสระไปประยุกต์ใช้ในการตรวจหาจำนวนพยางค์นั้น มีค่าประมาณ 50% ในกรณีที่จำนวนพยางค์ที่ตรวจหาได้มีค่าเท่ากับจำนวนพยางค์ที่ได้จากไฟล์กำกับเสียง และความถูกต้องจะมีค่าลดลงเรื่อยๆ เมื่อความผิดพลาดของจำนวนพยางค์ทั้งในแบบที่ขาดและแบบที่เกินมีจำนวนเพิ่มขึ้น การนำจำนวนพยางค์ที่ตรวจหาได้ไปใช้เป็นความรู้หนึ่งในการรู้จำเสียงพูด จะทำให้ความผิดพลาดที่เกิดขึ้นจากจำนวนที่ขาดซึ่งเกิดขึ้นจากการรู้จำเสียงพูดนั้นลดน้อยลง ถ้าความถูกต้องของจำนวนพยางค์ที่ตรวจหาได้มีค่ามาก ๆ แต่จากผลการทดลองที่ได้นั้น ความถูกต้องของการตรวจหาจำนวนพยางค์ที่เท่ากับจำนวนพยางค์ที่ระบุในไฟล์กำกับเสียงนั้นไม่สูงมากนัก ดังนั้น เมื่อเรานำระบบการตรวจหาสระนี้ไปประยุกต์ใช้ในการตรวจหาพยางค์ เพื่อหวังที่จะให้ความถูกต้องในการรู้จำเสียงพูดนั้นดีขึ้น จะพบว่าความมั่นใจที่จะไม่เกิดจำนวนที่ขาดในการรู้จำเสียงพูดอยู่ที่ประมาณ 0.5

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวเพียงจิต ดารีเย๊าะ เกิดเมื่อวันที่ 29 ธันวาคม พ.ศ. 2525 สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นและตอนปลายจากโรงเรียนเดชะปัตนยานุกูล สำเร็จการศึกษาระดับปริญญาบัณฑิต ในสาขาวิศวกรรมคอมพิวเตอร์ จากคณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ในปีการศึกษา 2547

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย