

ลักษณะสำคัญทางเสียงสำหรับการรู้จำเสียงวรรณยุกต์ไทยด้วยฮีดเดนคอนดิชันนัลแรนดอมฟิลด์

นายณัฐภูมิ เกิดเกียรติขจร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2555

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและเพิ่มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นเพิ่มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR) are the thesis authors' files submitted through the Graduate School.

ACOUSTIC FEATURES FOR HIDDEN CONDITIONAL RANDOM FIELDS-BASED
THAI TONE RECOGNITION

Mr. Natthawut Kertkeidkachorn

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 2012

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	ลักษณะสำคัญทางเสียงสำหรับการรู้จำเสียงวรรณยุกต์ไทย ด้วยฮิดเดนคอนดิชันนัลแรนดอมฟิลด์
โดย	นายณัฐวุฒิ เกิดเกียรติขจร
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุญยพุกกณะ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... คณะบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.เศรษฐา ปานงาม)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุญยพุกกณะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ชัย วุฒิวิวัฒน์ชัย)

ณัฐวุฒิ เกิดเกียรติชจร: ลักษณะสำคัญทางเสียงสำหรับการรู้จำเสียงวรรณยุกต์ไทยด้วยฮิดเดนคอนดิชันนัลแรนดอมฟิลด์. (ACOUSTIC FEATURES FOR HIDDEN CONDITIONAL RANDOM FIELDS-BASED THAI TONE RECOGNITION) อ. ที่ปริกษาวิทยานิพนธ์หลัก: ผศ.ดร.โปรดปราน บุญยพุกกณะ, อ. ที่ปริกษาวิทยานิพนธ์ร่วม: รศ. ดร.อติวงศ์ สุชาโต, 168 หน้า.

เสียงวรรณยุกต์มีบทบาทสำคัญในการระบุความหมายของคำในภาษาไทย เพื่อที่จะพัฒนาระบบรู้จำเสียงพูดในภาษาไทยให้มีความแม่นยำมากยิ่งขึ้นจึงจำเป็นต้องมีการสร้างและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทยที่มีประสิทธิภาพ ดังนั้นในวิทยานิพนธ์นี้จึงได้ทำการศึกษาและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทย โดยแบ่งการศึกษาเป็น 3 ส่วนได้แก่ การรับรู้เสียงวรรณยุกต์ภาษาไทย, การจำแนกเสียงวรรณยุกต์ภาษาไทย และการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ภาษาไทย

ในส่วนแรกได้ทำการศึกษาการรับรู้เสียงวรรณยุกต์ภาษาไทย โดยมุ่งเน้นการศึกษาไปที่ผลกระทบของความถี่มูลฐาน, รูปร่างของพลังงาน และ รูปร่างของสเปกตรัมที่มีผลต่อการรับรู้เสียงวรรณยุกต์ จากการทดลองพบว่าความถี่มูลฐานเพียงอย่างเดียวไม่เพียงพอที่จะทำให้การรับรู้เสียงวรรณยุกต์มีความถูกต้องเท่าเสียงต้นแบบ ในขณะที่เมื่อเพิ่มข้อมูลรูปร่างของพลังงาน ส่งผลให้การรับรู้เสียงหนึ่งพยางค์ได้รับความถูกต้องมากขึ้น และเมื่อเพิ่มข้อมูลทางสเปกตรัมโดยเฉพาะช่วงความถี่ฟอร์แมนต์ที่หนึ่งทำให้ความสามารถในการรับรู้เสียงวรรณยุกต์มีความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ เมื่อเทียบกับกรณีที่ไม่มีข้อมูลในช่วงฟอร์แมนต์ที่หนึ่ง

ในส่วนถัดมาได้เสนอการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยด้วยแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ทั้งในคำโดด และเสียงพูดต่อเนื่อง ซึ่งใช้ลักษณะเด่นเป็นความถี่มูลฐาน พบว่าการจำแนกเสียงวรรณยุกต์ภาษาไทยด้วยแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ให้ประสิทธิภาพสูงกว่าระบบอ้างอิงที่ดีที่สุดที่ใช้โครงข่ายประสาทเทียมเป็นตัวจำแนก โดยได้อัตราการลดลงของความผิดพลาดเป็น 10.58% สำหรับคำโดด และ 12.02% สำหรับเสียงพูดต่อเนื่อง นอกจากนี้ได้ทำการทดลองโดยการเพิ่มลักษณะเด่นอื่น สำหรับคำโดดได้ทำการเพิ่มค่าพลังงาน และสำหรับเสียงพูดต่อเนื่องได้ทำการเพิ่มค่าสเปกตรัมของสัญญาณเสียง พบว่าอัตราการลดลงของความผิดพลาดจะสูงสุดถึง 22.40% สำหรับคำโดด และ 13.90% สำหรับเสียงพูดต่อเนื่อง เมื่อเทียบกับการจำแนกเสียงวรรณยุกต์ที่ใช้แต่ความถี่มูลฐาน

ในส่วนสุดท้ายได้ทดลองประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ด้วยแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ในการรู้จำเสียงพูดการสะกดคำภาษาไทยพบว่า ระบบที่นำเสนอที่มีการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ให้ประสิทธิภาพดีกว่าระบบอ้างอิง ทำให้อัตราการลดลงของความผิดพลาดลดลงจากระบบอ้างอิงถึง 23.85%

ภาควิชา.....วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อ.....
 สาขาวิชา.....วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อ อ.ที่ปริกษาวิทยานิพนธ์หลัก.....
 ปีการศึกษา.....2555.....ลายมือชื่อ อ.ที่ปริกษาวิทยานิพนธ์ร่วม.....

5470202721 : MAJOR COMPUTER ENGINEERING

KEYWORDS : TONE RECOGNITION / HIDDEN CONDITIONAL RANDOM FIELDS / THAI TONE

NATTHAWUT KERTKEIDKACHORN: ACOUSTIC FEATURES FOR HIDDEN CONDITIONAL RANDOM FIELDS-BASED THAI TONE RECOGNITION. ADVISOR : ASST.PROF. PROADPRAN PUNYABUKKANA, Ph.D., CO-ADVISOR : ASSOC.PROF. ATIWONG SUCHATO, Ph.D., 168 pp.

In Thai, tones play an important role in differentiating the meaning of ambiguous words. In order to develop an accurate automatic Thai speech recognition system, it is necessary to build and develop an efficient Thai tone classification system. This thesis presents a study of Thai tone classification. It can be divided into three parts; 1) a study of Thai tone perception, 2) Thai tone classification and 3) the application of Thai tone classification system to Thai speech recognition system. Thai tone perceptual experiments were conducted in order to investigate the effects of acoustic cues, especially fundamental frequency, energy envelope, as well as spectral shape, to understand their effects to Thai tone perception. The findings suggest that fundamental frequency alone is not adequate to identify tone correctly while energy envelope makes a contribution to tone perception in mono-syllabic. Furthermore, spectral shape, especially in the vicinity of the first formant, contributes significantly to the participants' ability to correctly identify tones in all syllabic patterns. Moreover, we proposed a Thai tone classification system based on Hidden Conditional Random Fields (HCRF) in isolated word and continuous speech tasks. When acoustic features were limited to fundamental frequency, the results indicated that our HCRF-based system outperformed the baselines-Artificial Neural Network (ANN)-based and Hidden Markov Model (HMM)-based systems. The HCRF-based system provides error rate reduction of 10.58% and 12.02% for isolated word and continuous speech respectively. Furthermore, we explore other acoustic features in the HCRF-based system. For isolated word, we examine the energy of speech signal. For continuous speech, we explore the spectral information of signal. The results showed that incorporating extra acoustic features yielded better performance than fundamental frequency alone. The error rate reduction of 22.40% and 13.90% are achieved for isolated word and continuous speech respectively. Finally, we integrated the proposed HCRF-based Thai tone classification system into Thai speech recognition on Thai isolated spelling domain. We found that our integrated system significantly outperform the baseline, which is a HMM-based system, and was able to achieve an error rate reduction of 23.85%.

Department : Computer Engineering..... Student's Signature.....

Field of Study : Computer Engineering..... Advisor's Signature.....

Academic Year : 2012..... Co-advisor's Signature.....

กิตติกรรมประกาศ

ตลอดระยะเวลาการศึกษาและจัดทำวิทยานิพนธ์นี้ ผู้วิจัยได้รับทุนอุดหนุนการศึกษา ระดับบัณฑิตศึกษาจุฬาลงกรณ์มหาวิทยาลัยเพื่อเฉลิมฉลองวโรกาสที่พระบาทสมเด็จพระเจ้าอยู่หัวทรงเจริญพระชนมายุครบ 72 พรรษา ประจำปีการศึกษา 2554 ทำให้การดำเนินงานศึกษาวิจัยเป็นไปอย่างราบรื่น ข้าพเจ้าขอขอบพระคุณที่มอบโอกาสอันมีค่านี้ให้แก่ข้าพเจ้า

นอกจากนี้ข้าพเจ้าขอขอบพระคุณอาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุญพุกกณะ และ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม รองศาสตราจารย์ ดร.อดิวงค์ สุชาติ ที่กรุณาให้ความรู้คำแนะนำต่างๆ ตลอดจนโอกาสต่างๆ ที่มีคุณประโยชน์ต่อการทำวิทยานิพนธ์จนสำเร็จลุล่วง ข้าพเจ้าขอขอบคุณผู้ช่วยศาสตราจารย์ ดร. เศรษฐา ปานงาม และ อาจารย์ ดร.ชัย วุฒิวิวัฒน์ ที่กรุณาสละเวลาอันมีค่า และให้คำแนะนำอันเป็นประโยชน์ต่อการดำเนินการทำวิทยานิพนธ์

ข้าพเจ้าขอขอบคุณ หน่วยปฏิบัติการวิจัยวิทยาการมนุษยภาษา ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ: NECTEC ที่เอื้อเฟื้อทรัพยากร และเครื่องมือต่างๆ ที่จำเป็นต่อการศึกษาและทดลองในการทำวิทยานิพนธ์ และขอขอบคุณรุ่นพี่และเพื่อนร่วมงานในห้องปฏิบัติการระบบภาษาพูด ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำดีๆเสมอมา และสุดท้ายข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัว ของข้าพเจ้าที่ให้กำลังใจแก่ข้าพเจ้าในยามที่ท้อแท้ และสนับสนุนข้าพเจ้าในการจัดทำวิทยานิพนธ์นี้มาโดยตลอดจนทำให้วิทยานิพนธ์นี้สำเร็จลุล่วง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฎ
สารบัญภาพ.....	ฅ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	4
ขอบเขตของการวิจัย.....	4
วิธีการดำเนินการวิจัย.....	5
ประโยชน์ที่คาดว่าจะได้รับ	6
ผลงานตีพิมพ์จากวิทยานิพนธ์.....	7
ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์.....	7
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	9
แนวคิดและทฤษฎี.....	9
1. ทฤษฎีทางภาษาศาสตร์.....	9
1.1. ระบบเสียงในภาษาไทย.....	9
1.2. โครงสร้างพยางค์ในภาษาไทย.....	13
2. ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์สัญญาณเสียง.....	14
2.1. อัดสหัสสัมพันธ์.....	14
2.2. การคำนวณค่าพลังงานของสัญญาณ.....	15
2.3. การแปลงฟูริเยร์แบบช่วงเวลาสั้น.....	16
2.4. ค่าสัมประสิทธิ์การประมาณพหุเชิงเส้น.....	16
2.5. ค่าการประมาณพหุแบบอิงการรับฟังของมนุษย์.....	17

2.6. สัมประสิทธิ์ชีพสตรัมในย่านความถี่แมลง.....	18
3. ทฤษฎีที่เกี่ยวข้องกับการวัดผล.....	19
3.1. การทดสอบของแมคนีมาร์.....	19
3.2. ค่าความถูกต้อง.....	20
เอกสารและงานวิจัยที่เกี่ยวข้อง.....	20
งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาที่มีเสียงวรรณยุกต์.....	20
1. งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาไทย.....	21
2. งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาอื่นที่มีเสียงวรรณยุกต์.....	23
งานวิจัยที่เกี่ยวข้องกับแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์.....	26
บทที่ 3 การรับรู้เสียงวรรณยุกต์ภาษาไทย.....	30
การเตรียมสิ่งเร้า.....	31
1. การเตรียมสิ่งเร้าต้นแบบ.....	31
2. การเตรียมสิ่งเร้าที่มีการดัดแปลง.....	33
การทดลอง.....	37
1. การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่าง พลังงาน.....	37
2. การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม.....	39
ผลการทดลอง และสรุปผลการทดลอง.....	39
1. ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่าง พลังงาน.....	39
2. ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยที่แบ่งตามช่วงของสเปกตรัม.....	40
3. สรุปผล และอภิปรายผล.....	41
บทที่ 4 การจำแนกเสียงวรรณยุกต์ภาษาไทย.....	44
ฐานข้อมูลเสียง.....	44
1. ฐานข้อมูลเสียงคำโดด.....	44
2. ฐานข้อมูลเสียงพูดต่อเนื่อง.....	46
การสกัดลักษณะเด่นจากฐานข้อมูลเสียง.....	48
1. การเตรียมค่าความถี่มูลฐานจากฐานข้อมูลเสียง.....	48

1.1. การแบ่งพยางค์.....	48
1.2. การสกัดค่าความถี่มูลฐาน.....	50
1.3. การสมมุทค่าความถี่มูลฐาน.....	52
2. การเตรียมลักษณะเด่นจากฐานข้อมูลความถี่มูลฐานเพื่อใช้ในการเรียนรู้ของเครื่อง.....	53
2.1. การเปลี่ยนสเกลความถี่.....	53
2.2. เทคนิคการทำให้เป็นบรรทัดฐาน.....	54
การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน.....	55
1. การกำหนดชุดลักษณะเด่น.....	56
1.1. ชุดลักษณะเด่นสำหรับตัวแยกประเภท.....	56
1.2. ชุดลักษณะเด่นสำหรับตัวแยกประเภทที่มีความเป็นลำดับ.....	57
2. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด.....	59
2.1. การจัดเตรียมการทดลอง.....	59
2.2. ผลการทดลอง.....	61
2.3. วิเคราะห์และอภิปรายผล.....	62
3. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับเสียงพูดต่อเนื่อง.....	74
2.1. การจัดเตรียมการทดลอง.....	74
2.2. ผลการทดลอง.....	76
2.3. วิเคราะห์และอภิปรายผล.....	77
การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมที่ได้จากการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยมนุษย์.....	89
1. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงาน.....	89
1.1. การวิเคราะห์ลักษณะเด่น.....	90
1.2. การจัดเตรียมการทดลอง.....	100
1.3. ผลการทดลองและสรุปผลการทดลอง.....	101

2. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติม เป็นค่าสเปกตรัม.....	103
2.1. การวิเคราะห์ลักษณะเด่น.....	103
2.2. การจัดเตรียมการทดลอง.....	110
2.3. ผลการทดลองและสรุปผลการทดลอง.....	111
บทที่ 5 การประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ภาษาไทย.....	114
ฐานข้อมูลเสียงของการสะกดคำ.....	114
1. การออกแบบฐานข้อมูลเสียง.....	114
2. การเก็บรวบรวมฐานข้อมูลเสียง.....	115
3. การถอดเสียง.....	116
4. การสร้างพจนานุกรมการออกเสียง.....	117
การทดลองและผลการทดลอง.....	118
1. การทดลองบนระบบอ้างอิงในการรู้จำเสียงการสะกดคำภาษาไทย.....	118
2. การทดลองบนระบบที่นำเสนอที่มีการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ ในการรู้จำเสียงการสะกดคำภาษาไทย.....	120
วิเคราะห์ผลการทดลอง.....	125
บทที่ 6 สรุปผลการวิจัย และข้อเสนอแนะ.....	128
สรุปผลการวิจัย.....	128
การทดลองการรู้จำเสียงวรรณยุกต์ของมนุษย์.....	128
การจำแนกเสียงวรรณยุกต์.....	129
การประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูด.....	130
ข้อเสนอแนะ.....	132
รายการอ้างอิง.....	133
ภาคผนวก.....	139
ภาคผนวก ก ตารางเทียบหน่วยเสียงไทยที่ใช้ในวิทยานิพนธ์.....	140
ภาคผนวก ข คำที่เลือกใช้สำหรับการสะกดคำภาษาไทย.....	142
ประวัติผู้เขียนวิทยานิพนธ์.....	148

สารบัญตาราง

ตารางที่		หน้า
2.1	ตารางแสดงสัญลักษณ์ของหน่วยเสียงพยัญชนะในภาษาไทย[13].....	10
2.2	วิธีการ (manner)และฐาน(place)ในการออกเสียงในภาษาไทย [13].....	10
2.3	ตารางแสดงสัญลักษณ์ของหน่วยเสียงสระในภาษาไทย.....	11
2.4	โครงสร้างของพยางค์ในภาษาไทย.....	13
2.5	จำนวนความถูกต้องและความผิดพลาดของระบบก่อนการปรับปรุงและหลังการปรับปรุง.....	19
3.1	เสียงวรรณยุกต์ภาษาไทยของคำว่า “ลา” (/l a:/) 5 รูปแบบ.....	32
3.2	ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน.....	40
3.3	ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยแบ่งตามช่วงของสเปกตรัม..	41
3.4	คอนฟิวชันเมทริกซ์ของสิ่งเร้าต้นแบบในการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน.....	42
3.5	คอนฟิวชันเมทริกซ์ของสิ่งเร้าต้นแบบในการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม.....	42
4.1	ตารางแสดงคำที่ผู้พูดพูดในฐานข้อมูลเสียงคำโดด.....	45
4.2	ตารางแสดงจำนวนคู่หน่วยเสียงที่ครอบคลุมในฐานข้อมูลเสียงโลดัสชุดหน่วยเสียงสมดุล.....	47
4.3	ตารางแสดงรายละเอียดของแต่ละชุดข้อมูลในฐานข้อมูลเสียงโลดัส[12].....	47
4.4	ตารางแสดงจำนวนเสียงวรรณยุกต์ภาษาไทยในฐานข้อมูลเสียงโลดัสชุดหน่วยเสียงสมดุลที่มีการบันทึกเสียงที่เนคเทค.....	50
4.5	ตารางแสดงการตั้งค่าพารามิเตอร์ของแพทในการสกัดค่าความถี่มูลฐาน.....	52
4.6	ชุดลักษณะเด่นสำหรับตัวแยกประเภท.....	57
4.7	ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด.....	61
4.8	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียงวรรณยุกต์ สามัญ สำหรับคลังเสียงพูดคำโดด.....	66

ตารางที่	หน้า
4.9	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ เอก สำหรับคลังเสียงพูดคำโดด..... 66
4.10	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ โท สำหรับคลังเสียงพูดคำโดด..... 67
4.11	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ ตรี สำหรับคลังเสียงพูดคำโดด..... 67
4.12	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ จัตวา สำหรับคลังเสียงพูดคำโดด..... 68
4.13	คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดย ใช้โครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราอี อาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน..... 73
4.14	คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดย ใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราอีอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่า คะแนนมาตรฐาน..... 73
4.15	ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่ มูลฐานสำหรับเสียงพูดต่อเนื่อง..... 77
4.16	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ สามัญ สำหรับคลังเสียงพูดต่อเนื่อง..... 85
4.17	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ เอก สำหรับคลังเสียงพูดต่อเนื่อง..... 85
4.18	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ โท สำหรับคลังเสียงพูดต่อเนื่อง..... 86
4.19	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ ตรี สำหรับคลังเสียงพูดต่อเนื่อง..... 86
4.20	การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง วรรณยุกต์ จัตวา สำหรับคลังเสียงพูดต่อเนื่อง..... 87

ตารางที่	หน้า	
4.21	คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้โครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราอาร์บี และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน.....	88
4.22	คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน.....	88
4.23	ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงานในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดดเปรียบเทียบกับกรจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ความถี่มูลฐานสำหรับฐานข้อมูลเสียงคำโดด.....	101
4.24	คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF_E ในสเกลความถี่อัตราอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน.....	102
4.25	ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัมในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดด.....	112
4.26	ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัมในการจำแนกเสียงวรรณยุกต์ภาษาไทยในเสียงพูดต่อเนื่อง.....	112
4.27	คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF_PLP ในสเกลความถี่อัตราอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน.....	113
5.1	การสะกดคำในภาษาไทย.....	115

ตารางที่		หน้า
5.2	ผลการทดลองการรู้จำเสียงพูดสำหรับการสะกดคำไทยโดยมีการปรับจำนวนสถานะและมิกเจอร์.....	119
5.3	ความสับสนของคู่คำที่มีเสียงพยัญชนะและสระเดียวกันของระบบอ้างอิงและระบบที่เสนอ.....	126
5.4	ค่าที่มีการรู้จำที่ผิดพลาดมากกว่าร้อยละ40ของระบบอ้างอิงและระบบที่เสนอ	127

สารบัญภาพ

ภาพที่		หน้า
2.1	รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบหนึ่งพยางค์[14]..	13
2.2	สัญญาณเสียงที่มีลักษณะเป็นคาบและค่าอัตราสัมพันธ์ที่ระยะเวลาต่างๆ[16].	15
2.3	การหาความผิดพลาดจากการประมาณพัลลวะเชิงเส้น [20].....	17
2.4	ขั้นตอนในการหาค่าการประมาณพัลลวะแบบอิงการรับฟังของมนุษย์[21].....	18
2.5	5 ขั้นตอนในการหาค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล [18].....	18
3.1	รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบหนึ่งพยางค์.....	32
3.2	รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบสองพยางค์.....	33
3.3	รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบสามพยางค์.....	33
3.4	ตัวอย่างสัญญาณไซน์ซอซอดต์ที่ใช้ในการศึกษาลักษณะรูปร่างของความถี่มูลฐานที่สังเคราะห์ได้จากสิ่งเร้าต้นแบบของคำว่า “ลา” ในเสียงสามัญรูปแบบหนึ่งพยางค์.....	34
3.5	ตัวอย่างสัญญาณไซน์ซอซอดต์ที่ใช้ในการศึกษารูปร่างของพลังงานที่ห่อหุ้มที่สังเคราะห์ได้จากสิ่งเร้าต้นแบบของคำว่า “ลา” ในเสียงสามัญรูปแบบหนึ่งพยางค์.....	35
3.6	สเปกตรัมของสัญญาณเสียงสิ่งเร้าต้นแบบ ณ เวลาหนึ่งๆ.....	37
3.7	ส่วนต่อประสานผู้ใช้ของโปรแกรมที่ใช้สำหรับระบุเสียงวรรณยุกต์ภาษาไทยของผู้เข้าร่วมทดลอง.....	38
3.8	รูปร่างของความถี่มูลฐานของสิ่งเร้าต้นแบบโดยทางด้านซ้ายเป็นรูปร่างความถี่มูลฐานเสียงวรรณยุกต์เอก และทางด้านขวาเป็นรูปร่างความถี่มูลฐานเสียงวรรณยุกต์จัตวา.....	43
4.1	กระบวนการในการเตรียมค่าความถี่มูลฐานจากฐานข้อมูลเสียง.....	48
4.2	ตัวอย่างการแบ่งพยางค์ในฐานข้อมูลเสียงคำโดดโดยใช้โปรแกรมเวฟเซอร์เฟอร์	49
4.3	รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของฐานข้อมูลเสียงคำโดด.....	63
4.4	รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของฐานข้อมูลเสียงคำโดด.....	63

ภาพที่	หน้า
4.6	รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของ ฐานข้อมูลเสียงคำโดด..... 64
4.7	รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่ง วอร์นยุคต์ ของฐานข้อมูลเสียงคำโดด..... 64
4.8	รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอาร์ บีของฐานข้อมูลเสียงคำโดด..... 64
4.9	รูปร่างความเร่งของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของ ฐานข้อมูลเสียงคำโดด..... 65
4.10	รูปร่างความเร่งของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวอร์นยุคต์ ของฐานข้อมูลเสียงคำโดด..... 65
4.11	รูปร่างความเร่งของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอาร์บี ของฐานข้อมูลเสียงคำโดด..... 65
4.12	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิร์ตซ์ซึ่งไม่มีการทำให้เป็น บรรทัดฐานของฐานข้อมูลเสียงคำโดด..... 69
4.13	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิร์ตซ์ซึ่งมีการทำให้เป็นบรรทัด ฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงคำโดด..... 69
4.14	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิร์ตซ์ซึ่งมีการทำให้เป็นบรรทัด ฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงคำโดด..... 69
4.15	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวอร์นยุคต์ซึ่งไม่มีการทำให้ เป็นบรรทัดฐานของฐานข้อมูลเสียงคำโดด..... 70
4.16	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวอร์นยุคต์ซึ่งมีการทำให้เป็น บรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงคำโดด..... 70
4.17	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวอร์นยุคต์ซึ่งมีการทำให้เป็น บรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงคำโดด..... 70
4.18	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่อัตราอาร์บีซึ่งไม่มีการทำให้เป็น บรรทัดฐานของฐานข้อมูลเสียงคำโดด..... 71
4.19	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่อัตราอาร์บีซึ่งมีการทำให้เป็น บรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงคำโดด..... 71

ภาพที่	หน้า
4.21	รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของฐานข้อมูล เสียงพูดต่อเนื่อง..... 79
4.22	รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของ ฐานข้อมูลเสียงพูดต่อเนื่อง..... 79
4.23	รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอาร์บีของ ฐานข้อมูลเสียงพูดต่อเนื่อง..... 79
4.24	รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของ ฐานข้อมูลเสียงพูดต่อเนื่อง..... 80
4.25	รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่ง วรรณยุกต์ ของฐานข้อมูลเสียงพูดต่อเนื่อง..... 80
4.26	รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอาร์บี ของฐานข้อมูลเสียงพูดต่อเนื่อง..... 80
4.27	รูปร่างความเร่งของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของ ฐานข้อมูลเสียงพูดต่อเนื่อง..... 81
4.28	รูปร่างความเร่งของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ ของฐานข้อมูลเสียงพูดต่อเนื่อง..... 81
4.29	รูปร่างความเร่งของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอาร์บี ของฐานข้อมูลเสียงพูดต่อเนื่อง..... 81
4.30	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิร์ตซ์ซึ่งไม่มีการทำให้เป็น บรรทัดฐานของฐานข้อมูลเสียงพูดต่อเนื่อง..... 82
4.31	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิร์ตซ์ซึ่งมีการทำให้เป็นบรรทัด ฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงพูดต่อเนื่อง..... 82
4.32	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิร์ตซ์ซึ่งมีการทำให้เป็นบรรทัด ฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงพูดต่อเนื่อง..... 82
4.33	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุกต์ซึ่งไม่มีการทำให้ เป็นบรรทัดฐานของฐานข้อมูลเสียงพูดต่อเนื่อง..... 83
4.34	ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุกต์ซึ่งมีการทำให้เป็น บรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงพูดต่อเนื่อง..... 83

ภาพที่	หน้า
4.36	84
4.37	84
4.38	84
4.39	90
4.40	91
4.41	91
4.42	91
4.43	92
4.44	92
4.45	92
4.46	93
4.47	93
4.48	93
4.49	94
4.50	94
4.51	94

ภาพที่	หน้า	
4.53	แผนภาพกล่องของความแรงของควมถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 100.....	95
4.54	กราฟของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน.....	96
4.55	กราฟผลต่างของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน.....	97
4.56	กราฟความแรงของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน.....	97
4.57	แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 0.....	98
4.58	แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 25.....	98
4.59	แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 50.....	99
4.60	แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 75.....	99
4.61	แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 100.....	99
4.62	สเปกโตรแกรมของคำว่า “กอ” “ออ” และ “อู”	103
4.63	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์สามัญในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ.....	104
4.64	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์เอกในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ.....	104
4.65	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์โทในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ.....	105
4.66	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์ตรีในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ.....	105
4.67	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์จัตวาในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ.....	106
4.68	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์สามัญในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด.....	106
4.69	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์เอกในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด.....	107

ภาพที่	หน้า
4.71	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์ตรีในฐานข้อมูลเสียงพูดต่อเนื่อง จัดกลุ่มโดยพยัญชนะตัวสะกด..... 108
4.72	รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์จัตวาในฐานข้อมูลเสียงพูด ต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด..... 108
4.73	รูปร่างของความถี่มูลฐานของรูปร่างของเสียงวรรณยุกต์สามัญที่มีตัวสะกดเป็น แม่เอกกับเสียงวรรณยุกต์เอกที่มีตัวสะกดเป็นแม่.กา..... 109
5.1	โปรแกรมแสดงการสะกดคำให้ผู้พูดออกเสียง..... 116
5.2	โปรแกรมที่ผู้ถอดเสียงใช้ในการตรวจสอบความถูกต้องของการถอดเสียง..... 117
5.3	แผนภาพระบบที่นำเสนอในการรู้จำเสียงการสะกดคำในภาษาไทย..... 120
5.4	การคำนวณความน่าจะเป็นของเสียงวรรณยุกต์..... 122
5.5	กราฟความแม่นยำของการรู้จำเสียงพูดเมื่อปรับค่าน้ำหนัก w ในค่าที่ต่างๆกัน.. 124

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในหลายปีที่ผ่านมาเทคโนโลยีที่เกี่ยวข้องกับเสียงพูดถูกนำมาประยุกต์ใช้ในการการสื่อสารและโต้ตอบระหว่างมนุษย์กับคอมพิวเตอร์กันอย่างแพร่หลาย โดยเฉพาะอย่างยิ่งเทคโนโลยีระบบรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition System: ASRs) ได้มีการพัฒนาขึ้นอย่างต่อเนื่องและแพร่หลายจนมีความก้าวหน้าไปเป็นอย่างมาก ทำให้ระบบรู้จำเสียงพูดอัตโนมัติเข้ามามีบทบาทในชีวิตประจำวันมากยิ่งขึ้น เห็นได้จากการนำระบบรู้จำเสียงพูดอัตโนมัติมาประยุกต์ใช้งานในระบบต่างๆ เช่น ระบบผู้ช่วยอัจฉริยะ “สิริ” (Siri) [1] ระบบเขียนตามคำพูดอัตโนมัติ (Dragon Dictation) [2] และระบบสอบถามข้อมูลการจราจรผ่านทางโทรศัพท์เคลื่อนที่โดยใช้เสียง (Traffic Voice Information Services : TVIS) [3] เป็นต้น

สำหรับภาษาไทยได้มีการศึกษาและทดลองนำระบบรู้จำเสียงจากภาษาต่างประเทศเช่นภาษาอังกฤษ มาประยุกต์ใช้เพื่อสร้างระบบรู้จำเสียงพูดภาษาไทย แต่ผลลัพธ์ที่ได้กลับไม่ดีเท่าการรู้จำเสียงพูดในภาษาเหล่านั้น ปัจจัยหนึ่งที่ทำให้ระบบรู้จำเสียงพูดมีประสิทธิภาพลดลงเมื่อนำมาประยุกต์ใช้กับภาษาไทย เนื่องจากระบบรู้จำเสียงพูดส่วนใหญ่ที่นำมาประยุกต์ใช้จะถูกออกแบบมาสำหรับภาษาที่ไม่มีเสียงวรรณยุกต์เช่น ภาษาอังกฤษ เป็นต้น ซึ่งต่างกับภาษาไทยที่เป็นภาษาที่มีเสียงวรรณยุกต์ (Tonal Language) [4] เพราะเสียงพยัญชนะ (Consonants) และ เสียงสระ (Vowel) ไม่อาจแยกแยะความหมายของคำให้ชัดเจนได้ ยกตัวอย่างเช่น คำว่า คา /kha:/ สำหรับภาษาไทย ซึ่งมีเสียงวรรณยุกต์ 5 รูปแบบได้แก่ เสียงวรรณยุกต์สามัญ (Mid Tone), เสียงวรรณยุกต์เอก (Low Tone), เสียงวรรณยุกต์โท (Falling Tone), เสียงวรรณยุกต์ตรี (High Tone) และเสียงวรรณยุกต์จัตวา (Rising Tone) จะสามารถ ออกเสียง ได้เป็น คา (kha^0), ข่า (kha^1), ข๋า (kha^2), ค้า (kha^3) และ ขา (kha^4) ทั้ง 5 คำนี้มีความหมายที่แตกต่างกัน แต่หน่วยเสียง (Phoneme) ของพยัญชนะ และสระเป็นหน่วยเสียงเดียวกัน ซึ่งลักษณะนี้แตกต่างจากภาษาที่ไม่มีเสียงวรรณยุกต์ที่หน่วยเสียงของพยัญชนะ และสระจะสามารถใช้ในการระบุความหมายได้ ดังนั้น

จากตัวอย่างที่ยกขึ้นมาจะเห็นได้ว่าการจำแนกเสียงวรรณยุกต์ภาษาไทยมีความจำเป็นต่อการพัฒนาระบบรู้จำเสียงพูดภาษาไทยให้มีประสิทธิภาพ และความถูกต้องมากยิ่งขึ้น

การจำแนกเสียงวรรณยุกต์ภาษาไทย [5, 6, 7, 8] สามารถแบ่งการแก้ปัญหาด้วยเทคนิคในการเรียนรู้ของเครื่องได้ 2 แบบใหญ่ ได้แก่การจำแนกด้วยตัวจำแนกที่ไม่มีความเป็นลำดับ (Classifier) กับการจำแนกด้วยตัวจำแนกที่มีความเป็นลำดับ (Sequential Classifier) โดยในงานวิจัยที่ใช้การจำแนกด้วยตัวจำแนกที่ไม่มีความเป็นลำดับ [5, 6, 7] จะทำการเลือกลักษณะเด่นที่จะใช้ ซึ่งในงานวิจัยกลุ่มนี้จะเลือกลักษณะเด่นเป็น ความถี่มูลฐาน (Fundamental Frequency) และผลต่างของความถี่มูลฐาน (Delta of Fundamental Frequency) แต่เนื่องจากข้อมูลเสียงเป็นข้อมูลแบบอนุกรมเวลา (Time Series) ทำให้ต้องมีการแปลงอนุกรมเวลาให้อยู่ในรูปของเวกเตอร์ที่มีขนาดเฉพาะ เนื่องจากในอนุกรมเวลาข้อมูลแต่ละชุดนำเข้าจะมีความขนาดไม่เท่ากัน ดังนั้นจึงต้องมีการแปลงด้วยเทคนิคการเลือกลักษณะเด่น ซึ่งการแปลงข้อมูลให้อยู่ในรูปของเวกเตอร์ที่มีขนาดเฉพาะจะมีความซับซ้อนสูงในการหากฎเกณฑ์ที่เหมาะสมในการแปลงข้อมูล เพราะในบางครั้งการแปลงข้อมูลด้วยกฎเกณฑ์ที่ไม่เหมาะสม เวกเตอร์ที่ได้จะไม่สะท้อนลักษณะของอนุกรมเวลานั้นจริงและอาจจะทำให้ลักษณะเด่นบางอย่างสูญหายไประหว่างการแปลงข้อมูล [9] เพื่อหลีกเลี่ยงปัญหาดังกล่าวจึงมีงานวิจัยในการจำแนกเสียงวรรณยุกต์ที่ใช้ตัวจำแนกที่มีความเป็นลำดับ [8] โดยใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Models: HMMs) ในการจำแนกเสียงวรรณยุกต์ ซึ่งไม่จำเป็นต้องทำการแปลงลักษณะเด่นเป็นข้อมูลเวกเตอร์ขนาดเฉพาะ แต่อย่างไรก็ตามวิธีดังกล่าวยังพบว่ามีปัญหาเนื่องจากแบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองในกลุ่มเจเนอเรทีฟ (Generative model) ซึ่งแบบจำลองในกลุ่มนี้มีข้อเสีย เนื่องจากแบบจำลองมีการตั้งสมมุติฐานว่าแต่ละข้อมูลที่สังเกตได้ในลำดับของข้อมูลที่สังเกต (Observation Sequence) จะเป็นอิสระต่อกัน ซึ่งในความเป็นจริงแล้วสำหรับสัญญาณเสียงพูดนั้นจะมีความต่อเนื่อง (ความไม่เป็นอิสระต่อกัน) ของสัญญาณเสียงในแต่ละลำดับเวลา ส่งผลให้สมมุติฐานของแบบจำลองฮิดเดนมาร์คอฟนั้นไม่ถูกต้อง ดังนั้นในงานวิจัย [5, 6, 7] จึงเลือกใช้การจำแนกเสียงวรรณยุกต์ด้วยแบบจำลองโครงข่ายประสาทเทียม ถึงแม้แบบจำลองจะเป็นตัวจำแนกที่ไม่มีความเป็นลำดับ แต่

แบบจำลองดังกล่าวจัดอยู่ในแบบจำลองกลุ่มดิสคริมิเนทีฟ (Discriminative Model) ซึ่งไม่มีการตั้งสมมุติฐานที่เป็นอิสระต่อกัน

เมื่อไม่นานมานี้มีการคิดค้นแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ซึ่งเป็นแบบจำลองในกลุ่มดิสคริมิเนทีฟ และเป็นแบบจำลองที่มีความเป็นลำดับ โดยแบบจำลองดังกล่าวถูกเสนอเพื่อนำมาประยุกต์ใช้งานในการจำแนกและรู้จำหน่วยเสียง [10, 11] ซึ่งได้ผลความถูกต้องในการรู้จำหน่วยเสียงที่ดีกว่าแบบจำลองฮิดเดนมาร์คอฟ

จากที่มาดังกล่าวในตอนแรกวิทยานิพนธ์นี้จะเสนอฟีเจอร์ฟังก์ชัน (Feature Function) ที่เหมาะสมกับภาษาไทยโดยจะนิยามฟีเจอร์ฟังก์ชันจากลักษณะเฉพาะทางภาษาไทยรวมถึงจะทำการทดลองเพื่อวัดความเหมาะสมของฟีเจอร์ฟังก์ชันในการรู้จำเสียงพูดภาษาไทยอัตโนมัติ แต่หลังจากการทดลองเบื้องต้นพบว่าระยะเวลาในการเรียนรู้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์สำหรับหน่วยเสียง (Phoneme) ทั้งหมดในภาษาไทยนั้นสูงมากทำให้การปรับและการทดลองเลือกแต่ละฟีเจอร์ฟังก์ชันทำได้ยาก จึงได้ลดขนาดของปัญหาลงมาเป็นการจำแนกเสียงวรรณยุกต์ภาษาไทยแทน ซึ่งทำให้การเรียนรู้แบบจำลองอยู่ในช่วงเวลาที่เหมาะสม นอกจากนี้จากการทบทวนวรรณกรรมเพิ่มเติมพบว่าการนิยามฟีเจอร์ฟังก์ชันที่มีการนำเสนอใน [10, 11] เป็นฟีเจอร์ฟังก์ชันที่มีการนิยามดีแล้วกล่าวคือ การนิยามฟีเจอร์ไม่ได้ขึ้นกับภาษาใดๆ ซึ่งน่าจะสามารถนำมาประยุกต์ใช้งานกับภาษาไทยได้อยู่แล้ว ด้วยเหตุดังกล่าวจึงมีการปรับวัตถุประสงค์ในการวิจัยของวิทยานิพนธ์ให้มีความเหมาะสมมากขึ้น

ดังนั้นในวิทยานิพนธ์นี้จะนำแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์มาใช้ในการจำแนกเสียงวรรณยุกต์ภาษาไทย อีกทั้งจะพิจารณาลักษณะเด่นที่เหมาะสมในการจำแนกเสียงวรรณยุกต์ด้วยแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ และสุดท้ายนำเสนอการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ในการรู้จำเสียงพูดภาษาไทย เพื่อเป็นการแสดงให้เห็นถึงความสำคัญของปัญหาการจำแนกเสียงวรรณยุกต์ภาษาไทย

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและเสนอลักษณะเด่นที่มีประโยชน์ต่อการจำแนกเสียงวรรณยุกต์ภาษาไทยทั้งในคำโดด และเสียงพูดต่อเนื่อง
2. เพื่อศึกษาและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์
3. แสดงการประยุกต์ใช้งานการจำแนกเสียงวรรณยุกต์ภาษาไทยที่ใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ เพื่อการรู้จำเสียงพูดภาษาไทย

ขอบเขตของการวิจัย

1. วิธีการที่นำเสนอจะเป็นการจำแนกเสียงวรรณยุกต์สำหรับภาษาไทยทั้งคำโดดและเสียงพูดต่อเนื่อง
2. ในงานวิจัยนี้การศึกษาและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทยจะถือว่ารู้ขอบเขตของพยางค์ของเสียงวรรณยุกต์นั้นที่แน่ชัดก่อนการจำแนกเสียงวรรณยุกต์
3. ในการศึกษาและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ จะพิจารณาเฉพาะตำแหน่งเสียงวรรณยุกต์ที่สนใจจำแนกเท่านั้น โดยไม่สนใจบริบทข้างเคียง ทำนองเสียง และการเน้นหนักเบา
4. ในการเสนอลักษณะเด่นจะเสนอลักษณะเด่นที่เหมาะสมแก่การจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์เท่านั้น
5. ในการศึกษาและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์สำหรับเสียงพูดต่อเนื่องภาษาไทยจะใช้ฐานข้อมูลเสียงภาษาไทยโลตัส (LOTUS) [12] โดยจะเลือกใช้ชุดหน่วยเสียงสมดุล
6. ในการวัดผลการจำแนกจะเปรียบเทียบผลการจำแนกเสียงวรรณยุกต์กับตัวอ้างอิง (baseline) ที่ใช้โครงข่ายประสาทเทียม และ แบบจำลองฮิดเดนมาร์คอฟ
7. การแสดงการรู้จำเสียงพูดภาษาไทยที่มีการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ภาษาไทยด้วยแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ จะดำเนินการทดลองบนโดเมนการสะกดคำโดดภาษาไทย

วิธีการดำเนินการวิจัย

1. ขั้นตอนการศึกษาเบื้องต้น
 - 1.1. ศึกษาข้อมูลและงานวิจัยที่เกี่ยวข้องกับการจำแนกเสียงวรรณยุกต์ภาษาไทย
 - 1.2. ศึกษาข้อมูลและงานวิจัยที่เกี่ยวข้องกับการจำแนกเสียงวรรณยุกต์ภาษาอื่นๆ ที่มีเสียงวรรณยุกต์
 - 1.3. ศึกษาข้อมูลและงานวิจัยที่เกี่ยวกับแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์
 - 1.4. ศึกษาข้อมูลที่เกี่ยวข้องกับการสะกดคำในภาษาไทย
 - 1.5. วิเคราะห์สัญญาณเสียงเพื่อหาลักษณะเด่นที่มีความเกี่ยวข้องกับการจำแนกเสียงวรรณยุกต์ภาษาไทย
2. ขั้นตอนการทดลองศึกษาเบื้องต้น
 - 2.1. ศึกษาข้อมูลเกี่ยวกับการรับรู้เสียงวรรณยุกต์ภาษาไทยของมนุษย์เพื่อใช้ในการเลือกลักษณะเด่นสำหรับการจำแนกเสียงวรรณยุกต์ภาษาไทย
 - 2.2. ออกแบบการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยของมนุษย์ โดยนำลักษณะเด่นที่ได้ศึกษาในเบื้องต้นมาศึกษาเพิ่มเติม
 - 2.3. ทดลองและบันทึกผลการรับรู้เสียงวรรณยุกต์ภาษาไทยของมนุษย์
 - 2.4. วิเคราะห์และสรุปผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยของมนุษย์
3. ขั้นตอนการทดลองและทดสอบประสิทธิภาพ
 - 3.1. จัดเตรียมฐานข้อมูลเสียงคำโดดภาษาไทยสำหรับการจำแนกเสียงวรรณยุกต์ภาษาไทย
 - 3.2. พัฒนาระบบอ้างอิงในการจำแนกเสียงวรรณยุกต์ภาษาไทยทั้งคำโดดและเสียงพูดต่อเนื่อง
 - 3.3. พัฒนาระบบจำแนกเสียงวรรณยุกต์ภาษาไทยทั้งคำโดดและเสียงพูดต่อเนื่องโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์
 - 3.4. วิเคราะห์และสรุปผลการทดลอง
 - 3.5. เปรียบเทียบผลวิเคราะห์กับลักษณะเด่นที่ได้จากการรับรู้เสียงวรรณยุกต์ภาษาไทยของมนุษย์ เพื่อที่จะใช้ในการเพิ่มประสิทธิภาพการจำแนกเสียงวรรณยุกต์ภาษาไทย

- 3.6. ทดลองเพิ่มเติมโดยใช้ลักษณะเด่นที่ได้จากขั้นตอนก่อนหน้าโดยทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยทั้งคำโดดและเสียงพูดต่อเนื่อง
- 3.7. วิเคราะห์และสรุปผลการทดลอง
4. ขั้นตอนการออกแบบและทดลองระบบ
 - 4.1. จัดเตรียมฐานข้อมูลเสียงการสะกุดคำโดดในภาษาไทย
 - 4.2. พัฒนาระบบอ้างอิงในการรู้จำเสียงพูด การสะกุดคำที่เป็นคำโดดภาษาไทย
 - 4.3. พัฒนาการรู้จำแนกเสียงวรรณยุกต์ภาษาไทยบนฐานข้อมูลเสียงการสะกุดคำโดดในภาษาไทย
 - 4.4. พัฒนาการรู้จำเสียงพูดการสะกุดคำที่เป็นคำโดดในภาษาไทย ที่มีการนำการจำแนกเสียงวรรณยุกต์ภาษาไทยมาประยุกต์
 - 4.5. ทดลองและประเมินผลระบบ
 - 4.6. วิเคราะห์และสรุปผลการทดลอง
5. ขั้นตอนการสรุปผลและเรียบเรียงวิทยานิพนธ์

ประโยชน์ที่คาดว่าจะได้รับ

1. พบว่าพลังงานมีส่วนช่วยให้การรับรู้และการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดมีประสิทธิภาพมากขึ้น
2. พบว่าข้อมูลทางสเปกตรัมโดยเฉพาะอย่างยิ่งตั้งแต่ความถี่ฟอร์แมนต์ที่หนึ่ง (1st Formant) มีผลช่วยให้การรับรู้และการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องมีความถูกต้องมากขึ้น
3. ได้วิธีการพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทยด้วยแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่มีประสิทธิภาพมากขึ้น
4. สามารถนำการจำแนกเสียงวรรณยุกต์ภาษาไทยที่พัฒนาขึ้นมาไปประยุกต์ใช้กับการรู้จำเสียงพูดภาษาไทยได้
5. ได้วิธีการพัฒนาการรู้จำเสียงพูดการสะกุดคำในภาษาไทยที่มีความถูกต้องมากยิ่งขึ้น

ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตอบรับและตีพิมพ์ในบทความทางวิชาการ 2 หัวข้อ
เรื่องได้แก่

1. “Contribution of Spectral Shapes to Tone Perception” โดย “Natthawut Kertkeidkachorn”, “Surapol Vorapatratom”, “Sirinart Tangruamsub”, “Proadpran Punyabukkana” และ “Atiwong Suchato” ในงานประชุมวิชาการนานาชาติ “13th Annual Conference of the International Speech Communication Association” (Interspeech 2012) ณ Portland Oregon, USA ในวันที่ 9-13 กันยายน 2555
2. “THE CU-MFEC CORPUS FOR THAI AND ENGLISH SPELLING SPEECH RECOGNITION” โดย “Natthawut Kertkeidkachorn”, “Supadaech Chanjaradwichai”, “Teera Suri”, “Krerksak Likitsupin”, “Surapol Vorapatratom”, “Pawanrat Hirankan”, “Worasa Limpanadusadee”, “Supakit Chuetanapinyo”, “Kitanan Pitakpawatkul”, “Natnarong Puangsri”, “Nathacha Tangsirirat”, “Konlawachara Trakulsuk”, “Proadpran Punyabukkana” และ “Atiwong Suchato” ในงานประชุมวิชาการนานาชาติ “15th the Oriental Co-ordination and Standardization of Speech Databases and Assessment Conference” (O-COCOSDA 2012) ณ University of Macau Library มาเก๊า ในวันที่ 9-12 ธันวาคม 2555

ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

ในวิทยานิพนธ์เล่มนี้ได้แบ่งการนำเสนอออกเป็น 6 ส่วนดังต่อไปนี้

1. บทที่ 1 บทนำจะกล่าวถึงที่มาและความสำคัญของปัญหา เชื่อมโยงไปถึงวัตถุประสงค์ของการวิจัย ตลอดจนขอบเขตในการวิจัย ขั้นตอนในการวิจัย ประโยชน์ที่จะได้รับ รวมถึงผลงานที่ได้รับการเผยแพร่ และลำดับในการจัดเรียงเนื้อหาของวิทยานิพนธ์

2. บทที่ 2 จะกล่าวถึง 2 ส่วน ได้แก่ แนวคิดและทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยแนวคิดและทฤษฎีจะประกอบไปด้วย ทฤษฎีทางภาษาศาสตร์, ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์สัญญาณเสียง และทฤษฎีที่เกี่ยวข้องกับการวัดผล และในงานวิจัยที่เกี่ยวข้องจะประกอบด้วย งานวิจัยที่เกี่ยวข้องกับการจำแนกเสียงวรรณยุกต์ และงานวิจัยที่เกี่ยวข้องกับแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์
3. บทที่ 3 จะกล่าวถึงการทดลองการรับรู้เสียงวรรณยุกต์ของมนุษย์ โดยในบทนี้จะประกอบด้วย ความสำคัญของการทดลอง, การเตรียมสิ่งไว้, การทดลอง และผลการทดลองและสรุปผลการทดลอง
4. บทที่ 4 จะเป็นบทที่นำเสนอการจำแนกเสียงวรรณยุกต์ภาษาไทย โดยในบทได้แบ่งเนื้อหาออกเป็น 4 ส่วนหลักได้แก่ ส่วนฐานข้อมูลเสียง, ส่วนการสกัดลักษณะเด่นจากฐานข้อมูลเสียง, ส่วนการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน และ ส่วนการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมที่ได้จากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยมนุษย์
5. บทที่ 5 จะเป็นกล่าวถึงการนำการจำแนกเสียงวรรณยุกต์ไปประยุกต์ใช้กับการรู้จำเสียงพูด โดยในบทนี้จะแบ่งเนื้อหาเป็น การสร้างฐานข้อมูลเสียงของการสะกดคำภาษาไทย, การทดลองและผลการทดลองของระบบอ้างอิงและระบบที่นำเสนอที่มีการนำการจำแนกเสียงวรรณยุกต์ไปประยุกต์ใช้ และในส่วนสุดท้ายจะเป็นการวิเคราะห์ผลการทดลอง
6. บทที่ 6 จะเป็นการสรุปผลการวิจัย โดยจะสรุปผลการวิจัยแต่ละการทดลองของงานวิจัย และนำเสนอข้อสรุปรวมของงานวิจัย รวมไปถึงข้อเสนอแนะ

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

แนวคิดและทฤษฎี

สำหรับแนวคิดและทฤษฎีที่เกี่ยวข้องกับงานวิจัยนี้ประกอบด้วยสามส่วนหลักได้แก่ ทฤษฎีทางภาษาศาสตร์, ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์สัญญาณเสียง และทฤษฎีที่เกี่ยวข้องกับการวัดผล

1. ทฤษฎีทางภาษาศาสตร์

ทฤษฎีทางภาษาศาสตร์ที่เกี่ยวข้องจะประกอบด้วย 2 ส่วนได้แก่ ทฤษฎีของระบบเสียงในภาษา และ โครงสร้างพยางค์ในภาษาไทย

1.1. ระบบเสียงในภาษาไทย

ระบบเสียงในภาษาไทย [13] ประกอบด้วย 3 ส่วนได้แก่ เสียงพยัญชนะ, เสียงสระ, และเสียงวรรณยุกต์ โดยเสียงแต่ละประเภทประกอบด้วยหน่วยเสียงดังนี้ เสียงพยัญชนะประกอบด้วยหน่วยเสียง 21 หน่วยเสียง เสียงสระประกอบด้วยหน่วยเสียง 18 หน่วยเสียง และเสียงวรรณยุกต์ประกอบด้วยหน่วยเสียง 5 หน่วยเสียง

พยัญชนะในภาษาไทยนั้นมี 44 รูปได้แก่ ก ข ค ศ ฌ ง จ ฉ ช ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ล ว ศ ษ ส ห ฬ อ ฮ และประกอบด้วยหน่วยเสียง 21 หน่วยเสียง ดังที่แสดงในตารางที่ 2.1 โดยที่แต่ละหน่วยเสียงมีลักษณะวิธีการออกเสียงและฐานในการออกเสียงดังตารางที่ 2.2

ตารางที่ 2.1 ตารางแสดงสัญลักษณ์ของหน่วยเสียงพยัญชนะในภาษาไทย [13]

พยัญชนะ	หน่วยเสียง		พยัญชนะ	หน่วยเสียง	
	พยัญชนะต้น	พยัญชนะสะกด		พยัญชนะต้น	พยัญชนะสะกด
ก	/k/	/k/	บ	/b/	/p/
ข, ค, ช	/kh/	/k/	ป	/p/	/p/
ง	/ng/	/ng/	ผ, พ, ภ	/ph/	/p/
จ	/c/	/t/	ฝ, ฟ	/f/	/p/
ฉ, ฌ, ฌ	/ch/	/t/	ม	/m/	/m/
ซ, ศ, ษ, ส	/s/	/t/	ร	/r/	/n/
ญ, ย	/j/	/j/	ล, ฬ	/l/	/n/
ฎ, ฏ	/d/	/t/	ว	/w/	/w/
ฐ, ฑ	/t/	/t/	ห, ฮ	/h/	-
ฐ, ฑ, ฒ, ณ, ฑ	/th/	/t/	อ	/ʔ/	-
ณ, น	/n/	/n/			

ตารางที่ 2.2 วิธีการ (manner) และฐาน (place) ในการออกเสียงในภาษาไทย [13]

	ริมฝีปาก ทั้งสอง	ริมฝีปาก ล่างพื้นบน	ปุ่มเหงือก	หลังปุ่ม เหงือก	เพดาน แข็ง	เพดาน อ่อน	ผนังคอ
เสียงกัก	/p/ph/ b/		/t/th/d/			/k/kh/	/ʔ/
เสียงนาสิก	/m/		/n/			/ng/	
เสียงเสียด แทรก		/f/	/s/				/h/
เสียงผสม เสียดแทรก				/c/ch			
เสียงรัวลิ้น			/r/				
เสียงเปิด					/j/		
เสียงข้างลิ้น			/l/				

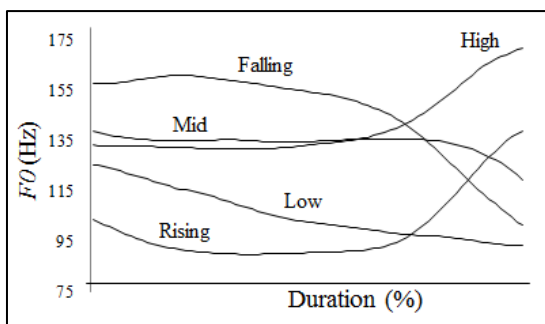
เสียงสระในภาษาไทยมีทั้งหมด 28 หน่วยเสียง ประกอบด้วยสระเสียงเดียว 18 หน่วยเสียง (เสียงของสระเดี่ยว ประกอบด้วยสระเสียงสั้น 9 หน่วยเสียง และสระเสียงยาว 9 หน่วยเสียง) สระประสม 6 หน่วยเสียง และ สระเกิน 4 หน่วยเสียง ดังในตารางที่ 2.3 สำหรับหน้าที่ของสระในภาษาไทยจะทำหน้าที่เป็นแกนกลางของพยางค์ ซึ่งจะไม่สามารถสร้างพยางค์ได้ถ้าหากขาดเสียงสระ นอกจากนั้นสระยังมีลักษณะพิเศษคือ จะไม่มีการกักเสียงไว้ซึ่งสามารถกล่าวได้ว่า ช่องทางการออกเสียง (Vocal Tract) จะเปิดตลอดช่วงเวลาที่ออกเสียงหน่วยเสียงสระนั้น

ตารางที่ 2.3 ตารางแสดงสัญลักษณ์ของหน่วยเสียงสระในภาษาไทย

ประเภทสระ	เสียงสั้น		เสียงยาว	
	รูปสระ	หน่วยเสียง	รูปสระ	หน่วยเสียง
สระเสียงเดียว	ะ	/a/	า	/aa/
	ิ	/i/	ี	/ii/
	ึ	/v/	ือ	/vv/
	ุ	/u/	ู	/uu/
	เะ	/e/	เ	/ee/
	แะ	/x/	แ	/xx/
	โะ	/o/	โ	/oo/
	เาะ	/@/	อ	/@@/
	เอะ	/q/	อ	/qq/
สระประสม	ัวะ	/ua/	ัว	/uua/
	เียะ	/va/	เีย	/vva/
	เือะ	/ia/	เือ	/iia/
สระเกิน	ำ	/am/		
	เ , เ	/aj/		
	เา	/aw/		

เสียงวรรณยุกต์เป็นองค์ประกอบหนึ่งของระบบเสียงในภาษาไทย ที่มีความสำคัญเป็นอย่างมาก [4,8,13] เพราะเมื่อเสียงวรรณยุกต์ต่างกันความหมายของคำก็ต่างกันเช่น คา (kha^0) , ข่า (kha^1) , ข่า (kha^2) , ค้า (kha^3) และ ขา (kha^4) แม้ว่าหน่วยเสียงพยัญชนะกับหน่วยเสียงสระทั้ง 5 คำที่ยกมาจะเหมือนกัน แต่หน่วยเสียงวรรณยุกต์ของทั้ง 5 คำนั้นแตกต่างกัน ส่งผลให้ความหมายของทั้ง 5 คำแตกต่างกันด้วย

สำหรับเสียงวรรณยุกต์ในภาษาไทยแบ่งออกเป็น 5 เสียงตามรูปร่างของความถี่มูลฐาน ดังแสดงในภาพที่ 2.1 ได้แก่ เสียงวรรณยุกต์สามัญ เสียงวรรณยุกต์เอก เสียงวรรณยุกต์โท เสียงวรรณยุกต์ตรี และเสียงวรรณยุกต์จัตวา แต่ละเสียงวรรณยุกต์จะมีลักษณะรูปร่างของความถี่มูลฐานตามชื่อที่ถูกตั้งไว้ กล่าวคือสำหรับเสียงวรรณยุกต์สามัญ (Mid Tone) ลักษณะของความถี่มูลฐานจะคงที่และอยู่ในระดับกลาง สำหรับเสียงวรรณยุกต์เอก (Low Tone) ลักษณะของความถี่มูลฐานจะคงที่และอยู่ในระดับต่ำ สำหรับเสียงวรรณยุกต์ตรี (High Tone) ลักษณะของความถี่มูลฐานจะคงที่และอยู่ในระดับสูง ส่วนเสียงวรรณยุกต์โท (Falling Tone) ลักษณะของความถี่มูลฐานจะขึ้นสูงในช่วงแรกและตกลงอย่างรวดเร็วในช่วงท้าย ในขณะที่เสียงวรรณยุกต์จัตวา (Rising Tone) ลักษณะของความถี่มูลฐานจะลดต่ำลงในช่วงแรกและกลับมาเพิ่มขึ้นอย่างเด่นชัดในช่วงหลัง ซึ่งจากลักษณะรูปร่างของความถี่มูลฐานดังกล่าว Luksaneeyanawin [4] ได้ทำการแบ่งกลุ่มของเสียงวรรณยุกต์ภาษาไทยตามการเปลี่ยนแปลงของความถี่มูลฐานออกเป็นสองกลุ่มหลักคือ กลุ่มเสียงวรรณยุกต์พลวัต (Static Tone) และกลุ่มเสียงวรรณยุกต์พลวัต (Dynamic Tone) ซึ่งในกลุ่มเสียงวรรณยุกต์พลวัตประกอบด้วย เสียงวรรณยุกต์ สามัญ เอก และตรี ในขณะที่กลุ่มเสียงวรรณยุกต์พลวัตประกอบด้วยเสียงวรรณยุกต์ โท และจัตวา เสียงวรรณยุกต์ในกลุ่มพลวัตจะมีอัตราการเปลี่ยนแปลงของความถี่มูลฐานที่คงที่ แต่เสียงวรรณยุกต์ในกลุ่มพลวัตจะมีการเปลี่ยนแปลงของความถี่มูลฐานอย่างรวดเร็ว



ภาพที่ 2.1 รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบหนึ่งพยางค์ [14]

1.2. โครงสร้างพยางค์ในภาษาไทย

โครงสร้างพยางค์ (Syllable Structure) ในภาษาไทย [15] จะประกอบด้วยส่วนที่เป็น ออนเซต (Onset) คือ พยางค์ต้นของพยางค์ ส่วนที่เป็นนิวเคลียส หรือแกนกลางของพยางค์ (Nucleus) และส่วนที่เป็นตัวสะกด (Coda) ซึ่งในบางครั้งจะพิจารณาส่วนที่เป็นแกนกลางของพยางค์กับตัวสะกดเป็นกลุ่มเดียวกันโดยเรียกใหม่ว่า ราม (Rhyme) ในตารางที่ 2.4 ได้แสดงโครงสร้างของพยางค์ในภาษาไทย

ตารางที่ 2.4 โครงสร้างของพยางค์ในภาษาไทย

โครงสร้างของพยางค์ในภาษาไทย		
ออนเซต	ราม	
	นิวเคลียส	ตัวสะกด
	T	
$C_i(C_f)$	$V(:)$	(C_f)

โดยที่ C_i แทนเสียงพยางค์ต้นที่เป็นไปได้, V แทนเสียงสระ, $:$ แทนความยาวของสระเสียงยาว, C_f แทนเสียงตัวสะกด, $()$ หมายถึงจะมีสิ่งที่อยู่ในวงเล็บก็ได้ หรือไม่มีก็ได้ สำหรับ (C_i) จะเป็นไปได้แค่ 3 สามพยางค์เท่านั้นได้แก่ “ร” “ล” และ “ว” ซึ่งเมื่อเกิดรูปแบบที่มีเป็น C_iC_i จะเรียกรูปแบบนี้ว่าคำควบกล้ำ และ T แทนเสียงวรรณยุกต์ ซึ่งจะเห็นได้ว่าเสียงวรรณยุกต์เป็นเสียงที่ครอบคลุมตลอดทั้งพยางค์ ดังนั้นในการจะวิเคราะห์เสียงวรรณยุกต์จึงควรพิจารณาขอบเขตของพยางค์ให้ชัดเจนก่อน

2. ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์สัญญาณเสียง

ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์สัญญาณเสียงประกอบด้วย อัตสหสัมพันธ์ (Autocorrelation), การคำนวณค่าพลังงานของสัญญาณ (Energy), การแปลงฟูริเยร์แบบช่วงเวลาสั้น (Short-Time Fourier Transform), ค่าสัมประสิทธิ์การประมาณพหุเชิงเส้น (Linear Prediction Coefficients: LPC), ค่าการประมาณพหุเชิงเส้นแบบอิงการรับฟังของมนุษย์ (Perceptual Linear Predictive: PLP), และค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล (Mel-Frequency Cepstral Coefficients: MFCC)

2.1. อัตสหสัมพันธ์

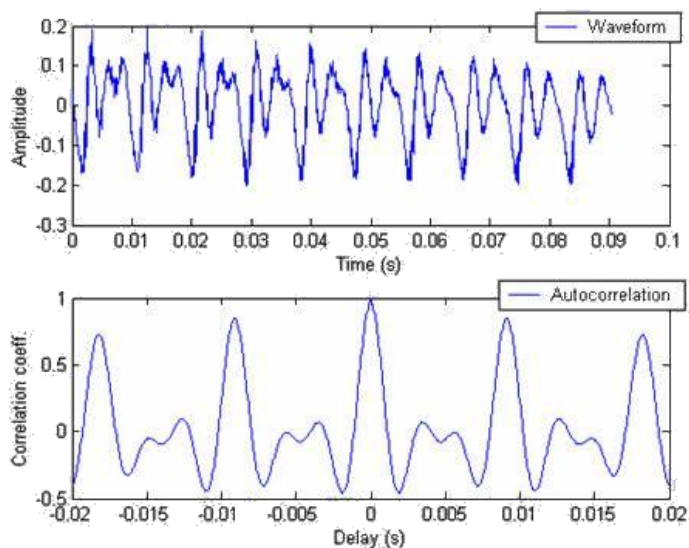
อัตสหสัมพันธ์ [16] เป็นวิธีการคำนวณความเป็นคาบของสัญญาณเพื่อที่จะนำไปใช้ในการคำนวณค่าความถี่มูลฐาน (Fundamental Frequency) โดยในการคำนวณค่าอัตสหสัมพันธ์ทำได้จากการทำสหสัมพันธ์แบบไขว้ (Cross-Correlation) ซึ่งเป็นการนำสัญญาณ 2 สัญญาณมาเปรียบเทียบกับเพื่อหาความคล้ายคลึงของสัญญาณทั้งสอง โดยสัญญาณหนึ่งจะมีการเอียงกับอีกสัญญาณหนึ่งโดยการเร่งหรือหน่วงสัญญาณไปในช่วงเวลาหนึ่ง สำหรับอัตสหสัมพันธ์นั้นสัญญาณที่นำมาใช้เปรียบเทียบกับนั้นเป็นสัญญาณเดียวกัน

ค่าอัตสหสัมพันธ์ $autocorr(m, k)$ ของสัญญาณเสียงที่กรอบเวลา (Frame) ที่ m เมื่อหน่วงเวลาไปเป็นระยะเวลา k สามารถคำนวณได้จากสมการที่ (2.1)

$$autocorr(m, k) = \frac{1}{N\sigma^2} \sum_{n=1}^N [s(n)w(m-n) - \mu][s(n+k)w(m-n+k) - \mu] \quad (2.1)$$

โดยที่ $s(n)$ คือสัญญาณเสียงตำแหน่งที่ n และ $w(m)$ คือฟังก์ชันหน้าต่าง (window function) ที่มีขนาดความกว้าง N ส่วน μ และ σ คือค่าเฉลี่ยและค่าความแปรปรวนของ $s(n)$ ตามลำดับ

ซึ่งจะพบว่าค่าอัตสหสัมพันธ์จะมีค่าสูงเมื่อมีการหน่วงเวลาหรือเอียงเวลาออกไปเท่ากับคาบของสัญญาณเสียงนั้น ดังแสดงในภาพที่ 2.2



ภาพที่ 2.2 สัญญาณเสียงที่มีลักษณะเป็นคาบและค่าอัตโนมัติที่ระยะเวลาต่างๆ [16]

2.2. การคำนวณค่าพลังงานของสัญญาณ

ค่าพลังงานของสัญญาณเสียง [17] เป็นหนึ่งในค่าที่นิยมนำมาใช้ในการวิเคราะห์สัญญาณเสียงซึ่งในการคำนวณค่าพลังงานของสัญญาณเสียงโดยทั่วไปสามารถคำนวณได้ตามสมการที่ 2.2

$$E = \sum s[n]^2 \quad (2.2)$$

โดยที่ E คือค่าพลังงานที่คำนวณได้ และ $s(n)$ คือสัญญาณเสียงตำแหน่งที่ n แต่เนื่องจากในการวิเคราะห์สัญญาณจริงเราวิเคราะห์สัญญาณในช่วงที่มีกรอบของสัญญาณเวลาสั้นๆ ดังนั้นในการคิดค่าพลังงานจึงต้องพิจารณาผลของกรอบของสัญญาณเวลาด้วย ดังนั้นจะสามารถคำนวณค่าพลังงานในช่วงกรอบสัญญาณเวลาได้ตามสมการที่ 2.3

$$E(m) = \sum_{n=0}^N w(m)s(m-n)]^2 \quad (2.3)$$

โดยที่ $E(m)$ คือค่าพลังงานที่คำนวณได้ของสัญญาณเสียงที่กรอบเวลาที่ m และ $s(n)$ คือสัญญาณเสียงตำแหน่งที่ n และ $w(m)$ คือฟังก์ชันหน้าต่างที่มีขนาดความกว้าง N

2.3. การแปลงฟูรีเยร์แบบช่วงเวลาสั้น

การแปลงฟูรีเยร์แบบช่วงเวลาสั้น [18] เป็นการแปลงฟูรีเยร์แบบหนึ่งที่ได้รับค่านิยมในการวิเคราะห์สเปกตรัมของสัญญาณเสียง ซึ่งสามารถแสดงได้ในรูปสมการที่ (2.4)

$$X_{STFT}(e^{j\omega}, n) = \sum x[n-m]w[m]e^{-j\omega m} \quad (2.4)$$

โดยที่ $x(n)$ คือ สัญญาณเสียงพูดที่ไม่ต่อเนื่อง (Discrete), $w(m)$ คือฟังก์ชันหน้าต่าง และ m, n คือตัวแปรทางเวลาที่มีค่าไม่ต่อเนื่อง ส่วนค่า ω คือค่าความถี่ที่มีความต่อเนื่อง

จากสมการที่ (2.5) เมื่อมีการสุ่มตัวอย่างสัญญาณ (Sampling) $X_{STFT}(e^{j\omega}, n)$ ที่มีความถี่เท่ากันที่ n จะได้สมการใหม่ดังต่อไปนี้

$$X_{STFT}(k, n) = \sum_{m=0}^{R-1} x[n-m]w[m]e^{-j\frac{2\pi}{N}kn} \quad (2.5)$$

โดยที่ ค่าของ k อยู่ระหว่าง 0 ถึง $n-1$ ในกรณีนี้จุดที่ใช้ในการแปลงฟูรีเยร์เป็นสัญญาณไม่ครบคาบ และขนาดของ “ฟังก์ชันหน้าต่าง” มีค่ามากกว่า R ($m \geq R$)

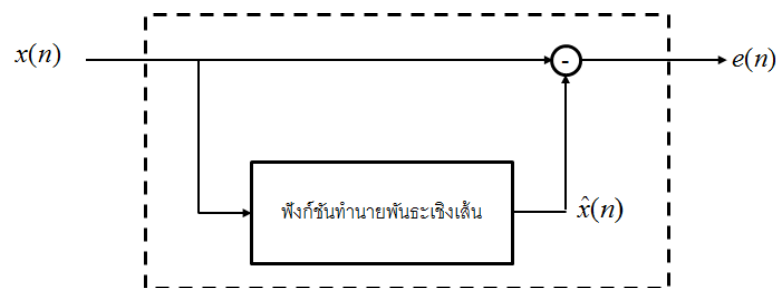
การแสดงค่าสเปกตรัมของการแปลงฟูรีเยร์ดังกล่าวจะเรียกว่า สเปกโตรแกรม (spectrogram) ซึ่งแบ่งออกได้เป็น 2 ประเภทใหญ่ ๆ ได้แก่ แถบกว้าง (Wideband) ซึ่งจะเห็นการเปลี่ยนแปลงเวลาได้ดี ในขณะที่แถบแคบ (Narrowband) จะเห็นการเปลี่ยนแปลงความถี่ได้ดี

2.4. ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

การคำนวณค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น [19] สามารถคำนวณได้โดยการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ทำให้เกิดความผิดพลาดในการประมาณน้อยที่สุด เนื่องจากการในการประมาณเชิงเส้นจะประมาณค่าของสัญญาณตำแหน่งปัจจุบันตามสมการที่

$$\hat{x}_i(n) \approx \sum_{k=1}^p a_k x_i(n-k) \quad (2.6)$$

โดยที่ $x_i(n)$ คือ ค่าของสัญญาณตำแหน่งที่ n , $\hat{x}_i(n)$ คือ ค่าของสัญญาณเสียงที่ได้จากการประมาณพัลส์เชิงเส้นตำแหน่งที่ n a_k คือค่าสัมประสิทธิ์การประมาณพัลส์เชิงเส้นลำดับที่ k จากทั้งหมด P อันดับและในการคำนวณหาค่าสัมประสิทธิ์การประมาณพัลส์เชิงเส้นที่เหมาะสมก็จะคำนวณโดยให้เกิดความผิดพลาดในการหาตำแหน่งปัจจุบันน้อยที่สุด โดยที่ในการคำนวณค่าสัมประสิทธิ์การประมาณพัลส์เชิงเส้นสามารถคำนวณความผิดพลาดได้ดังภาพที่ 2.3

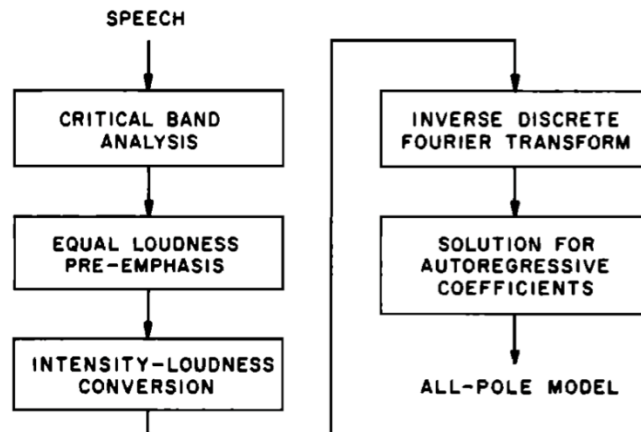


ภาพที่ 2.3 การหาความผิดพลาดจากการประมาณพัลส์เชิงเส้น [20]

โดยที่ $x(n)$ คือ สัญญาณจริง, $\hat{x}(n)$ คือ ค่าของสัญญาณเสียงที่ได้จากการประมาณพัลส์เชิงเส้น และ $e(n)$ คือ คลาดเคลื่อนจากสัญญาณเสียงที่ประมาณกับสัญญาณเสียงจริง

2.5. ค่าการประมาณพัลส์แบบอิงการรับฟังของมนุษย์

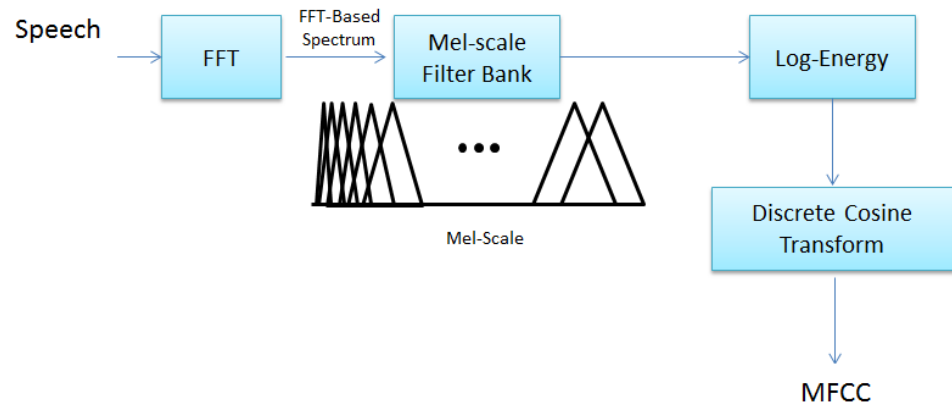
ในการคำนวณค่าการประมาณพัลส์แบบอิงการรับฟังของมนุษย์ [21] จะทำโดยการนำสัญญาณเสียงมาเข้าสู่ชุดกรองความถี่ที่เหมาะสมต่อการได้ยิน และในกระบวนการถัดไปจะเป็นการทำพีริเอมิฟาสซิสของความดัง และใช้กฎในการกำหนดความเข้มของเสียงเพื่อให้เข้ากับการรับรู้เสียงของมนุษย์ก่อนที่จะทำการแปลงอินเวอร์สฟูริเยร์และคำนวณการประมาณพัลส์แบบอิงการรับฟังของมนุษย์ ดังในภาพที่ 2.4



ภาพที่ 2.4 ขั้นตอนในการหาค่าการประมาณพารามิเตอร์แบบอิงการรับฟังของมนุษย์[21]

2.6. สัมประสิทธิ์เซปสตรีมในย่านความถี่เมล

ในการคำนวณค่าสัมประสิทธิ์เซปสตรีมในย่านความถี่เมล [18] มีวิธีการคำนวณ โดยเริ่มจากการนำสัญญาณมาทำการแปลงฟูริเยร์ และใช้ชุดตัวกรองความถี่สเกลเมลในการกรองความถี่หลังจากกระบวนการดังกล่าวจะทำการคำนวณค่าลอกริทึมของพลังงาน และทำการแปลงดิสครีตโคไซน์ก็จะได้ค่าสัมประสิทธิ์เซปสตรีมในย่านความถี่เมล ดังในภาพที่ 2.5



ภาพที่ 2.5 ขั้นตอนในการหาค่าสัมประสิทธิ์เซปสตรีมในย่านความถี่เมล [18]

3. ทฤษฎีที่เกี่ยวข้องกับการวัดผล

ทฤษฎีที่เกี่ยวข้องกับการวัดผลจะประกอบด้วย 2 ส่วนได้แก่ การทดสอบของแมคเนียร์ (McNemar's test) และค่าความถูกต้อง (Correctness)

3.1. การทดสอบของแมคเนียร์

การทดสอบของแมคเนียร์ [22] เป็นการทดสอบเพื่อวัดความแตกต่างของระบบก่อนการปรับปรุง และหลังการปรับปรุง เพื่อพิจารณาว่าระบบทั้งสองมีความแตกต่างกันทางนัยสำคัญทางสถิติหรือไม่ โดยในการทดสอบของแมคเนียร์จะใช้วิธีคำนวณต่อไปนี้ สมมุติให้ความถูกต้องของระบบก่อนการปรับปรุง และระบบหลังการปรับปรุง เป็นดังตารางที่ 2.5

ตารางที่ 2.5 จำนวนความถูกต้องและความผิดพลาดของระบบก่อนการปรับปรุงและหลังการปรับปรุง

		ระบบหลังการปรับปรุง	
		ถูกต้อง	ผิดพลาด
ระบบก่อนการปรับปรุง	คำตอบ		
	ถูกต้อง	n_{00}	n_{01}
ปรับปรุง	ผิดพลาด	n_{10}	n_{11}

ในการทดสอบด้วยการทดสอบของแมคเนียร์จะใช้ค่า P ในการทดสอบและพิจารณาค่า P นั้นหาค่า P น้อยกว่าค่าระดับนัยสำคัญ (Significance Level: α) ที่ตั้งไว้ แสดงว่าระบบทั้งสองมีความแตกต่างกันทางนัยสำคัญทางสถิติที่ระดับนัยสำคัญ α ซึ่งในการคำนวณค่า P สามารถคำนวณได้จากสมการ 2.7

$$P = \begin{cases} 2 \sum_{m=N_{10}}^k \binom{k}{m} \left(\frac{1}{2}\right)^k ; n_{10} > k/2 \\ 2 \sum_{m=0}^{N_{10}} \binom{k}{m} \left(\frac{1}{2}\right)^k ; n_{10} < k/2 \end{cases} \quad (2.7)$$

โดยค่า k จะหาได้จากจำนวนความถูกต้องเมื่อระบบก่อนปรับปรุงตอบถูก แต่ระบบหลังปรับปรุงตอบผิดรวมกับจำนวนความถูกต้องเมื่อระบบก่อนปรับปรุงตอบผิด แต่ระบบหลังปรับปรุงตอบถูก ($n_{10} + n_{01}$)

3.2. ค่าความถูกต้อง

ในการวัดผลการทดลองจะใช้วิธีในการวัดค่าความถูกต้องคำนวณได้จากการสมการ 2.8

$$Correct(\%) = \frac{W}{N} \times 100\% \quad (2.8)$$

โดยที่ *Correct* คือ เปอร์เซนต์ความถูกต้องของระบบ, W คือ จำนวนคำที่ระบุถูกต้อง และ N คือจำนวนคำทั้งหมด

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องจะประกอบด้วย งานวิจัยที่เกี่ยวข้องกับการจำแนกเสียงวรรณยุกต์ในภาษาที่มีเสียงวรรณยุกต์ และงานวิจัยที่เกี่ยวข้องกับแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์

งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาที่มีเสียงวรรณยุกต์

งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาที่มีเสียงวรรณยุกต์จะแบ่งเป็น 2 ส่วนได้แก่ งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาไทย และงานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาอื่นที่มีเสียงวรรณยุกต์

1. งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาไทย

สำหรับงานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาไทยจะมีงานวิจัย [5, 6, 7, 8, 23] โดยเริ่มจากในปี 1998 Tungthangthum [8] ได้เสนอการรู้จำเสียงวรรณยุกต์ในภาษาไทย โดยการนำรูปร่างของความถี่มูลฐาน (Fundamental Frequency: F0) ซึ่งสกัดได้โดยการใช้อัตสหสัมพันธ์ เพื่อนำไปเป็นข้อมูลขาเข้าสำหรับแบบจำลองฮิดเดนมาร์คอฟ จากนั้นจะทำการรู้จำลักษณะรูปร่าง (Contour) ของความถี่มูลฐาน เพื่อที่จะบอกว่าสัญญาณเสียงที่เข้ามานั้นมีเสียงวรรณยุกต์อะไร ซึ่งจากการทดลอง Tungthangthum [8] พบว่าเสียงวรรณยุกต์กับเสียงสระเป็นอิสระต่อกัน กล่าวคือ ข้อมูลของเสียงวรรณยุกต์ที่สกัดได้จากสระตัวหนึ่งสามารถนำไปบอกเสียงวรรณยุกต์ของสระตัวอื่นได้

จากนั้นในปี 2002 Thubthong [5] ได้เสนอการรู้จำเสียงวรรณยุกต์คำโดด โดยใช้การรวมโครงข่ายประสาทเทียม ในขั้นแรกเขาได้เสนอชุดของลักษณะเด่นที่ใช้ค่าความถี่มูลฐาน และผลต่างของความถี่มูลฐานมา 3 ชุด และใช้โครงข่ายประสาทเทียม 1 โครงข่ายในการจำแนกเสียงวรรณยุกต์ โดยก่อนการจำแนกเสียงวรรณยุกต์ ความถี่มูลฐานที่นำมาใช้เป็นลักษณะเด่นในแต่ละชุด จะถูกทำให้เป็นบรรทัดฐานเดียวกันด้วยค่าเฉลี่ยของความถี่มูลฐาน จากนั้นในการทดลอง ใช้ชุดลักษณะเด่นทั้งสามชุดเพื่อที่จะจำแนกเสียงวรรณยุกต์ในฐานข้อมูลที่เขาออกแบบมาพบว่า ชุดลักษณะเด่นที่ประกอบด้วยความถี่มูลฐาน 2 ตำแหน่ง ได้แก่ตำแหน่งเริ่มต้นและตำแหน่งสุดท้าย และผลต่างของความถี่มูลฐาน 4 ตำแหน่งได้แก่ตำแหน่งที่ 20%, 40%, 60% และ 80% ของระยะเวลา จะให้ผลในการจำแนกเสียงวรรณยุกต์เฉลี่ยสูงที่สุดเป็น 97.12% ในกรณีที่ฐานข้อมูลเรียนรู้และทดสอบเป็นฐานข้อมูลเดียวกัน และได้ผลเฉลี่ยเป็น 94.29% เมื่อกรณีที่ฐานข้อมูลเรียนรู้และทดสอบเป็นฐานข้อมูลคนละชุดกัน และในการทดลองถัดมาได้ทำการสร้างโครงข่ายประสาทเทียม 3 โครงข่าย โดยแต่ละโครงข่ายจะเลือกใช้ลักษณะเด่นคนละอันจากลักษณะเด่นที่เขาเสนอ ซึ่งจะได้โครงข่ายประสาทเทียม 3 ตัวจากนั้นจะทำการจำแนกเสียงวรรณยุกต์ โดยใช้การคำนวณความเป็นไปได้จากทั้ง 3 โครงข่าย จากการทดลองเขาพบว่าการรวมกันของโครงข่ายประสาทเทียมที่ดีที่สุดคือ การโหวตโดยเชื่อผลส่วนมากซึ่งจะได้ผลเป็น 96.34% ในชุดเรียนรู้เป็นชุดเดียวกัน และได้ผล 95.76% เมื่อชุดทดสอบและชุดเรียนรู้เป็นคนละชุดกัน

ในปีเดียวกัน Thubthong [6] ได้เสนอการศึกษาการสร้างแบบจำลองเสียงวรรณยุกต์ภาษาไทย โดยในการทดลองเขาได้ศึกษาชุดลักษณะเด่น 3 ชุด เพื่อหาลักษณะเด่นที่เหมาะสมกับการจำแนกเสียงวรรณยุกต์ภาษาไทย ซึ่งลักษณะเด่นทั้ง 3 ชุด ได้แก่ ชุดที่หนึ่งค่าสัมประสิทธิ์สมการถดถอยดีกรี 4 ชุดที่สอง คือ ความถี่มูลฐาน 2 ตำแหน่งได้แก่ตำแหน่งเริ่มต้นและตำแหน่งสุดท้าย และผลต่างของความถี่มูลฐาน 5 ตำแหน่งได้แก่ตำแหน่งที่ 0%, 25%, 50%, 75% และ 100% ของระยะเวลา ซึ่งจะเห็นได้ว่าในชุดนี้จะคล้ายชุดลักษณะเด่นที่ดีที่สุดที่เขาเสนอไปก่อนหน้านี้ [5] แต่ตำแหน่งของช่วงที่นำเสนอแตกต่างกัน และชุดที่สามคือ ความถี่มูลฐาน และผลต่างของความถี่มูลฐาน 5 ตำแหน่งได้แก่ตำแหน่งที่ 0%, 25%, 50%, 75% และ 100% ของระยะเวลา รวมเป็น 10 ค่า อีกทั้งเขายังศึกษาโดยการปรับความถี่มูลฐานไปเป็นความถี่สเกลต่างๆ อันได้แก่ ความถี่สเกลกึ่งวรรณยุกต์ (Semi-Tone) และ ความถี่สเกลความถี่อัตราอาร์บี (Equivalent Rectangular Bandwidth rate: ERB-rate) นอกจากนี้ในการทดลองของเขายังมีการทดลองการปรับค่าวิธีการทำให้เป็นบรรทัดฐานจากเดิมที่ทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของความถี่มูลฐานสำหรับการทดลองนี้เขาได้เพิ่มการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานเข้าไปด้วย ซึ่งจากการทดลองของเขาพบว่าสเกลความถี่อัตราอาร์บีให้ผลดีที่สุด ในขณะที่การทำให้เป็นบรรทัดฐานที่ดีที่สุดคือ การทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน

ต่อมาในปี 2004 Tan [7] ได้ทำการทดลองเช่นเดียวกับ Thubthong [6] แต่ทดลองบนฐานข้อมูลเสียงที่ต่างกัน ซึ่งจากการทดลองของ Tan [7] พบว่า การปรับความถี่มูลฐานไปเป็นความถี่สเกลกึ่งวรรณยุกต์ และใช้การทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของความถี่มูลฐานกลับให้ผลในการจำแนกเสียงวรรณยุกต์มีประสิทธิภาพสูงที่สุดเป็น 72.21% ซึ่งขัดแย้งกับ Thubthong [6] นอกจากนี้จากการทดลองของ Tan [7] เมื่อเขาสร้างโครงข่ายประสาทเทียมเพิ่มเติมโดยการแบ่งข้อมูลตามเสียงพยัญชนะตัวสะกดและสร้างตัวจำแนกเสียงวรรณยุกต์ด้วยข้อมูลจากตัวสะกดที่เขาแยก และในการทดสอบก็นำเสียงพูดไปจำแนกเสียงวรรณยุกต์ตามตัวสะกดของเสียงนั้น พบว่าประสิทธิภาพในการจำแนกเสียงวรรณยุกต์มีประสิทธิภาพที่สูงขึ้นเป็น 77.13% เขาจึงสรุปว่าเสียงตัวสะกดมีส่วนในการช่วยจำแนกเสียงวรรณยุกต์เช่นเดียวกับการศึกษาของ [24]

สุดท้ายในปี 2009 Maleerat [23] ได้เสนอการจำแนกเสียงวรรณยุกต์คำโดด โดยพัฒนาต่อจาก Thubthong [5] ซึ่งเขาได้ใช้ชุดลักษณะเด่นทั้ง 3 ที่ Thubthong [5] เสนอ แต่เขาได้รวมลักษณะเด่นทั้ง 3 เป็นลักษณะเด่นชุดเดียวกันแทนที่จะใช้แต่ละชุด ซึ่งจากการทดลองของเขาพบว่าได้ความถูกต้องในการจำแนกเสียงวรรณยุกต์คำโดดเป็น 91.4% แต่อย่างไรก็ตามเมื่อพิจารณาลักษณะเด่นที่เขาพัฒนาต่อจากที่ Thubthong [5] เสนอในงานวิจัยของเขา จะพบว่าไปคล้ายกับลักษณะเด่นที่ Thubthong [6] เสนอในอีกงานวิจัยหนึ่ง แต่ต่างกันที่ตำแหน่งของระยะเวลาที่เลือกมาใช้เท่านั้น

2. งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาอื่นที่มีเสียงวรรณยุกต์

งานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในภาษาอื่นที่มีเสียงวรรณยุกต์ประกอบด้วยงานวิจัย [25, 26, 27, 28, 29, 30] เริ่มจากในปี 1995 Lee [25] ได้เสนอการรู้จำเสียงวรรณยุกต์คำโดดในภาษาจีนสำเนียงกวางตุ้งโดยเลือกใช้ลักษณะเด่นจากระดับความสัมพันธ์ของระดับเสียง (Relative Pitch Level) การเปลี่ยนแปลงของระดับเสียง (Temporal Pitch Variation) และช่วงเวลา (Duration) ดังนั้นในการทดลองของเขาจึงประกอบด้วยค่าความถี่มูลฐานในตำแหน่งเริ่มต้นและตำแหน่งจบซึ่งสอดคล้องกับระดับความสัมพันธ์ของระดับเสียง และ ใช้การเปลี่ยนแปลงภาพรวมของระดับเสียงโดยการคำนวณระดับเสียงตลอดทั้งเสียงพูดเพื่อใช้ในการแทนการเปลี่ยนแปลงของระดับเสียง และในส่วนของช่วงเวลาจะใช้ช่วงเวลาของการเกิดเสียงวรรณยุกต์ และอัตราการลดลงของพลังงาน (Energy Drop Rate) เนื่องจากช่วงเวลาและพลังงานมีความสัมพันธ์กัน [25] จากนั้นเขาได้ใช้การทำให้เป็นบรรทัดฐานเดียวกันก่อนการรู้จำเสียงวรรณยุกต์เพื่อลดความแตกต่างของช่วงเวลาและบุคคล จากนั้นในการทดลองได้ใช้โครงข่ายประสาทเทียมในการจำแนกเสียงวรรณยุกต์ โดยจากการทดลองพบว่าจากการจำแนกเสียงวรรณยุกต์ของคนคนเดียวได้ความถูกต้องที่ 89% และในการจำแนกเสียงวรรณยุกต์ที่ไม่ขึ้นกับบุคคลได้ความถูกต้อง 87.6%

จากนั้นในปี 1998 Jian [26] เสนอการจำแนกเสียงวรรณยุกต์ของภาษาไต้หวันโดยใช้ระดับความสูงต่ำของเสียงและการเคลื่อนที่ของพลังงาน เนื่องจากเมื่อเขาทำการวิเคราะห์ลักษณะ

ของความถี่มูลฐานของเสียงวรรณยุกต์ในภาษาไต้หวันแล้วพบว่าถึงแม้จะมีความเป็นไปได้ในการแยกแยะเสียงวรรณยุกต์ออกจากกันด้วยค่าความถี่มูลฐานเพียงอย่างเดียว แต่ถ้าพิจารณาบางคู่เสียงวรรณยุกต์ โดยเขายกตัวอย่างเสียงตกสั้น (Short Falling) กับเสียงตกยาว (Long Falling) พบว่าทั้งสองมีความคล้ายกันมากดังนั้นเขาจึงเสนอการจำแนกคู่เสียงเหล่านั้นโดยใช้การเคลื่อนที่ของพลังงาน โดยเขาเลือกการวิเคราะห์ค่าพลังงานด้วยการใช้แบบจำลอง ADSR (Attack, Decay Sustain and Release) ในการวิเคราะห์เมื่อนำการเคลื่อนที่ของพลังงานในช่วงเริ่ม Decay มาใช้พบว่าสามารถนำมาช่วยในการวิเคราะห์และแยกแยะเสียงวรรณยุกต์ในภาษาไต้หวันได้

ในปี 2004 Tian [27] เสนอการรู้จำเสียงวรรณยุกต์ภาษาจีนโดยใช้แบบจำลองเกาส์เซียนมิกเจอร์ และแบบจำลองฮิดเดนมาร์คอฟ โดยใช้ลักษณะเด่นที่แตกต่างกัน สำหรับแบบจำลองฮิดเดนมาร์คอฟเขาเลือกใช้ลักษณะเด่นเป็นค่าความถี่มูลฐานและผลต่างของความถี่มูลฐาน แต่ในขณะที่แบบจำลองเกาส์เซียนมิกเจอร์เลือกใช้ลักษณะเด่น 2 ประเภทคือ ค่าสัมประสิทธิ์ของสมการถดถอยของทั้งความถี่มูลฐาน กับ ค่าสัมประสิทธิ์ของสมการถดถอยของทั้งความถี่มูลฐานซึ่งมีการแบ่งส่วนเป็นส่วนๆ 4 ส่วน จากการทดลองแรกของเขาพบว่าค่าสัมประสิทธิ์ของสมการถดถอยของทั้งความถี่มูลฐานซึ่งมีการแบ่งส่วนเป็นส่วนๆ 4 ส่วนที่ใช้แบบจำลองเกาส์เซียนมิกเจอร์ ให้ผลดีที่สุด คือ 85.21% สำหรับการจำแนกเสียงวรรณยุกต์ที่ขึ้นกับผู้พูด และ 65.91% สำหรับการจำแนกเสียงวรรณยุกต์ที่ไม่ขึ้นกับผู้พูด ดังนั้นเขาจึงสรุปในการทดลองแรกว่ารูปร่างที่แน่นอนของเสียงวรรณยุกต์ภาษาจีน ไม่มีความจำเป็นต่อการรู้จำเสียงวรรณยุกต์ นอกจากนี้จากการทดลองยังพบว่าได้ผลดีกว่าแบบจำลองฮิดเดนมาร์คอฟที่ผลการรู้จำเพียง 77.18% สำหรับการจำแนกเสียงวรรณยุกต์ที่ขึ้นกับผู้พูด และ 62.50% สำหรับการจำแนกเสียงวรรณยุกต์ที่ไม่ขึ้นกับผู้พูด และในการทดลองที่สองเขาได้พัฒนาระบบที่ใช้แบบจำลองฮิดเดนมาร์คอฟ โดยเติมลักษณะเด่นค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล เข้าไปเพื่อใช้ในการสร้างแบบจำลองเสียงวรรณยุกต์จากเดิมที่มีเพียง 5 แบบจำลอง เป็น 12,824 แบบจำลอง พบว่าได้ความถูกต้องเพิ่มเป็น 86.03% สำหรับการจำแนกเสียงวรรณยุกต์ที่ขึ้นกับผู้พูด และ 81.01% สำหรับการจำแนกเสียงวรรณยุกต์ที่ไม่ขึ้นกับผู้พูด และเมื่อพิจารณากับการทดลองก่อนหน้าของเขาพบว่าได้ผลที่ดีกว่าในการทดลองแรก ดังนั้นเขาจึงสรุปว่าการแมพระหว่างหนึ่งเสียงวรรณยุกต์ต่อหนึ่งแบบจำลองไม่เพียงพอต่อการ

รู้จำเสียงวรรณยุกต์ ดังนั้นการเพิ่มจำนวนแบบจำลองของแต่ละเสียงวรรณยุกต์จะช่วยเพิ่มประสิทธิภาพในการรู้จำเสียงวรรณยุกต์เป็นอย่างมาก

ต่อจากนั้นในปี 2008 Quang [28] ได้เสนอการรู้จำเสียงวรรณยุกต์ภาษาเวียดนามโดยใช้แบบจำลองฮิดเดนมาร์คอฟ ซึ่งใช้ลักษณะเด่นได้แก่ ความถี่มูลฐานของสัญญาณเสียง ผลต่างของความถี่มูลฐานของสัญญาณเสียง และความเร่งของความถี่มูลฐานของสัญญาณเสียง รวมถึงค่าพลังงานในช่วงสั้นอันได้แก่ ค่าพลังงาน ผลต่างของพลังงานและความเร่งของพลังงาน ในการจำแนกเสียงวรรณยุกต์ อีกทั้งเขายังพิจารณาการทำให้เป็นบรรทัดฐานด้วยวิธีต่างๆ ซึ่งจากการทดลองเขาพบว่าลอการิทึมของความถี่มูลฐานที่ทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานให้ความแม่นยำในการรู้จำเสียงวรรณยุกต์สูงสุดโดยได้ให้ความแม่นยำที่ 70.44 % นอกจากนี้เมื่อเขาแบ่งข้อมูลเรียนรู้และทดสอบตามแต่ละเพศพบว่าระบบของเขาให้ความถูกต้องเพิ่มขึ้นเป็น 72.83%

ในปีเดียวกัน Quang [29] เสนอการนำข้อมูลของเสียงวรรณยุกต์มาใช้ในการพัฒนาการรู้จำเสียงพูดของภาษาเวียดนาม โดยที่ในการสร้างแบบจำลองของเสียงวรรณยุกต์ภาษาเวียดนาม เขาได้เลือกใช้แบบจำลองฮิดเดนมาร์คอฟในการสร้างแบบจำลองเสียงวรรณยุกต์โดยลักษณะเด่นที่เขานำมาใช้ในการสร้างแบบจำลองเสียงวรรณยุกต์ได้แก่ ความถี่มูลฐานของสัญญาณเสียง ผลต่างของความถี่มูลฐานของสัญญาณเสียง และความเร่งของความถี่มูลฐานของสัญญาณเสียง รวมถึงค่าพลังงาน ผลต่างของพลังงานและความเร่งของพลังงาน จากนั้นเขาทำการรวมแบบจำลองเสียงวรรณยุกต์ กับระบบที่ใช้ในการรู้จำเสียงพูดซึ่งประกอบด้วยแบบจำลองทางเสียงและแบบจำลองทางภาษา โดยในการรวมระบบเขาได้มีการพิจารณาคำตอบของระบบที่นำเสนอจากแบบจำลองทางเสียง แบบจำลองทางภาษา และแบบจำลองเสียงวรรณยุกต์ โดยมีการให้นำหนักของแบบจำลองในค่าต่างๆ จากการทดลองของเขาสามารถลดความผิดพลาดจากระบบลงเมื่อยังไม่มีการนำแบบจำลองทางเสียงวรรณยุกต์มาประยุกต์ใช้ได้ถึง 28.6%

ในปี 2008 Hongxiu [30] ได้นำแบบจำลองคอนดิชันนอลแรนดอมฟิลด์มาสร้างแบบจำลองวรรณยุกต์ของภาษาจีนโดยข้อมูลนำเข้าเป็นข้อมูลเกี่ยวกับคำศัพท์และฉันทลักษณ์ (Lexical and Prosodic) ของภาษาจีน จากนั้นนำแบบจำลองวรรณยุกต์ดังกล่าวไปช่วยในการ

ตรวจสอบความเป็นไปได้ในการรู้จำเสียงพูด โดยใช้วิธีการคิดคะแนนใหม่ ซึ่งจะมีการคำนึงถึงการเกิดเสียงวรรณยุกต์เข้าไปด้วยในการหาคำตอบในการรู้จำเสียงที่ดีที่สุด โดยเมื่อระบบรู้จำเสียงพูดได้คำตอบ N คำตอบที่ดีที่สุดออกมาก็ให้นำคะแนนของแต่ละคำที่ได้ไปหาคะแนนใหม่โดยเพิ่มคะแนนของแบบจำลองวรรณยุกต์เข้าไป ซึ่งทำให้ผลความถูกต้องในการรู้จำเสียงพูดมากยิ่งขึ้นถึง 7.6%

งานวิจัยที่เกี่ยวข้องกับแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์

งานวิจัยที่เกี่ยวข้องกับแบบจำลองคอนดิชันนัลแรนดอมฟิลด์ [31] เริ่มมาจาก ในปี 2001 Lafferty [31] เป็นผู้ทำการเสนอ แบบจำลองคอนดิชันนัลแรนดอมฟิลด์สำหรับการแก้ปัญหาเซกเมนต์และการระบุประเภทของสายข้อมูล (Segment and Label Sequential Data) ซึ่งแบบจำลองดังกล่าวเป็นแบบจำลองดิสคริมิเนทีฟ แบบจำลองประเภทนี้สามารถแก้ปัญหาที่เป็นข้อดีของแบบจำลองฮิดเดนมาร์คอฟที่เป็นแบบเจเนเนอเรทีฟได้กล่าวคือ แบบจำลองฮิดเดนมาร์คอฟได้มีการตั้งสมมุติฐานความไม่ขึ้นต่อกัน แต่ในความเป็นจริงข้อมูลเสียงพูดจากแต่ละช่วงเวลามีความสัมพันธ์กันอย่างมากทำให้สมมุติฐานดังกล่าวไม่ถูกต้อง นอกจากนี้แบบจำลองคอนดิชันนัลแรนดอมฟิลด์ยังมีข้อได้เปรียบเมื่อเทียบกับแบบจำลองแบบดิสคริมิเนทีฟอื่นๆ ซึ่งประสบกับปัญหาป้ายที่ลำเอียง (Label Bias Problem) คือ การระบุประเภทข้อมูลที่ไม่ยุติธรรมในกรณีที่สถานะใดสถานะหนึ่งมีเส้นทางในไปยังอีกสถานะหนึ่งที่มีค่าน้อยกว่าการค้นหาด้วยวิเทอร์บี (Viterbi Search) ก็จะทำให้การเลือกสถานะที่มีเส้นทางที่น้อยกว่าเพราะการกระจายตัวของความน่าจะเป็นภายในมากกว่า สถานะที่มีเส้นทางมาก ซึ่งกรณีดังกล่าวเกิดจากการทำให้เป็นบรรทัดฐาน (Normalize) ค่าความน่าจะเป็นด้วยค่าบรรทัดฐานในกลุ่ม (Local Normalize) แต่สำหรับแบบจำลองคอนดิชันนัลแรนดอมฟิลด์ จะใช้ค่าบรรทัดฐานครอบคลุม (Global Normalize) ซึ่งทำให้ปัญหาดังกล่าวไม่เกิดขึ้น ซึ่งจากการทดลอง Lafferty พบว่าเมื่อข้อมูลมีความสัมพันธ์ที่ขึ้นต่อกันมากยิ่งขึ้นผลการรู้จำของแบบจำลองคอนดิชันนัลแรนดอมฟิลด์ประสบความสำเร็จมากกว่า

Lafferty [31] ได้เสนอสมการของแบบจำลองคอนดิชันนัลแรนดอมฟิลด์ไว้ดังสมการที่ (2.9)

$$p_{\theta}(y|x) \propto \exp\left(\sum_{e \in E, k} \lambda_k f_k(e, y|_e, x) + \sum_{v \in V, k} \mu_k g_k(v, y|_v, x)\right) \quad (2.9)$$

โดยที่	y	คือ สายข้อมูลที่ต้องการระบุประเภท
	x	คือ สายข้อมูลขาเข้า
	k	คือ หมายเลขของพีเจอร์ฟังก์ชัน
	λ, μ	คือ พารามิเตอร์ของแบบจำลอง
	e	คือ เส้นทางการเปลี่ยนสถานะหนึ่งไปอีกสถานะหนึ่ง ซึ่งเป็นสมาชิกของเส้นทางทั้งหมด (E) ของแบบจำลอง
	v	คือ การเกิดสถานะนั้นๆ จาก (V)

ในเวลาต่อมา Gunawardana [10] ได้เสนอแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์สำหรับการจำแนกหน่วยเสียง (Phone Classification) ซึ่งแบบจำลองดังกล่าวได้ทำการเพิ่มสถานะที่ซ่อนเร้น (Hidden State) เข้าไปในแบบจำลองคอนดิชันนัลแรนดอมฟิลด์ เพื่อจัดการกับตัวแปรที่ซ่อนเร้น (Hidden Variable) เนื่องจากการเปลี่ยนแปลงพลังงานและสเปกตรัมของสัญญาณเสียงพูดมีการเปลี่ยนแปลงตลอดเวลาในหนึ่งหน่วยเสียง อีกทั้งแบบจำลองดังกล่าวยังสามารถเพิ่มความซับซ้อนของพีเจอร์ฟังก์ชัน (Feature Function) ได้ แต่ในงานวิจัยดังกล่าวไม่ได้นำข้อได้เปรียบนี้มาใช้ แต่จำกัดพีเจอร์ฟังก์ชันให้มีทอพอโลยี (Topology) เช่นเดียวกับแบบจำลองฮิดเดนมาร์คอฟ และลักษณะเด่นที่ใช้ในการทดลอง เขาได้ทำการเลือกเพียงแค่ค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมลและอนุพันธ์ของอันดับที่ 1 และ 2 ของค่าดังกล่าวในแต่ละกรอบสัญญาณ (Frame) เพื่อเปรียบเทียบประสิทธิภาพกับแบบจำลองฮิดเดนมาร์คอฟ ซึ่งผลในการจำแนกหน่วยเสียงที่ได้ พบว่าความผิดพลาดที่น้อยที่สุดในการจำแนกหน่วยเสียงที่ได้จากแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ คือ 21.7% ส่วนความผิดพลาดที่น้อยที่สุดในการจำแนกหน่วยเสียงที่ได้จากแบบจำลองฮิดเดนมาร์คอฟ คือ 24.6%

Gunawardana [10] ได้เสนอสมการของแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ไว้ดังสมการที่ (2.10)

$$p(w|o;\lambda) = \frac{1}{z(o;\lambda)} \sum_{s \in W} \exp\{\lambda \cdot f(w,s,o)\} \quad (2.10)$$

โดย	w	คือ สายข้อมูลที่ต้องการระบุประเภท
	o	คือ สายข้อมูลขาเข้า
	λ	คือ พารามิเตอร์ของแบบจำลอง
	$z(o;\lambda)$	คือ ฟังก์ชันครอบคลุมที่ทำให้เป็นบรรทัดฐานเดียวกัน
	$f(w,s,o)$	คือ ฟังก์ชัน
	s	คือ ตัวแปรที่ซ่อนอยู่ในการระบุประเภท (w)

การกำหนดฟังก์ชันในการสร้างแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ เป็นขั้นตอนสำคัญในการใช้งานแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ เนื่องจากฟังก์ชัน [31] คือ ฟังก์ชันที่อธิบายความสัมพันธ์ของสเปซของข้อมูลที่สังเกตได้ กับข้อมูลที่ต้องการระบุประเภท Gunawardana [10] เสนอฟังก์ชันสำหรับแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ โดยได้กำหนดทอพอโลยีของแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ให้เป็นแบบเดียวกับแบบจำลองฮิดเดนมาร์คอฟ ดังนั้นจึงได้ฟังก์ชันตามสมการที่ 2.11 – 2.15

หลังจาก Gunawardana [10] เสนอแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ในปี 2005 ต่อมาในปี 2009 Sung [11] ได้นำเสนองานต่อจาก Gunawardana [10] โดยการนำแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์มาจำหน่วยเสียง (Phone Recognition) ซึ่งงานที่ Gunawardana [10] เสนอเมื่อปี 2005 เป็นเพียงการจำแนกหน่วยเสียง โดยที่ความแตกต่างระหว่างการจำและการจำแนก คือ การจำแนกหน่วยเสียงจะรู้ขอบเขตของหน่วยเสียงที่ชัดเจน ซึ่งผิดกับการจำหน่วยเสียงจะไม่รู้ขอบเขตของหน่วยเสียง

$$f_w^{(LM)}(w, s, o) = \delta(w = w') \quad \forall w' \quad (2.11)$$

$$f_{ss'}^{(Tr)}(w, s, o) = \sum_{t=1}^T \delta(s_{t-1} = s) \delta(s_t = s') \quad \forall s, s' \quad (2.12)$$

$$f_s^{(Occ)}(w, s, o) = \sum_{t=1}^T \delta(s_t = s) \quad \forall s \quad (2.13)$$

$$f_s^{(M1)}(w, s, o) = \sum_{t=1}^T \delta(s_t = s) o_t \quad \forall s \quad (2.14)$$

$$f_s^{(M2)}(w, s, o) = \sum_{t=1}^T \delta(s_t = s) o_t^2 \quad \forall s \quad (2.15)$$

- โดยที่ w คือ สิ่งที่ต้องการระบุประเภท
 o คือ สายข้อมูลขาเข้า
 s คือ สายของสถานะที่ซ่อนเร้น ประกอบด้วย $\{s_1, s_2, s_3, \dots, s_t\}$
 T คือ ความยาวของสายข้อมูลขาเข้า
 $\delta(\cdot)$ คือ ฟังก์ชันการกระตุ้น (Activation Function)

Sung [11] ได้ทำการใช้ตัวแปรที่ซ่อนเร้น 2 ตัวในแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ เพื่อแก้ไขปัญหาขอบเขตของหน่วยเสียงและการเปลี่ยนแปลงของสัญญาณเสียง ตัวแปรที่ซ่อนเร้นตัวแรกสำหรับการพิจารณาหน่วยเสียงย่อย (Subphones) และตัวแปรที่ซ่อนเร้นอีกตัวหนึ่งไว้สำหรับพิจารณามิกซ์เจอร์คอมโพเนนต์ (Mixture Component) ซึ่งเปรียบได้กับ เกาส์เซียนมิกซ์เจอร์ (Gaussian Mixture) ของแบบจำลองฮิดเดนมาร์คอฟ แต่มีความแตกต่างกันที่ค่าการกระจายตัวจะไม่ถูกทำให้เป็นบรรทัดฐานเดียวกัน (Normalize) ซึ่งทำให้การสร้างแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ง่ายกว่าการสร้างแบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ที่เสนอโดย Sung [11] จากการทดลองพบว่า แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ มีความผิดพลาดในการรู้จำหน่วยเสียงพูดที่น้อยที่สุด เมื่อเทียบกับการรู้จำหน่วยเสียงพูดด้วยแบบจำลองฮิดเดนมาร์คอฟ โดยที่แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์มีความผิดพลาดในการรู้จำหน่วยเสียงเพียง 28.3%

บทที่ 3

การรับรู้เสียงวรรณยุกต์ภาษาไทย

ในบทนี้จะนำเสนอการศึกษาการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยมนุษย์ เนื่องจาก การศึกษาการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยมนุษย์เป็นปัจจัยสำคัญในการสร้างและพัฒนา ระบบการจำแนกเสียงวรรณยุกต์ภาษาไทยอัตโนมัติของคอมพิวเตอร์ เนื่องจากเป็นที่รู้กันดีว่า มนุษย์สามารถแยกแยะเสียงวรรณยุกต์ได้ดีกว่าคอมพิวเตอร์ ดังนั้นการศึกษาการรับรู้เสียง วรรณยุกต์ภาษาไทยโดยมนุษย์ จะสามารถเพิ่มพูนองค์ความรู้พื้นฐานทางด้านลักษณะของเสียงที่ มนุษย์ใช้สำหรับแยกแยะเสียงวรรณยุกต์ภาษาไทย เพื่อที่จะนำองค์ความรู้นั้นไปประยุกต์ใช้ในการ พัฒนาระบบการจำแนกเสียงวรรณยุกต์ภาษาไทยให้มีประสิทธิภาพมากยิ่งขึ้น

ในการศึกษาการรับรู้เสียงวรรณยุกต์นั้น ในงานวิจัยส่วนใหญ่ [32, 33, 34, 35] จะเน้น การศึกษาความสัมพันธ์ระหว่างการเปลี่ยนแปลงความถี่มูลฐาน (Fundamental Frequency) หรือ ระดับเสียง (Pitch) ของสัญญาณเสียงพูด กับ เสียงวรรณยุกต์ ซึ่งผลที่ตามมาคือมีการนำความถี่ มูลฐาน หรือระดับเสียงไปเป็นลักษณะเด่นในการสร้างแบบจำลองเพื่อใช้ในการจำแนกเสียง วรรณยุกต์ [24, 25, 36] แม้ว่าจะได้ผลลัพธ์ และความถูกต้องที่น่าพอใจในระดับหนึ่ง แต่ก็ยังไม่ สามารถเทียบเท่ากับมนุษย์ได้ โดยเฉพาะอย่างยิ่งในกรณีที่เสียงพูดต่อเนื่อง ดังนั้นการทดลอง การรับรู้เสียงวรรณยุกต์โดยลักษณะเด่นอื่นจึงเป็นสิ่งจำเป็น โดยในการทดลองนี้จะศึกษาลักษณะ รูปร่างของสเปกตรัมและพลังงานที่มีผลต่อการรับรู้เสียงวรรณยุกต์ภาษาไทยของผู้ที่ใช้ภาษาไทย เป็นภาษาหลัก (Thai Native Speaker)

ขั้นตอนในการศึกษาและทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยผู้ที่ใช้ภาษาไทย เป็นภาษาหลักประกอบไปด้วย การเตรียมสิ่งเร้า (Stimuli) การทดลอง ผลการทดลอง และสรุปผล การทดลอง

การเตรียมสิ่งเร้า

การเตรียมสิ่งเร้าสำหรับการทดลอง ประกอบด้วยสองส่วน ได้แก่ การเตรียมสิ่งเร้าต้นแบบ และการเตรียมสิ่งเร้าที่มีการดัดแปลงจากสิ่งเร้าต้นแบบ

1. การเตรียมสิ่งเร้าต้นแบบ

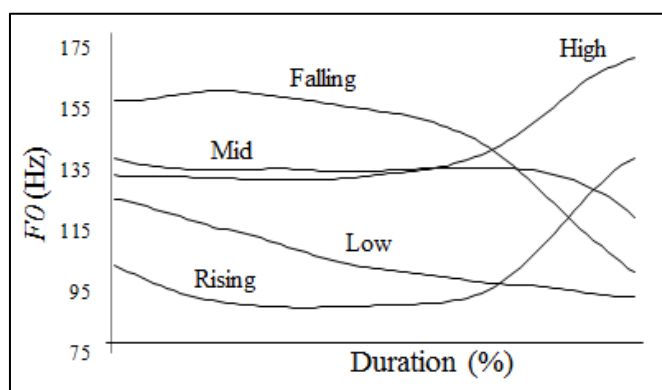
เสียงที่จะนำมาเป็นสิ่งเร้าต้นแบบเป็นเสียงที่อัดจากผู้พูดที่ใช้ภาษาไทยเป็นภาษาหลัก และไม่มีปัญหาในการออกเสียงภาษาไทย กล่าวคือ ผู้พูดสามารถออกเสียงภาษาไทยได้อย่างชัดเจนและถูกต้อง โดยการระบวนการบันทึกเสียงจะดำเนินการในห้องบันทึกเสียงที่ปราศจากเสียงรบกวน ไมโครโฟนที่ใช้ในการบันทึกเสียงเป็นไรต์เอ็นทีวันเอ ไมโครโฟนคอนเดนเซอร์ (Rode NT1-A Condenser Microphone) โดยไมโครโฟนจะวางห่างจากปากของผู้พูดประมาณ 15 เซนติเมตร สัญญาณเสียงจะถูกบันทึกที่ความถี่สุ่ม 44.1 กิโลเฮิร์ตซ์ และบันทึกในรูปแบบเอ็มเอสเวฟ (MS wave) โดยใช้ 16 บิต พัลส์เอม (Pulse Code Modulation; PCM) ในการเข้ารหัส

รูปแบบพยางค์ที่นำมาใช้ในการสร้างสิ่งเร้ามี 3 รูปแบบ ได้แก่ รูปแบบหนึ่งพยางค์ (Mono-Syllabic), รูปแบบสองพยางค์ (Bi-Syllabic) และรูปแบบสามพยางค์ (Tri-Syllabic) โดยแต่ละรูปแบบจะใช้คำว่า “ลา” (/l a:/) เป็นคำพื้นฐาน โดยในรูปแบบหนึ่งพยางค์ ผู้พูดจะต้องพูดคำว่า “ลา” (/l a:/) ให้ครบทั้ง 5 เสียงวรรณยุกต์ภาษาไทยดังแสดงในตาราง 3.1 ดังนั้นในรูปแบบหนึ่งพยางค์จะมีสิ่งเร้า 5 อัน สำหรับรูปแบบสองพยางค์ จะเป็นการพูดคำว่า “ลาลา” (/l a: l a:/) โดย “ลา” แต่ละตัวจะเป็นไปได้ 5 รูปแบบเสียงวรรณยุกต์ภาษาไทยเช่นเดียวกับรูปแบบหนึ่งพยางค์ ทำให้รูปแบบที่สองจะมีสิ่งเร้าทั้งสิ้น 25 อัน สำหรับรูปแบบสามพยางค์ จะเป็นการพูดคำว่า “ลาลาลา” (/l a: l a: l a:/) โดย “ลา” แต่ละตัวจะเป็นไปได้ 5 รูปแบบเสียงวรรณยุกต์ภาษาไทยเช่นเดียวกับรูปแบบหนึ่งพยางค์และสองพยางค์ ดังนั้นในรูปแบบสามพยางค์จะมีสิ่งเร้าที่เป็นไปได้ทั้งหมด 125 รูปแบบ รวมทั้งหมดจะได้สิ่งเร้าต้นแบบทั้งหมด 155 อัน และในการบันทึกเสียงแต่ละครั้งจะมีการเว้นช่วงเสียงเงียบ (silence) ทั้งต้นและท้ายสิ่งเร้าอย่างชัดเจน

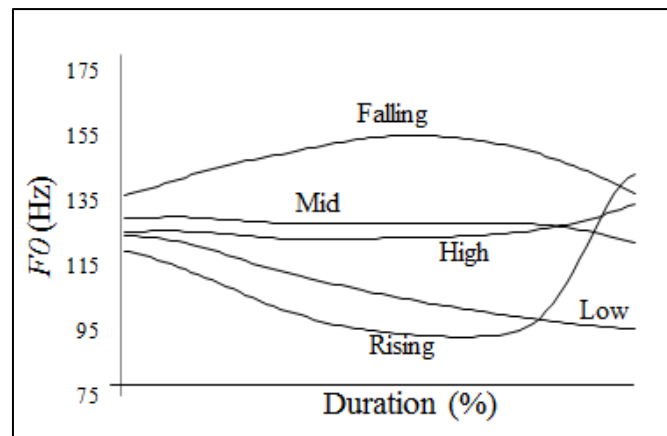
ตารางที่ 3.1 เสียงวรรณยุกต์ภาษาไทยของคำว่า “ลา” (/l a:/) 5 รูปแบบ

รูปเขียน	เสียงอ่าน	รูปหน่วยเสียง
ลา	lā:	l-aa-z ⁰
หล่า	là:	l-aa-z ¹
ล่ำ	lâ:	l-aa-z ²
ล้า	lá:	l-aa-z ³
หลา	lǎ:	l-aa-z ⁴

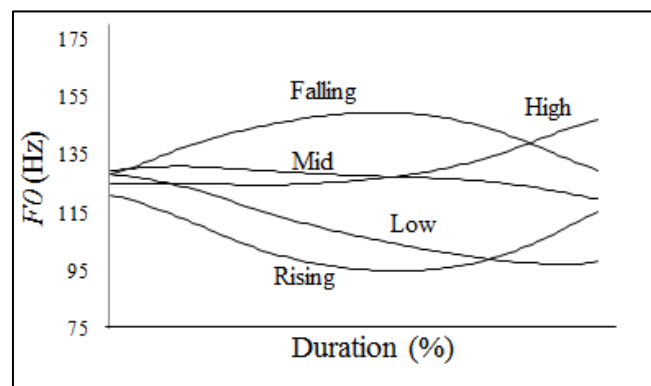
ในภาพที่ 3.1 – 3.3 เป็นภาพแสดงการเปลี่ยนแปลงรูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยทั้ง 5 เสียง ซึ่งได้จากการคำนวณค่าเฉลี่ยของความถี่มูลฐานของสิ่งเร้าในแต่ละช่วงเวลาในสเกลเปอร์เซ็นต์ โดยในภาพที่ 3.1 จะเป็นภาพการเปลี่ยนแปลงรูปร่างของความถี่มูลฐานของรูปแบบหนึ่งพยางค์ และภาพที่ 3.2 จะเป็นภาพการเปลี่ยนแปลงรูปร่างของความถี่มูลฐานของรูปแบบสองพยางค์ และภาพที่ 3.3 จะเป็นภาพการเปลี่ยนแปลงรูปร่างของความถี่มูลฐานของรูปแบบสามพยางค์ตามลำดับ ซึ่งแกนนอนของแต่ละภาพจะแทนช่วงเวลาก่อนเป็นสเกลเปอร์เซ็นต์ตั้งแต่ 0% ถึง 100% โดยที่ตำแหน่ง 0% คือจุดเริ่มต้นของค่าความถี่มูลฐานของแต่ละเสียงวรรณยุกต์ภาษาไทย และที่ตำแหน่ง 100% คือ จุดสิ้นสุดของความถี่มูลฐานของแต่ละเสียงวรรณยุกต์ภาษาไทย สำหรับแกนตั้งของแต่ละภาพจะเป็นระดับของความถี่ ซึ่งมีหน่วยเป็นเฮิรตซ์



ภาพที่ 3.1 รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบหนึ่งพยางค์



ภาพที่ 3.2 รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบสองพยางค์

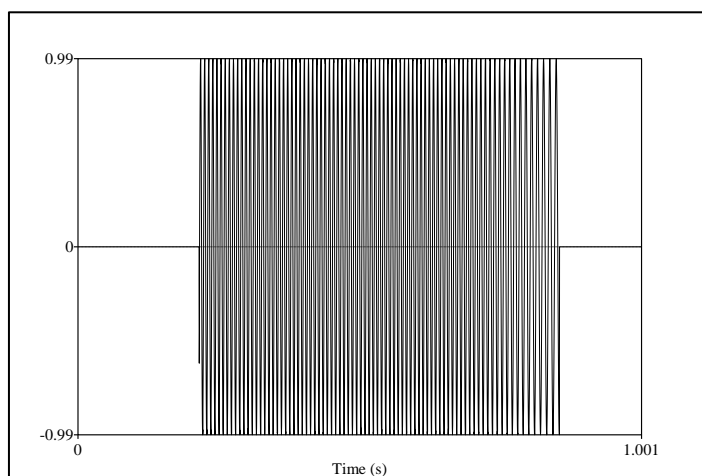


ภาพที่ 3.3 รูปร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยในรูปแบบสามพยางค์

2. การเตรียมสิ่งเร้าที่มีการดัดแปลง

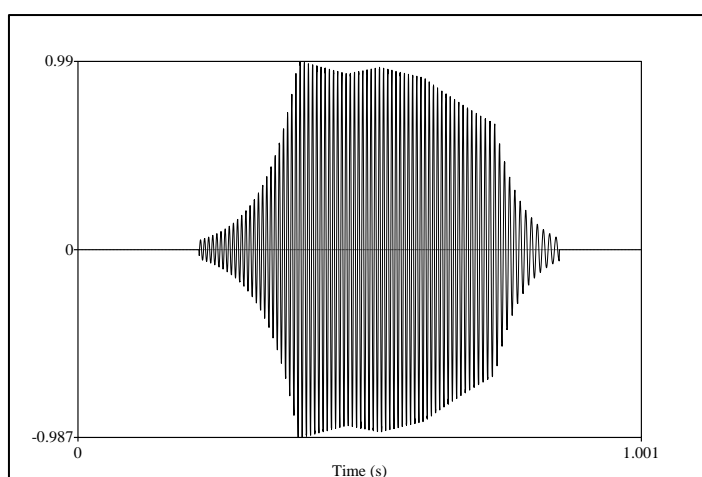
ในส่วนนี้จะเป็นการจัดเตรียมสิ่งเร้าที่มีการดัดแปลง เพื่อใช้ในการศึกษาผลที่เกิดจากลักษณะทางเสียงต่างๆ ต่อการรับรู้เสียงวรรณยุกต์ภาษาไทย โดยมีการดัดแปลงสิ่งเร้าต้นแบบตามขั้นตอนวิธีต่างๆ เพื่อให้ได้สิ่งเร้าที่มีการดัดแปลงที่สามารถนำไปใช้ในการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทย โดยลักษณะทางเสียงที่สนใจในการทดลองการรับรู้เสียงวรรณยุกต์ ได้แก่ 1) ลักษณะรูปร่างของความถี่มูลฐาน 2) รูปร่างของพลังงานที่ห่อหุ้ม (Energy Envelope) ของสัญญาณเสียงตลอดหน่วยพยางค์นั้น และ 3) ส่วนประกอบของสเปกตรัมของสัญญาณเสียงที่แบ่งช่วงคร่าวๆ ตามโครงสร้างกลุ่มความถี่สั้นพ้องของช่องเสียง (Formant's Structures)

ในการเตรียมสิ่งเร้าตัดแปลงสำหรับศึกษาลักษณะรูปร่างของความถี่มูลฐานจะเริ่มจากการสกัดค่าของความถี่มูลฐานจากสัญญาณเสียงของสิ่งเร้าต้นแบบโดยใช้เครื่องมือที่ชื่อว่า “PRAAT” [37] ในการสกัดค่าตามรอยของความถี่มูลฐาน (Fundamental Frequency Tracks) ออกมา และในขั้นตอนถัดมาจะนำค่าความถี่มูลฐานที่สกัดได้จากสัญญาณเสียงของสิ่งเร้าต้นแบบไปใช้ในการสังเคราะห์สัญญาณไซน์ซออยด์ (Sinusoidal Signals) ที่เป็นไปตามค่าของความถี่มูลฐานที่สกัดได้จากสิ่งเร้าต้นแบบ และเพื่อที่จะเน้นการศึกษาความสัมพันธ์เฉพาะความถี่มูลฐานกับการรับรู้เสียงวรรณยุกต์ภาษาไทย ค่าแอมพลิจูด (Amplitude) ของสัญญาณไซน์ซออยด์ จะถูกกำหนดให้เป็นค่าคงที่ตลอดช่วงที่ไม่ใช่ช่วงเสียงเงียบ (non-silent) ของสิ่งเร้าต้นแบบ และในบางกรณี PRAAT [37] ไม่สามารถคำนวณค่าของความถี่มูลฐานได้ในบางเฟรมของสิ่งเร้าต้นแบบ เนื่องจากผลกระทบของลักษณะการเปล่งเสียงของเส้นเสียงส่งผลให้การคำนวณในบางเฟรมเกิดการสูญหายหรือเกิดการคำนวณที่คลาดเคลื่อน ดังนั้นเพื่อชดเชยผลกระทบดังกล่าวเทคนิคในการประมาณค่า (Interpolation Techniques) และเทคนิคการลู่ (Smoothing Techniques) [38] จึงได้ถูกนำมาประยุกต์ใช้ ซึ่งผลลัพธ์ที่ได้คือสิ่งเร้าที่มีการตัดแปลงกลุ่มที่หนึ่ง เพื่อที่จะใช้ในการศึกษาลักษณะรูปร่างของความถี่มูลฐาน โดยตั้งชื่อสิ่งเร้าที่มีการตัดแปลงกลุ่มนี้ว่า กลุ่มความถี่มูลฐาน (Fo set) ซึ่งในภาพที่ 3.4 แสดงตัวอย่างของสัญญาณไซน์ซออยด์ที่สังเคราะห์ได้ในกลุ่มนี้



ภาพที่ 3.4 ตัวอย่างสัญญาณไซน์ซออยด์ที่ใช้ในการศึกษาลักษณะรูปร่างของความถี่มูลฐานที่สังเคราะห์ได้จากสิ่งเร้าต้นแบบของคำว่า “ลา” ในเสียงสามัญรูปแบบหนึ่งพยางค์

สำหรับในกลุ่มที่สองจะเป็นการตัดแปลงเพิ่มเติมจากสิ่งเร้าที่มีการตัดแปลงในกลุ่มความถี่มูลฐาน เพื่อที่จะศึกษารูปร่างของพลังงานที่ห่อหุ้มของสัญญาณเสียงตลอดหน่วยพยางค์นั้น ซึ่งขั้นตอนในการตัดแปลง จะทำการตัดแปลงขนาดของแอมพลิจูดของสิ่งเร้าที่มีการตัดแปลงในกลุ่มความถี่มูลฐาน จากเดิมที่มีค่าคงที่ให้มีค่าเป็นไปตามรูปร่างห่อหุ้มแอมพลิจูดของรูปคลื่น (Waveforms) ในสิ่งเร้าต้นแบบที่ตรงกันกับสิ่งเร้าที่มีการตัดแปลงในกลุ่มความถี่มูลฐาน โดยผลลัพธ์ที่ได้ของสัญญาณไซนูซอยด์นอกจากจะได้ลักษณะรูปร่างของความถี่มูลฐานแล้ว ยังได้รูปร่างของพลังงานที่ห่อหุ้มของสัญญาณเสียงตลอดหน่วยพยางค์นั้นด้วย ดังนั้นจึงตั้งชื่อสิ่งเร้าที่มีการตัดแปลงกลุ่มนี้ว่า กลุ่มความถี่มูลฐานและพลังงาน (Fo+E set) ดังภาพที่ 3.5 ที่แสดงตัวอย่างของสัญญาณไซนูซอยด์ในกลุ่มความถี่มูลฐานและพลังงานได้



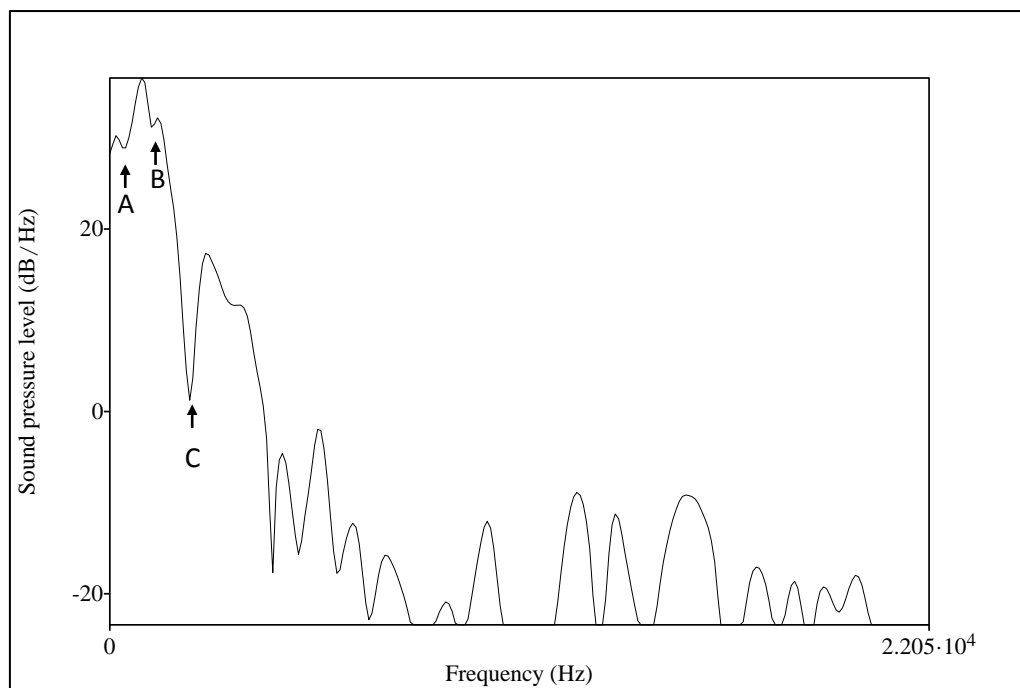
ภาพที่ 3.5 ตัวอย่างสัญญาณไซนูซอยด์ที่ใช้ในการศึกษารูปร่างของพลังงานที่ห่อหุ้มที่สังเคราะห์ได้จากสิ่งเร้าต้นแบบของคำว่า “ลา” ในเสียงสามัญรูปแบบหนึ่งพยางค์

สำหรับการตัดแปลงสิ่งเร้าต้นแบบในลำดับถัดไปจะเป็นการตัดแปลงสิ่งเร้าต้นแบบเพื่อศึกษาส่วนประกอบของสเปกตรัมของสัญญาณเสียงที่แบ่งช่วงคร่าวๆ ตามโครงสร้างกลุ่มความถี่สั้นพ้องของช่องเสียงที่มีผลต่อการรับรู้เสียงวรรณยุกต์ภาษาไทย โดยในกลุ่มนี้จะแบ่งออกเป็น 3 เซตย่อยได้แก่ เซตที่หนึ่งเป็นเซตของสิ่งเร้าตัดแปลงที่ไม่มีองค์ประกอบของฟอร์แมนต์ เซตที่สองเป็นเซตของสิ่งเร้าตัดแปลงที่มีเฉพาะโครงสร้างกลุ่มความถี่สั้นพ้องของช่องเสียงที่หนึ่ง หรือ ความ

ถี่พอร์แมนต์ที่หนึ่ง และเซตที่สามเป็นเซตของสิ่งเร้าตัดแปลงที่มีทั้งความถี่พอร์แมนต์ที่หนึ่งและความถี่พอร์แมนต์ที่สอง

ในเซตที่หนึ่งสิ่งเร้าต้นแบบจะถูกกรองช่วงความถี่ต่ำด้วยวงจรกรองสัญญาณความถี่ต่ำ (Low Pass Filter) เพื่อที่จะตัดช่วงที่ความถี่สูง หรือกล่าวอีกนัยหนึ่งคือ การกรององค์ประกอบของพอร์แมนต์ต่างๆ ทิ้งไป โดยในเซตนี้วงจรกรองสัญญาณความถี่ต่ำจะถูกตั้งค่าไว้ให้กรองความถี่ที่ต่ำกว่าครึ่งหนึ่งของค่าเฉลี่ยพอร์แมนต์ที่หนึ่งที่คำนวณจากทุกๆ สิ่งเร้าต้นแบบ ซึ่งการกรองความถี่ในลักษณะนี้สามารถอนุมานเป็นการคำนวณอย่างสังเขปของจุดต่ำสุดสัมพัทธ์ (Local Minimum) ระหว่างความถี่มูลฐานกับความถี่พอร์แมนต์ที่หนึ่ง โดยที่ตำแหน่งดังกล่าวสามารถพิจารณาได้ตามภาพที่ 3.6 ซึ่งเป็นภาพที่แสดงสเปกตรัมของสัญญาณเสียงสิ่งเร้าต้นแบบ ณ เวลาหนึ่ง โดยที่ในเซตนี้จะกรองความถี่ที่สูงกว่าตำแหน่ง A ทิ้ง ซึ่งตำแหน่ง A เป็นจุดต่ำสุดสัมพัทธ์ระหว่างความถี่มูลฐานกับความถี่พอร์แมนต์ที่หนึ่ง โดยค่าที่ตั้งไว้จากการคำนวณครึ่งหนึ่งของค่าเฉลี่ยของพอร์แมนต์ที่หนึ่งคือ 450 เฮิร์ตซ์ ดังนั้นจึงตั้งชื่อสิ่งเร้าที่มีการตัดแปลงเซตนี้ว่า “Lo_Fo” นอกจากนี้สำหรับในกรณีนี้สามารถตั้งค่าประมาณของครึ่งหนึ่งของค่าเฉลี่ยของพอร์แมนต์ที่หนึ่งเป็นค่าคงตัวได้เนื่องจากมีสมมุติฐานที่ว่าค่าความถี่พอร์แมนต์ของพยางค์เดียวกัน และผู้พูดคนเดียวกันในการบันทึกเสียงครั้งเดียวกันจะมีค่าประมาณที่ใกล้เคียงกันดังนั้นจึงสามารถกรองความถี่ได้ด้วยค่าความถี่คงที่

สำหรับในอีกสองเซตที่เหลือจะใช้แนวคิดเดียวกันกับเซตที่หนึ่ง เพียงแต่ตำแหน่งที่จะใช้ในการกรองความถี่เปลี่ยนไปตามสิ่งที่ต้องการศึกษา โดยเซตที่สองตั้งชื่อว่า “Lo_F1/F2” โดยเซตนี้เป็นการหาค่าเฉลี่ยจุดกึ่งกลางระหว่างความถี่พอร์แมนต์ที่หนึ่งกับความถี่พอร์แมนต์ที่สอง ซึ่งเมื่อพิจารณาจาก ภาพที่ 3.6 จะได้ตำแหน่งที่ B โดยจากการประมาณการอย่างสังเขป ของตำแหน่งนี้จะได้ค่าประมาณ 1,200 เฮิร์ตซ์ และสำหรับเซตสุดท้ายเป็นเซต “Lo_F2/F3” เซตนี้จะเป็นการหาค่าเฉลี่ยจุดกึ่งกลางระหว่างความถี่พอร์แมนต์ที่สองกับความถี่พอร์แมนต์ที่สาม ซึ่งจากการประมาณค่าจะได้ค่าของตำแหน่งนี้ที่ความถี่ 2,100 เฮิร์ตซ์ ซึ่งหากพิจารณาจากภาพที่ 3.6 ก็คือตำแหน่งที่ C



ภาพที่ 3.6 สเปกตรัมของสัญญาณเสียงสังเร้าต้นแบบ ณ เวลาหนึ่งๆ

การทดลอง

ในการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยประกอบไปด้วย การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน และการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม

1. การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน

การทดลองที่หนึ่งจะเป็นการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน โดยจุดประสงค์ของการทดลองนี้เพื่อที่จะศึกษาผลกระทบของความถี่มูลฐานและรูปร่างของพลังงานต่อการรับรู้เสียงวรรณยุกต์ ซึ่งในการทดลองแบ่งเป็นสองส่วน ได้แก่ การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานเพียงอย่างเดียว กับ การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน

ในการทดลองได้เลือกผู้เข้าร่วมทดลองเป็นผู้ใช้ภาษาไทยเป็นภาษาหลักจำนวนทั้งหมด 15 คน ประกอบด้วย ผู้ชาย 11 คน และผู้หญิง 4 คน ซึ่งแต่ละคนไม่มีประวัติในการมีปัญญา

ทางการได้ยิน แต่เพื่อความถูกต้องของการทดลองผู้เข้าร่วมทดลองแต่ละคนจะถูกทดสอบความสามารถในการแยกแยะเสียงวรรณยุกต์ภาษาไทยก่อนเริ่มทำการทดลอง โดยผู้ทำการทดลองจะต้องออกเสียงวรรณยุกต์ต่างๆ กัน และถามผู้เข้าร่วมทดลองว่าเสียงวรรณยุกต์ที่ออกเสียงมาคือเสียงวรรณยุกต์อะไร และให้ผู้เข้าร่วมทดลองผันเสียงวรรณยุกต์ให้ผู้ทำการทดลองฟังก่อนเริ่มการทดลอง ซึ่งจากการทดสอบคัดกรองไม่พบผู้มีปัญหาในการได้ยินหรือมีปัญหาแยกแยะเสียงวรรณยุกต์ โดยอายุของผู้เข้าร่วมทดลองอยู่ตั้งแต่ 21 ถึง 27 ปี และได้อายุเฉลี่ยที่ 22.5 ปี

ผู้เข้าร่วมทดลองแต่ละคนจะได้ยินสิ่งเร้าจากคอมพิวเตอร์ผ่านทางชุดหูฟัง OKER Dynamic Headphone OE-2688M.V. หลังจากนั้นผู้เข้าร่วมทดลองจะต้องทำการระบุเสียงวรรณยุกต์ภาษาไทยที่ได้ยินจากสิ่งเร้าว่าเป็นเสียงวรรณยุกต์ใดผ่านส่วนต่อประสานผู้ใช้ (User Interface) ดังแสดงในภาพที่ 3.7 โดยผู้เข้าร่วมทดลองสามารถฟังสิ่งเร้าได้อีกรอบจนกว่าจะได้คำตอบที่แน่นอน แล้วจึงทำการระบุเสียงวรรณยุกต์ภาษาไทยในช่องที่ตรงกับสิ่งเร้าที่ได้ยิน และกดปุ่มถัดไปเพื่อยืนยันคำตอบและทำการฟังสิ่งเร้าอันถัดไปจนครบแบบทดสอบ

ภาพที่ 3.7 ส่วนต่อประสานผู้ใช้ของโปรแกรมที่ใช้สำหรับระบุเสียงวรรณยุกต์ภาษาไทยของผู้เข้าร่วมทดลอง

สำหรับสิ่งเร้าที่ผู้เข้าร่วมการทดลองจะได้ฟังในการทดลองนี้จะมี 3 ชุด ได้แก่ ชุดสิ่งเร้าต้นแบบ ชุดสิ่งเร้าที่มีการดัดแปลงกลุ่ม Fo+E และ ชุดสิ่งเร้าที่มีการดัดแปลงกลุ่ม Fo ตามลำดับ โดยแต่ละชุดมีการสุ่มลำดับที่ไม่เหมือนกันสำหรับผู้เข้าร่วมทดลองแต่ละคน และในแต่ละชุดจะประกอบด้วย สิ่งเร้า 5 อันจากรูปแบบหนึ่งพยางค์ สิ่งเร้า 25 อันจากรูปแบบสองพยางค์ และสิ่งเร้า

125 อันจากรูปแบบสามพยางค์ ดังนั้น สิ่งเร้าในแต่ละเซตที่ผู้เข้าร่วมทดลองแต่ละคนจะได้ยินจะมีทั้งหมด เซตละ 165 อัน

2. การทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม

ในการทดลองที่สอง เป็นศึกษาการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม เพื่อที่จะศึกษาผลกระทบในแต่ละช่วงสเปกตรัมต่อการรับรู้เสียงวรรณยุกต์ โดยในการทดลองนี้ใช้ผู้พูดภาษาไทยเป็นภาษาหลักทั้งหมด 11 คน เป็นผู้ชาย 9 คน และผู้หญิงอีก 2 ซึ่งอายุเฉลี่ยของกลุ่มนี้อยู่ที่ 23.5 ปี สำหรับผู้เข้าร่วมทดลองในการทดลองนี้ได้ผู้ชาย 5 คน จากการทดลองที่แล้ว และอีกคนที่เหลืออีก 6 คนเป็นคนผู้เข้าร่วมทดลองใหม่ ซึ่งก่อนเริ่มการทดลองจะมีการทดสอบความสามารถในการแยกแยะเสียงวรรณยุกต์ภาษาไทย และใช้อุปกรณ์และโปรแกรม รวมถึงวิธีการเดียวกับการทดลองที่หนึ่ง

สิ่งเร้าที่ใช้ในการทดลองนี้ประกอบด้วยสิ่งเร้าต้นแบบ และสิ่งเร้าในกลุ่มที่ใช้ในการศึกษา ส่วนประกอบของสเปกตรัมของสัญญาณเสียงที่แบ่งช่วงคร่าวๆ ตามโครงสร้างกลุ่มความถี่สั้นพ้องของช่องเสียงซึ่งประกอบไปด้วย 3 เซตได้แก่ เซต Lo_F1, เซต Lo_F1/F2 set และเซต Lo_F2/F3

ผลการทดลอง และสรุปผลการทดลอง

เนื้อหาส่วนนี้จะแบ่งออกเป็นสามส่วนได้แก่ ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน, ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม และสรุปผล และอภิปรายผลทั้งสองการทดลอง

1. ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน

ในตารางที่ 3.2 แสดงผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน จากผลการทดลองพบว่าผู้เข้าร่วมทดลองสามารถที่จะระบุเสียงวรรณยุกต์ภาษาไทยของสิ่งเร้าต้นแบบในรูปแบบหนึ่งพยางค์ได้ความถูกต้องสูงสุด และค่อยๆ ลดลงเรื่อยๆ เมื่อจำนวนรูปแบบพยางค์ค่อยๆ เพิ่มขึ้นมา ซึ่งจากผลการทดลองในกลุ่ม Fo และ Fo+E ก็พบแนวโน้มในลักษณะเดียวกัน กล่าวคือ ความถูกต้องในการระบุเสียงวรรณยุกต์ภาษาไทยในสิ่งเร้า

จะสูงสุด เมื่อสิ่งเร้าอยู่ในรูปแบบหนึ่งพยางค์ และจะค่อยๆ ลดลงเมื่อจำนวนพยางค์เพิ่มขึ้น และจากการทดลองเป็นที่เห็นอย่างชัดเจนว่าผู้เข้าร่วมทดลองสามารถที่จะระบุเสียงวรรณยุกต์ภาษาไทยของสิ่งเร้าต้นแบบได้ความถูกต้องมากที่สุด และมีความแตกต่างกันทางนัยทางสถิติอย่างชัดเจนเมื่อทำการทดสอบของแมคเนียร์กับกลุ่มอื่น เมื่อกำหนดค่าแอลฟา (α) ไว้ที่ 0.05 ทุกๆ รูปแบบพยางค์ แต่อย่างไรก็ตามพบว่าผลการทดลองในกลุ่ม Fo และ Fo+E ไม่ได้มีความแตกต่างกันทางนัยทางสถิติที่อย่างชัดเจนเมื่อทำการทดสอบด้วยการทดสอบของแมคเนียร์ ที่กำหนดค่าแอลฟาไว้ที่ 0.05 ยกเว้นในกรณีรูปแบบหนึ่งพยางค์ที่พบความแตกต่างกันทางนัยทางสถิติระหว่าง กลุ่ม Fo และ Fo+E อย่างชัดเจนด้วยการทดสอบของแมคเนียร์ ที่กำหนดค่าแอลฟาไว้ที่ 0.05 ซึ่งจากผลจากการทดลองพบว่าเมื่อจำนวนรูปแบบพยางค์เพิ่มขึ้นรูปร่างของพลังงานไม่สามารถช่วยในการระบุเสียงวรรณยุกต์ภาษาไทยให้มีความถูกต้องเพิ่มขึ้นได้ นอกจากนี้ในกรณีที่รูปแบบพยางค์เพิ่มขึ้นจะสามารถสังเกตเห็นความแตกต่างระหว่างสิ่งเร้าที่มีการดัดแปลงในกลุ่ม Fo และ Fo+E กับ สิ่งเร้าต้นแบบอย่างชัดเจน จึงสรุปได้ว่าการจำแนกเสียงวรรณยุกต์ภาษาไทยยังมีลักษณะเด่นอื่น ที่ผู้ใช้ภาษาไทยเป็นภาษาหลักใช้ช่วยในการแยกแยะเสียงวรรณยุกต์ภาษาไทย

ตารางที่ 3.2 ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงาน

	รูปแบบหนึ่งพยางค์	รูปแบบสองพยางค์	รูปแบบสามพยางค์
สิ่งเร้าต้นแบบ	100.0%	94.9%	90.6%
Fo+E	97.3%	85.5%	80.0%
Fo	92.0%	86.1%	80.6%

2. ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยที่แบ่งตามช่วงของสเปกตรัม

ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยที่แบ่งตามช่วงของสเปกตรัมแสดงอยู่ในตารางที่ 3.3 ซึ่งจากผลการทดลองพบว่า ในรูปแบบหนึ่งพยางค์ได้รับการระบุเสียงวรรณยุกต์ภาษาไทยได้อย่างมีความถูกต้องมากที่สุด ในขณะที่เซต Lo_F1 ในรูปแบบสามพยางค์ได้ความถูกต้องน้อยที่สุด ซึ่งจากการทดสอบของแมคเนียร์ ที่กำหนดค่าแอลฟาไว้ที่ 0.05 พบว่า เซต

Lo_F1 มีความแตกต่างทางนัยทางสถิติกับเซตอื่น และสิ่งเร้าต้นแบบอย่างชัดเจน และจากตารางที่ 3.3 จะเห็นได้อย่างชัดเจนว่าผู้เข้าร่วมทดลองไม่สามารถระบุสิ่งเร้าที่มีการตัดแปลงในเซต Lo_F1 ได้ดีเท่าสิ่งเร้าในเซตอื่น ซึ่งหมายความว่าลักษณะทางเสียงที่สำคัญได้แก่ โครงสร้างกลุ่มความถี่สั้นพ้องของช่องเสียง หรือโครงสร้างของความถี่ฟอร์แมนต์ ได้ถูกรองออกไปด้วยวงจรกรองสัญญาณความถี่ต่ำ จากผลการทดลองยังสอดคล้องกับการศึกษาของ Punyayodhin [39] ว่า โครงสร้างของความถี่ฟอร์แมนต์มีความสัมพันธ์ และมีอิทธิพลต่อรูปร่างเสียงวรรณยุกต์ภาษาไทย และยังเป็นสิ่งที่สังเกตได้ว่า การกรองความถี่ที่สูงกว่าความถี่ฟอร์แมนต์ที่หนึ่งออกไปไม่ส่งผลต่อการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยผู้ใช้ภาษาไทยเป็นภาษาหลัก ดังจะเห็นผลได้จากในตารางที่ 3.3

ตารางที่ 3.3 ผลการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยแบ่งตามช่วงของสเปกตรัม

	รูปแบบหนึ่งพยางค์	รูปแบบสองพยางค์	รูปแบบสามพยางค์	ค่าเฉลี่ย
สิ่งเร้าต้นแบบ	100.0%	96.0%	92.3%	94.0%
Lo_F2/F3	100.0%	97.8%	91.1%	94.0%
Lo_F1/F2	100.0%	97.8%	91.9%	94.5%
Lo_F1	94.0%	86.8%	81.6%	84.1%

3. สรุปผล และอภิปรายผล

จากการทดลองเมื่อวิเคราะห์คอนฟิวชันเมทริกซ์ของสิ่งเร้าต้นแบบในการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานกับรูปร่างพลังงานในตารางที่ 3.4 และวิเคราะห์คอนฟิวชันเมทริกซ์ของสิ่งเร้าต้นแบบในการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัมในตารางที่ 3.5 จะเห็นได้ว่าคู่สิ่งเร้าที่มีความสับสนมากที่สุดคือคู่เสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตวา และเมื่อพิจารณารูปร่างของความถี่มูลฐานของสิ่งเร้าต้นแบบของเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตวาในภาพที่ 3.8 ซึ่งเป็นการพล็อตทุกรูปแบบพยางค์ของเสียงวรรณยุกต์เอกและเสียงวรรณยุกต์จัตวาที่มีทั้งหมดในสิ่งเร้าต้นแบบที่ใช้ในการทดลองนี้ จะสังเกตได้ว่ารูปร่างของความถี่มูลฐานของทั้งสองเสียงวรรณยุกต์มีลักษณะใกล้เคียงกันมาก นอกจากนี้คอนฟิว

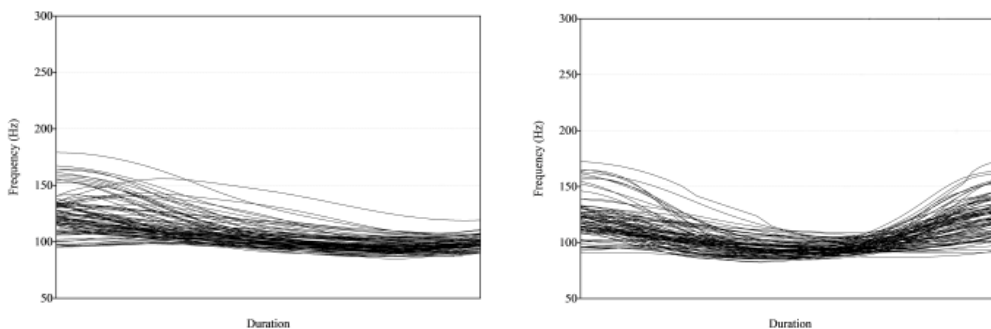
ชั้นเมทริกซ์ของสิ่งเราที่มีการดัดแปลงอื่นมีลักษณะ และแนวโน้มไปในทิศทางเดียวกันกับสิ่งเราต้นแบบ กล่าวคือคู่สิ่งเราที่มีความสับสนมากที่สุดคือคู่เสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตวา เช่นเดียวกันกับคอนฟิวชั้นเมทริกซ์ของสิ่งเราต้นแบบ

ตารางที่ 3.4 คอนฟิวชั้นเมทริกซ์ของสิ่งเราต้นแบบในการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทย โดยความถี่มูลฐานกับรูปร่างพลังงาน

พุด \ รับรู้	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	380	2	1	1	0	98.96%
เอก	14	355	2	4	13	91.49%
โท	3	10	373	7	2	94.43%
ตรี	8	1	31	355	5	88.75%
จัตวา	5	30	2	3	343	89.56%
ทั้งหมด						92.62%

ตารางที่ 3.5 คอนฟิวชั้นเมทริกซ์ของสิ่งเราต้นแบบในการรับรู้เสียงวรรณยุกต์ภาษาไทยแบ่งตามช่วงของสเปกตรัม

พุด \ รับรู้	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	259	0	1	0	0	99.62%
เอก	9	241	1	1	8	92.69%
โท	7	1	248	4	0	95.38%
ตรี	4	1	12	242	1	93.08%
จัตวา	0	27	0	1	232	89.23%
ทั้งหมด						94.00%



ภาพที่ 3.8 รูปร่างของความถี่มูลของสิ่งเร้าต้นแบบโดยทางด้านซ้ายเป็นรูปร่างความถี่มูลฐานเสียงวรรณยุกต์เอก และทางด้านขวาเป็นรูปร่างความถี่มูลฐานเสียงวรรณยุกต์จัตวา

จากการศึกษาลักษณะทางเสียงที่มีผลต่อการรับรู้เสียงวรรณยุกต์ภาษาไทยของผู้ใช้ภาษาไทยเป็นภาษาหลัก พบว่าการใช้ค่าความถี่มูลฐานเพียงอย่างเดียวไม่เพียงพอต่อการรับรู้เสียงวรรณยุกต์ภาษาไทยได้ดีพอ ดังนั้นลักษณะทางเสียงอื่นอันได้แก่ รูปร่างของพลังงาน และข้อมูลทางสเปกตรัมจึงถูกนำมาพิจารณาร่วมด้วย ซึ่งพบว่ารูปร่างของพลังงานมีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ภาษาไทยเฉพาะในกรณีรูปแบบหนึ่งพยางค์ แต่ข้อมูลทางสเปกตรัม (Spectral Information) โดยเฉพาะในช่วงความถี่ฟอร์แมนต์ที่หนึ่งมีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ภาษาไทยของผู้ใช้ภาษาไทยเป็นภาษาหลักในทุกกรณีรูปแบบพยางค์ ดังนั้นการศึกษาเพิ่มเติมเกี่ยวกับข้อมูลทางสเปกตรัมจึงสามารถช่วยให้การจำแนกเสียงวรรณยุกต์ในภาษาไทยได้ความถูกต้องมากขึ้น

บทที่ 4

การจำแนกเสียงวรรณยุกต์ภาษาไทย

การจำแนกเสียงวรรณยุกต์ภาษาไทยจะประกอบด้วย สี่ส่วนหลัก ได้แก่ ส่วนของฐานข้อมูลเสียง (Speech Corpora) ส่วนของการสกัดลักษณะเด่นจากฐานข้อมูลเสียง ส่วนของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน และส่วนของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมที่ได้จากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยมนุษย์

ฐานข้อมูลเสียง

สำหรับฐานข้อมูลเสียงที่ใช้ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยจะใช้ฐานข้อมูลสองฐานข้อมูล ได้แก่ฐานข้อมูลเสียงคำโดด (Isolated Word Speech Corpus) และฐานข้อมูลเสียงพูดต่อเนื่อง (Continuous Speech Corpus)

1. ฐานข้อมูลเสียงคำโดด

ฐานข้อมูลเสียงคำโดดถูกใช้ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยเพื่อศึกษาผลกระทบในการจำแนกเสียงวรรณยุกต์ภาษาไทย เมื่อฐานข้อมูลเสียงเป็นลักษณะคำโดด โดยกระบวนการในการสร้างฐานข้อมูลเสียงคำโดดเริ่มจากการเก็บข้อมูลเสียงจากผู้ใช้ภาษาไทยเป็นภาษาหลัก ซึ่งสำหรับฐานข้อมูลเสียงนี้ได้ทำการเก็บเสียงจากผู้พูดที่เป็นผู้ใช้ภาษาไทยเป็นภาษาหลักทั้งหมด 12 คน เป็นชาย 6 คน และ หญิง 6 คน โดยที่ช่วงอายุของคนเหล่านี้อยู่ที่ 21-22 ปี ผู้พูดแต่ละคนจะต้องพูดคำโดดทั้งหมด 22 คำ โดยในแต่ละคำจะต้องออกเสียงให้ครบทั้งห้าวรรณยุกต์ ได้แก่ เสียงวรรณยุกต์สามัญ, วรรณยุกต์เอก, วรรณยุกต์โท, วรรณยุกต์ตรี และวรรณยุกต์จัตวา ตามลำดับ ดังแสดงในตารางที่ 4.1 และระหว่างเสียงแต่ละคำจะออกเสียงห่างกันประมาณ 1 วินาทีเป็นอย่างน้อย โดยในช่วงที่เว้นระยะห่างนั้นจะเป็นช่วงเสียงเงียบ ดังนั้นในฐานข้อมูลเสียงคำโดดนี้จะประกอบไปด้วยจำนวนเสียงวรรณยุกต์ภาษาไทยทั้งหมด 1,320 เสียง ซึ่งแบ่งตามแต่ละชนิดเสียงวรรณยุกต์ได้เท่ากับ 264 เสียงต่อหนึ่งเสียงวรรณยุกต์

ตารางที่ 4.1 ตารางแสดงคำที่ผู้พูดพูดในฐานะข้อมูลเสียงคำโดด

คำ	เสียงวรรณยุกต์				
	สามัญ	เอก	โท	ตรี	จัตวา
1 อี	zī:	zì:	zî:	zí:	zǐ:
2 อู	zū:	zù:	zû:	zú:	zǔ:
3 เออ	zī̄:	zì̄:	zî̄:	zī̄:	zǐ̄:
4 อา	zā:	zà:	zâ:	zá:	zǎ:
5 แอ	zē:	zè:	zê:	zé:	zě:
6 ดา	dā:	dà:	dâ:	dá:	dǎ:
7 คา	k ^h ā:	k ^h à:	k ^h â:	k ^h á:	k ^h ǎ:
8 ที	t ^h ī:	t ^h ì:	t ^h î:	t ^h í:	t ^h ǐ:
9 แม	mē:	mè:	mê:	mé:	mě:
10 โห	hō:	hò:	hô:	hó:	hǒ:
11 รอ	rō:	rò:	rô:	ró:	rǒ:
12 เตี่ย	tī:a	tì:a	tî:a	tí:a	tǐ:a
13 เสือ	zuū:a	zù:a	zû:a	zú:a	zǔ:a
14 บอก	bō:k	bò:k	bô:k	bó:k	bǒ:k
15 รูป	rū:p	rù:p	rû:p	rú:p	rǔ:p
16 ปาด	pā:t	pà:t	pâ:t	pá:t	pǎ:t
17 รีม	rī:m	rì:m	rî:m	rí:m	rǐ:m
18 แจง	tɕē:ŋ	tɕè:ŋ	tɕê:ŋ	tɕé:ŋ	tɕě:ŋ
19 ยาว	jā:w	jà:w	jâ:w	já:w	jǎ:w
20 gay	ŋā:j	ŋà:j	ŋâ:j	ŋá:j	ŋǎ:j
21 กาน	kā:n	kà:n	kâ:n	ká:n	kǎ:n
22 แบน	bē:n	bè:n	bê:n	bé:n	bě:n

กระบวนการในการบันทึกเสียงทั้งหมดจะดำเนินการในห้องบันทึกเสียง โดยไมโครโฟนที่ใช้ในการบันทึกเสียงเป็น ไมโครโฟนทางเดียว (Shure SM58 Unidirectional Microphone) โดยไมโครโฟนจะวางห่างจากปากผู้พูดประมาณ 15 เซนติเมตร และสัญญาณเสียงจะถูกบันทึกด้วย

โปรแกรมคิวเบส (Cubase) [40] โดยสัญญาณเสียงจะถูกบันทึกที่ความถี่สุ่ม 44.1 กิโลเฮิรตซ์ และบันทึกในรูปแบบเอ็มเอสเวฟ (MS wav) โดยใช้การเข้ารหัส 16 บิต พีซีเอ็ม และเพื่อเป็นการควบคุมคุณภาพของสัญญาณเสียงให้เป็นไปตามเงื่อนไขสภาพแวดล้อมสะอาด (Clean Environment) จึงมีการตั้งเงื่อนไขคือ เมื่อสัญญาณมีค่าอัตราสัญญาณเสียงต่อคลื่นรบกวน (Signal to Noise Ratio: SNR) น้อยกว่า 30 dB จะต้องมีการบันทึกเสียงใหม่ อีกทั้งถ้าสัญญาณเสียงพูดเกิดการขริบ (Clip) กล่าวคือสัญญาณที่บันทึกมีค่ามากเกินไปจนขอบเขตที่จะบันทึกจนโดนตัดทิ้งก็ต้องทำการบันทึกเสียงใหม่เช่น และเมื่อเสร็จสิ้นกระบวนการบันทึกเสียง ข้อมูลเสียงจะถูกนำมาแบ่งเป็นส่วนๆ ให้สอดคล้องกับการออกเสียงในข้อมูลเสียง โดยใช้ผู้เชี่ยวชาญในการแบ่งอย่างระมัดระวัง

2. ฐานข้อมูลเสียงพูดต่อเนื่อง

สำหรับฐานข้อมูลเสียงพูดต่อเนื่องที่จะใช้ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยเพื่อศึกษาลักษณะและผลกระทบเมื่อทำการจำแนกเสียงวรรณยุกต์ภาษาไทยบนฐานข้อมูลเสียงพูดต่อเนื่อง คือ ฐานข้อมูลเสียงขนาดใหญ่สำหรับระบบรู้จำเสียงพูดต่อเนื่องภาษาไทย “โลดัส” (Large Vocabulary Thai Continuous Speech Recognition Corpus: LOTUS) [12] ซึ่งเป็นฐานข้อมูลเสียงพูดภาษาไทยขนาดใหญ่ที่ได้รับการออกแบบ และพัฒนาให้ได้ตามมาตรฐานสากล อีกทั้งฐานข้อมูลเสียงโลดัสยังถูกนำไปประยุกต์ใช้งานเพื่อสร้างโปรแกรมประยุกต์มากมาย อาทิเช่น ไอส์ปีช (ISpeech) [41] ซึ่งเป็นเครื่องช่วยประมวล (engine) ในการสร้างระบบรู้จำเสียงพูดภาษาไทย และ โปรแกรมประยุกต์แปลงเสียงพูดภาษาไทยสู่ภาษาอื่นๆ ในโครงการยูสตาร์ (Universal Speech Translation Advanced Research: U-STAR) [42] นอกจากนี้ฐานข้อมูลเสียงโลดัสยังถูกนำไปใช้ในงานวิจัยต่างๆ [17, 42] มากมายในการศึกษาลักษณะทางเสียงในภาษาไทย รวมถึงการสร้างแบบจำลองทางเสียงเพื่อใช้สร้างระบบรู้จำเสียงพูดภาษาไทย

ฐานข้อมูลเสียงโลดัสประกอบด้วยข้อมูลเสียงพูด 4 ชุดย่อยได้แก่ ชุดหน่วยเสียงสมมูล (Phonetically Distribution set: PD), ชุดฝึกฝน (Training set: TR), ชุดทดสอบเพื่อพัฒนา (Development Test set: DT) ชุดทดสอบเพื่อประเมิน (Evaluation Test set: ET) สำหรับการ

สร้างระบบรู้จำเสียงพูดชุดหน่วยเสียงสมมูลจะถูกใช้ในการสร้างแบบจำลองฐานเสียงขั้นต้น เนื่องจากชุดหน่วยเสียงสมมูลเป็นเสียงที่ขนาดเล็กที่สุดที่มีความครอบคลุมหน่วยเสียงทั้งหมดในภาษาไทย โดยจำนวนคู่หน่วยเสียงที่ครอบคลุมเป็นไปตามตารางที่ 4.2 ชุดฝึกฝนจะใช้ในการเรียนรู้ลักษณะของหน่วยเสียงย่อยเพิ่มเติมด้วยจำนวนคำศัพท์ในภาษาไทยเป็น 5,000 คำ ชุดทดสอบเพื่อพัฒนาใช้ในการปรับแต่งค่าพารามิเตอร์ต่างๆของระบบรู้จำเสียงพูดให้มีความเหมาะสม และชุดทดสอบเพื่อประเมินใช้ในการวัดความสามารถของระบบรู้จำเสียงพูด โดยในตารางที่ 4.3 ได้แสดงรายละเอียดเพิ่มเติมของแต่ละชุด

ตารางที่ 4.2 ตารางแสดงจำนวนคู่หน่วยเสียงที่ครอบคลุมในฐานข้อมูลเสียงโลตัสชุดหน่วยเสียงสมมูล

คู่หน่วยเสียง	จำนวนรูปแบบ
C_iV	583
VC_f	157
C_fC_i	329
VC_i	559
ทั้งหมด	1,628

ตารางที่ 4.3 ตารางแสดงรายละเอียดของแต่ละชุดข้อมูลในฐานข้อมูลเสียงโลตัส [12]

	PD	TR	DT	ET
จำนวนประโยค	801	3,007	500	500
จำนวนคำศัพท์	2,269	5,000	1,622	1,630
จำนวนคำ	7,847	55,504	8,076	8,290
จำนวนผู้พูดที่เนคเทค	48	24	12	12
จำนวนผู้พูดที่มหาวิทยาลัยสงขลานครินทร์	100	60	20	20
จำนวนผู้พูดที่มหาวิทยาลัยเทคโนโลยีมหานคร	100	60	20	20

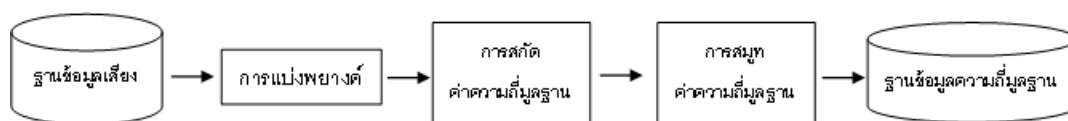
ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยจะตั้งอยู่บนสมมุติฐานที่ว่าขอบเขตของพยางค์แต่ละตัวจะถูกกำหนดมาก่อนที่จะทำการจำแนกเสียงวรรณยุกต์ภาษาไทย ส่งผลให้ในการทดลองเลือกใช้แต่ชุดหน่วยเสียงสมมูลที่มีการบันทึกเสียงที่เนคเทคซึ่งมีผู้พูดทั้งหมด 48 คน เนื่องจากในชุดหน่วยเสียงสมมูลที่มีการบันทึกเสียงที่เนคเทค เป็นชุดเสียงเพียงชุดเดียวที่มีการแบ่งส่วนของหน่วยเสียงอย่างถูกต้อง รวมถึงมีการปรับแนวเวลามาให้อย่างชัดเจน โดยนักภาษาศาสตร์ ดังนั้นในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยจะทำการทดลองบนชุดหน่วยเสียงสมมูลที่มีการบันทึกเสียงที่เนคเทคเท่านั้น

การสกัดลักษณะเด่นจากฐานข้อมูลเสียง

การสกัดลักษณะเด่นจากฐานข้อมูลเสียงประกอบด้วยขั้นตอนหลักดังต่อไปนี้ การเตรียมค่าความถี่มูลฐานจากฐานข้อมูลเสียง, การเตรียมลักษณะเด่นจากฐานข้อมูลความถี่มูลฐานเพื่อใช้ในการเรียนรู้ของเครื่อง (Feature Extraction)

1. การเตรียมค่าความถี่มูลฐานจากฐานข้อมูลเสียง

กระบวนการในการเตรียมค่าความถี่มูลฐานจากฐานข้อมูลเสียงเป็นไปตามภาพที่ 4.1 โดยเริ่มจากการนำฐานข้อมูลเสียงมาแบ่งพยางค์ จากนั้นสกัดค่าความถี่มูลฐานตามแต่ละช่วงพยางค์ และในขั้นตอนสุดท้ายจะเป็นการสมทค่าความถี่มูลฐาน



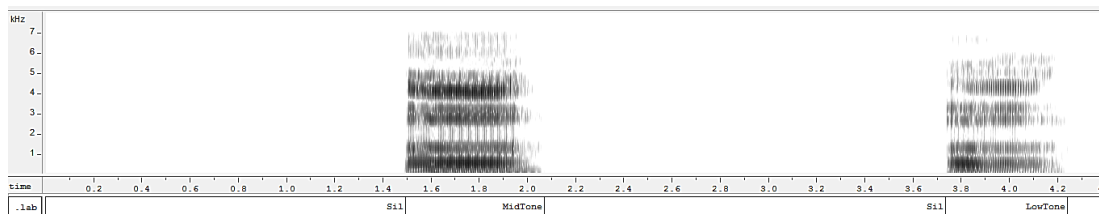
ภาพที่ 4.1 กระบวนการในการเตรียมค่าความถี่มูลฐานจากฐานข้อมูลเสียง

1.1. การแบ่งพยางค์

ในการจำแนกเสียงวรรณยุกต์ภาษาไทยขอบเขตของพยางค์เป็นสิ่งจำเป็นในการที่จะคำนวณหาลักษณะเด่นของเสียงวรรณยุกต์ในแต่ละพยางค์ เนื่องจากเสียงวรรณยุกต์หนึ่งเสียงจะ

ขึ้นกับพยางค์หนึ่งพยางค์ ดังนั้นการแบ่งพยางค์ที่มีความถูกต้องก็จะทำให้ได้ผลลัพธ์ในการจำแนกเสียงวรรณยุกต์ที่สูงขึ้น

ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทย เพื่อความถูกต้องในการทดลองจะตั้งสมมุติฐานที่ว่ารู้ขอบเขตของพยางค์แต่ละพยางค์อย่างชัดเจน โดยสำหรับฐานข้อมูลเสียงคำโดดการแบ่งส่วนพยางค์จะถูกแบ่งตามพยางค์โดยใช้ผู้เชี่ยวชาญในการแบ่งขอบเขตทางสัทศาสตร์ (Phonetic) โดยใช้โปรแกรมเวฟเซอร์เฟอร์ (WaveSurfer) [43] ในการพิจารณาแผนภาพเสียงหรือสเปกโตรแกรม (Spectrogram) และทำการระบุเสียงวรรณยุกต์ตรงแถบการถอดเสียงในขอบเขตของพยางค์ดังกล่าว ดังแสดงในภาพที่ 4.2



ภาพที่ 4.2 ตัวอย่างการแบ่งพยางค์ในฐานข้อมูลเสียงคำโดดโดยใช้โปรแกรมเวฟเซอร์เฟอร์

สำหรับฐานข้อมูลเสียงพูดต่อเนื่องโลดส์ในการแบ่งพยางค์ แม้ว่าในชุดหน่วยเสียงสมมูลที่มีการบันทึกเสียงที่เนคเทค จะมีการแบ่งส่วนของหน่วยเสียง รวมถึงมีการปรับแนวเวลามาโดยนักภาษาศาสตร์ แต่ไม่ได้มีการระบุขอบเขตของเสียงวรรณยุกต์มาให้อย่างชัดเจนดังนั้น ก่อนการทดลองจึงจำเป็นต้องระบุหน่วยเสียงในกระบวนการถอดเสียง (Transcription) ก่อน โดยขั้นตอนในการระบุเสียงวรรณยุกต์ภาษาลงในกรออดเสียงทำได้โดยการจับคู่ กลุ่มหน่วยเสียงของการถอดเสียง กับเสียงวรรณยุกต์ของกลุ่มหน่วยเสียงดังกล่าว และทำการแปลงกลุ่มหน่วยเสียงทั้งหมดให้เป็นการถอดเสียงในรูปแบบเสียงวรรณยุกต์ ซึ่งการกระบวนการดังกล่าวจะทำให้เราได้รับการถอดเสียงที่มีหน่วยเสียงเป็นเสียงวรรณยุกต์ แต่เนื่องจากการพิจารณาเสียงวรรณยุกต์ของกลุ่มหน่วยเสียงของแต่ละกลุ่มจำเป็นต้องมีฐานข้อมูลข้อความที่บอกว่าข้อความ มีวิธีการอ่านอย่างไร ประกอบด้วยเสียงวรรณยุกต์ใดก่อน ซึ่งในฐานข้อมูลเสียงโลดส์มีฐานข้อมูลข้อความอยู่แล้ว แต่ฐานข้อมูลข้อความนั้นยังไม่มี ความถูกต้องสมบูรณ์ดังนั้น ก่อนการนำมาใช้จึงได้มีการปรับให้ฐานข้อมูลข้อความมีความสมบูรณ์มากขึ้นก่อน หลังจากการจับคู่ กลุ่มหน่วยเสียงของการถอด

เสียง กับเสียงวรรณยุกต์ของกลุ่มหน่วยเสียงจะได้จำนวนของเสียงวรรณยุกต์ที่จะใช้ในการทดลอง เป็นไปตามตารางที่ 4.4

ตารางที่ 4.4 ตารางแสดงจำนวนเสียงวรรณยุกต์ภาษาไทยในฐานข้อมูลเสียงโลดส์ชุดหน่วยเสียง สมดุลที่มีการบันทึกเสียงที่เนคเทค

เสียงวรรณยุกต์	จำนวนเสียงวรรณยุกต์
สามัญ	8,662
เอก	5,277
โท	4,966
ตรี	4,202
จัตวา	2,150
ทั้งหมด	25,257

1.2. การสกัดค่าความถี่มูลฐาน

การสกัดค่าความถี่มูลฐานในการทดลองถือเป็นส่วนสำคัญในการทดลองการจำแนกเสียงวรรณยุกต์เพราะในงานที่เกี่ยวข้องกับการจำแนกเสียงวรรณยุกต์ ความถี่มูลฐานถือเป็นลักษณะเด่นสำคัญที่สุดที่ใช้ในการระบุเสียงวรรณยุกต์ โดยในการศึกษานี้เราจะใช้วิธีอัตโนมัติในการสกัดค่าความถี่มูลฐานออกจากฐานข้อมูลเสียงที่ถูกแบ่งช่วงเวลาอย่างชัดเจนจากขั้นตอนการแบ่งพยางค์

ในการสกัดค่าความถี่มูลฐานจะใช้วิธีอัตโนมัติ โดยเครื่องมือในการช่วยสกัดค่าความถี่มูลฐานคือ “แพรท” (PRAAT) [37] ซึ่งเป็นเครื่องมือที่นิยมใช้ในการวิเคราะห์ทางสัทศาสตร์กันอย่างแพร่หลาย โดยเราจะใช้วิธีอัตโนมัติที่มีการปรับปรุงของแพรท [37] ในการสกัดค่าความถี่มูลฐาน ซึ่งคำนวณได้สมการที่ 4.1 สมการของวิธีอัตโนมัติ

$$r_A(\tau) \equiv \int x(t)x(t+\tau)dt \quad (4.1)$$

โดยที่ τ คือ ช่วงล่าช้าของเวลา และ $r_A(\tau)$ คือ ฟังก์ชันของอัตโนมัติสัมพันธ์ของช่วงล่าช้าของเวลา τ และ $x(t)$ คือฟังก์ชันของสัญญาณเสียงพูดที่เวลา t และ $x(t+\tau)$ คือฟังก์ชันของสัญญาณเสียงพูดที่เวลา $t+\tau$. เนื่องจากการทำอัตโนมัติสัมพันธ์ตามสมการที่ 4.1 จะถูกผลกระทบจากกรอบของสัญญาณ หรือ ผลกระทบจากวินโดว (Window Effect) เพราะสัญญาณจะถูกตัดแบ่งมาคิดตามช่วงเวลา ดังนั้นเพื่อลดผลกระทบดังกล่าว Boersma [44] ซึ่งเป็นผู้พัฒนาแพททิจได้เสนอวิธีในการคำนวณอัตโนมัติสัมพันธ์ โดยลดผลของกรอบสัญญาณเวลาเพื่อให้การคำนวณค่าความถี่มูลฐานมีความถูกต้องมากยิ่งขึ้นโดยการ ทำอัตโนมัติสัมพันธ์ของกรอบสัญญาณแล้วนำไปหารกับการทำอัตโนมัติสัมพันธ์ของสัญญาณเวลาเป็นไปตามสมการที่ 4.2

$$r_X(\tau) \approx \frac{r_A(\tau)}{r_W(\tau)} \quad (4.2)$$

โดยที่ τ คือ ช่วงล่าช้าของเวลา และ $r_X(\tau)$ คือ ฟังก์ชันอัตโนมัติสัมพันธ์ของที่มีการลดผลของกรอบสัญญาณเวลาที่ช่วงล่าช้าของเวลา τ และ $r_A(\tau)$ คือ ฟังก์ชันของอัตโนมัติสัมพันธ์ของสัญญาณเสียงที่ช่วงล่าช้าของเวลา τ และ $r_W(\tau)$ ฟังก์ชันของอัตโนมัติสัมพันธ์ของกรอบสัญญาณที่ช่วงล่าช้าของเวลา τ

ในการสกัดค่าความถี่มูลฐานจากแพททิจจำเป็นต้องมีการศึกษาและปรับค่าพารามิเตอร์ที่เหมาะสมซึ่งจากการทบทวนวรรณกรรม พบว่า Quang [28] ได้ทำการศึกษาการตั้งค่าพารามิเตอร์ของแพททิจในการสกัดค่าความถี่มูลฐานสำหรับภาษาเวียดนาม ซึ่งจากการศึกษาพบว่าวิธีการตั้งค่าที่ให้ความถูกต้องและแม่นยำที่สุดมีการกำหนดค่าพารามิเตอร์เป็นไปตามตารางที่ 4.5

ตารางที่ 4.5 ตารางแสดงการตั้งค่าพารามิเตอร์ของแพรทในการสกัดค่าความถี่มูลฐาน

พารามิเตอร์	การตั้งค่า
วิธีการ (Method)	ac
เวลาถัดไป (Time step) (วินาที)	0.01
พื้นของความถี่ (Pitch floor) (เฮิรตซ์)	100.0
เพดานของความถี่ (Pitch ceiling) (เฮิรตซ์)	400.0
ความแม่นยำสูง (Very accurate)	True
ขีดแบ่งเสียงเงียบ (Silence Threshold) (โวลต์)	0.02
ขีดแบ่งโชมะ (Voice Threshold) (โวลต์)	0.45

ดังนั้นในการสกัดค่าความถี่มูลฐานจากแพรทจะใช้พารามิเตอร์ในการตั้งค่าแพรทเช่นเดียวกับที่ Quang [28] ใช้

1.3. การสมูทค่าความถี่มูลฐาน

จากการสกัดค่าความถี่มูลฐานจากแพรทถึงแม้จะมีความถูกต้องอยู่ในระดับสูงเนื่องจากเป็นการวิเคราะห์ช่วงสั้นของสัญญาณของความถี่มูลฐาน แต่อย่างไรก็ตามพบว่ายังคงมีข้อผิดพลาดได้เนื่องจากผลของโครงสร้างความถี่ฟอร์แมนต์ที่มารบกวนการสกัดค่า ดังนั้นการสมูทความถี่มูลฐานจึงมีความจำเป็นในการลดผลกระทบจากโครงสร้างความถี่ฟอร์แมนต์ [7] ซึ่งในการสมูทค่าความถี่มูลฐานจะใช้ฟังก์ชันสมูทของแพรทซึ่งเป็นการสมูทสัญญาณด้วยเกาส์เซียน (Gaussian) โดยนำสัญญาณเสียงทางเวลามาทำการสังวัตนาการ (Convolution) กับ ด้วยเกาส์เซียนตามสมการที่ 4.3 ซึ่งการสังวัตนาการในข้อมูลเชิงเวลาคือการคูณกันของข้อมูลในทางเชิงความถี่ตามสมการที่ 4.4

$$G(f) = e^{-\left(\frac{f}{b}\right)^2} \quad (4.3)$$

$$F_0'(f) = F_0(f) \times G(f) \quad (4.4)$$

โดย $G(f)$ คือ เกาส์เซียนฟังก์ชันของความถี่ f และ b คือ ค่าของความกว้างแถบความถี่ (bandwidth) ของตัวกรองสัญญาณ และ $F_0(f)$ คือฟังก์ชันของสัญญาณในโดเมนความถี่ f ก่อนการสุ่ม และ $F_0'(f)$ คือฟังก์ชันของสัญญาณในโดเมนความถี่ f หลังการสุ่ม

ในการสังวัตนาการระหว่างเกาส์เซียน กับฟังก์ชันของสัญญาณเวลาในโดเมนความถี่ในส่วนของเสียงอโหมะ (Unvoiced) จะถูกประมาณค่าด้วยการประมาณค่าในช่วงเชิงเส้น (Linear Interpolation) เพื่อที่จะให้ได้ค่าความถี่มูลฐานตลอดช่วงสัญญาณก่อนการทำสังวัตนาการ และหลังจากกระบวนการสังวัตนาการค่าความถี่มูลฐานในช่วงดังกล่าวจะถูกนำออกเมื่อการสังวัตนาการเรียบร้อยแล้ว และสำหรับพารามิเตอร์ค่าของตัวของความกว้างแถบความถี่ของตัวกรองสัญญาณของแพท ได้กำหนดให้มีค่า 10 เฮิร์ตซ์ ซึ่งเป็นค่าโดยปริยาย (Default) ของโปรแกรม

2. การเตรียมลักษณะเด่นจากฐานข้อมูลความถี่มูลฐานเพื่อใช้ในการเรียนรู้ของเครื่อง

หลังจากได้ฐานข้อมูลความถี่มูลฐานจากฐานข้อมูลเสียงพูด ในลำดับถัดไปจะเป็นการเตรียมลักษณะเด่นจากฐานข้อมูลความถี่มูลฐานเพื่อใช้ในการเรียนรู้ของเครื่อง ซึ่งประกอบด้วย การเปลี่ยนสเกลความถี่ (Frequencies Scales) และเทคนิคการทำให้เป็นบรรทัดฐาน (Normalization Techniques)

2.1. การเปลี่ยนสเกลความถี่

ในการเตรียมลักษณะเด่นจากฐานข้อมูลความถี่มูลฐานเพื่อใช้ในการเรียนรู้ของเครื่องมีการทดลองเปลี่ยนสเกลความถี่จากเฮิร์ตซ์ เนื่องจากในการทดลองของ Thubthong [6] กับ Tan [7] ได้มีการทดลองเปลี่ยนสเกลความถี่จากเฮิร์ตซ์ ไปเป็นความถี่สเกลกึ่งวรรณยุกต์ (Semi-Tone) ซึ่งเป็นความถี่ที่นิยมใช้ในทางดนตรี ซึ่งสามารถคำนวณได้จากสมการ 4.5 [24]

$$S(f) = 69 + 12 \log_2 \left| \frac{f}{440} \right| \quad (4.5)$$

โดยที่ f คือความถี่ในหน่วยเฮิร์ตซ์ และ $S(f)$ คือฟังก์ชันในการแปลงความถี่ในสเกลเฮิร์ตซ์เป็นความถี่ในสเกลกึ่งวรรณยุกต์ และ นอกจากนี้ในการทดลองของทั้งสองยังได้มีการ

ทดลองเปลี่ยนสเกลความถี่จากเฮิรตซ์ ไปเป็นอีกสเกลความถี่ คือ สเกลความถี่อัตราอาร์บี (Equivalent Rectangular Bandwidth rate: ERB-rate) ซึ่งเป็นสเกลที่สะท้อนการแสดงผลทางพฤติกรรมทางเสียง (Psychoacoustic) โดยสามารถคำนวณได้ตามสมการที่ 4.6 [24]

$$ERB(f) = 11.17 \ln \left| \frac{f + 312}{f + 14675} \right| + 43.0 \quad (4.6)$$

โดยที่ f คือความถี่ในหน่วยเฮิรตซ์ และ $ERB(f)$ คือฟังก์ชันในการแปลงความถี่ในสเกลเฮิรตซ์เป็นความถี่ในสเกลความถี่อัตราอาร์บี ซึ่งจากการศึกษาของทั้งสองพบว่าการเปลี่ยนความถี่ในสเกลเฮิรตซ์ไปเป็นความถี่สเกลกึ่งวรรณยุกต์ และความถี่ในสเกลความถี่อัตราอาร์บี ส่งผลให้ความแม่นยำในการจำแนกเสียงวรรณยุกต์สูงขึ้น แต่อย่างไรก็ตามพบว่าทั้งสองได้ข้อสรุปของการเปลี่ยนสเกลที่ไม่สอดคล้องกัน กล่าวคือ Thubthong ได้ข้อสรุปว่าการเปลี่ยนสเกลความถี่เฮิรตซ์ไปเป็นความถี่ในสเกลความถี่อัตราอาร์บีให้ความถูกต้องสูงสุด ในทางกลับกัน Tan ได้ข้อสรุปว่าการเปลี่ยนสเกลความถี่เฮิรตซ์ไปเป็นความถี่สเกลกึ่งวรรณยุกต์ให้ความถูกต้องสูงสุด ดังนั้นในการจำแนกเสียงวรรณยุกต์ภาษาไทยจึงมีการศึกษาทั้งสองสเกลความถี่ดังกล่าวว่าสเกลใดมีความเหมาะสม

2.2. เทคนิคการทำให้เป็นบรรทัดฐาน

เนื่องจากโดยทั่วไปความถี่มูลฐานของเสียงพูดเป็นสิ่งที่ขึ้นกับลักษณะเฉพาะของแต่ละบุคคล ยกตัวอย่างเช่นโดยปกติช่วงของความถี่มูลฐานเสียงของผู้ชายจะค่าเฉลี่ยอยู่ที่ 125 เฮิรตซ์ ซึ่งเป็นที่แน่นอนว่าช่วงของความถี่มูลฐานเสียงของผู้หญิงจะมีค่ามากกว่าค่า โดยปกติความถี่มูลฐานเฉลี่ยของผู้หญิงจะอยู่ที่ 225 เฮิรตซ์ ดังนั้นเพื่อลดความห่างกันของผู้พูดเพศชาย และหญิง รวมถึงผู้พูดแต่ละคนเทคนิคการทำให้เป็นบรรทัดฐาน จึงเป็นสิ่งจำเป็น โดยในการศึกษานี้จะใช้เทคนิคการทำให้เป็นบรรทัดฐานสองแบบที่ Thubthong และ Tan [6, 7] ได้แก่ เทคนิคการทำให้เป็นบรรทัดฐานด้วยเทคนิคคะแนนมาตรฐาน (Z-Score Normalization Technique) โดยเป็นไปตามสมการ 4.7

$$F_0'(x) = \frac{F_0(x) - \bar{F}_0}{S.D._{F_0}} \quad (4.7)$$

โดยที่ $F_0'(x)$ คือ ฟังก์ชันของความถี่มาตรฐานหลังจากการทำให้เป็นบรรทัดฐาน และ $F_0(x)$ คือ ฟังก์ชันของความถี่มาตรฐานก่อนจากการทำให้เป็นบรรทัดฐาน และ \bar{F}_0 ค่าเฉลี่ยเลขคณิตของความถี่มาตรฐาน และ $S.D._{F_0}$ คือค่าเบี่ยงเบนมาตรฐานของความถี่มาตรฐาน และเทคนิคการทำให้เป็นบรรทัดฐานอีกวิธีหนึ่ง ได้แก่ เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ย (Mean Normalization Technique) ซึ่งเป็นไปตามสมการ 4.8

$$F'_0(x) = \frac{F_0(x) - \bar{F}_0}{F_0} \quad (4.8)$$

โดยที่ $F_0'(x)$ คือ ฟังก์ชันของความถี่มาตรฐานหลังจากการทำให้เป็นบรรทัดฐาน และ $F_0(x)$ คือ ฟังก์ชันของความถี่มาตรฐานก่อนจากการทำให้เป็นบรรทัดฐาน และ \bar{F}_0 ค่าเฉลี่ยเลขคณิตของความถี่มาตรฐาน โดยทั้งสองเทคนิคการทำให้เป็นบรรทัดฐานจะคำนวณค่า \bar{F}_0 และ ค่า $S.D._{F_0}$ จากแฟ้มเสียงในแต่ละแฟ้ม เนื่องจากในทางปฏิบัติจะไม่สามารถคำนวณหาค่า \bar{F}_0 และ ค่า $S.D._{F_0}$ ของผู้พูดแต่ละคนได้จริงเนื่องจากในการประมวลผลแฟ้มเสียงจะถูกประมวลผลทีละแฟ้มเสียง ดังนั้นเพื่อที่จะสามารถนำวิธีการทำให้เป็นบรรทัดฐานมาใช้ในการจำแนกเสียงวรรณยุกต์ได้ จึงทำการคำนวณ ค่า \bar{F}_0 และ ค่า $S.D._{F_0}$ แต่ละแฟ้มเสียง และกระบวนการทำให้เป็นบรรทัดฐานก็จะเกิดขึ้นภายในแต่ละแฟ้มเสียงแทนที่จะคำนวณจากแฟ้มเสียงของผู้พูดคนนั้นทั้งหมด

การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มาตรฐาน

ในส่วนของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มาตรฐาน จุดประสงค์เพื่อที่จะศึกษาแบบจำลองที่เหมาะสมแก่การจำแนกเสียงวรรณยุกต์ เมื่อลักษณะเด่นถูกจำกัดไว้ที่ความถี่มาตรฐานเท่านั้น ซึ่งในการทดลองจะประกอบด้วยกำหนดยุทธลักษณะเด่น และการทดลองย่อยสองการทดลอง ได้แก่ การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มาตรฐานสำหรับคำโดด และ การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มาตรฐานสำหรับเสียงพูดต่อเนื่อง

1. การกำหนดชุดลักษณะเด่น

ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานมีการนำเทคนิคในการเรียนรู้ของเครื่องมาใช้ด้วยกันสามเทคนิค ได้แก่ โครงข่ายประสาทเทียม, แบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ซึ่งจำเป็นต้องการเลือกลักษณะเด่นที่มีความเหมาะสมกับแต่ละตัวแยกประเภท (Classifier) โดยในการกำหนดลักษณะเด่นจะแบ่งเป็นสองส่วนได้แก่ชุดลักษณะเด่นสำหรับตัวแยกประเภทซึ่งจะนำไปใช้กับโครงข่ายประสาทเทียม และชุดลักษณะเด่นสำหรับตัวแยกประเภทที่มีความเป็นลำดับ (Sequential Classifier) โดยจะนำไปใช้กับแบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์

1.1. ชุดลักษณะเด่นสำหรับตัวแยกประเภท

สำหรับชุดลักษณะเด่นสำหรับตัวแยกประเภทจำเป็นต้องมีการพิจารณาลักษณะเด่นอย่างรอบคอบเนื่องจากข้อมูลเสียงเป็นข้อมูลอนุกรมเวลา (Time Series) ซึ่งโดยปกติแล้วตัวแยกประเภทจะรับข้อมูลเวกเตอร์ความยาวเฉพาะเข้ามาแล้วทำการจำแนกประเภทของข้อมูลเวกเตอร์สำหรับในกรณีข้อมูลเสียงพูดซึ่งเป็นอนุกรมเวลาจะไม่มีทางนำข้อมูลเวกเตอร์เข้าที่มีขนาดเฉพาะได้แน่นอน ดังนั้นในการแก้ปัญหาจึงจำเป็นต้องมีการแปลงลักษณะเด่นทางเสียงที่เป็นอนุกรมเวลาให้อยู่ในรูปแบบที่เป็นเวกเตอร์ที่มีขนาดเฉพาะโดยผ่านกระบวนการคัดเลือกลักษณะเด่น (Feature Selection) โดยลักษณะเด่นในชุดลักษณะเด่นสำหรับตัวแยกประเภทจะใช้ลักษณะเด่นที่เสนอโดย Thubthong [6] ซึ่งประกอบด้วยลักษณะเด่น 3 ชุดได้แก่ ชุดค่าสัมประสิทธิ์ของสมการถดถอยดีกรี 3 (3rd order Polynomial Regression Coefficients: PRC) ซึ่งคำนวณได้ตามสมการที่ 4.9

$$y = ax^3 + bx^2 + cx + d \quad (4.9)$$

โดยที่ a , b , c และ d ค่าสัมประสิทธิ์ของสมการถดถอยดีกรี 3 ซึ่งคำนวณได้จากการนำความถี่มูลฐานมาหาค่าสัมประสิทธิ์ในสมการ 4.9 สำหรับชุดลักษณะเด่นที่สองและสาม คือ ชุดลักษณะเด่นของค่าความถี่มูลฐาน และผลต่างของความถี่มูลฐาน หรือความชันของความถี่มูลฐาน โดยค่าของความถี่มูลฐานสามารถหาได้จากการคำนวณในสมการถดถอยดีกรีสาม และ

ผลต่างของความถี่มูลฐาน หรือความชันของความถี่มูลฐานสามารถคำนวณได้จากการหาอนุพันธ์ (Derivative) ของสมการ 4.9 เช่นกัน โดยสำหรับลักษณะเด่นในชุดที่สองจะใช้ความชันของความถี่มูลฐานที่ตำแหน่ง 0%, 25%, 50%, 75% และ 100% ของระยะเวลาทั้งหมดของความถี่มูลของเสียงวรรณยุกต์ภาษาไทยแต่ละตัว รวมกับค่าความถี่มูลฐานที่ตำแหน่งเริ่มต้น 0% ของระยะเวลาทั้งหมด และตำแหน่งสิ้นสุด 100% ของระยะเวลาทั้งหมด โดยเรียกลักษณะเด่นชุดนี้ว่า “F2_dF5”

สำหรับลักษณะเด่นในชุดที่สามมีลักษณะเด่นเดียวกันกับลักษณะเด่นชุด F2_dF5 ซึ่งประกอบด้วยความชันของความถี่มูลฐานที่ตำแหน่ง 0%, 25%, 50%, 75% และ 100% ของระยะเวลาทั้งหมด และค่าความถี่มูลฐานที่ตำแหน่งเริ่มต้น และตำแหน่งสิ้นสุด แต่จะมีการเพิ่มลักษณะเด่นอีก 3 ได้แก่ค่าความถี่มูลฐานที่ตำแหน่ง 25%, 50% และ 75% ของระยะเวลาทั้งหมดของความถี่มูลของเสียงวรรณยุกต์ภาษาไทยแต่ละตัวส่งผลให้ในชุดที่สามมีลักษณะเด่นทั้งหมด 10 ตัวโดยตั้งชื่อลักษณะเด่นชุดนี้ว่า “F5_dF5”

ดังนั้นสามารถสรุปชุดลักษณะเด่นสำหรับตัวแยกประเภททั้งสามชุดได้แก่ ชุด PRC, ชุด F2_dF5 และชุด F5_dF5 ได้ตามตารางที่ 4.6

ตารางที่ 4.6 ชุดลักษณะเด่นสำหรับตัวแยกประเภท

ชื่อชุดลักษณะเด่น	ลักษณะเด่น
PRC	a, b, c, d
F2_dF5	$F_0, F_{100},$ $dF_0, dF_{25}, dF_{50}, dF_{75}, dF_{100}$
F5_dF5	$F_0, F_{25}, F_{50}, F_{75}, F_{100}$ $dF_0, dF_{25}, dF_{50}, dF_{75}, dF_{100}$

1.2. ชุดลักษณะเด่นสำหรับตัวแยกประเภทที่มีความเป็นลำดับ

เนื่องจากตัวแยกประเภทที่มีความเป็นลำดับสามารถจำแนกข้อมูลเสียงซึ่งเป็นอนุกรมเวลาได้โดยไม่ต้องทำการแปลงลักษณะเด่นเป็นเวกเตอร์ที่มีขนาดเฉพาะ ส่งผลให้การเลือกลักษณะเด่นสามารถทำได้ง่ายขึ้น สำหรับชุดลักษณะเด่นสำหรับตัวแยกประเภทที่มีความเป็นลำดับ จะใช้ลักษณะเด่นสองชุดได้แก่ ลักษณะเด่นชุดที่หนึ่งประกอบด้วยค่าความถี่มูลฐานและ

ผลต่างของความถี่มูลฐาน (delta) ซึ่งผลต่างของความถี่มูลฐานสามารถคำนวณได้ตามสมการ 4.10 และตั้งชื่อลักษณะเด่นชุดนี้คือ “F_dF”

$$d_t = \begin{cases} \frac{f_{t+\theta} - f_{t-\theta}}{2\theta}, & \theta < t < T - \theta \\ f_{t+1} - f_t, & t < \theta \\ f_t - f_{t-1}, & t \geq T - \theta \end{cases} \quad (4.10)$$

โดย d_t คือผลต่างของความถี่มูลฐานที่เวลา t ใดๆ และ θ คือ ระยะระหว่างความถี่มูลฐานทั้งสองตำแหน่ง และ f_t คือความถี่มูลฐานที่เวลา t ใดๆ T คือความยาวของจำนวนข้อมูลความถี่มูลฐานต่อหนึ่งเสียงวรรณยุกต์ สำหรับลักษณะเด่นชุดที่สอง ประกอบด้วยค่าความถี่มูลฐานและผลต่างของความถี่มูลฐาน (delta) ซึ่งผลต่างของความถี่มูลฐานสามารถคำนวณเช่นเดียวกันกับลักษณะเด่นชุด F_dF แต่มีการเพิ่มลักษณะเด่นเพิ่มเติมคือผลต่างของผลต่างของความถี่มูลฐาน หรือความเร่ง (Acceleration) ซึ่งสามารถคำนวณได้จากสมการ 4.11

$$a_t = \begin{cases} \frac{d_{t+\theta} - d_{t-\theta}}{2\theta}, & \theta < t < T - \theta \\ d_{t+1} - d_t, & t < \theta \\ d_t - d_{t-1}, & t \geq T - \theta \end{cases} \quad (4.11)$$

โดย a_t คือความเร่งของความถี่มูลฐานที่เวลา t ใดๆ และ θ คือ ระยะระหว่างความถี่มูลฐานทั้งสองตำแหน่ง และ d_t คือผลต่างของความถี่มูลฐานที่เวลา t ใดๆ ซึ่งคำนวณได้จากสมการ 4.10 T คือความยาวของจำนวนข้อมูลความถี่มูลฐานต่อหนึ่งเสียงวรรณยุกต์ ซึ่งจากในงานวิจัยรู้จำเสียงพูดพบว่า การเพิ่มความเร่งเข้าของค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมลเข้าไปในแบบจำลองฮิดเดนมาร์คอฟพบว่า ได้ความถูกต้องและแม่นยำที่สูงขึ้น อีกทั้งในการรู้จำเสียงวรรณยุกต์ในภาษาที่มีเสียงวรรณยุกต์ที่ไม่ใช่ภาษาไทย [28, 29] ก็ได้มีการนำความเร่งมาประยุกต์ในการจำแนกเสียงวรรณยุกต์ด้วย ซึ่งจะตั้งชื่อลักษณะเด่นชุดนี้ว่า “F_dF_aF”

2. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด

ในส่วนของ การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดจะประกอบไปด้วย การจัดเตรียมการทดลอง, ผลการทดลอง, วิเคราะห์และอภิปรายผล

2.1. การจัดเตรียมการทดลอง

สำหรับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดจะใช้ตัวแยกประเภทได้แก่ โครงข่ายประสาทเทียม และแบบจำลองฮิดเดนมาร์คอฟ เป็นระบบอ้างอิงตามการศึกษาของ Thubthong [6] Tan [7] และ Tungthangthum [8] ตามลำดับ นอกจากนี้ในการศึกษานี้ได้นำแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์มาใช้เป็นตัวแยกประเภทของระบบที่จะเสนอโดยตัวแยกประเภทแต่ละตัวได้มีกำหนดค่าพารามิเตอร์ดังต่อไปนี้

สำหรับโครงข่ายประสาทเทียมได้นำคลังโปรแกรม (Library) สำเร็จรูปที่ชื่อ โครงข่ายประสาทเทียมรวดเร็ว (Fast Artificial Neural Network: FANN) [45] มาใช้ในการเรียนรู้โครงข่ายประสาทเทียม และได้กำหนดค่าพารามิเตอร์ของโปรแกรมตาม [24] โดยเลือกอัลกอริทึมการแพร่ย้อนกลับเป็นอัลกอริทึมในการเรียนรู้ของเครื่อง และกำหนดจำนวนรอบในการคำนวณที่ 2,000 รอบ ด้วยอัตราในการเรียนรู้ที่ 0.0001 และค่าโมเมนตัมเท่ากับ 0.9 และในการกำหนดค่าน้ำหนักเริ่มต้นให้สุ่มค่าอยู่ในช่วงระหว่าง -1.0 ถึง 1.0 และกำหนดให้มีจำนวนบัพ (node) ของข้อมูลขาเข้าตามจำนวนลักษณะเด่นในแต่ละชุดลักษณะเด่น ซึ่งสำหรับโครงข่ายประสาทเทียมมีชุดลักษณะเด่นได้แก่ PCR, F2_dF5 และ F5_dF5 ซึ่งมีลักษณะเด่นจำนวน 4, 7 และ 10 ตามลำดับ ซึ่งในแต่ละชุดจะมีการทำให้เป็นบรรทัดฐานสองแบบได้แก่ การทำให้เป็นบรรทัดฐานโดยค่าเฉลี่ย และการทำให้เป็นบรรทัดฐานโดยค่าคะแนนมาตรฐาน ดังนั้นในการทดลองด้วยตัวจำแนกนี้จะมีการทดลองโดยไม่มีการทำให้เป็นบรรทัดฐาน และการทำให้เป็นบรรทัดฐานอีก 2 แบบดังที่กล่าวอีกทั้งในการทดลองยังมีการเปลี่ยนสเกลความถี่ด้วย โดยมีการเปลี่ยนเป็น สเกลความถี่กึ่งวรรณยุกต์ และ สเกลความถี่อัตราฮาร์ปี นอกจากนี้ในชั้นประสาทที่ซ่อนอยู่ (Hidden Layer) ได้กำหนดจำนวนฮิดเดนบัพ (Hidden Node) ไว้ที่ 10 บัพ และสำหรับข้อมูลขาออกได้กำหนดจำนวนบัพไว้ทั้งสิ้น 5 บัพตามจำนวนเสียงวรรณยุกต์ในภาษาไทยทั้ง 5 เสียง

สำหรับแบบจำลองฮิดเดนมาร์คอฟได้นำ เซทที่เค (Hidden Markov Model Toolkit: HTK) [46] ในการเรียนรู้แบบจำลอง โดยกำหนดแบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองฮิดเดนมาร์คอฟ จากซ้ายไปขวา 5 สถานะ และเรียนรู้แบบจำลองเสียงวรรณยุกต์ในรูปแบบไม่ขึ้นกับบริบท (Context-Independent) โดยมีจำนวนหน่วยเสียง 5 หน่วยเสียงตามเสียงวรรณยุกต์ในภาษาไทย โดยชุดลักษณะเด่นที่เลือกใช้เป็นชุดลักษณะเด่น F_dF และ F_dF_aF ซึ่งมีขนาดของเวกเตอร์ลักษณะเด่น 2 และ 3 ตามลำดับ และกำหนดส่วนประกอบของเกาส์เซียนที่ 1 มิกเจอร์ และเป็นเมทริกซ์แปรปรวนทแยงมุม (Diagonal Covariance Matrix) โดยลักษณะเด่นทั้งสองชุดได้มีการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน อีกทั้งยังมีการเปลี่ยนสเกลความถี่เป็นสเกลความถี่เฮิรตซ์, สเกลความถี่กึ่งวรรณยุกต์ และ สเกลความถี่อัตราอาร์บีด้วย

ในการเรียนรู้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ได้ใช้คลังโปรแกรมสำเร็จรูปที่เอชซีอาร์เอฟ (Hidden Conditional Random Fields Library: hCRF Lib) [47] ในการเรียนรู้ โดยเลือกแบบจำลองของคลังโปรแกรมเอชซีอาร์เอฟเป็นแบบจำลองเกาส์เซียนฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ซึ่งอิมพลีเมนต์ตามแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่เสนอโดย Gunawardana [10] ในการเลือกใช้แบบจำลองนี้ได้มีการปรับพารามิเตอร์ต่างๆ โดยเลือกอัลกอริทึมในการเรียนรู้แบบจำลองเป็น Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) และมีการกำหนดแอสสองแคช (L2 cache) ด้วย และในตัวแบบจำลองกำหนดให้มีสถานะฮิดเดนภายในทั้ง 3 สถานะ และกำหนดกรอบหน้าต่าง หรือวินโดวส์ เป็น 0 นอกจากนี้มีการกำหนดค่าน้ำหนักเริ่มต้นให้สุ่มค่าอยู่ในช่วงระหว่าง -1.0 ถึง 1.0 และเริ่มสถานะแบบจำลองด้วยเกาส์เซียน ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดได้กำหนดจำนวนในการเรียนรู้ไว้ที่ 500 รอบและใช้ลักษณะเด่นชุดเด่น F_dF และ F_dF_aF ซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานในการเรียนรู้แบบจำลอง และทดลองบนสเกลความถี่เฮิรตซ์, สเกลความถี่กึ่งวรรณยุกต์ และ สเกลความถี่อัตราอาร์บี

นอกจากนี้ในการทดลองได้ใช้วิธีการตรวจสอบไขว้ (K-fold Cross Validation) โดยได้กำหนดค่า k เท่ากับ 6 สำหรับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด ซึ่งแบ่งข้อมูลเป็น 6 ส่วนโดยการสุ่ม โดยในแต่ละกลุ่มจะประกอบด้วยผู้พูดเพศชายหนึ่งคน และเพศหญิงหนึ่งคน รวมทั้งหมด 6 กลุ่ม โดยข้อมูล 5 กลุ่มจะถูกนำไปเรียนรู้ และข้อมูล

อีกกลุ่มจะถูกใช้ในการตรวจสอบ โดยกระบวนการตรวจสอบจะดำเนินไปเรื่อยๆ จนกระทั่งข้อมูล ทั้ง 6 กลุ่มโดนตรวจสอบเรียบร้อยแล้ว

2.2. ผลการทดลอง

ในส่วนนี้จะเสนอผลการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน สำหรับคำโดดซึ่งใช้ตัวแยกประเภท และการกำหนดค่าพารามิเตอร์ต่างๆ ตามที่เสนอไปโดยใน ตารางที่ 4.7 แสดงผลความแม่นยำในการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน สำหรับคำโดด

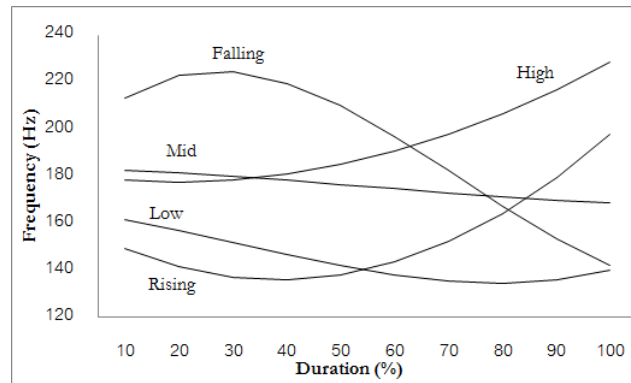
ตารางที่ 4.7 ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด

ตัวแยกประเภท	ชุดลักษณะเด่น	เทคนิคการทำให้เป็นบรรทัดฐานด้วย	ความแม่นยำ (%)		
			สเกลความถี่		
			เฮิร์ตซ์	กึ่งวรรณยุกต์	อัตราอีอาร์บี
โครงข่ายประสาทเทียม	PCR	-	69.77	65.91	67.95
	F2_dF5		57.65	77.27	80.68
	F5_dF5		63.11	81.14	81.29
	F2_dF5	ค่าเฉลี่ย	91.67	91.14	92.12
	F5_dF5		92.88	92.12	93.11
	F2_dF5	คะแนนมาตรฐาน	91.74	91.89	92.27
	F5_dF5		92.95	93.11	93.48
แบบจำลองฮิดเดนมาร์คคอฟ	F_dF	คะแนน	85.45	85.45	85.91
	F_dF_aF	มาตรฐาน	86.21	86.36	86.29
แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์	F_dF	คะแนน	92.88	93.03	93.48
	F_dF_aF	มาตรฐาน	93.33	94.09	94.24

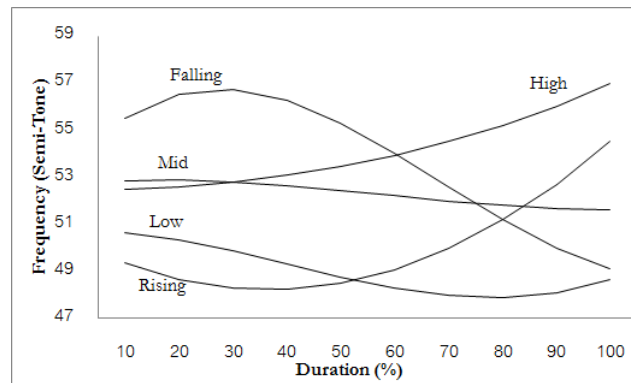
2.3. วิเคราะห์และอภิปรายผล

เมื่อพิจารณารูปร่างของความถี่มูลฐาน ผลต่างของความถี่มูล และความเร่งของความถี่มูลฐานในสเกลความถี่ที่ต่างกันดังแสดงในภาพที่ 4.3 ถึง 4.11 พบว่าลักษณะรูปร่างของความถี่มูลฐาน ผลต่างของความถี่มูล และความเร่งของความถี่มูลฐานในสเกลความถี่เฮิรตซ์, กิ่งวรรณยุกต์ และอัตราฮาร์โมนีของฐานข้อมูลเสียงคำโดดมีรูปร่างลักษณะที่คล้ายกัน แต่กลับให้ผลการจำแนกเสียงวรรณยุกต์ที่แตกต่างกันเป็นอย่างมาก

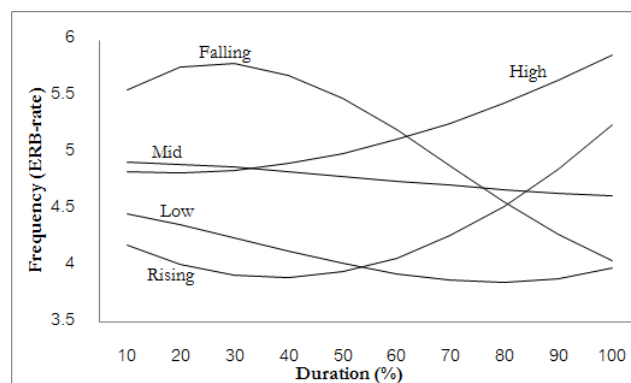
เมื่อพิจารณาผลการทดลองเมื่อใช้สเกลความถี่ที่ต่างกันในตัวแยกประเภทโครงข่ายประสาทเทียมที่ไม่มีการทำให้เป็นบรรทัดฐานพบว่าความแม่นยำในการจำแนกเสียงวรรณยุกต์มีความแตกต่างกันเป็นอย่างมาก และพบว่าสเกลความถี่อัตราฮาร์โมนีให้ความถูกต้องในการจำแนกเสียงสูงที่สุด เมื่อพิจารณาข้อมูลความถี่มูลฐานในแต่ละสเกลของแต่ละเสียงวรรณยุกต์ที่แสดงในตารางที่ 4.8 ถึง 4.12 พบว่าค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของสเกลความถี่อัตราฮาร์โมนีมีค่าน้อยที่สุดในทุกๆ ตาราง เมื่อเทียบกับอีกสองสเกล และพบว่าสเกลความถี่กิ่งวรรณยุกต์มีค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของความถี่มูลฐานเป็นลำดับรองลงมา และสเกลความถี่เฮิรตซ์มีค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานที่มากที่สุด ซึ่งจากข้อมูลดังกล่าวเมื่อพิจารณาจากมุมมองของวิธีการเรียนรู้ของเครื่องพบว่าลักษณะเด่นที่มีขนาดใหญ่กว่าส่งผลให้การคำนวณฟังก์ชันต้นทุน (Cost Function) มีความซับซ้อน และช้ากว่าการคำนวณฟังก์ชันต้นทุนที่ลักษณะเด่นมีขนาดเล็ก [48] ดังนั้นในการทดลองจึงพบว่าสเกลความถี่อัตราฮาร์โมนีให้ความถูกต้องในการจำแนกเสียงสูงที่สุด ซึ่งสอดคล้องกับเทคนิคของการเรียนรู้ด้วยเครื่อง



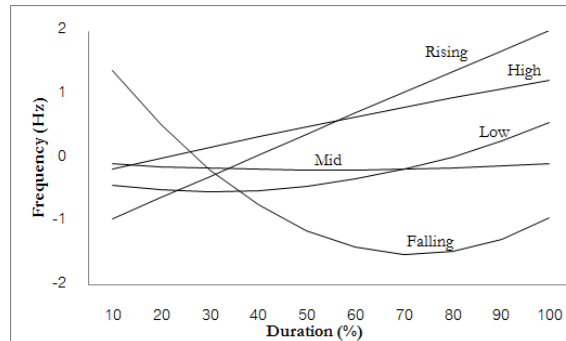
ภาพที่ 4.3 รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิรตซ์ของฐานข้อมูลเสียงคำโดด



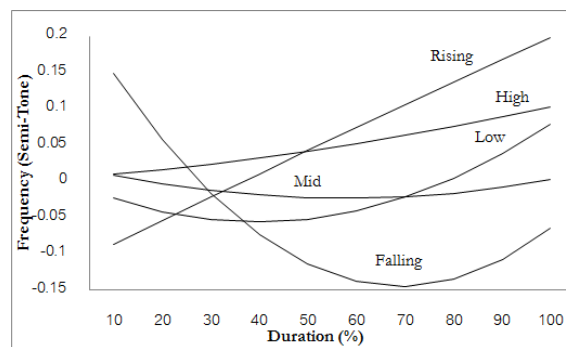
ภาพที่ 4.4 รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของฐานข้อมูลเสียงคำโดด



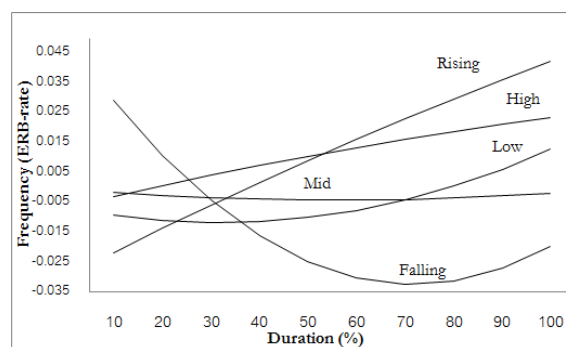
ภาพที่ 4.5 รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอีอาร์บีของฐานข้อมูลเสียงคำโดด



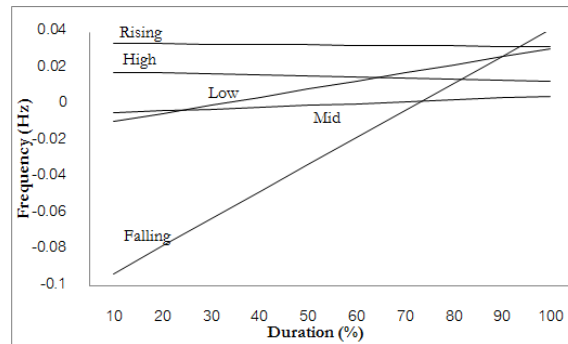
ภาพที่ 4.6 รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิร์ตซ์ของฐานข้อมูลเสียงคำโดด



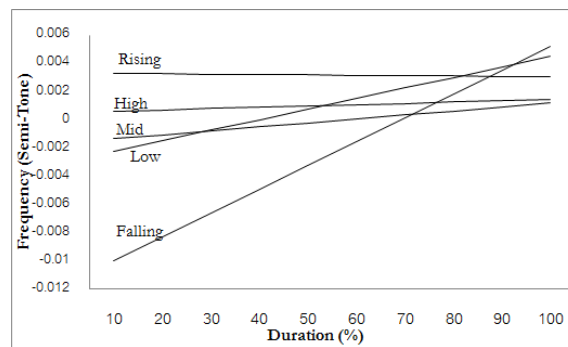
ภาพที่ 4.7 รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของฐานข้อมูลเสียงคำโดด



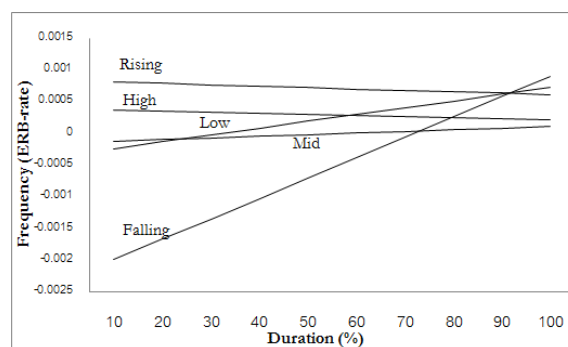
ภาพที่ 4.8 รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอีอาร์บีของฐานข้อมูลเสียงคำโดด



ภาพที่ 4.9 รูปร่างความถี่ของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิรตซ์ของฐานข้อมูลเสียงคำโดด



ภาพที่ 4.10 รูปร่างความถี่ของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของฐานข้อมูลเสียงคำโดด



ภาพที่ 4.11 รูปร่างความถี่ของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอีอาร์บีของฐานข้อมูลเสียงคำโดด

ตารางที่ 4.8 การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ สามัญ สำหรับคลังเสียงพูดคำโดด

ช่วงเวลาที่ย่อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	182.59	54.71	52.84	5.38	4.92	1.19
N = 20	181.48	54.19	52.87	5.20	4.90	1.17
N = 40	178.33	53.41	52.64	5.21	4.84	1.17
N = 60	174.62	52.33	52.21	5.25	4.75	1.15
N = 80	171.17	51.92	51.80	5.32	4.68	1.15
N = 100	168.80	55.22	51.64	5.55	4.62	1.21

ตารางที่ 4.9 การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ เอก สำหรับคลังเสียงพูดคำโดด

ช่วงเวลาที่ย่อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	161.73	50.05	50.67	5.40	4.46	1.13
N = 20	157.03	48.85	50.34	5.32	4.36	1.11
N = 40	146.68	43.74	49.31	5.05	4.13	1.01
N = 60	138.03	37.68	48.29	4.56	3.93	0.89
N = 80	134.64	34.60	47.88	4.24	3.85	0.82
N = 100	140.10	44.66	48.67	5.10	3.98	1.01

ตารางที่ 4.10 การกระจายตัวของข้อมูลความถี่มาตรฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ โท สำหรับคลังเสียงพูดคำโดด

ช่วงเวลาที่ย่อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	213.36	60.10	55.51	5.02	5.56	1.22
N = 20	222.72	62.97	56.53	4.89	5.76	1.26
N = 40	219.51	67.38	56.26	5.41	5.69	1.36
N = 60	197.03	66.04	54.03	5.85	5.21	1.38
N = 80	167.23	54.98	51.19	5.54	4.57	1.21
N = 100	142.03	42.21	49.09	4.91	4.04	0.97

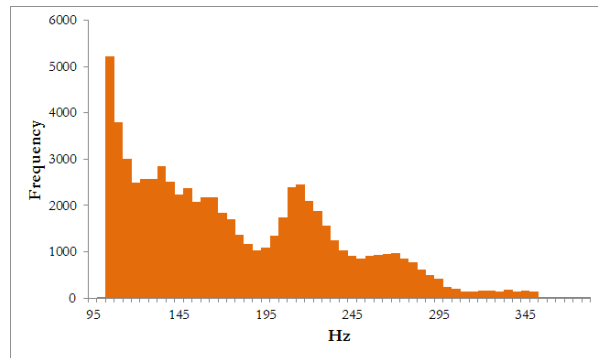
ตารางที่ 4.11 การกระจายตัวของข้อมูลความถี่มาตรฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ ตรี สำหรับคลังเสียงพูดคำโดด

ช่วงเวลาที่ย่อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	178.31	49.95	52.48	5.11	4.84	1.10
N = 20	177.42	48.50	52.61	4.77	4.83	1.06
N = 40	180.82	47.14	53.10	4.55	4.91	1.03
N = 60	190.76	48.80	53.94	4.47	5.12	1.04
N = 80	206.80	55.85	55.22	4.68	5.44	1.14
N = 100	228.52	72.67	57.00	5.47	5.86	1.42

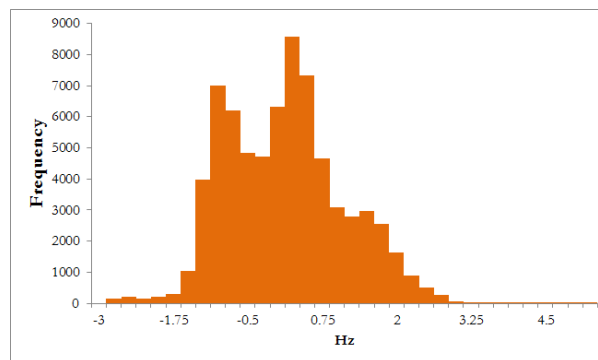
ตารางที่ 4.12 การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ จัตุวา สำหรับคลังเสียงพูดคำโดด

ช่วงเวลาที่ย่อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	149.49	42.97	49.37	5.00	4.19	0.99
N = 20	141.62	39.38	48.66	4.67	4.01	0.92
N = 40	135.94	35.25	48.23	4.35	3.89	0.84
N = 60	143.54	35.43	49.07	4.18	4.07	0.83
N = 80	164.24	42.40	51.19	4.43	4.53	0.95
N = 100	197.89	68.92	54.55	6.03	5.25	1.44

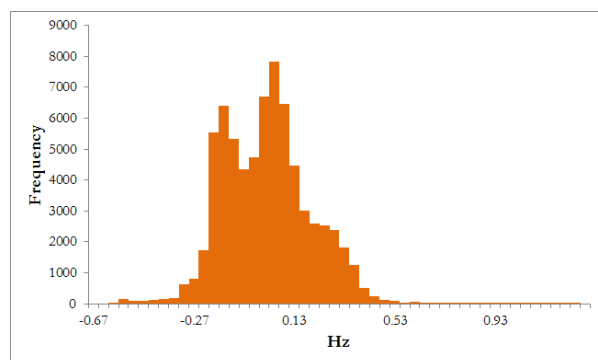
เมื่อพิจารณาฮิสโทแกรม (Histogram) ของความถี่มูลฐานรวมทั้งหมดทุกเสียงวรรณยุกต์
ในภาพที่ 4.12 ถึง 4.20 พบว่าการกระจายตัวของความถี่มูลฐานทั้งสามสเกลมีการกระจายตัวที่สูง
มากเพราะเนื่องมาจากความถี่มูลฐานเป็นสิ่งที่ขึ้นกับผู้พูดแต่ละคน ซึ่งจากการทดลองพบว่าเมื่อไม่
มีการทำให้เป็นบรรทัดฐานเดียวกันจะทำให้ความถูกต้องในการจำแนกลดลงอย่างมีนัยสำคัญ
และส่งผลให้สเกลความถี่อัตราฮิอาร์บีมีความถูกต้องในการจำแนกสูงสุดอย่างเห็นได้ชัด แต่เมื่อมี
การทำให้เป็นบรรทัดฐานเดียวกันพบว่า สเกลความถี่มีบทบาทน้อยลงเป็นอย่างมากเห็นได้จากผล
การทดลองในตารางที่ 4.7 เมื่อเปลี่ยนสเกลความถี่ในการจำแนกที่มีการทำให้เป็นบรรทัดฐาน
เดียวกันผลที่ได้ไม่แตกต่างกันมากซึ่งสอดคล้องกับเทคนิคในการเรียนรู้ด้วยเครื่อง โดยเมื่อมีการ
ทำให้เป็นบรรทัดฐานเดียวกันจะส่งผลให้ปัญหาคำนวณฟังก์ชันต้นทุนหมดไป [48] แต่อย่างไรก็
ตามการเปลี่ยนสเกลจากเฮิร์ตซ์ไปเป็นสเกลความถี่อื่นส่งผลให้ความถูกต้องในการจำแนกเสียง
วรรณยุกต์ดีขึ้นกว่าการใช้ความถี่สเกลเฮิร์ตซ์



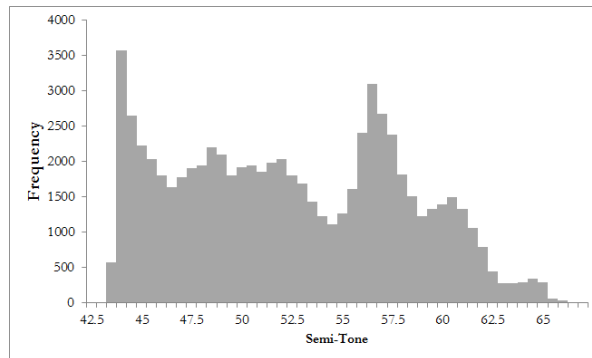
ภาพที่ 4.12 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิรตซ์ซึ่งไม่มีการทำให้เป็นบรรทัดฐานของฐานข้อมูลเสียงคำโดด



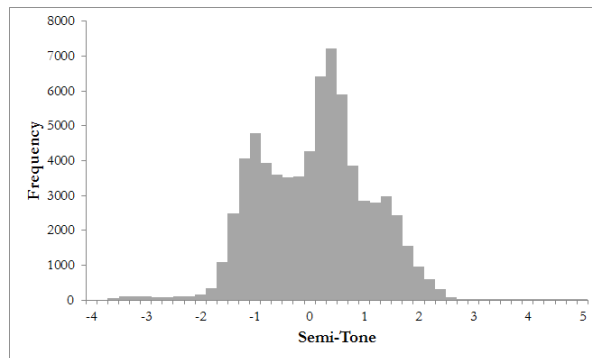
ภาพที่ 4.13 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิรตซ์ซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงคำโดด



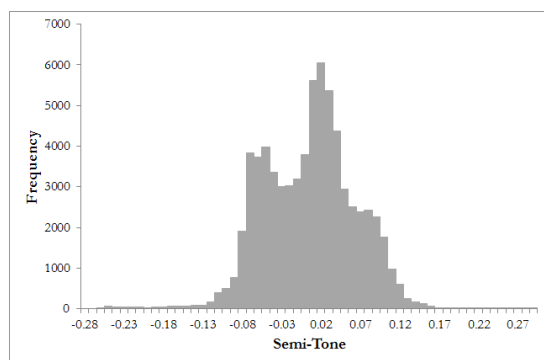
ภาพที่ 4.14 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิรตซ์ซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงคำโดด



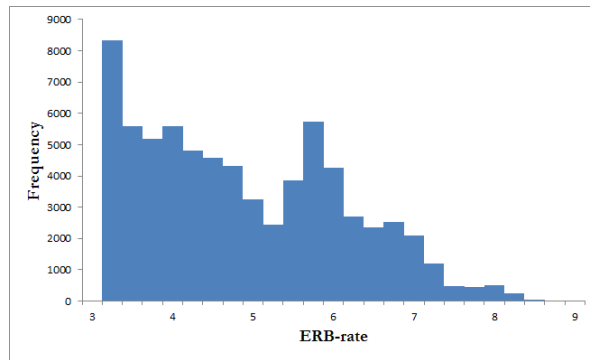
ภาพที่ 4.15 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุคที่ไม่มีกาทำให้เป็นบรรทัดฐานของฐานข้อมูลเสียงคำโดด



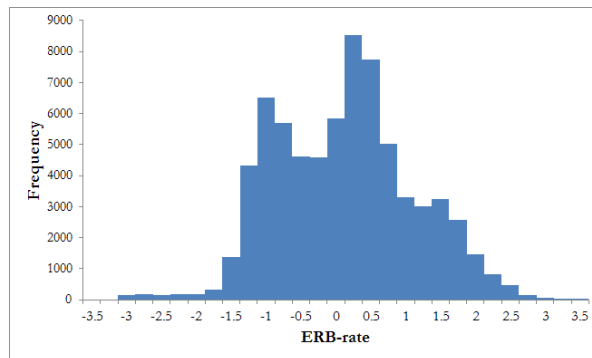
ภาพที่ 4.16 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุคที่มีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงคำโดด



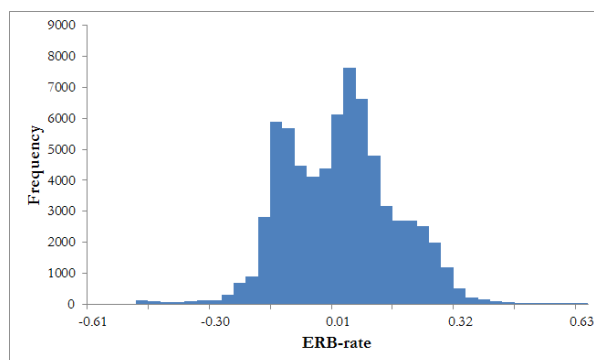
ภาพที่ 4.17 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุคที่มีการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงคำโดด



ภาพที่ 4.18 ฮิสโทแกรมของความถี่มาตรฐานในสเกลความถี่อัตราปีซึ่งไม่มีการทำให้เป็นบรรทัดฐานของฐานข้อมูลเสียงคำโดด



ภาพที่ 4.19 ฮิสโทแกรมของความถี่มาตรฐานในสเกลความถี่อัตราปีซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงคำโดด



ภาพที่ 4.20 ฮิสโทแกรมของความถี่มาตรฐานในสเกลความถี่อัตราปีซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงคำโดด

เมื่อพิจารณาฮิสโทแกรมของความถี่มูลฐานที่มีการทำให้เป็นบรรทัดฐานเดียวกันพบว่า การกระจายตัวมีความเป็นระเบียบมากขึ้น และผลก็สอดคล้องกับผลการจำแนกเสียงวรรณยุกต์ ภาษาไทยที่มีการทำให้เป็นบรรทัดฐานเดียวกัน ซึ่งผลที่ได้มีความถูกต้องมากยิ่งขึ้น และเมื่อพิจารณาฮิสโทแกรมของความถี่มูลฐานที่มีการทำให้เป็นบรรทัดฐานเดียวกันด้วยค่าเฉลี่ย และค่ามาตรฐานพบว่าทั้งสองมีความคล้ายกัน นอกจากนี้เมื่อใช้เทคนิคการทำให้เป็นบรรทัดฐานเดียวกัน ด้วยวิธีทั้งสองส่งผลให้ความถูกต้องในการจำแนกมีค่ามากขึ้น โดยจากการทดลองพบว่า การทำให้เป็นบรรทัดฐานเดียวกันด้วยค่ามาตรฐานจะให้ผลในการจำแนกที่ดีกว่าซึ่งสอดคล้องกับ Thubthong [6] ที่ทำการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยได้ผลดีที่สุดเมื่อใช้การทำให้เป็นบรรทัดฐานเดียวกันด้วยค่ามาตรฐาน ดังนั้นสำหรับการจำแนกเสียงวรรณยุกต์ภาษาไทยพบว่าการเปลี่ยนสเกลความถี่จากเฮิรตซ์ไปเป็นสเกลความถี่อ็อกตาอัวร์บีให้ผลในการรู้จำที่สูงขึ้น และเมื่อรวมกับการใช้เทคนิคการทำให้เป็นบรรทัดฐานเดียวกันด้วยค่ามาตรฐานจะให้ผลในการจำแนกเสียงวรรณยุกต์ดีที่สุด

นอกจากนี้จากการทดลองพบว่าระบบที่นำเสนอโดยใช้แบบจำลองฮิดเดนคอนดิชันนัล แรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_{dF}_{aF} ในสเกลความถี่อ็อกตาอัวร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานให้ความแม่นยำในการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดสูงที่สุดที่ร้อยละ 94.17 ซึ่งสูงกว่าระบบอ้างอิงที่เป็นโครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น $F5_{dF5}$ ในสเกลความถี่อ็อกตาอัวร์บี และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานที่เป็นระบบอ้างอิงที่ดีที่สุด ซึ่งระบบดังกล่าวได้ความแม่นยำที่ร้อยละ 93.48 ซึ่งได้ความผิดพลาดที่ลดลง (Error Reduction) ถึงร้อยละ 10.58 และในตารางที่ 4.13 กับ 4.14 แสดงตารางคอนฟิวชันเมทริกซ์ของทั้งสองระบบ

ตารางที่ 4.13 คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดยใช้
โครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราอาร์บี และใช้เทคนิค
การทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน

จำแนก เสียง	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	244	17	3	0	0	92.42%
เอก	10	240	1	2	11	90.91%
โท	6	0	256	2	0	96.97%
ตรี	1	0	0	256	7	96.97%
จัตวา	5	13	1	7	238	90.15%
ทั้งหมด						93.48%

ตารางที่ 4.14 คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดยใช้
แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราอาร์บี
และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน

จำแนก เสียง	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	246	11	2	4	1	93.18%
เอก	13	240	1	0	10	90.91%
โท	3	0	259	2	0	98.11%
ตรี	0	0	1	258	5	97.73%
จัตวา	0	19	1	3	241	91.29%
ทั้งหมด						94.24%

ซึ่งจากตารางคอนฟิวชันเมทริกซ์ของทั้งสองระบบพบว่าเกิดความสับสนกันระหว่างคู่ของเสียงวรรณยุกต์สามัญกับเสียงวรรณยุกต์เอก และคู่ของเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตวาเป็นคู่ที่มีความสับสนมากที่สุดสองอันดับแรก ส่งผลให้ความถูกต้องในการรู้จำเสียงวรรณยุกต์เอกมีค่าน้อยที่สุด

3. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับเสียงพูดต่อเนื่อง

ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับเสียงพูดต่อเนื่องจะประกอบไปด้วยสามส่วนได้แก่ การจัดเตรียมการทดลอง, ผลการทดลอง, วิเคราะห์และอภิปรายผล เช่นเดียวกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด

3.1. การจัดเตรียมการทดลอง

การจัดเตรียมการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับเสียงพูดต่อเนื่องจะใช้ตัวแยกประเภทสามประเภทได้แก่ โครงข่ายประสาทเทียม แบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ เช่นเดียวกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด โดยได้นำ โครงข่ายประสาทเทียม และแบบจำลองฮิดเดนมาร์คอฟ มาเป็นระบบอ้างอิงตามการศึกษาของ Thubthong [6] Tan [7] และ Tungthangthum [8] ตามลำดับเช่นเดียวกัน และได้มีกำหนดค่าพารามิเตอร์ของตัวแยกประเภทแต่ดังต่อไปนี้

สำหรับการเรียนรู้โครงข่ายประสาทเทียมได้นำคลังโครงข่ายประสาทเทียมรวดเร็วเช่นเดียวกับการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดมาใช้ในการทดลองได้กำหนดค่าพารามิเตอร์ของโปรแกรมตามการทดลองการจำแนกวรรณยุกต์พูดในฐานข้อมูลเสียงต่อเนื่อง [7, 24] ซึ่งมีการกำหนดค่าเริ่มต้นเหมือนกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด โดยเลือกอัลกอริทึมการแพร่ย้อนกลับเป็นอัลกอริทึมในการเรียนรู้ของเครื่อง และกำหนดจำนวนรอบในการคำนวณที่ 2,000 รอบ ด้วยอัตราในการเรียนรู้ที่ 0.0001 และค่าโมเมนตัมเท่ากับ 0.9 และในการกำหนดค่าน้ำหนักเริ่มต้นให้สุ่มค่าอยู่

ในช่วงระหว่าง -1.0 ถึง 1.0 และกำหนดให้มีจำนวนบัพ ของข้อมูลขาเข้าตามจำนวนลักษณะเด่น ในแต่ละชุดลักษณะเด่น ซึ่งสำหรับโครงข่ายประสาทเทียมมีชุดลักษณะเด่นได้แก่ PCR, F2_dF5 และ F5_dF5 ซึ่งมีลักษณะเด่นจำนวน 4, 7 และ 10 ตามลำดับ ในแต่ละชุดจะมีการทำให้เป็นบรรทัดฐานสองแบบได้แก่ การทำให้เป็นบรรทัดฐานโดยค่าเฉลี่ย และการทำให้เป็นบรรทัดฐานโดยค่าคะแนนมาตรฐาน ดังนั้นในการทดลองด้วยตัวจำแนกนี้จะมีการทดลองโดยไม่มี การทำให้เป็นบรรทัดฐาน และการทำให้เป็นบรรทัดฐานอีก 2 แบบดังที่กล่าว อีกทั้งในการทดลองยังมีการเปลี่ยนสเกลความถี่ด้วย โดยมีการเปลี่ยนเป็น สเกลความถี่เฮิรตซ์, สเกลความถี่กึ่งวรรณยุกต์ และ สเกลความถี่อัตรารีอาร์บี นอกจากนี้ในชั้นประสาทที่ซ่อนอยู่ ได้กำหนดจำนวนฮิดเดนบัพ ไว้ที่ 10 บัพ และสำหรับข้อมูลขาออกได้กำหนดจำนวนบัพไว้ทั้งสิ้น 5 บัพตามจำนวนเสียงวรรณยุกต์ใน ภาษาไทยทั้ง 5 เสียง

สำหรับแบบจำลองฮิดเดนมาร์คอฟได้นำเอชทีเคในการเรียนรู้แบบจำลองกำหนดการตั้ง ค่าพื้นฐานเช่นเดียวกันกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน สำหรับคำโดด ซึ่งกำหนดค่าดังต่อไปนี้ โดยกำหนดแบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองฮิดเดนมาร์คอฟจากซ้ายไปขวา 5 สถานะ และเรียนรู้แบบจำลองเสียงวรรณยุกต์ในรูปแบบไม่ขึ้นกับ บริบท โดยมีจำนวนหน่วยเสียงห้าหน่วยเสียงตามเสียงวรรณยุกต์ในภาษาไทย โดยชุดลักษณะเด่น ที่เลือกใช้เป็นชุดลักษณะเด่น F_dF และ F_dF_aF ซึ่งมีขนาดของเวกเตอร์ลักษณะเด่น 2 และ 3 ตามลำดับ และกำหนดส่วนประกอบของเกาส์เซียนที่ หนึ่ง มิกเจอร์และเป็นเมทริกซ์แปรปรวน ทแยงมุมโดยลักษณะเด่นทั้งสองชุดได้มีการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน อีกทั้งยังมีการเปลี่ยนสเกลความถี่เป็นสเกลความถี่เฮิรตซ์, สเกลความถี่กึ่งวรรณยุกต์ และ สเกลความถี่ อัตรารีอาร์บีด้วย

ในการเรียนรู้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ได้ใช้คลังโปรแกรมสำเร็จรูปที่ เอชซีอาร์เอฟเช่นเดียวกันกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน สำหรับคำโดด ในการเลือกแบบจำลองของคลังโปรแกรมเอชซีอาร์เอฟเป็นแบบจำลองเกาส์เซียน ฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ และเลือกอัลกอริทึมในการเรียนรู้แบบจำลองเป็น L-BFGS อีกทั้งมีการกำหนดแอลสองแคชด้วย และในตัวแบบจำลองกำหนดให้มีสถานะฮิดเดนภายในทั้ง 3 สถานะ และกำหนดกรอบหน้าต่าง หรือวินโดวส์ เป็น 0 นอกจากนี้มีการกำหนดค่าน้ำหนักเริ่มต้น

ให้สุ่มค่าอยู่ในช่วงระหว่าง -1.0 ถึง 1.0 และเริ่มสถานะแบบจำลองด้วยเกาส์เซียน ในการทดลอง การจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดได้กำหนดจำนวนในการ เรียนรู้ไว้ที่ 2,000 รอบและใช้ลักษณะเด่นชุดเด่น F_{dF} และ F_{dF_aF} ซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานในการเรียนรู้แบบจำลอง และทดลองบนสเกลความถี่เฮิรตซ์, สเกล ความถี่กึ่งวรรณยุกต์ และ สเกลความถี่อัตราอาร์บี

ในการทดลองได้เลือกใช้วิธีการตรวจสอบไขว้เช่นเดียวกับการทดลองการจำแนกเสียง วรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด โดยได้กำหนดค่า k เท่ากับ 4 สำหรับการ ทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับเสียงพูดต่อเนื่อง ซึ่งแบ่ง ข้อมูลเป็น 4 ส่วนโดยการสุ่ม โดยในแต่ละกลุ่มจะประกอบด้วยผู้พูดเพศชาย 6 คน และผู้พูดเพศ หญิง 6 คน รวมทั้งสิ้น 12 คน โดยข้อมูล 3 กลุ่มจะถูกนำไปเรียนรู้ และข้อมูลอีกกลุ่มจะถูกใช้ในการ ตรวจสอบ โดยกระบวนการตรวจสอบจะดำเนินไปเรื่อยๆ จนกระทั่งข้อมูลทั้ง 4 กลุ่มโดน ตรวจสอบเรียบร้อยแล้ว

3.2. ผลการทดลอง

ในส่วนนี้จะเสนอผลการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน บนฐานข้อมูลเสียงพูดต่อเนื่องโลดส์ ซึ่งใช้ตัวแยกประเภท และการกำหนดค่าพารามิเตอร์ต่างๆ โดยในตารางที่ 4.15 จะเป็นการแสดงผลความแม่นยำจากการทดลอง

ตารางที่ 4.15 ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับเสียงพูดต่อเนื่อง

ตัวแยกประเภท	ชุดลักษณะเด่น	เทคนิคการทำให้เป็นบรรทัดฐานด้วย	ความแม่นยำ (%)		
			สเกลความถี่		
			เฮิรตซ์	กึ่งวรรณยุกต์	อัตราฮิอาร์บี
โครงข่ายประสาทเทียม	PCR	-	47.32	49.17	47.40
	F2_dF5		46.15	55.66	58.09
	F5_dF5		51.71	54.86	58.24
	F2_dF5	ค่าเฉลี่ย	64.95	64.11	64.99
	F5_dF5		65.58	65.26	66.69
	F2_dF5	คะแนน	65.63	66.00	65.89
	F5_dF5	มาตรฐาน	66.20	66.67	67.05
แบบจำลองฮิดเดนมาร์คอฟ	F_dF	คะแนน	65.00	65.11	65.17
	F_dF_aF	มาตรฐาน	65.31	65.58	65.67
แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์	F_dF	คะแนน	69.61	69.67	70.17
	F_dF_aF	มาตรฐาน	70.32	70.62	71.01

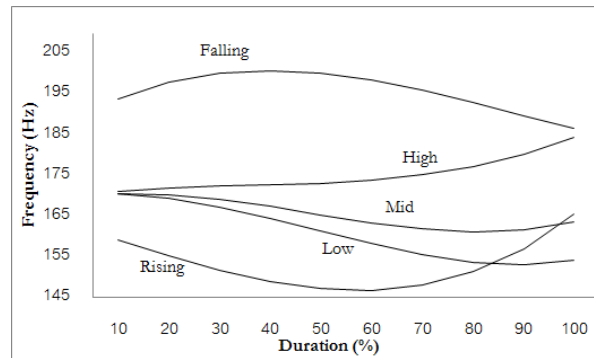
3.3. วิเคราะห์และอภิปรายผล

ในการทดลองพบว่าระบบที่นำเสนอโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราฮิอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานให้ความแม่นยำในการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องได้ความแม่นยำสูงสุดที่ร้อยละ 71.4 ซึ่งเมื่อเทียบกับระบบอ้างอิงที่เป็นโครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราฮิอาร์บี และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานที่เป็นระบบอ้างอิงที่ดีที่สุด พบว่าระบบที่นำเสนอให้ความถูกต้องในการการจำแนกสูงกว่าถึงร้อยละ 3.96 ซึ่งเมื่อคำนวณเป็นความผิดพลาดที่ลดลงจะเท่ากับ ร้อยละ 12.02

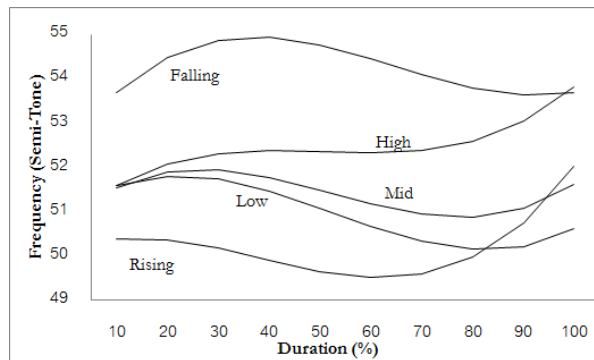
ในการวิเคราะห์การทดลอง เมื่อพิจารณารูปร่างของความถี่มูลฐาน ผลต่างของความถี่มูล และความเร่งของความถี่มูลฐานในสเกลความถี่ที่ต่างกันดังแสดงในภาพที่ 4.21 ถึง 4.29 พบว่าลักษณะรูปร่างของความถี่มูลฐาน ผลต่างของความถี่มูล และความเร่งของความถี่มูลฐานในสเกลความถี่เฮิรตซ์, กิ่งวรรณยุกต์ และอัตราอ็อกตาฟี่ของฐานข้อมูลเสียงพูดต่อเนื่องพบว่ามีรูปร่างลักษณะที่คล้ายกัน แต่กลับให้ผลการจำแนกเสียงวรรณยุกต์ที่ต่างกัน

จะเห็นได้ชัดในการทดลองจำแนกด้วยตัวแยกประเภท โครงข่ายประสาทเทียมเมื่อไม่มีการทำให้เป็นบรรทัดฐาน แต่ความแตกต่างไม่เท่ากับการจำแนกเสียงวรรณยุกต์สำหรับคำโดด ซึ่งเมื่อพิจารณาข้อมูลพบว่าขนาดข้อมูลของสเกลความถี่เฮิรตซ์ และสเกลความถี่กึ่งวรรณยุกต์ในตารางที่ 4.16 - 4.20 พบว่าขนาดของค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานมีขนาดเล็กลงในเสียงพูดต่อเนื่อง ทำให้ค่าของข้อมูลมีขนาดเล็กลง ส่งผลให้ความซับซ้อนในการคำนวณฟังก์ชันต้นทุนลดลง แต่อย่างไรก็ตามสเกลความถี่อ็อกตาฟี่ก็ยังคงให้ความถูกต้องสูงสุดเมื่อไม่มีการทำให้เป็นบรรทัดฐานเนื่องจากค่าของข้อมูลเล็กสุดตามแนวคิดของการเรียนรู้ของเครื่อง [48]

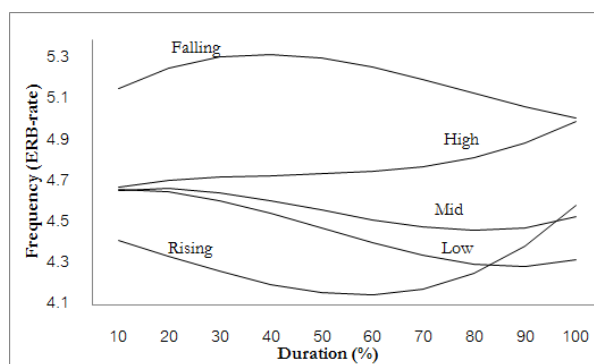
เมื่อพิจารณาฮิสโทแกรมของความถี่มูลฐานพบว่าในก่อนการทำให้เป็นบรรทัดฐานเดียวกันข้อมูลจะถูกแบ่งออกเป็นรูปภูเขาสองรูปซึ่งอนุมานได้ว่ารูปภูเขาที่มีความถี่มูลฐานต่ำเป็นเสียงของกลุ่มผู้ชาย และรูปภูเขาอีกรูปที่มีคสามถี่มูลฐานสูงจะเป็นกลุ่มของเสียงผู้หญิง ดังแสดงในรูป 4.30, 4.33 และ 4.36 ซึ่งลักษณะดังกล่าวส่งผลต่อการรู้จำเสียงพูดที่ไม่ขึ้นกับบุคคล ดังนั้นการทำให้เป็นบรรทัดฐานเดียวกันจะทำให้การกระจายตัวมีลักษณะเป็นรูปภูเขาดังรูปที่ 4.31, 4.32, 4.34, 4.35, 4.37 และ 4.38 ซึ่งมีความเป็นระเบียบมากขึ้นและลดความแตกต่างระหว่างผู้พูดลง ซึ่งจากผลการทดลองก็สอดคล้องกับการกระจายตัวของความถี่มูลฐาน กล่าวคือ ในการจำแนกเสียงวรรณยุกต์มีความถูกต้องสูงขึ้นไม่ว่าจะใช้วิธีการทำให้เป็นบรรทัดฐานแบบ ด้วยค่าเฉลี่ย หรือค่ามาตรฐาน



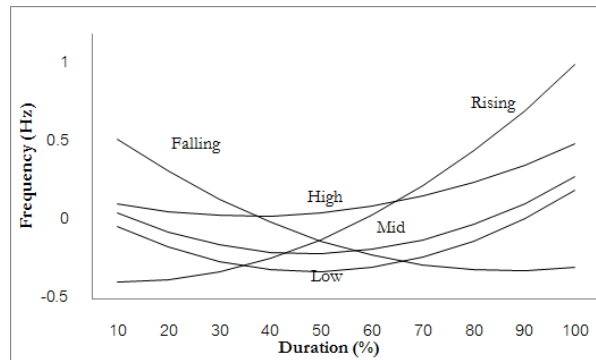
ภาพที่ 4.21 รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิรตซ์ของฐานข้อมูลเสียงพูดต่อเนื่อง



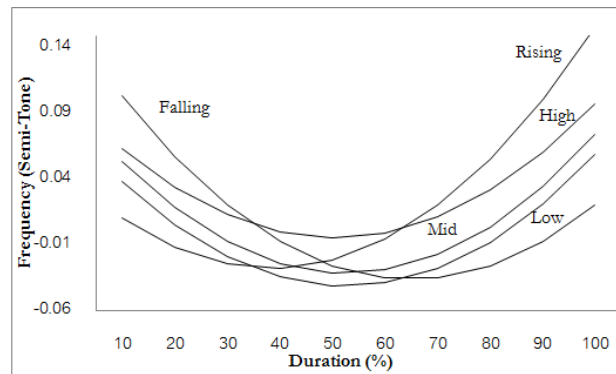
ภาพที่ 4.22 รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ ของฐานข้อมูลเสียงพูดต่อเนื่อง



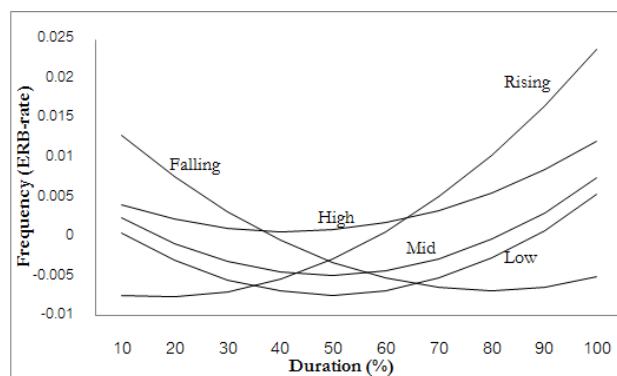
ภาพที่ 4.23 รูปร่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอีอาร์บีของฐานข้อมูลเสียงพูดต่อเนื่อง



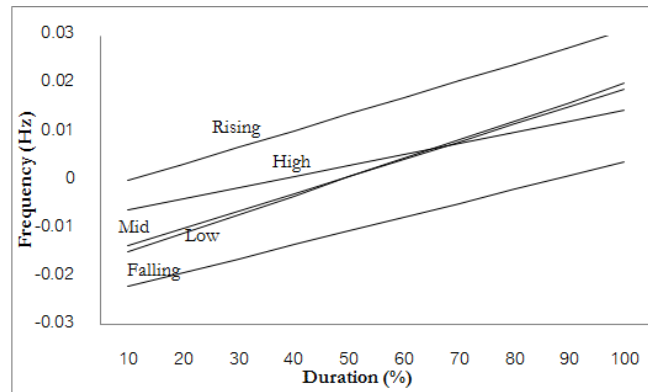
ภาพที่ 4.24 รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิรตซ์ของฐานข้อมูลเสียงพูดต่อเนื่อง



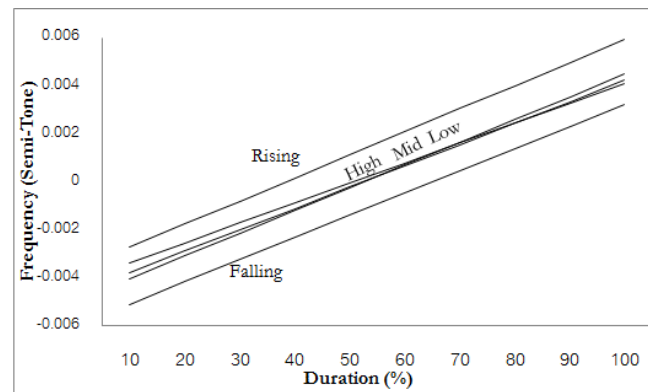
ภาพที่ 4.25 รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของฐานข้อมูลเสียงพูดต่อเนื่อง



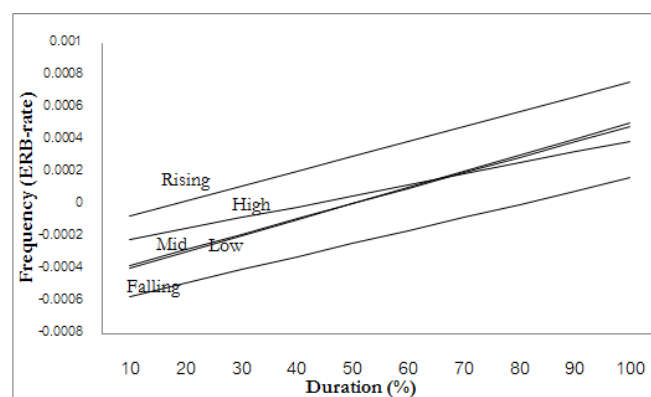
ภาพที่ 4.26 รูปร่างของผลต่างของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราฮาร์ปีของฐานข้อมูลเสียงพูดต่อเนื่อง



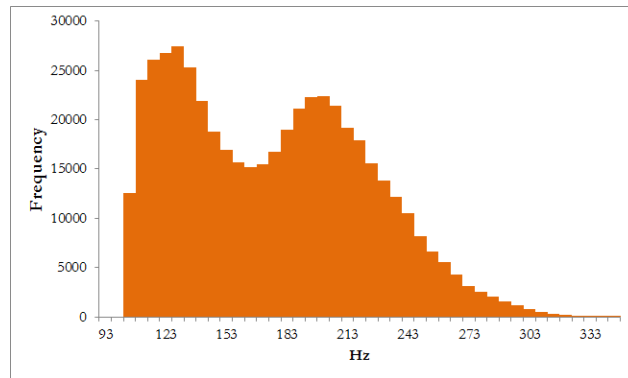
ภาพที่ 4.27 รูปร่างความถี่ของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่เฮิรตซ์ของฐานข้อมูลเสียงพูดต่อเนื่อง



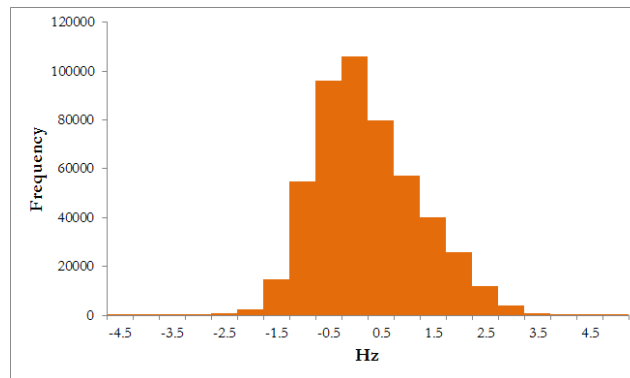
ภาพที่ 4.28 รูปร่างความถี่ของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่กึ่งวรรณยุกต์ของฐานข้อมูลเสียงพูดต่อเนื่อง



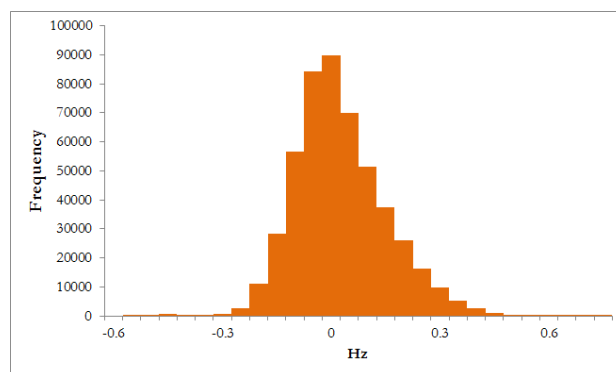
ภาพที่ 4.29 รูปร่างความถี่ของความถี่มูลฐานในสเกลความถี่ในสเกลความถี่อัตราอีอาร์บีของฐานข้อมูลเสียงพูดต่อเนื่อง



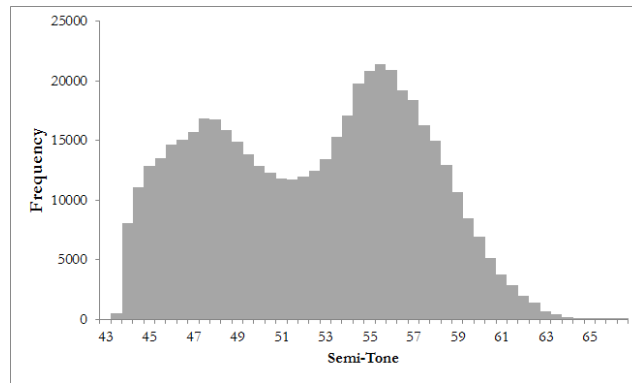
ภาพที่ 4.30 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิรตซ์ซึ่งไม่มีการทำให้เป็นบรรทัดฐานของฐานข้อมูลเสียงพูดต่อเนื่อง



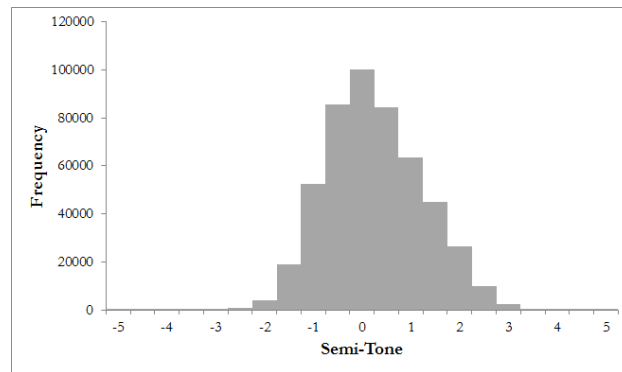
ภาพที่ 4.31 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิรตซ์ซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงพูดต่อเนื่อง



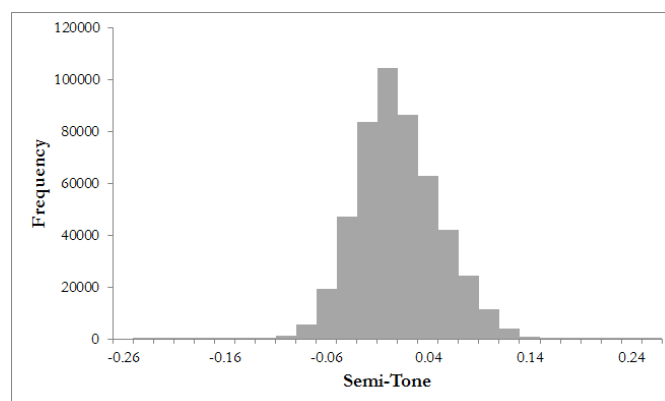
ภาพที่ 4.32 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่เฮิรตซ์ซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงพูดต่อเนื่อง



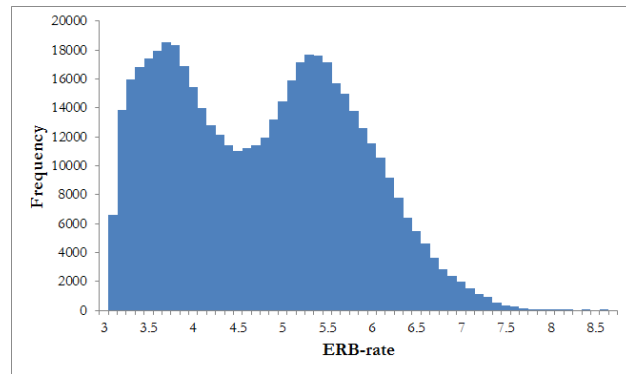
ภาพที่ 4.33 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุคที่ไม่มีทำให้เป็นบรรทัดฐานของฐานข้อมูลเสียงพูดต่อเนื่อง



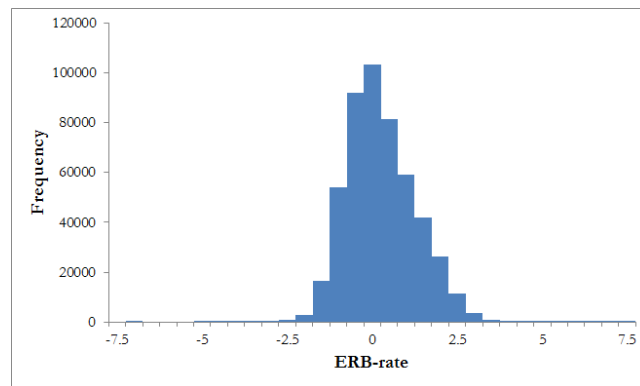
ภาพที่ 4.34 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุคที่มีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงพูดต่อเนื่อง



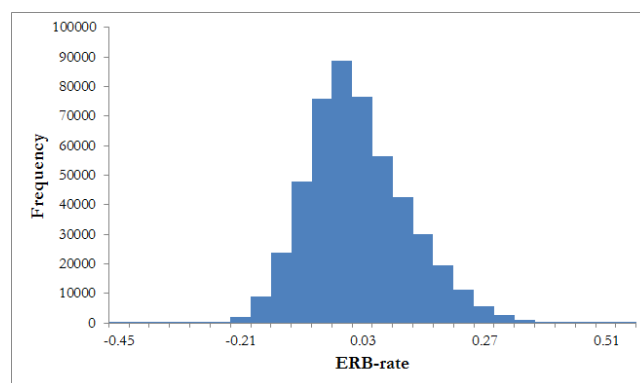
ภาพที่ 4.35 ฮิสโทแกรมของความถี่มูลฐานในสเกลความถี่กึ่งวรรณยุคที่มีการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงพูดต่อเนื่อง



ภาพที่ 4.36 ฮิสโทแกรมของความถี่มาตรฐานในสเกลความถี่อัตราปีซึ่งไม่มีการทำให้เป็นบรรทัดฐานของฐานข้อมูลเสียงพูดต่อเนื่อง



ภาพที่ 4.37 ฮิสโทแกรมของความถี่มาตรฐานในสเกลความถี่อัตราปีซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงพูดต่อเนื่อง



ภาพที่ 4.38 ฮิสโทแกรมของความถี่มาตรฐานในสเกลความถี่อัตราปีซึ่งมีการทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ยของฐานข้อมูลเสียงพูดต่อเนื่อง

ตารางที่ 4.16 การกระจายตัวของข้อมูลความถี่มาตรฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ สามัญ สำหรับคลังเสียงพูดต่อเนื่อง

ช่วงเวลาที่ ร้อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	170.46	46.93	51.54	5.12	4.66	1.05
N = 20	170.33	46.28	51.89	4.78	4.67	1.03
N = 40	167.43	45.22	51.78	4.83	4.61	1.02
N = 60	163.41	43.56	51.19	4.69	4.52	0.99
N = 80	161.15	42.77	50.89	4.68	4.47	0.97
N = 100	163.54	44.75	51.61	5.28	4.53	1.02

ตารางที่ 4.17 การกระจายตัวของข้อมูลความถี่มาตรฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ เอก สำหรับคลังเสียงพูดต่อเนื่อง

ช่วงเวลาที่ ร้อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	170.55	46.95	51.59	5.16	4.67	1.05
N = 20	169.52	45.90	51.81	4.80	4.65	1.02
N = 40	164.53	43.86	51.46	4.74	4.55	0.99
N = 60	158.24	41.53	50.68	4.57	4.40	0.95
N = 80	153.78	40.15	50.16	4.54	4.30	0.92
N = 100	154.24	40.96	50.62	5.03	4.32	0.95

ตารางที่ 4.18 การกระจายตัวของข้อมูลความถี่มาตรฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ โท สำหรับคลังเสียงพูดต่อเนื่อง

ช่วงเวลาที่ ร้อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	193.76	53.82	53.68	5.27	5.16	1.15
N = 20	197.95	53.84	54.48	4.83	5.26	1.14
N = 40	200.88	55.22	54.93	4.98	5.33	1.16
N = 60	198.44	56.45	54.45	5.05	5.26	1.19
N = 80	192.95	58.28	53.79	5.34	5.14	1.24
N = 100	186.71	62.70	53.69	6.41	5.01	1.35

ตารางที่ 4.19 การกระจายตัวของข้อมูลความถี่มาตรฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ ตรี สำหรับคลังเสียงพูดต่อเนื่อง

ช่วงเวลาที่ ร้อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	171.13	46.61	51.59	5.17	4.68	1.04
N = 20	171.99	45.57	52.08	4.70	4.71	1.01
N = 40	172.79	44.69	52.38	4.68	4.74	1.00
N = 60	173.91	44.09	52.32	4.50	4.76	0.98
N = 80	177.19	44.76	52.59	4.51	4.82	0.99
N = 100	184.48	48.85	53.83	5.31	5.00	1.07

ตารางที่ 4.20 การกระจายตัวของข้อมูลความถี่มูลฐานในสเกลต่างๆ ณ เวลาต่างๆ ของเสียง
วรรณยุกต์ จัตุวา สำหรับคลังเสียงพูดต่อเนื่อง

ช่วงเวลาที่ย่อย ละ N ของเวลา ทั้งหมด	สเกลความถี่					
	เฮิร์ตซ์		กึ่งวรรณยุกต์		อัตราฮิอาร์บี	
	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>	\bar{X}	<i>S.D.</i>
N = 0	159.24	42.75	50.38	5.03	4.42	0.98
N = 20	155.38	39.96	50.36	4.50	4.34	0.92
N = 40	149.01	35.91	49.91	4.28	4.20	0.84
N = 60	146.75	33.00	49.52	3.92	4.15	0.77
N = 80	151.37	33.58	49.97	3.90	4.26	0.78
N = 100	165.65	41.35	52.02	4.89	4.59	0.94

เมื่อพิจารณาการกระจายตัวของความถี่มูลฐานที่มีการทำให้เป็นบรรทัดฐานเดียวกันทั้ง 2 วิธี พบว่ารูปร่างการกระจายตัวมีความคล้ายกัน แต่จากการทดลองพบว่าทำให้เป็นบรรทัดฐานเดียวกันด้วยค่ามาตรฐานจะให้ผลในการจำแนกที่ดีกว่าซึ่งสอดคล้องกับ Thubthong [6] ที่ทำการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยได้ผลดีที่สุดเมื่อใช้การทำให้เป็นบรรทัดฐานเดียวกันด้วยค่ามาตรฐาน ดังนั้นสำหรับการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องพบว่าการเปลี่ยนสเกลความถี่จากเฮิร์ตซ์ไปเป็นสเกลความถี่อัตราฮิอาร์บีให้ผลในการรู้จำที่สูงขึ้น และเมื่อรวมกับการใช้เทคนิคการทำให้เป็นบรรทัดฐานเดียวกันด้วยค่ามาตรฐานจะให้ผลในการจำแนกเสียงวรรณยุกต์ที่ดีที่สุด

เมื่อพิจารณาการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้โครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราฮิอาร์บี และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน พบว่าเกิดความสับสนกันระหว่างคู่ของเสียงวรรณยุกต์สามัญกับเสียงวรรณยุกต์เอก และคู่ของเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตุวาเป็นคู่ที่มีความสับสนมากที่สุดสองอันดับแรก ส่งผลให้ความถูกต้องในการรู้จำเสียงวรรณยุกต์เอกมีค่าน้อยที่สุดเช่นเดียวกับการจำแนกเสียงวรรณยุกต์สำหรับค่าโดดดังแสดงในตารางที่ 4.21 สำหรับ

การจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานพบว่าเสียงวรรณยุกต์จัดทำให้ความถูกต้องน้อยที่สุด แทนที่จะเป็นเสียงวรรณยุกต์เอกดังแสดงในตารางที่ 4.22

ตารางที่ 4.21 คอนฟิวชันเมตริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้โครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราอาร์บี และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน

จำแนก เสียง	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	6532	1110	459	392	169	75.41%
เอก	1865	2838	305	131	138	53.78%
โท	392	129	4080	354	11	82.16%
ตรี	975	139	552	2327	209	55.38%
จัตวา	406	320	9	257	1158	53.86%
ทั้งหมด						67.05%

ตารางที่ 4.22 คอนฟิวชันเมตริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน

จำแนก เสียง	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	6557	1075	449	438	143	75.70%
เอก	825	3889	158	212	193	73.70%
โท	548	177	3755	473	13	75.61%
ตรี	751	389	386	2519	157	59.95%
จัตวา	224	471	15	225	1215	56.51%
ทั้งหมด						71.01%

จากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ความถี่มูลฐานทั้งสองการทดลองได้ผลการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดด และการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องพบว่าทั้งสองการทดลองวิธีที่นำเสนอโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราออร์บิและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานให้ผลการทดลองในการจำแนกเสียงวรรณยุกต์ภาษาไทยสูงสุด อีกทั้งยังมีความถูกต้องมากกว่าระบบอ้างอิง [6, 7, 8] ซึ่งใช้โครงข่ายประสาทเทียม และแบบจำลองฮิดเดนมาร์คอฟ

การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมที่ได้จากการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยมนุษย์

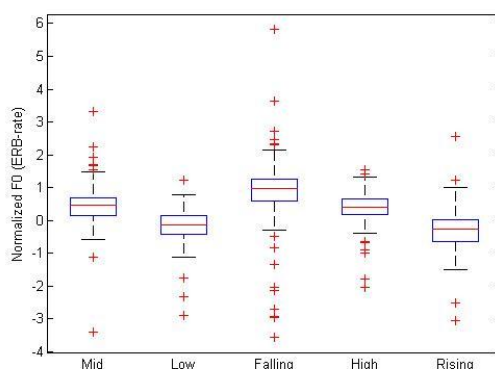
ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมที่ได้จากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยมนุษย์จะเป็นการทำการทดลองต่อจากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน ซึ่งในการทดลองนี้จะประกอบด้วยการทดลองย่อยสองการทดลองเช่นเดียวกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐาน ได้แก่ การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงาน และการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัม

1. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงาน

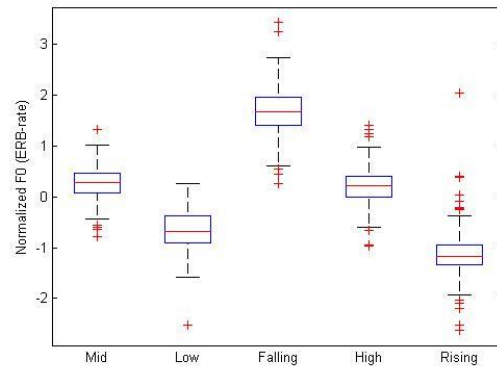
ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงาน มีวัตถุประสงค์เพื่อศึกษาการนำค่าพลังงานมาใช้ในการจำแนกเสียงวรรณยุกต์ในคำโดด โดยตั้งสมมุติฐานว่าค่าพลังงานจะสามารถเพิ่มประสิทธิภาพในการจำแนกเสียงวรรณยุกต์ได้ เนื่องจากการทดลองการรับรู้เสียงวรรณยุกต์พบว่าค่าพลังงานมีส่วนช่วยให้การรับรู้เสียงวรรณยุกต์มีความแม่นยำมากยิ่งขึ้น โดยรายละเอียดในการทดลองประกอบด้วยการวิเคราะห์ลักษณะเด่น, การจัดเตรียมการทดลอง, ผลการทดลองและสรุปผลการทดลอง

1.1. การวิเคราะห์ลักษณะเด่น

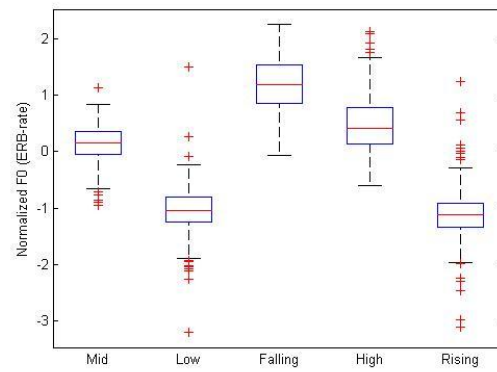
จากฐานข้อมูลเสียงคำโดด เมื่อพิจารณาแผนภาพกล่อง (Box Plot) ของความถี่มูลฐานผลต่างของความถี่มูลฐาน และความเร่งในตำแหน่งเวลาที่ 0, 25, 50, 75 และ 100 ของระยะเวลา ดังแสดงในภาพที่ 4.39 – 4.53 พบว่า แผนภาพกล่องของเสียงวรรณยุกต์เอกจะมีส่วนทับซ้อนกับเสียงวรรณยุกต์สามัญ และเสียงวรรณยุกต์จัตวาเกือบทั้งหมดส่งผลให้เสียงวรรณยุกต์เอกสร้าง ความสับสนในการจำแนกเสียงวรรณยุกต์สามัญกับเสียงวรรณยุกต์จัตวา อีกทั้งส่งผลให้เสียงวรรณยุกต์เอกมีความถูกต้องในการรู้จำแนกเสียงวรรณยุกต์น้อยที่สุด ซึ่งสอดคล้องกับคอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดยใช้โครงข่ายประสาทเทียมที่ใช้ชุดลักษณะเด่น F5_dF5 ในสเกลความถี่อัตราฮิอาร์ปี และใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน และคอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดด โดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF ในสเกลความถี่อัตราฮิอาร์ปีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน ในตารางที่ 4.13 – 4.14



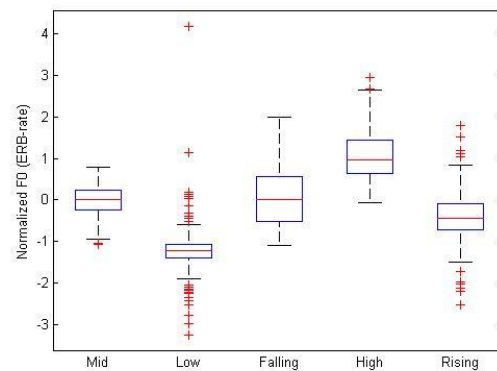
ภาพที่ 4.39 แผนภาพกล่องของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 0



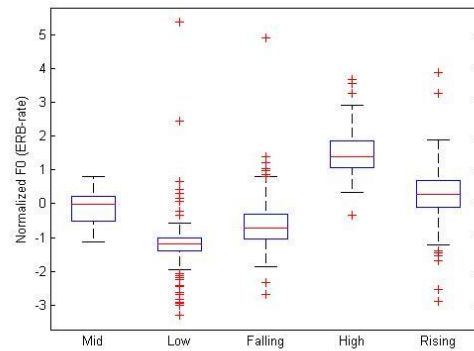
ภาพที่ 4.40 แผนภาพกล่องของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 25



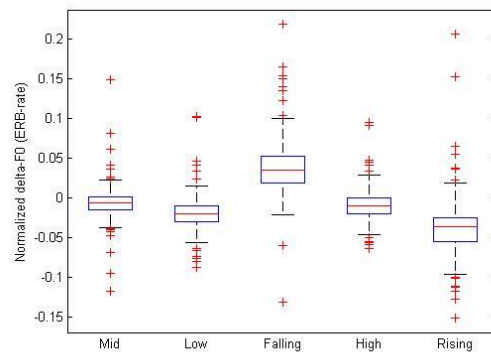
ภาพที่ 4.41 แผนภาพกล่องของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 50



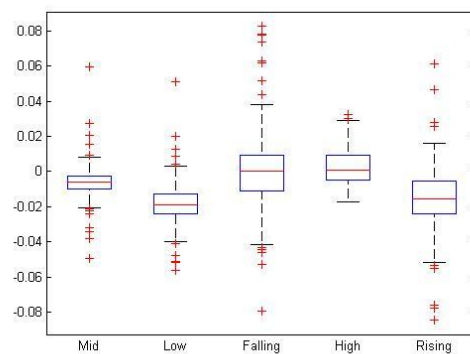
ภาพที่ 4.42 แผนภาพกล่องของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 75



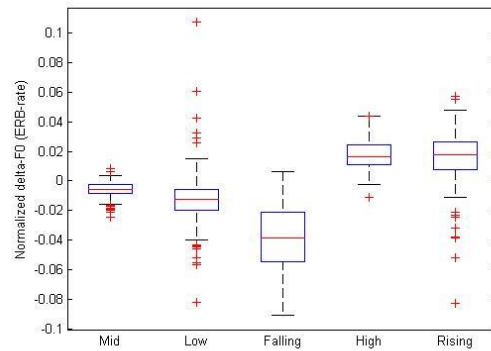
ภาพที่ 4.43 แผนภาพกล่องของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 100



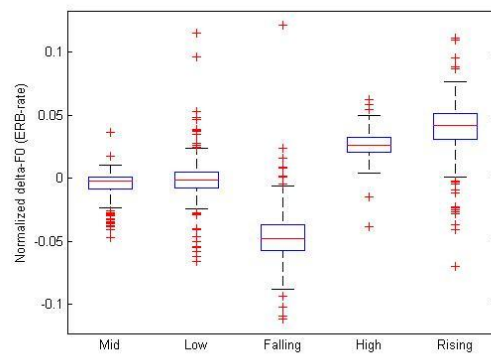
ภาพที่ 4.44 แผนภาพกล่องของผลต่างของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 0



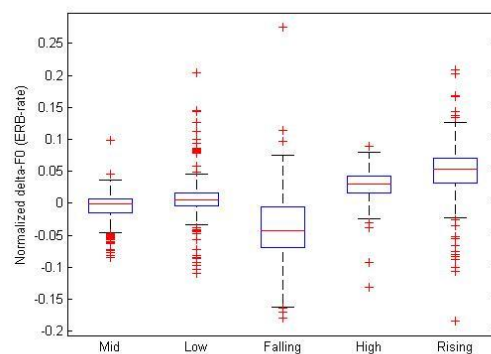
ภาพที่ 4.45 แผนภาพกล่องของผลต่างของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 25



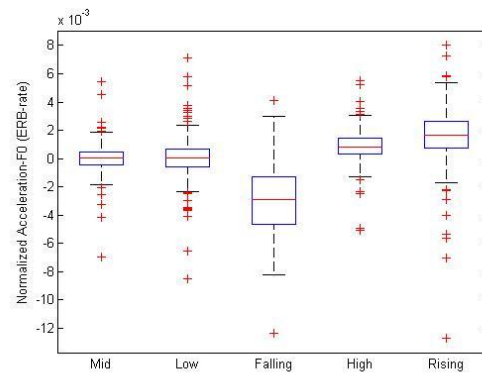
ภาพที่ 4.46 แผนภาพกล่องของผลต่างของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 50



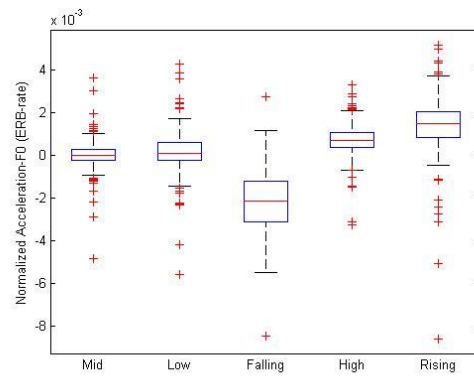
ภาพที่ 4.47 แผนภาพกล่องของผลต่างของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 75



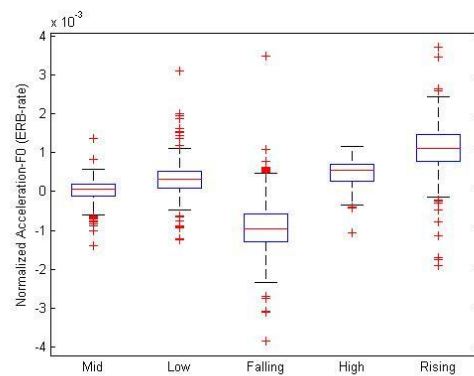
ภาพที่ 4.48 แผนภาพกล่องของผลต่างของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 100



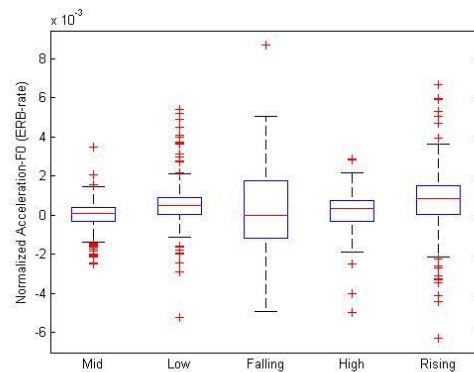
ภาพที่ 4.49 แผนภาพกล่องของความเร่งของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 0



ภาพที่ 4.50 แผนภาพกล่องของความเร่งของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 25

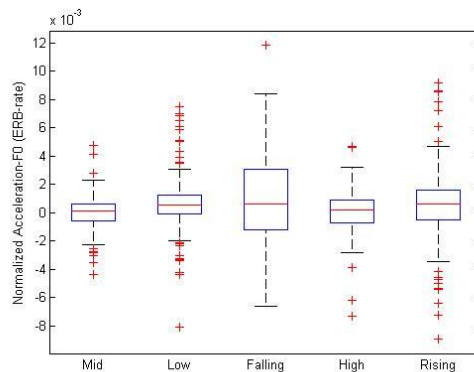


ภาพที่ 4.51 แผนภาพกล่องของความเร่งของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย
ที่ตำแหน่ง 50



ภาพที่ 4.52 แผนภาพกล่องของความเร่งของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย

ที่ตำแหน่ง 75

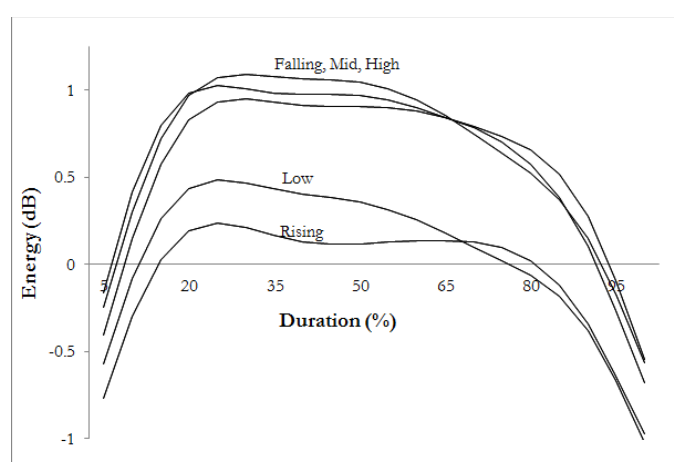


ภาพที่ 4.53 แผนภาพกล่องของความเร่งของความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย

ที่ตำแหน่ง 100

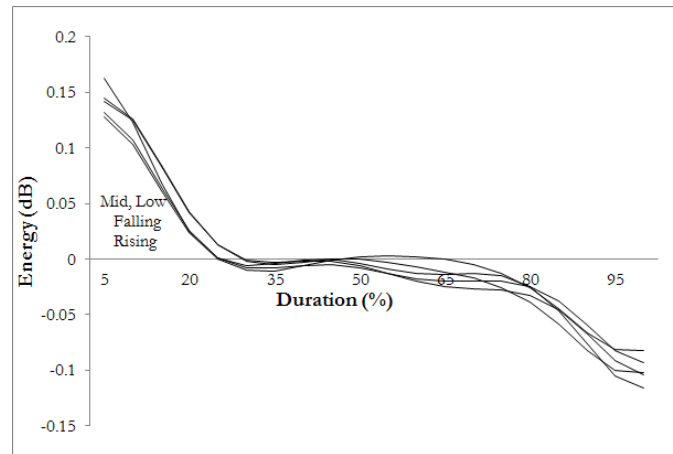
ดังนั้นการหาลักษณะเด่นเพิ่มเติมจึงจำเป็นเพื่อช่วยให้การจำแนกเสียงวรรณยุกต์ภาษาไทยมีประสิทธิภาพมากยิ่งขึ้น โดยเมื่อพิจารณาจากองค์ความรู้พื้นฐานในการรับรู้เสียงวรรณยุกต์ภาษาไทยของผู้ใช้ภาษาไทยเป็นภาษาหลักพบว่ารูปร่างพลังงานมีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ในรูปแบบหนึ่งพยางค์ทำให้ผู้เข้าร่วมทดลองสามารถแยกแยะเสียงวรรณยุกต์ภาษาไทยได้มีประสิทธิภาพขึ้นอย่างมีนัยสำคัญ ดังนั้นในลักษณะเด่นสำหรับคำโดดจะพิจารณาพลังงานเป็นค่าลักษณะเด่นเพิ่มเติมเป็นหลัก เพื่อที่จะใช้ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงานในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดด

เมื่อพิจารณารูปร่างของพลังงานที่ถูกทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของฐานข้อมูลเสียงพูดคำโดดพบว่าค่าพลังงานสามารถช่วยแยกแยะความสับสนของเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์สามัญ และเสียงวรรณยุกต์เอกกับจัตวาได้ดังแสดงในภาพที่ 4.54 ซึ่งตรงตามวัตถุประสงค์ที่ต้องการที่จะช่วยในการพัฒนาการจำแนกเสียงวรรณยุกต์เอก เสียงวรรณยุกต์สามัญ และเสียงวรรณยุกต์จัตวา

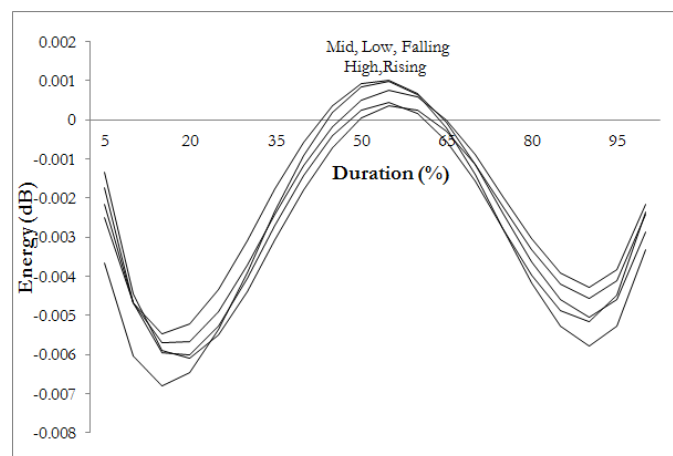


ภาพที่ 4.54 กราฟของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน

แต่เมื่อพิจารณาผลต่างของพลังงานและความเร่งของพลังงานพบว่ารูปร่างและค่าของทั้งสองของทั้งห้าเสียงวรรณยุกต์มีความคล้ายคลึงกันและทับซ้อนกันเป็นอย่างมากดังแสดงในภาพที่ 4.55 – 4.56 ส่งผลให้ในการพิจารณาลักษณะเด่นเพิ่มเติมสำหรับฐานข้อมูลเสียงคำโดดจะใช้เฉพาะค่าพลังงานของข้อมูลเสียง



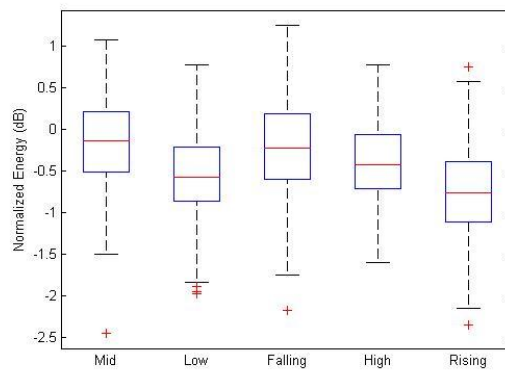
ภาพที่ 4.55 กราฟผลต่างของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐาน
ด้วยค่ามาตรฐาน



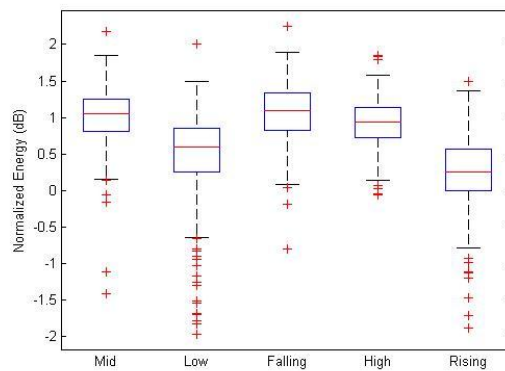
ภาพที่ 4.56 กราฟความเร่งของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐาน
ด้วยค่ามาตรฐาน

นอกจากนี้เมื่อพิจารณาแผนภาพกล่องของพลังงานของแต่ละเสียงวรรณยุกต์ในตำแหน่งเวลาที่ 0, 25, 50, 75 และ 100 ของระยะเวลาดังแสดงในภาพที่ 4.57 – 4.61 ซึ่งจากแผนภาพกล่องจะเห็นได้ว่าค่าเฉลี่ย และกลุ่มข้อมูลของเสียงวรรณยุกต์สามัญ เสียงวรรณยุกต์เอก และเสียงวรรณยุกต์จัตวา มีความแตกต่างกัน ในขณะที่เสียงวรรณยุกต์สามัญ เสียงวรรณยุกต์โท และเสียงวรรณยุกต์ตรี มีความใกล้เคียงกัน ซึ่งเป็นไปตามกราฟของพลังงานของฐานข้อมูลเสียงพูดคำโดดที่ถูกทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน ซึ่งจากจุดนี้จึงตั้งสมมุติฐานได้ว่าค่าพลังงานจะ

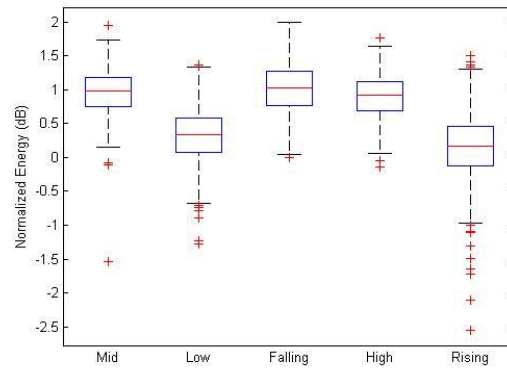
ช่วยในการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับฐานข้อมูลเสียงคำโดดได้ดีขึ้น กล่าวคือจะช่วยลดความสับสนระหว่างคู่เสียงวรรณยุกต์สามัญกับเสียงวรรณยุกต์เอก และ คู่เสียงเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตวาได้



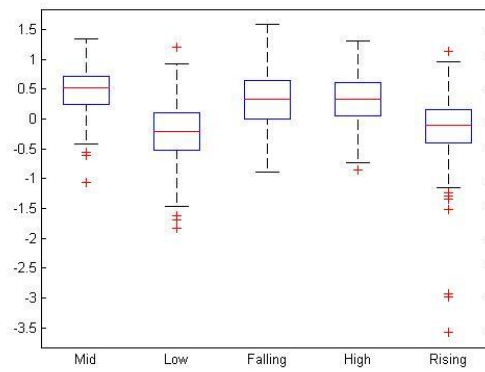
ภาพที่ 4.57 แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 0



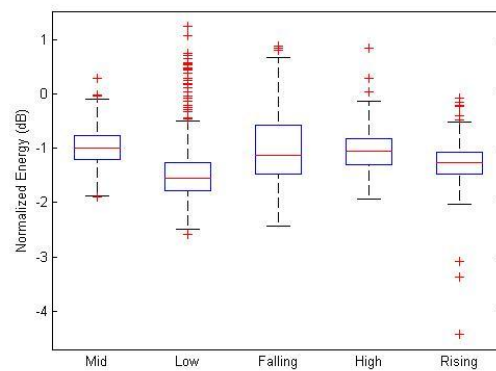
ภาพที่ 4.58 แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 25



ภาพที่ 4.59 แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 50



ภาพที่ 4.60 แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 75



ภาพที่ 4.61 แผนภาพกล่องของพลังงานของเสียงวรรณยุกต์ภาษาไทยที่ตำแหน่ง 100

1.2. การจัดเตรียมการทดลอง

ในการจัดเตรียมการทดลองสำหรับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงานในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดดจะใช้ระบบที่นำเสนอที่ดีที่สุดในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ความถี่มูลฐานในการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดซึ่งคือระบบที่ใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_{dF}_{aF} ในสเกลความถี่อัตราฮาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานเป็นระบบอ้างอิง และทำการเพิ่มลักษณะเด่นคือพลังงานเข้าไปในชุดลักษณะเด่น F_{dF}_{aF} โดยจะตั้งชื่อชุดลักษณะเด่นที่มีการเพิ่มพลังงานเข้าไปเป็น " $F_{dF}_{aF}_E$ " ซึ่งในการทดลองโดยชุดลักษณะเด่นใหม่ชุดนี้ จะใช้ความถี่มูลฐานในสเกลความถี่อัตราฮาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานเช่นเดียวกับระบบอ้างอิง แต่จะเพิ่มพลังงานเป็นลักษณะเด่นเติมโดยค่าของพลังงานสามารถที่จะสกัดได้จากแฟ้มเสียงแต่ละแฟ้มโดยใช้แพทในการสกัดค่าพลังงานโดยกำหนดกรอบของสัญญาณเวลาเป็น 0.01 วินาที และกำหนดการตั้งค่าของแบบจำลอง โดยมีการปรับพารามิเตอร์โดยเลือกอัลกอริทึมในการเรียนรู้แบบจำลองเป็น L-BFGS และในตัวแบบจำลองกำหนดให้มีสถานะฮิดเดนภายในทั้ง 3 สถานะ และกำหนดกรอบหน้าต่าง หรือวินโดวส์ เป็น 0 นอกจากนี้มีการกำหนดค่าน้ำหนักเริ่มต้นให้สุ่มค่าอยู่ในช่วงระหว่าง -1.0 ถึง 1.0 และเริ่มสถานะแบบจำลองด้วยเกาส์เซียน ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดได้กำหนดจำนวนในการเรียนรู้ไว้ที่ 500 รอบเช่นเดียวกับการทดลองก่อนหน้า

นอกจากนี้ในการทดลองได้ใช้วิธีการตรวจสอบไขว้โดยได้กำหนดค่า k เท่ากับ 6 สำหรับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดด ซึ่งแบ่งข้อมูลเป็น 6 ส่วนโดยการสุ่ม โดยในแต่ละกลุ่มจะประกอบด้วยผู้พูดเพศชายหนึ่งคน และเพศหญิงหนึ่งคน รวมทั้งหมด 6 กลุ่ม โดยข้อมูล 5 กลุ่มจะถูกนำไปเรียนรู้ และข้อมูลอีกกลุ่มจะถูกใช้ในการตรวจสอบ โดยกระบวนการตรวจสอบจะดำเนินไปเรื่อยๆ จนกระทั่งข้อมูลทั้ง 6 กลุ่มโดนตรวจสอบเรียบร้อยแล้วเช่นเดียวกับการทดลองก่อนหน้าเช่นกัน

1.3. ผลการทดลองและสรุปผลการทดลอง

จากการทดลองจะได้ผลการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงานในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดดเปรียบเทียบกับการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ความถี่มูลฐานในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดดดังแสดงในตารางที่ 4.23 ซึ่งพบว่าเมื่อเพิ่มค่าพลังงานเข้าไปเป็นลักษณะเด่นในชุด "F_dF_aF_E" ให้ผลความถูกต้องมากขึ้นโดยได้ความถูกต้องในการจำแนกเป็น ร้อยละ 95.53 โดยเมื่อคิดเป็นความผิดพลาดที่ลดลงจากระบบอ้างอิงจะลดลงถึงร้อยละ 22.40 ซึ่งเป็นไปตามสมมุติฐานที่เมื่อเพิ่มลักษณะเด่น คือ พลังงานจะสามารถช่วยให้ผลในการจำแนกเสียงวรรณยุกต์ได้ความถูกต้องมากขึ้น

ตารางที่ 4.23 ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าพลังงานในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดดเปรียบเทียบกับการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ความถี่มูลฐานสำหรับฐานข้อมูลเสียงคำโดด

ชุดลักษณะเด่น	ความแม่นยำ (%)
F_dF_aF	94.24
F_dF_aF_E	95.53

และเมื่อพิจารณาคอนฟิวชันเมตริกซ์การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF_E ในตารางที่ 4.24 จะเห็นได้ว่าความถูกต้องในการจำแนกเสียงวรรณยุกต์เอก และเสียงวรรณยุกต์จัตวา อีกทั้งความสับสนในคู่ของเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์สามัญ และเสียงวรรณยุกต์เอกกับเสียงวรรณยุกต์จัตวา มีค่าลดลงทั้งคู่เมื่อเทียบกับการใช้ชุดลักษณะเด่น F_dF_aF ในตารางที่ 4.22 ซึ่งเป็นไปตามสมมุติฐานของการทดลอง แต่ก็พบว่ามีความสับสนในกลุ่มระหว่างเสียงวรรณยุกต์สามัญ เสียงวรรณยุกต์โท และเสียงวรรณยุกต์ตรี มีความสับสนเพิ่มขึ้นเล็กน้อย ซึ่งเมื่อกลับไปพิจารณากราฟของพลังงานในภาพที่ 4.54 จะเห็นได้ว่าค่าพลังงานของเสียงวรรณยุกต์สามัญ เสียงวรรณยุกต์โท และเสียงวรรณยุกต์ตรีมีความใกล้เคียงกัน และยังเห็น

ชัดเจนเมื่อพิจารณาแผนภาพกล่องของพลังงานในรูปที่ 4.58 – 4.62 โดยจะเห็นได้ว่าค่าเฉลี่ยและการกระจายตัวของพลังงานมีลักษณะเดียวกันส่งผลให้การเพิ่มลักษณะเด่นพลังงานสร้างความสับสนเพิ่มขึ้นในกลุ่มเสียงวรรณยุกต์ทั้งสามกลุ่มนี้ แต่อย่างไรก็ตามการเพิ่มลักษณะเด่นดังกล่าวก็ให้ผลการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยรวมมีความถูกต้องที่สูงขึ้น

ตารางที่ 4.24 คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับคำโดดโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น $F_{dF_aF_E}$ ในสเกลความถี่อัตราอีอาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน

จำแนก เสียง	สามัญ	เอก	โท	ตรี	จัตวา	ความถูกต้อง
สามัญ	250	9	2	1	2	94.70%
เอก	10	247	1	0	6	93.56%
โท	2	0	256	5	1	96.97%
ตรี	2	0	1	260	1	98.48%
จัตวา	1	12	0	3	248	93.94%
ทั้งหมด						95.53%

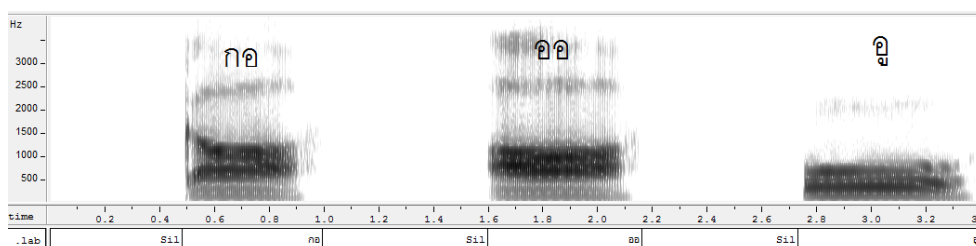
จากการทดลองเพิ่มลักษณะเด่นสำหรับการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดดพบว่า การเพิ่มลักษณะเด่นพลังงานส่งผลให้การจำแนกเสียงวรรณยุกต์โดยรวมมีค่าสูงขึ้น ซึ่งสอดคล้องกับการรับรู้เสียงวรรณยุกต์ภาษาไทยของมนุษย์ ถึงแม้ว่าการเพิ่มลักษณะเด่นพลังงานอาจสร้างความสับสนในกลุ่มเสียงวรรณยุกต์บางกลุ่มเพิ่มขึ้นเล็กน้อย แต่เมื่อพิจารณาผลลัพธ์ในภาพรวมพบว่าการเพิ่มลักษณะเด่นพลังงานสามารถลดความผิดพลาดลงจากระบบอ้างอิงได้ถึงร้อยละ 22.40

2. การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัม

สำหรับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัม มีจุดประสงค์ที่จะศึกษาข้อมูลทางสเปกตรัมว่าจะมีส่วนช่วยเพิ่มประสิทธิภาพในการจำแนกเสียงวรรณยุกต์เช่นเดียวกับการรับรู้เสียงวรรณยุกต์หรือไม่ ซึ่งในการศึกษาจะศึกษาบนพื้นฐานข้อมูลเสียงคำโดดและฐานข้อมูลเสียงพูดต่อเนื่อง เพราะจากการทดลองการรับรู้เสียงวรรณยุกต์พบว่าข้อมูลทางสเปกตรัมมีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ภาษาไทยให้มีความถูกต้องที่สูงในทั้งคำโดดและเสียงพูดต่อเนื่อง โดยรายละเอียดในการทดลองประกอบด้วย การวิเคราะห์ลักษณะเด่น, การจัดเตรียมการทดลอง และผลการทดลองและสรุปผล

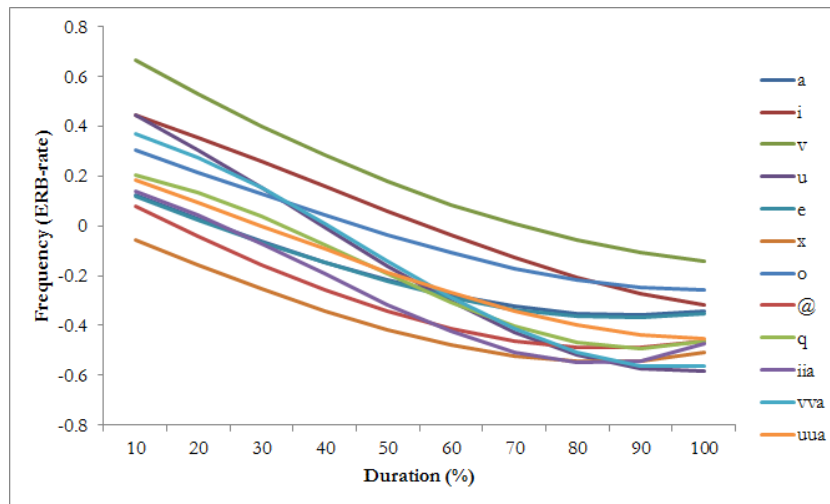
2.1. การวิเคราะห์ลักษณะเด่น

จากการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทยโดยมนุษย์พบว่าสเปกตรัมของสัญญาณเสียงพูดมีบทบาทเป็นอย่างมากในการช่วยให้มนุษย์สามารถที่จะแยกแยะเสียงวรรณยุกต์ภาษาไทยได้ดีขึ้น และมีความถูกต้องสูงขึ้นอย่างมีนัยสำคัญ ซึ่งเมื่อพิจารณาสเปกตรัมจากอีกมุมมองหนึ่งอาจจะพิจารณาลักษณะของสเปกตรัมโดยแบ่งตามกลุ่มของเสียงพยัญชนะ หรือเสียงสระ เนื่องจากลักษณะของสเปกตรัมเกี่ยวข้องกับเสียงพยัญชนะ และเสียงสระ กล่าวคือสเปกตรัมของเสียงของพยัญชนะ หรือสระกลุ่มเดียวจะมีลักษณะที่คล้ายคลึงกัน ยกตัวอย่างเช่น ภาพที่ 4.62 เมื่อพิจารณา สเปกโตรแกรมของคำว่า “กอ” “ออ” และ “อู” จะเห็นได้ว่าในช่วงกลางสระออของคำว่า “กอ” กับ “ออ” จะมีรูปร่างคล้ายกัน แต่จะแตกต่างกันในช่วงต้นเนื่องจากพยัญชนะต้นแตกต่างกัน และเมื่อพิจารณา “ออ” กับ “อู” พบว่ารูปร่างช่วงสระแตกต่างกันชัดเจนในขณะที่ช่วงขึ้นต้นจะคล้ายกัน เนื่องจากเป็นพยัญชนะต้นเดียวกันเป็นต้น

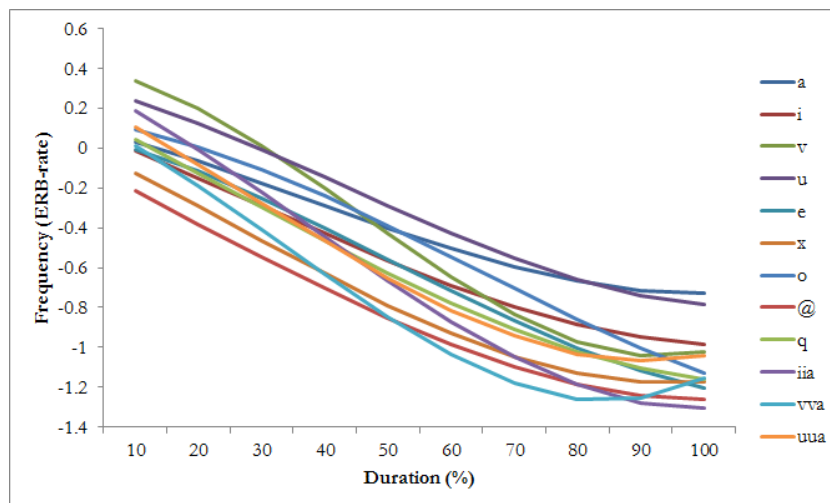


ภาพที่ 4.62 สเปกโตรแกรมของคำว่า “กอ” “ออ” และ “อู”

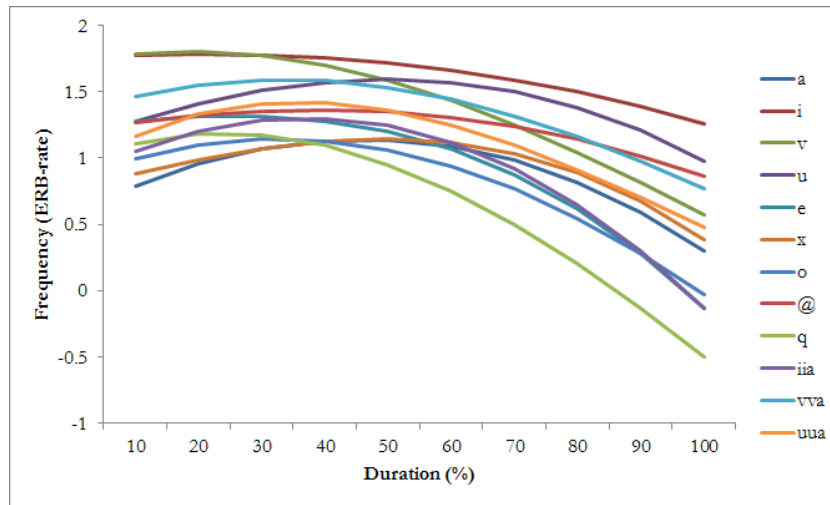
ดังนั้นจากสาเหตุที่กล่าวมาจึงทำการพิจารณารูปร่างของความถี่มูลฐานโดยแบ่งตามกลุ่มของสระ และตัวสะกดตามเสียงวรรณยุกต์ต่างๆ ดังแสดงในภาพที่ 4.63 – 4.72 จากภาพจะเห็นได้ว่ามีความแตกต่างกันเองภายในกลุ่มเป็นอย่างมาก



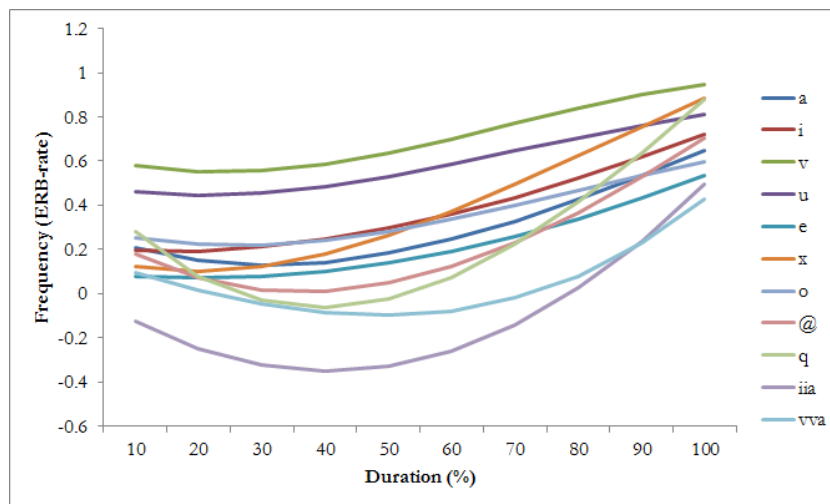
ภาพที่ 4.63 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์สามัญในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ



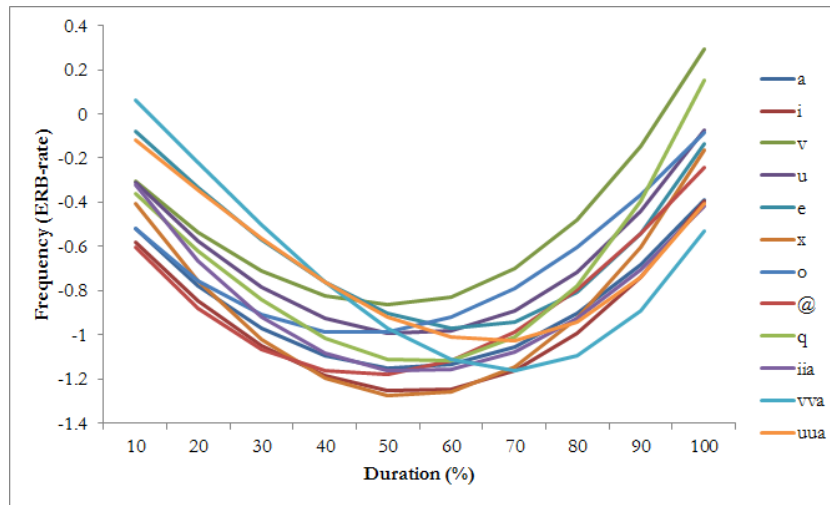
ภาพที่ 4.64 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์เอกในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ



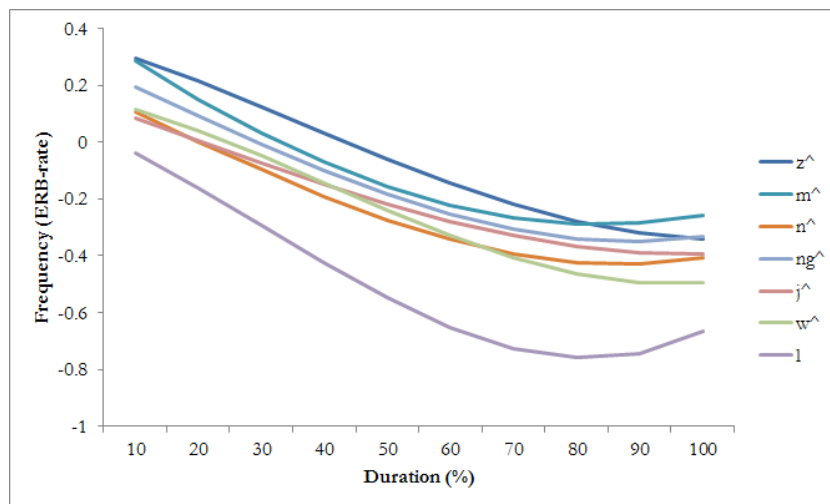
ภาพที่ 4.65 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์โทในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ



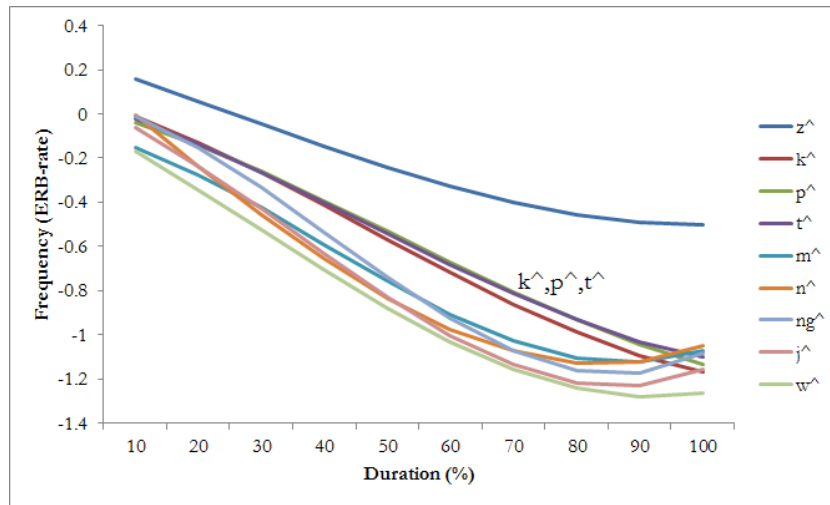
ภาพที่ 4.66 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์ตรีในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยสระ



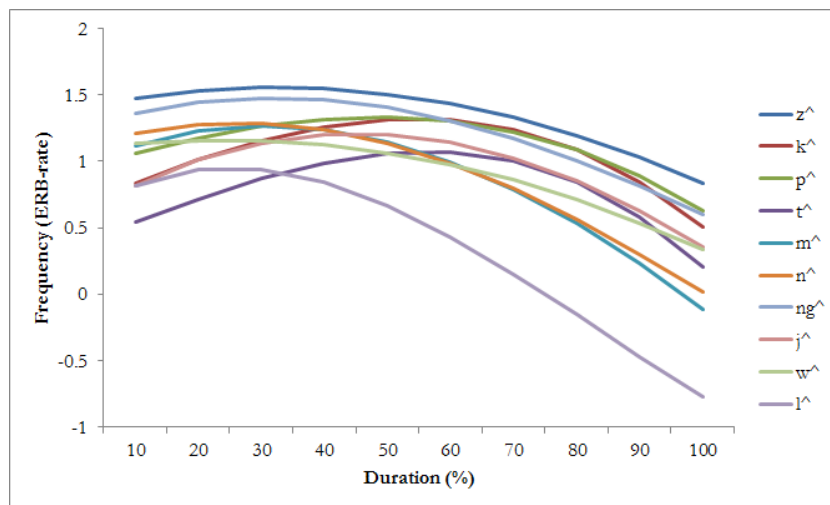
ภาพที่ 4.67 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์จัตวาในฐานข้อมูลเสียงพูดต่อเนื่องจัด
กลุ่มโดยสระ



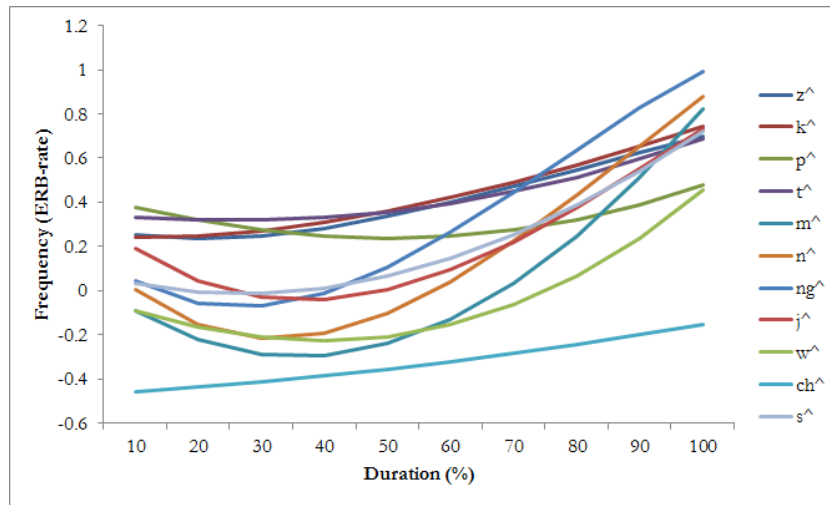
ภาพที่ 4.68 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์สามัญในฐานข้อมูลเสียงพูดต่อเนื่องจัด
กลุ่มโดยพยัญชนะตัวสะกด



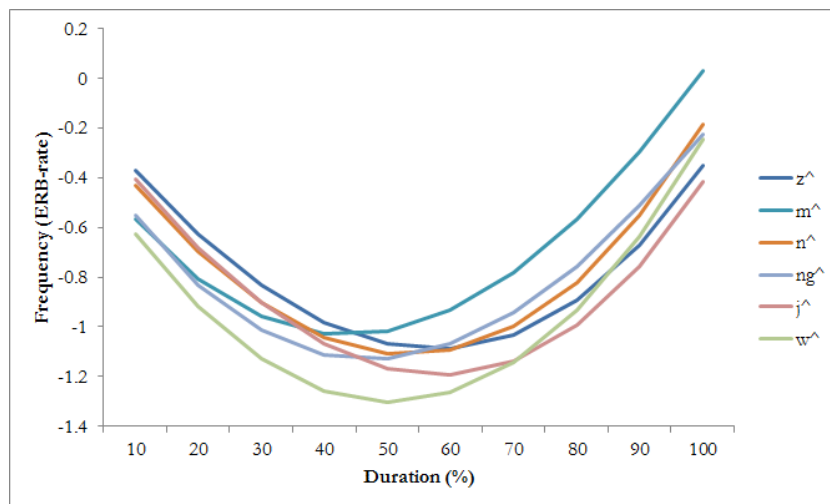
ภาพที่ 4.69 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์เอกในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด



ภาพที่ 4.70 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์โทในฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด



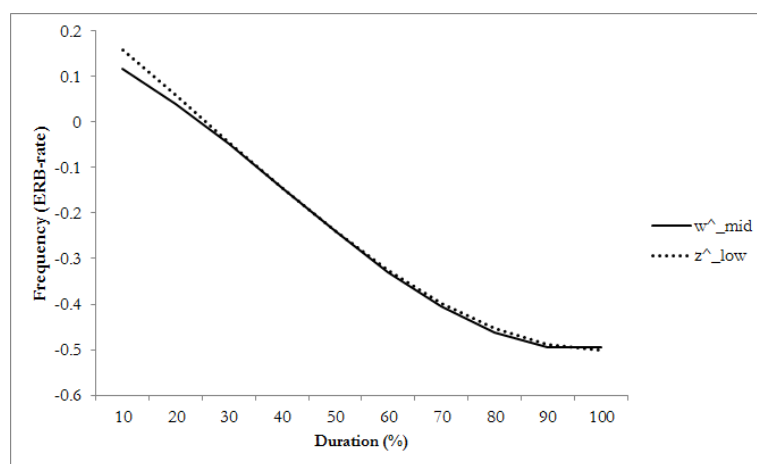
ภาพที่ 4.71 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์ตรีพื้นฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด



ภาพที่ 4.72 รูปร่างของความถี่มูลฐานของเสียงวรรณยุกต์จัตวาพื้นฐานข้อมูลเสียงพูดต่อเนื่องจัดกลุ่มโดยพยัญชนะตัวสะกด

นอกจากนี้เมื่อพิจารณาอย่างละเอียดจะพบว่าแม้ว่าจะจัดกลุ่มด้วยสระหรือเสียงพยัญชนะตัวสะกดก็พบว่ารูปร่างเฉลี่ยของเสียงวรรณยุกต์ในกลุ่มเดียวกันมีความแตกต่างกันพอสมควร และในบางเส้นของความถี่มูลฐานในเสียงวรรณยุกต์หนึ่งก็ไปทับซ้อนกับเส้นของความถี่มูลฐานในอีกเสียงวรรณยุกต์หนึ่ง เช่นในเสียงวรรณยุกต์สามัญในกลุ่มที่มีความถี่ต่ำ อาจจะไปปนกับกลุ่มที่มีความถี่สูงของเสียงวรรณยุกต์เอก ยกตัวอย่างเช่น เมื่พล็อตรูปร่างความถี่

มูลฐานของเสียงวรรณยุกต์สามัญที่มีตัวสะกดเป็นแม่กอกและเสียงวรรณยุกต์เอกที่มีตัวสะกดเป็นแม่ก.กา ดังแสดงในภาพที่ 4.73 จะเห็นได้ว่ามีลักษณะทับซ้อนกันอย่างชัดเจน ซึ่งแน่นอนว่าลักษณะดังกล่าวจะส่งผลให้การจำแนกเสียงเสียงวรรณยุกต์มีความผิดพลาด ดังนั้นหากไม่มีการหาลักษณะเด่นเพิ่มเติมที่จะช่วยแยกความแตกต่างในลักษณะนี้ได้จะทำให้การจำแนกเสียงวรรณยุกต์มีความแม่นยำที่ลดลง



ภาพที่ 4.73 รูปร่างของความถี่มูลฐานของรูปร่างของเสียงวรรณยุกต์สามัญที่มีตัวสะกดเป็นแม่กอกกับเสียงวรรณยุกต์เอกที่มีตัวสะกดเป็นแม่ก.กา

นอกจากนี้จากการทดลองของ Tan [7] เมื่อทำการแบ่งการจำแนกเสียงวรรณยุกต์ภาษาไทยต่อเนื่อง โดยการจัดกลุ่มข้อมูลเสียงพูดตามพยัญชนะตัวสะกดของแต่ละเสียงวรรณยุกต์ด้วยมือ และนำข้อมูลแต่ละกลุ่มไปสร้างตัวจำแนกเฉพาะตัวของแต่ละกลุ่มขึ้นมา จากการทดลองของ Tan พบว่าได้ความถูกต้องโดยรวมในการจำแนกเสียงวรรณยุกต์ภาษาไทยมากขึ้นจากเดิมที่มีการใช้ตัวจำแนกเพียงตัวเดียว แต่อย่างไรก็ตามวิธีนี้มีข้อจำกัดคือจำเป็นต้องแบ่งข้อมูลตามพยัญชนะตัวสะกดด้วย ก่อนการนำไปทดลองซึ่งจะทำให้ไม่สามารถนำไปประยุกต์ใช้งานจริงได้ เนื่องจากในความเป็นจริงสัญญาณเสียงที่นำเข้ามาจะไม่ได้ระบุกลุ่มของพยัญชนะตัวสะกดมาให้

จากเหตุผลที่ว่าจากการทดลองการรับรู้เสียงวรรณยุกต์ของมนุษย์พบว่าสเปกตรัมมีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ภาษาไทย และเมื่อพิจารณาความแตกต่างระหว่างรูปร่างของความถี่มูลฐานในแต่ละกลุ่มเสียงสระ และเสียงพยัญชนะตัวสะกด ดังที่แสดงให้ดูก่อนหน้านี้ อีกทั้งจากการทดลองของ Tan [7] ที่ได้ทำการทดลองแบ่งกลุ่มการจำแนกตามพยัญชนะตัวสะกดแล้วได้

ความถูกต้องที่สูงขึ้น ดังนั้นในการทดลองนี้จึงเลือกที่จะพิจารณาสเปกตรัมของสัญญาณเสียงเป็นลักษณะเด่นเพิ่มเติมในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทย โดยลักษณะเด่นที่จะนำมาใช้แทนสเปกตรัมของสัญญาณเสียงที่จะนำมาทดลองมี 3 แบบ คือ ค่าสัมประสิทธิ์การประมาณพันธ์เชิงเส้น, ค่าการประมาณพันธ์แบบอิงการรับฟังของมนุษย์ และค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล ซึ่งทั้ง 3 ค่าเป็นค่าที่นิยมนำมาใช้ในการแทนรูปร่างของ สเปกตรัม

2.2. การจัดเตรียมการทดลอง

ในการจัดเตรียมการทดลองสำหรับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัมจะเลือกใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_{dF}_{aF} ในสเกลความถี่อัตรารีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซึ่งมีประสิทธิภาพสูงสุดจากการทดลองก่อนหน้านี้เป็นระบบอ้างอิงบนทั้งสองฐานข้อมูล

สำหรับลักษณะเด่นที่เพิ่มของสเปกตรัมที่จะนำมาใช้ในการทดลองเพิ่มเติมได้แก่ ค่าสัมประสิทธิ์การประมาณพันธ์เชิงเส้น, ค่าการประมาณพันธ์แบบอิงการรับฟังของมนุษย์ และค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล เพื่อที่จะหาว่าค่าใดจะมีความเหมาะสมที่จะใช้ในการจำแนกเสียงวรรณยุกต์ภาษาไทยมากที่สุด โดยในการสกัดค่าทั้ง 3 จะใช้เครื่องมือคือ เอชทีเค ในการช่วยสกัดค่า โดยกำหนดอัตราร่วมของแฟ้มเสียงที่ 1.6 กิโลเฮิร์ตซ์ และใช้กรอบสัญญาณเวลาแฮมมิง (Hamming) ที่มีความกว้าง 25 มิลลิวินาที ซึ่งจะขยับทุกๆ 10 มิลลิวินาที และกำหนดค่าสัมประสิทธิ์พรีเอมฟาซิส (Pre-emphasis Coefficient) เป็น 0.97 และกำหนดค่าสำหรับค่าสัมประสิทธิ์การประมาณพันธ์เชิงเส้นกำหนด เป็น LPC_D_A_E กำหนดค่าสำหรับค่าการประมาณพันธ์แบบอิงการรับฟังของมนุษย์ เป็น PLP_D_A_E กำหนดค่าสำหรับค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมลเป็น MFCC_D_A_E ซึ่งหมายความว่า จะมีการนำค่าอัตราการเปลี่ยนแปลง และความเร่งของค่าทั้ง 3 มาใช้ด้วย และกำหนดดีกรีของสัมประสิทธิ์แต่ละค่าเป็น 12 จากผลการกำหนดดังกล่าวจะได้เวกเตอร์ของค่าสัมประสิทธิ์การประมาณพันธ์เชิงเส้น, ค่าการประมาณพันธ์แบบอิงการรับฟังของมนุษย์ และค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล ที่มีขนาด 39 มิติ และมีการนำค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานที่ได้จากชุดฝึกฝนไปใช้ในการทำ

ให้เป็นบรรทัดฐานด้วยค่ามาตรฐานของทั้งชุดเรียนรู้ และชุดที่จะนำไปทดสอบ ของแต่ละลักษณะเด่นที่จะเพิ่มเข้าไปเพื่อลดความแตกต่างของสเปกตรัมระหว่างผู้พูด [10]

ในการทดลองนี้จะทำการเพิ่มลักษณะเด่นเข้าไปในชุดลักษณะเด่นของระบบอ้างอิงคือ F_{dF_aF} ซึ่งเมื่อเพิ่มค่าสัมประสิทธิ์การประมาณพหุเชิงเส้นกำหนดจากชุดลักษณะเด่นของระบบอ้างอิงจะตั้งชื่อชุดลักษณะเด่นนี้ว่า " $F_{dF_aF_LPC}$ " เมื่อเพิ่มค่าการประมาณพหุแบบอิงการรับฟังของมนุษย์จากชุดลักษณะเด่นของระบบอ้างอิงจะตั้งชื่อชุดลักษณะเด่นนี้ว่า " $F_{dF_aF_PLP}$ " และเมื่อเพิ่มค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมลจากชุดลักษณะเด่นของระบบอ้างอิงจะตั้งชื่อชุดลักษณะเด่นนี้ว่า " $F_{dF_aF_MFCC}$ " โดยในการทดลองจะใช้ชุดลักษณะเด่น F_{dF_aF} , $F_{dF_aF_LPC}$, $F_{dF_aF_PLP}$ และ $F_{dF_aF_MFCC}$ ซึ่งจะมีการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานและค่าความถี่มูลฐานในสเกลความถี่อัตราอาร์บี

การทดลองนี้จะเลือกอัลกอริทึมในการเรียนรู้แบบจำลองเป็น L-BFGS และในตัวแบบจำลองกำหนดให้มีสถานะฮิดเดนภายในทั้ง 3 สถานะ และกำหนดกรอบหน้าต่าง หรือ วินโดวส์ เป็น 0 นอกจากนี้มีการกำหนดค่าน้ำหนักเริ่มต้นให้สุ่มค่าอยู่ในช่วงระหว่าง -1.0 ถึง 1.0 และเริ่มสถานะแบบจำลองด้วยเกาส์เซียน ในการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยความถี่มูลฐานสำหรับคำโดดได้กำหนดจำนวนในการเรียนรู้ไว้ที่ 2,000 รอบเช่นเดียวกับการทดลองก่อนหน้า

ในการทดลองได้เลือกใช้วิธีการตรวจสอบไขว้เช่นเดียวกับการทดลองก่อนหน้า โดยเลือกค่า k เท่ากับ 4 เช่นเดียวกัน ซึ่งข้อมูลจะถูกแบ่งเป็น 4 ส่วนโดยการสุ่ม โดยในแต่ละกลุ่มจะประกอบด้วยผู้พูดเพศชาย 6 คน และผู้พูดเพศหญิง 6 คน รวมทั้งสิ้น 12 คน โดยข้อมูล 3 กลุ่มจะถูกนำไปเรียนรู้ และข้อมูลอีกกลุ่มจะถูกใช้ในการตรวจสอบจนข้อมูลทุกชุดที่แบ่งถูกตรวจสอบทั้งหมดเช่นเดียวกับการทดลองก่อนหน้า

2.3. ผลการทดลองและสรุปผลการทดลอง

จากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัม ได้ทำการทดลองเพิ่มกับชุดลักษณะเด่นอีก 3 ชุด ได้แก่ F_{dF_aF} , $F_{dF_aF_LPC}$, $F_{dF_aF_PLP}$ และ $F_{dF_aF_MFCC}$ บนสองฐานข้อมูลเสียง ได้ผลการทดลองดังแสดงอยู่ในตารางที่ 4.25 สำหรับคำโดด และตารางที่ 4.26 สำหรับเสียงพูดต่อเนื่อง

ตารางที่ 4.25 ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้
ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัมในการจำแนกเสียงวรรณยุกต์ภาษาไทยในคำโดด

ชุดลักษณะเด่น	ความแม่นยำ (%)
F_dF_aF	94.24
F_dF_aF_LPC	90.68
F_dF_aF_PLP	91.74
F_dF_aF_MFCC	91.14

ตารางที่ 4.26 ผลความแม่นยำของการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยโดยใช้
ลักษณะเด่นเพิ่มเติมเป็นค่าสเปกตรัมในการจำแนกเสียงวรรณยุกต์ภาษาไทยในเสียงพูดต่อเนื่อง

ชุดลักษณะเด่น	ความแม่นยำ (%)
F_dF_aF	71.01
F_dF_aF_LPC	72.08
F_dF_aF_PLP	75.04
F_dF_aF_MFCC	74.91

จากผลการทดลองการจำแนกเสียงวรรณยุกต์โดยการนำค่าสเปกตรัม ซึ่งแทนที่ด้วยค่าสัมประสิทธิ์การประมาณพหุระเชิงเส้น, ค่าการประมาณพหุระแบบอิงการรับฟังของมนุษย์ และค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล สำหรับฐานข้อมูลเสียงคำโดดพบว่าความแม่นยำในการจำแนกเสียงวรรณยุกต์มีค่าลดลง ซึ่งปัจจัยหนึ่งอาจจะมาจากจำนวนข้อมูลในการเรียนรู้ที่มีจำนวนน้อยเกินไป และไม่ครอบคลุมเพียงพอ ในขณะที่สำหรับฐานข้อมูลเสียงพูดต่อเนื่องพบว่าชุดลักษณะเด่นที่มีค่าการประมาณพหุระแบบอิงการรับฟังของมนุษย์จะให้ความถูกต้องมากที่สุดในการทดลอง โดยได้ความถูกต้องที่ร้อยละ 75.04 ซึ่งเมื่อคิดเป็นร้อยละของข้อผิดพลาดที่ลดลงจากระบบอ้างอิงจะเท่ากับ 13.90 และชุดลักษณะเด่นที่มีค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล ให้ความถูกต้องรองลงมา โดยมีความแตกต่างกับชุดลักษณะเด่นที่มีค่าการประมาณพหุระแบบอิงการรับฟังของมนุษย์เล็กน้อย ในขณะที่ค่าสัมประสิทธิ์การประมาณพหุระเชิงเส้นทำให้ความ

ถูกต้องเพิ่มขึ้นจากระบบอ้างอิงเล็กน้อย เมื่อพิจารณาคอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อซึ่งใช้ชุดลักษณะเด่น F_dF_aF_PLP ในตารางที่ 4.26 พบว่าความถูกต้องในการจำแนกเสียงวรรณยุกต์ส่วนใหญ่มีความถูกต้องเพิ่มขึ้นเป็นอย่างมากเมื่อเทียบกับระบบอ้างอิง แต่อย่างไรก็ตามพบว่าความถูกต้องในการจำแนกเสียงวรรณยุกต์บางตัวมีความถูกต้องลดลงเล็กน้อยจากระบบอ้างอิง ซึ่งเมื่อพิจารณาพบว่าความถูกต้องในการแยกเสียงวรรณยุกต์เอกออกจากเสียงวรรณยุกต์สามัญมีความถูกต้องลดลงส่งผลความถูกต้องในการจำแนกเสียงวรรณยุกต์เอกมีความถูกต้องลดลง แต่เมื่อพิจารณาการแยกเสียงวรรณยุกต์สามัญออกจากเสียงวรรณยุกต์เอกมีค่าเพิ่มขึ้น ซึ่งเมื่อพิจารณาค่ารวมในการที่จำแนกเสียงวรรณยุกต์เอกเป็นวรรณยุกต์สามัญ และวรรณยุกต์สามัญเป็นวรรณยุกต์เอก พบว่าความผิดพลาดในการจำแนกของชุดลักษณะเด่นที่นำเสนอมีค่า โดยรวมลดลง

ตารางที่ 4.27 คอนฟิวชันเมทริกซ์ของการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่อง โดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ที่ใช้ชุดลักษณะเด่น F_dF_aF_PLP ในสเกลความถี่อัตราฮาร์บีและใช้เทคนิคการทำให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐาน

จำแนก เสียง	สามัญ	เอก	โท	ตรี	จัตวา	ความ ถูกต้อง
สามัญ	6903	822	347	445	145	79.69%
เอก	1001	3857	67	155	197	73.09%
โท	383	108	4101	362	12	82.58%
ตรี	525	383	346	2774	174	66.02%
จัตวา	281	388	18	144	1319	61.35%
ทั้งหมด						75.04%

บทที่ 5

การประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ภาษาไทย

การประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ภาษาไทยจะนำมาประยุกต์ใช้งานกับการรู้จำเสียงพูดภาษาไทยในโดเมนการสะกดคำภาษาไทยที่เป็นคำโดด เนื่องจากการสะกดคำภาษาไทยนั้นจะมีคำบางคำที่มีเสียงพยัญชนะ และสระเหมือนกันแต่ต่างกันว่าเสียงวรรณยุกต์เช่น ขอ กับ คอ ทั้งสองคำมีเสียงพยัญชนะ และสระที่เหมือนกันแต่จะต่างกันว่าเสียงวรรณยุกต์เท่านั้น นอกจากนี้การสะกดคำภาษาไทยผู้สะกดจะสะกดอย่างไรก็ได้ซึ่งไม่มีหลักไวยากรณ์ที่ชัดเจน ส่งผลให้แบบจำลองทางภาษา (Language Model) ไม่สามารถทำให้การรู้จำเสียงพูดมีประสิทธิภาพที่สูงขึ้นได้ ดังนั้นเนื้อหาในบทนี้จะนำเสนอการประยุกต์การจำแนกเสียงวรรณยุกต์ภาษาไทยเพื่อเพิ่มความสามารถในการรู้จำการสะกดคำภาษาไทย ซึ่งประกอบด้วย ฐานข้อมูลเสียงของการสะกดคำ, การทดลองและผลการทดลอง และการวิเคราะห์ผลการทดลอง

ฐานข้อมูลเสียงของการสะกดคำ

การเก็บรวบรวมฐานข้อมูลเสียงการสะกดคำในภาษาไทยประกอบไปด้วยขั้นตอนในการออกแบบฐานข้อมูลเสียง การเก็บรวบรวมฐานข้อมูลเสียง การถอดเสียง และการสร้างพจนานุกรมการออกเสียง

1. การออกแบบฐานข้อมูลเสียง

การออกแบบฐานข้อมูลเสียงจะคำนึงถึงหลักในการสะกดคำภาษาไทย ซึ่งตามหลักภาษาไทยการสะกดคำจะประกอบด้วยการสะกด 4 ประเภท ได้แก่การสะกดพยัญชนะ การสะกดสระ การสะกดเครื่องหมายเสียงวรรณยุกต์ในภาษาไทย และการสะกดเครื่องหมายอื่นๆ ซึ่งรูปแบบในการสะกดแต่ละชนิดจะเป็นไปได้ดังแสดงในตารางที่ 5.1 ซึ่งมีผลให้จำนวนการสะกดพยัญชนะมีทั้ง 90 คำการสะกดสระมี 30 คำ การสะกดเครื่องหมายวรรณยุกต์มี 8 คำ และการสะกดเครื่องหมายอื่น ๆ มี 7 คำ ซึ่งทั้งหมดรวมเป็น 135 คำ โดยคำที่ใช้ในการสะกดคำภาษาไทยทั้งหมดจะถูกแสดงอยู่ในภาคผนวก ข.

ตารางที่ 5.1 การสะกดคำในภาษาไทย

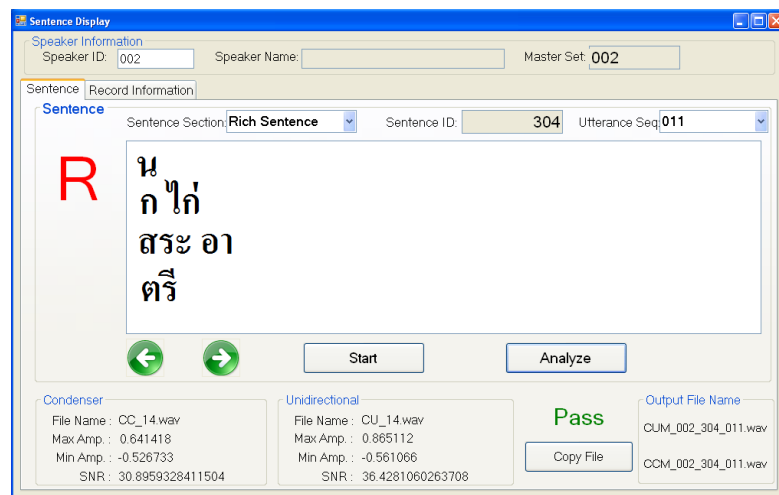
ประเภท	รูปแบบ	ตัวอย่าง
พยัญชนะ	ชื่อพยัญชนะ	กอ
	ชื่อพยัญชนะ + ส่วนขยาย	กอ ไก่
สระ	ชื่อสระ	อา
	“สระ” + ชื่อสระ	สระอา
เครื่องหมายเสียง วรรณยุกต์ในภาษาไทย	ชื่อวรรณยุกต์	เอก
	“ไม้” + ชื่อวรรณยุกต์	ไม้เอก
เครื่องหมายอื่นๆ	ชื่อเครื่องหมาย	ไม้ยมก

2. การเก็บรวบรวมฐานข้อมูลเสียง

ในการเก็บรวบรวมฐานข้อมูลเสียงพูดได้เชิญผู้พูดมาจำนวน 50 คน ซึ่งเป็นชาย 25 คน และหญิง 25 คนในการบันทึกเสียง โดยที่เกณฑ์อายุอยู่ในช่วง 17 ถึง 40 ปี และผู้พูดทุกคนเป็นผู้ที่ใช้ภาษาไทยเป็นหลักและเป็นสำเนียงกลางซึ่งเป็นสำเนียงมาตรฐานของภาษาไทย ในการบันทึกเสียงจะทำการบันทึกเสียงทั้งหมดภายในห้องอัดเสียง โดยเลือกใช้ไมโครโฟนไรต์เอ็นทีวันเอ ไมโครโฟนคอนเดนเซอร์ เป็นไมโครโฟนสำหรับการบันทึกเสียง และตั้งห่างจากผู้พูดประมาณ 20 ซม.

การบันทึกเสียงได้ใช้ซอฟต์แวร์ Cubase [40] ในการบันทึกเสียงโดยสัญญาณเสียงจะถูกบันทึกที่ความถี่สุ่ม 44.1 กิโลเฮิร์ตซ์ และบันทึกในรูปแบบเอ็มเอสเวฟโดยใช้การเข้ารหัส 16 บิตพีซีเอ็ม และสำหรับข้อความที่จะให้ผู้พูดจะถูกแสดงผ่านโปรแกรมในภาพที่ 5.1 ซึ่งผู้พูดจะพูดตามข้อความที่ปรากฏและระหว่างข้อความผู้พูดจะต้องเว้นการพูดช่วงเวลาหนึ่งเพื่อเป็นการลดความต่อเนื่องในการออกเสียง

สำหรับขั้นตอนการบันทึกเสียง ผู้ทำการบันทึกเสียงจะเริ่มให้สัญญาณผู้พูด เพื่อให้ผู้พูดเตรียมพร้อม โปรแกรมจะทำการนับถอยหลัง 3 วินาที โดยในระหว่างนั้น โปรแกรมจะทำการบันทึกเสียงในช่วงเวลา 3 วินาทีนี้เพื่อใช้เป็นส่วนอ้างอิงในการคำนวณสัญญาณรบกวนจากสภาพแวดล้อม หลังจากนั้นจะมีข้อความปรากฏให้ผู้พูดออกเสียง หลังจากการบันทึกเสียง ข้อมูลเสียงจะถูกตรวจสอบ ถ้าอัตราส่วนสัญญาณต่อสิ่งรบกวนมีค่าน้อยกว่า 30 dB ผู้พูดจะต้องพูดประโยคที่แสดงอยู่ใหม่ รวมถึงถ้าค่าแอมพลิจูดมากกว่า 0.99 (เกิดการคลิก) หรือสัญญาณเสียงมีแอมพลิจูดน้อยกว่า 0.4 ก็จะต้องบันทึกเสียงใหม่เช่นกัน และเมื่อผู้พูดสามารถบันทึกเสียงได้ตามเงื่อนไขดังกล่าว ผู้บันทึกจะทำการกดปุ่มถัดไปเพื่อเริ่มบันทึกชุดคำถัดไป



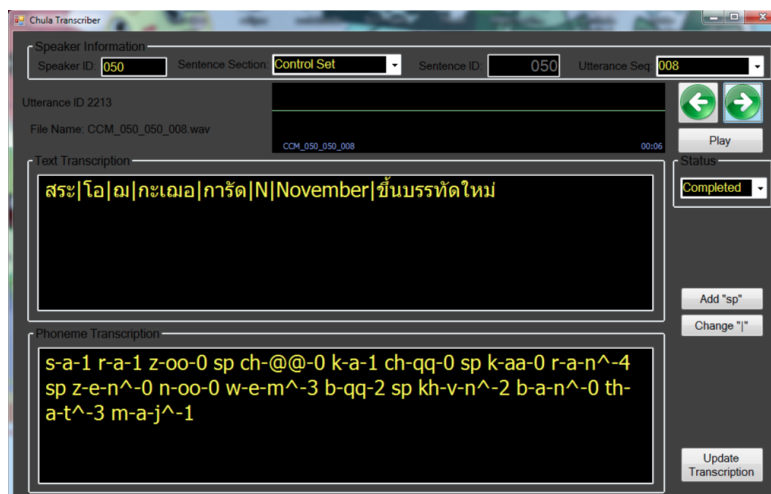
ภาพ 5.1 โปรแกรมแสดงการสะกดคำให้ผู้พูดออกเสียง

นอกจากนี้ในการบันทึกเสียงของผู้พูด หากผู้พูดออกเสียงได้ไม่ตรงตามการออกเสียงที่ถูกต้องของคำนั้นๆ ผู้พูดก็จะถูกขอให้บันทึกเสียงคำนั้นๆ ใหม่อีกครั้งจนกว่าจะถูกต้อง

3. การถอดเสียง

สำหรับการถอดเสียงจะถูกเตรียมข้อความโดยผู้ที่ใช้ภาษาไทยเป็นภาษาหลักและกระบวนการถอดเสียงจะให้ผู้ที่ใช้ภาษาไทยเป็นภาษาหลักมาฟังเสียงพูดผ่านทางโปรแกรมดังแสดงในภาพ 5.2 โดยผู้ถอดเสียงจะทำการตรวจสอบและแก้ไขเสียงที่ถอดให้ตรงกับเสียงที่ผู้ถอด

เสียงได้ยิน ถ้าหากเกิดความผิดพลาดหรือความไม่สอดคล้องกันระหว่างเสียงที่บันทึกมากับเสียงที่ถอดได้ผู้ถอดเสียงจะทำการแก้ไขให้มีความสามารถคล้องกัน



ภาพ 5.2 โปรแกรมที่ผู้ถอดเสียงใช้ในการตรวจสอบความถูกต้องของการถอดเสียง

ข้อผิดพลาดที่อาจเกิดขึ้นได้แก่ การเว้นวรรค การพูดคำเกิน การพูดคำขาด และการพูดคำผิด หากในกระบวนการถอดเสียงพบสิ่งเหล่านี้ จะต้องถูกดำเนินการแก้ไขให้ถูกต้องตามเสียงต้นแบบ

4. การสร้างพจนานุกรมการออกเสียง

การสร้างพจนานุกรมการออกเสียงสำหรับการสะกดคำไทยจะใช้สัญลักษณ์แทนหน่วยเสียงในภาษาไทยดังแสดงในภาคผนวก ก. ซึ่งเป็นหน่วยเสียงมาตรฐานที่นิยมใช้กันอย่างแพร่หลายใน [12, 49] โดยสัญลักษณ์หน่วยเสียงที่ใช้มีจำนวน 75 หน่วยเสียงซึ่งแทนเสียงพยัญชนะ และสระในภาษาไทย นอกจากนี้ในแต่ละพยางค์จะถูกระบุเสียงวรรณยุกต์ไว้ต่อท้ายเพื่อบ่งบอกเสียงวรรณยุกต์ของคำนั้นๆ โดย 0 แทนเสียงวรรณยุกต์สามัญ 1 แทนเสียงวรรณยุกต์เอก 2 แทนเสียงวรรณยุกต์โท 3 แทนเสียงวรรณยุกต์ตรี 4 แทนเสียงวรรณยุกต์จัตวา

ในการสร้างพจนานุกรมเบื้องต้นจะใช้โปรแกรมแปลงรูปเขียนเป็นหน่วยเสียง (Grapheme to Phoneme, G2P) ในการแปลงหน่วยเสียงขั้นต้น และใช้คนในการตรวจสอบและแก้ไขในขั้นสุดท้ายอีกรอบ

โดยสรุปฐานข้อมูลเสียงสำหรับการถอดคำโตภาษาไทยมีรายละเอียดดังต่อไปนี้ จำนวนผู้พูดทั้งหมด 50 คน อายุเฉลี่ยที่ 21.73 โดยที่เป็นเพศชาย 25 คนและเพศหญิง 25 คน และได้จำนวนแฟ้มเสียงทั้งสิ้น 2,250 แฟ้มเสียง ซึ่งมีความยาว ประมาณ 5 ชั่วโมง และคำศัพท์ทั้งหมด 135 คำ ดังแสดงอยู่ในภาคผนวก ข.

การทดลองและผลการทดลอง

การทดลองจะแบ่งออกเป็น 2 การทดลองคือ การทดลองบนระบบอ้างอิงในการรู้จำเสียง การถอดคำภาษาไทย และการทดลองบนระบบที่นำเสนอที่มีการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงการถอดคำภาษาไทย

1. การทดลองบนระบบอ้างอิงในการรู้จำเสียงการถอดคำภาษาไทย

ระบบอ้างอิงของการรู้จำเสียงในโดเมนของการถอดคำในภาษาไทยจะใช้แบบจำลองฮิดเดนมาร์คอฟซึ่งเป็นแบบจำลองที่นิยมนำมาใช้ในการรู้จำเสียงพูดอย่างแพร่หลาย [50] ในการรู้จำเสียงด้วยแบบจำลองฮิดเดนมาร์คอฟจะกำหนดขอบเขตในการเรียนรู้เป็นระดับคำ เนื่องจากถ้ากำหนดขอบเขตในการเรียนรู้เป็นระดับหน่วยเสียงย่อย ในการสร้างแบบจำลองจะต้องเพิ่มจำนวนหน่วยเสียงย่อยจากที่ปกติมีอยู่ 75 หน่วยเสียงเป็น 375 หน่วยเสียง เนื่องจากต้องคำนึงถึงเสียงวรรณยุกต์ด้วย ส่งผลให้จำนวนตัวเลือกในการรู้จำเสียงมีมากเกินไป อีกทั้งในการรู้จำการถอดคำมีคำศัพท์จำกัดเพียงแค่ 135 คำ ดังนั้นการจะเรียนรู้แบบจำลองฮิดเดนมาร์คอฟในระดับหน่วยเสียงนั้นอาจจะได้ผลลัพธ์ไม่ดี ในการทดลองจึงเลือกการเรียนรู้แบบจำลองฮิดเดนมาร์คอฟในระดับคำแทน

ในการทดลองกำหนดให้แบบจำลองฮิดเดนมาร์คอฟมีทอพอโลยีจากซ้ายไปขวาโดยจำนวนสถานะของแบบจำลองเป็นไปได้ตั้งแต่ 3 สถานะถึง 8 สถานะเพื่อใช้ในการหาจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมในการรู้จำเสียงในระดับคำ นอกจากนี้มีการกำหนดเกาส์เซียนมิกเจอร์ของแต่ละสถานะจาก 1 มิกเจอร์ ถึง 32 มิกเจอร์ เพื่อหาจำนวนเกาส์เซียนมิกเจอร์ที่เหมาะสมกับการทดลอง และใช้เมทริกซ์แปรปรวนทแยงในการแทนค่าที่สังเกตได้ที่เกิดจากแบบจำลองฮิดเดนมาร์คอฟ สำหรับลักษณะเด่นที่จะใช้ในการทดลองจะเลือกใช้ค่า

สัมประสิทธิ์เซปสตรีมในย่านความถี่ 39 มิตี โดยกำหนดกรอบสัญญาณเป็นแฮมมิงวินโดว์ที่มีขนาด 25 มิลลิวินาที และมีการขยับกรอบของสัญญาณทุกๆ 10 มิลลิวินาที และกำหนดค่าพีเอ็ม ฟาซิส เป็น 0.97 และใช้จำนวนช่องสัญญาณของชุดกรองสัญญาณเป็น 24 ช่องสัญญาณ

ในการทดลองข้อมูลเสียงพูดของผู้พูด 50 คนได้ถูกแบ่งอย่างสุ่ม แต่ควบคุมให้มีความสมดุลกันระหว่างเพศชายและเพศหญิงในชุดเรียนรู้และชุดทดสอบ โดยข้อมูลของผู้พูด 40 คน (ชาย 20 คน, หญิง 20 คน) เป็นข้อมูลสำหรับที่ใช้ในการเรียนรู้ และข้อมูลอีก 10 คน (ชาย 5 คน, หญิง 5 คน) เป็นข้อมูลที่ใช้สำหรับทดสอบ ซึ่งได้ผลการทดลองเป็นไปตามตารางที่ 5.2

ตารางที่ 5.2 ผลการทดลองการรู้จำเสียงพูดสำหรับการสะกดคำไทยโดยมีการปรับจำนวนสถานะและมิกเจอร์

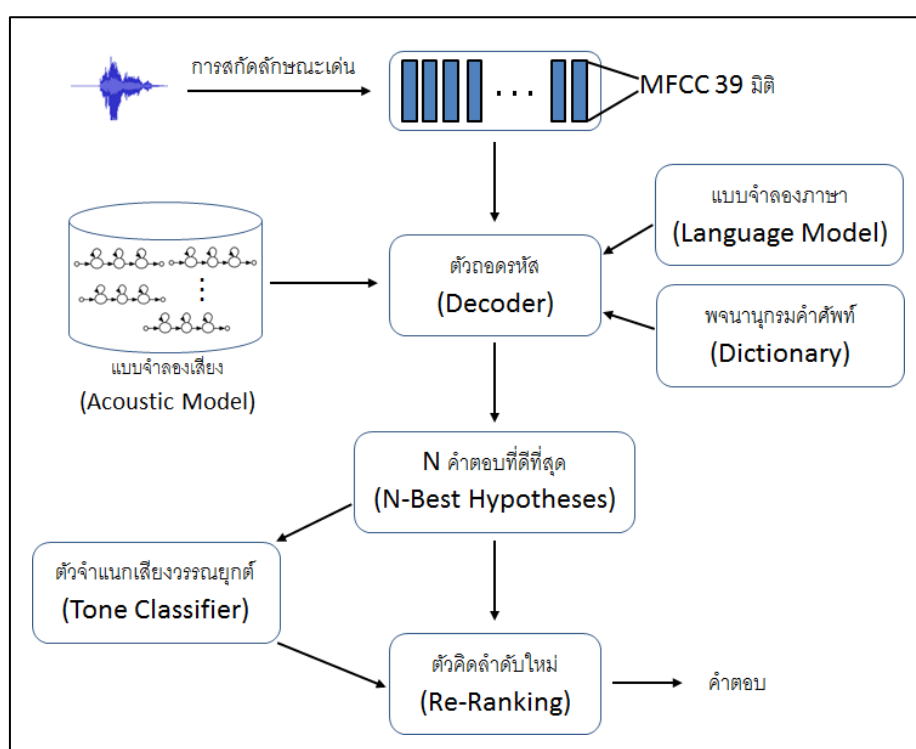
สถานะ \ มิกเจอร์	ความแม่นยำ (%)					
	1	2	4	8	16	32
3	39.56	45.26	63.26	67.93	70.30	68.59
4	52.15	60.00	74.96	75.63	79.33	76.22
5	58.96	67.56	80.52	81.26	81.41	81.63
6	60.59	76.00	82.22	81.41	83.04	81.33
7	62.30	76.67	81.70	82.00	84.15	81.78
8	65.26	77.26	81.26	80.74	82.96	80.30

จากผลการทดลองพบว่า การเพิ่มจำนวนสถานะของแบบจำลองฮิดเดนมาร์คคอฟและการเพิ่มเกาส์เซียนมิกเจอร์มีส่วนเพิ่มประสิทธิภาพในการรู้จำเสียงพูดเป็นอย่างมากดังที่แสดง แต่เมื่อเพิ่มถึงค่าหนึ่งประสิทธิภาพก็จะเริ่มลดลง ซึ่งพบว่าจำนวนสถานะของแบบจำลองฮิดเดนมาร์คคอฟเท่ากับ 7 สถานะกับจำนวนเกาส์เซียนมิกเจอร์ 16 มิกเจอร์ ให้ผลสูงสุดในการรู้จำเสียงพูดการสะกดคำในภาษาไทย โดยได้ความแม่นยำที่ ร้อยละ 84.15

2. การทดลองบนระบบที่นำเสนอที่มีการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงการสะกดคำภาษาไทย

ระบบที่นำเสนอของการรู้จำเสียงการสะกดคำภาษาไทยสำหรับคำโดดจะเป็นการพัฒนาเพิ่มเติมต่อจากระบบข้างอิงด้วยการนำความน่าจะเป็นในการจำแนกเสียงวรรณยุกต์ภาษาไทยมาช่วยในการพิจารณาคะแนนใหม่ (Re-Scoring) โดยระบบที่นำเสนอจะมีภาพรวมดังแสดงในภาพที่

5.3



ภาพที่ 5.3 แผนภาพระบบที่นำเสนอในการรู้จำเสียงการสะกดคำในภาษาไทย

ในการรู้จำเสียงพูดจำเป็นต้องสร้างแบบจำลองเสียง แบบจำลองภาษาและพจนานุกรมคำศัพท์ จากการทดลองของระบบข้างอิงที่ผ่านมาได้การกำหนดค่าพารามิเตอร์การเรียนรู้แบบจำลองที่ดีที่สุด โดยเลือกใช้ค่าสัมประสิทธิ์เชปสตรัมในย่านความถี่ 39 มิติเป็นลักษณะเด่นที่ตั้งจำนวนช่องสัญญาณของชุดกรองสัญญาณเป็น 24 และกำหนดกรอบสัญญาณเป็น แฮมมิงวินโดว โดยที่มีขนาด 25 มิลลิวินาที และกรอบสัญญาณจะเคลื่อนที่ไป 10 มิลลิวินาทีทุกครั้ง

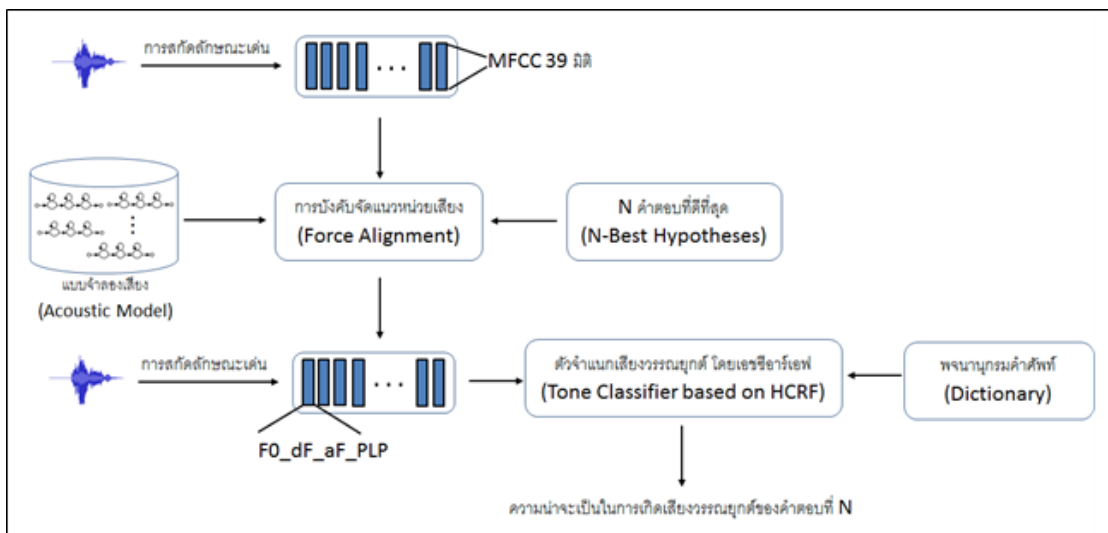
ช่วงเวลา อีกทั้งตั้งค่าพรีเอ็มฟาซิส เป็น 0.97 และกำหนดจำนวนสถานะของแบบจำลองเท่ากับ 7 สถานะ โดยในแต่ละสถานะกำหนดให้มีจำนวนเกาส์เซียนมิกเจอร์เท่ากับ 16 มิกเจอร์

สำหรับแบบจำลองภาษาที่ใช้ในการรู้จำเสียงพูดมีการกำหนดหลักไวยากรณ์เป็นการสะกดคำโดดหนึ่งคำ ซึ่งคำศัพท์ที่สามารถพูดได้จะอยู่ในพจนานุกรมคำศัพท์ โดยในพจนานุกรมมีจำนวนคำศัพท์ทั้งหมด 135 คำ

เมื่อได้แบบจำลองทางเสียง แบบจำลองทางภาษา และพจนานุกรมคำศัพท์แล้วจะนำแบบจำลองดังกล่าวมาใช้ในการรู้จำเสียงพูดโดยการนำสัญญาณเสียงใหม่ที่ต้องการจะรู้จำมาสกัดลักษณะเด่นเป็นค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่ 39 มิติ จากนั้นจะนำแบบจำลองทั้งสอง และพจนานุกรมคำศัพท์รวมถึงสัญญาณเสียงที่ถูกแปลงเป็นลักษณะเด่นเข้าสู่ตัวถอดรหัส ซึ่งสำหรับการทดลองนี้ตัวถอดรหัสจะสร้างสมมุติฐานของคำตอบที่ดีที่สุดที่เป็นไปได้ออกมา N คำตอบ (N-Best Hypotheses) ซึ่งในกรณีนี้พิจารณาคำตอบที่ดีที่สุดเพียงคำตอบเดียว (N = 1) คำตอบจะได้ผลลัพธ์เหมือนกับการทดลองก่อนหน้านี้ แต่เนื่องจากต้องการพิจารณาคำตอบอื่นที่เป็นไปได้มากกว่าคำตอบเดียว โดยใช้ความน่าจะเป็นจากการจำแนกเสียงวรรณยุกต์มาช่วยในการพิจารณาคำตอบใหม่อีกครั้ง จึงกำหนดให้ N มีค่ามากกว่า 1 ซึ่งหากพิจารณาค่า N เท่ากับ 20 พบว่าคำตอบในสมมุติฐานทั้ง 20 สมมุติฐานที่ตั้งไว้จะครอบคลุมคำตอบที่ถูกต้องทั้งหมดไว้แล้ว กล่าวได้ว่าคำตอบที่ถูกต้องทั้งหมดในการรู้จำเสียงพูดจะอยู่ในสมมุติฐานในลำดับที่ไม่เกินสมมุติฐานที่ดีที่สุดที่ 20 จึงไม่มีความจำเป็นที่จะต้องกำหนดค่าของ N ที่มีมากกว่านี้ ดังนั้นในการทดลองจึงกำหนดค่า N เท่ากับ 20

เมื่อได้คำตอบที่ดีที่สุด 20 คำตอบ คำตอบเหล่านั้นจะถูกนำไปใช้ในการคำนวณความน่าจะเป็นในการเกิดเสียงวรรณยุกต์ของคำตอบนั้นๆ ซึ่งมีกระบวนการเป็นไปตาม ภาพที่ 5.4 โดยจะเริ่มจากการหาขอบเขตของพยางค์ก่อนที่จะนำไปคำนวณหาคะแนนความน่าจะเป็นของเสียงวรรณยุกต์ เนื่องจากการจำแนกเสียงวรรณยุกต์จำเป็นต้องทราบขอบเขตของพยางค์ก่อน จึงเลือกใช้การปรับแนวข้อความอัตโนมัติด้วยแบบจำลองฮิดเดนมาร์คอฟซึ่งเรียนรู้แบบจำลองในระดับหน่วยจากข้อมูลชุดฝึกสอน 40 คน โดยกำหนดจำนวนสถานะของแบบจำลองฮิดเดน

มาร์คอฟเป็น 7สถานะ ด้วยจำนวนเกาส์เซียนมิกเจอร์เท่ากับ 16 มิกเจอร์ซึ่งเป็นเมทริกซ์ความแปรปรวนทแยงมุม และใช้ลักษณะเด่นเป็น ค่าสัมประสิทธิ์ซีพสตรัมในย่านความถี่ 39 มิติ และกำหนดกรอบสัญญาณเป็น แฮมมิงวินโดว โดยที่มีขนาด 25 มิลลิวินาที และกรอบสัญญาณจะเคลื่อนที่ไป 10 มิลลิวินาทีทุกครึ่งช่วงเวลา อีกทั้งตั้งค่าพรีเอมฟาซิส เป็น 0.97 ในการจัดเรียงแนวข้อความอัตโนมัติจะทำการนำคำตอบที่ได้จาก N คำตอบที่ดีที่สุดมาจัดเรียงแนวข้อความทีละคำตอบ และเมื่อได้จัดเรียงแนวข้อความเรียบร้อยแล้วผลลัพธ์ที่ได้คือขอบเขตของพยางค์ที่จะนำไปจำแนกเสียงวรรณยุกต์ต่อไป



ภาพที่ 5.4 การคำนวณความน่าจะเป็นของเสียงวรรณยุกต์

หลังจากได้ช่วงระยะเวลาของพยางค์ที่ต้องการจะจำแนกเสียงวรรณยุกต์มาแล้วในลำดับถัดไปจะเป็นการสกัดลักษณะเด่นจากการนำสัญญาณเสียงพูดสำหรับการจำแนกเสียงวรรณยุกต์ ซึ่งในการทดลองนี้จะใช้ลักษณะเด่นชุด F_dF_aF_PLP โดยเลือกใช้ความถี่มูลฐานในสเกลความถี่อัตราอ็อกทาว์ที่มีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานจากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่องก่อนนี้ เนื่องจากในการทดลองการรับรู้เสียงวรรณยุกต์ของมนุษย์พบว่าหากการเสียงวรรณยุกต์ที่มากกว่าหนึ่งพยางค์ สเปกตรัมจะมีส่วนช่วยเป็นอย่างมากในการจำแนกเสียงวรรณยุกต์ภาษาไทยได้ดีขึ้น และได้แสดงให้เห็นในการจำแนกเสียงวรรณยุกต์ต่อเนื่องภาษาไทยด้วยแบบจำลองฮิดเดนคอนดิชันัลแรนดอมฟิลด์แล้วว่าเมื่อมีการเพิ่มค่าของสเปกตรัมจะช่วยให้ประสิทธิภาพในการจำแนกเสียงวรรณยุกต์มีประสิทธิภาพสูงขึ้น

เป็นอย่างมาก ซึ่งสำหรับการสะกดคำในภาษาไทยจะมีการสะกดมากกว่าหนึ่งพยางค์ จึงเลือกที่จะใช้ลักษณะเด่นชุดนี้ในการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับการสะกดคำในภาษาไทย

ตัวจำแนกเสียงวรรณยุกต์ด้วยแบบจำลองฮิดเดนคอนดิชันัลแรนดอมฟิลด์ได้รับการเรียนรู้จากชุดข้อมูลฝึกสอน 40 คน ที่ได้ทำการแบ่งเช่นเดียวกับการทดลองในระบบอ้างอิง โดยในการเรียนรู้ได้มีการใช้เทคนิคปรับแนวข้อความอัตโนมัติเพื่อหาขอบเขตของพยางค์ และใช้ผู้เชี่ยวชาญในการปรับตำแหน่งข้อมูลให้มีความถูกต้องก่อนจะนำข้อมูลในแต่ละช่วงเวลามาเรียนรู้แบบจำลองซึ่งใช้ลักษณะเด่นชุด F_dF_aF_PLP ในสเกลความถี่อัตราฮาร์โมนิกที่มีการใช้การทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน โดยกำหนดค่าพารามิเตอร์ต่างๆเช่นเดียวกับการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับเสียงพูดต่อเนื่อง ซึ่งเมื่อนำตัวจำแนกที่มีการเรียนรู้แล้วไปใช้ในการจำแนกข้อมูลทดสอบที่เหลืออีก 10 คน โดยที่ได้มีการจัดแบ่งช่วงของข้อความไว้เป็นอย่างดีแล้วได้ผลลัพธ์ความถูกต้องในการจำแนกเสียงวรรณยุกต์ที่ร้อยละ 87.79 ซึ่งเป็นที่น่าพอใจ ดังนั้นจะใช้ตัวจำแนกเสียงวรรณยุกต์นี้ในการทดลอง

ในการคำนวณความน่าจะเป็นของเสียงวรรณยุกต์จะทำโดยการนำสัญญาณเสียงในช่วงเวลาที่ปรับแนวข้อความเรียบร้อยแล้วตรงกับช่วงของเสียงวรรณยุกต์ของสมมุติฐานที่ N โดยจะนำลักษณะเด่นในช่วงเวลานั้นและพจนานุกรมคำศัพท์ที่ใช้อ้างอิงเสียงวรรณยุกต์ มาเป็นข้อมูลขาเข้าให้กับตัวจำแนกเสียงวรรณยุกต์ด้วยแบบจำลองฮิดเดนคอนดิชันัลแรนดอมฟิลด์ หลังจากนั้นแบบจำลองจะคำนวณความน่าจะเป็นในการเกิดเสียงวรรณยุกต์ที่กำหนดออกมา

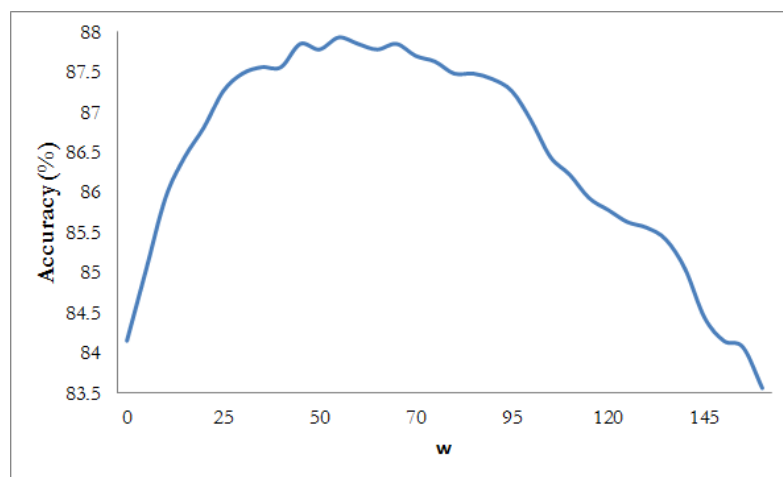
หลังจากได้คำตอบที่ดีที่สุด N คำตอบ และได้ความน่าจะเป็นในการเกิดเสียงวรรณยุกต์ที่กำหนดของทั้ง N คำตอบ ระบบที่เสนอจะทำการคิดคะแนนใหม่ โดยนำค่าความน่าจะเป็นในการเกิดเสียงวรรณยุกต์นั้นมารวมกับความน่าจะเป็นในการเกิดของแต่ละคำตอบใน N คำตอบ โดยสุดท้ายระบบจะเลือกคำตอบที่มีความน่าจะเป็นรวมมากที่สุดโดยเป็นไปตามสมการ 5.1

$$W = \arg \max_w (\log(P(O|W)) + \log(P(W)) + w \log(P(T|W))) \quad (5.1)$$

โดยที่ W คือคำศัพท์ในพจนานุกรมซึ่งมีทั้งหมด 135 คำ O คือสายของข้อมูลที่สังเกตได้ (Observation Sequence) T คือสายของข้อมูลของเสียงวรรณยุกต์ และ $P(O|W)$ คือความ

น่าจะเป็นในการเกิดสายข้อมูลชุดที่สังเกตได้เมื่อกำหนดให้เกิดค่า W และ $P(W)$ คือความน่าจะเป็นของการเกิดค่า W ซึ่งความน่าจะเป็นทั้งสองส่วนนี้จะได้จากระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟ สำหรับ $P(T|W)$ คือความน่าจะเป็นในการเกิดของเสียงวรรณยุกต์เมื่อกำหนดค่าเป็น W ซึ่งในส่วนนี้ได้จากการคำนวณความน่าจะเป็นในการเกิดเสียงวรรณยุกต์ด้วยแบบจำลองฮิดเดนคอนดิชันัลแรนดอมฟิลด์ แต่เนื่องจากความสำคัญระหว่างทั้งสองแบบจำลองมีค่าต่างกัน ดังนั้นจึงกำหนดน้ำหนักที่จะใช้ในการถ่วงความเป็นไปได้รวมโดยใช้ w ซึ่งเป็นค่าน้ำหนักที่ใช้ในการถ่วงค่าความสำคัญของเสียงวรรณยุกต์ เพื่อที่จะให้ได้ประสิทธิภาพที่ดีที่สุดในการรู้จำเสียงพูดการสะกดคำโดดในภาษาไทย ในการทดลองจึงมีการลองปรับค่า w เพื่อหาค่าที่มีความเหมาะสมและให้ประสิทธิภาพที่ดีที่สุด และในการเลือกคำตอบจะเลือกคำตอบที่ให้ผลลัพธ์ของความน่าจะเป็นรวมที่มีค่ามากที่สุดเป็นคำตอบสำหรับการรู้จำเสียงพูด

จากการทดลองการรู้จำเสียงพูดการสะกดคำสำหรับคำโดดด้วยระบบที่นำเสนอ เมื่อมีการปรับน้ำหนักที่ใช้ในการถ่วงค่าความสำคัญของเสียงวรรณยุกต์ตามค่าต่าง พบว่าเป็นไปดังภาพที่ 5.5



ภาพที่ 5.5 กราฟความแม่นยำของการรู้จำเสียงพูดเมื่อปรับค่าน้ำหนัก w ในค่าที่ต่าง ๆ กัน

จากการทดลองพบว่าเมื่อปรับค่าน้ำหนัก w ให้มีค่าเท่ากับ 52 จะได้ผลลัพธ์ในการรู้จำเสียงพูดการสะกดคำโดดภาษาไทยมากที่สุดโดยได้ความแม่นยำในการรู้จำเสียงพูดเท่ากับร้อยละ 87.93 ซึ่งเมื่อเทียบกับความผิดพลาดที่ลดลงจากระบบอ้างอิงจะเท่ากับร้อยละ 23.85 และเมื่อพิจารณากราฟความแม่นยำจะพบว่าในช่วงแรกเมื่อกำหนดให้ค่าน้ำหนัก w มีค่ามากขึ้น ความ

แม่นยำในการรู้จำเสียงจะสูงขึ้น จนได้ความแม่นยำมากที่สุดเมื่อค่าน้ำหนัก w มีค่าเท่ากับ 52 หลังจากนั้นในช่วงหลังความแม่นยำในการรู้จำเสียงพูดจะค่อยๆลดหลั่นลงไปตามลำดับ

วิเคราะห์ผลการทดลอง

ในการทดลองได้ใช้การทดสอบของแมคนีมาร์ในการพิจารณาความแตกต่างของประสิทธิภาพของระบบว่าระบบมีประสิทธิภาพต่างกันขนาดใด โดยเมื่อทำการทดสอบของแมคนีมาร์พบว่าระบบที่นำเสนอมีประสิทธิภาพการรู้จำเสียงภาษาไทยสำหรับการสะกดคำโดดที่ดีว่าอย่างมีนัยสำคัญทางสถิติที่ระดับค่า p เท่ากับ 0.0001

นอกจากนี้จากการทดลองเมื่อพิจารณาค่าน้ำหนัก w พบว่าการปรับค่า w เป็นสิ่งจำเป็นเมื่อลองพิจารณาความน่าจะเป็นในการเกิดคำ W ($P(W)$) และ ความน่าจะเป็นในการเกิดสายข้อมูลชุดที่สังเกตได้เมื่อกำหนดให้เกิดคำ W ($P(O|W)$) จะพบว่าโอกาสในการเกิดทั้งสองน้อยกว่าโอกาสที่จะเกิดความน่าจะเป็นในการเกิดของเสียงวรรณยุกต์เมื่อกำหนดคำเป็น W ($P(T|W)$) เนื่องจากตัวเลือกของเสียงวรรณยุกต์มีเพียง 5 ตัว ในขณะที่ตัวเลือกของคำที่จะเกิดขึ้นทั้งหมดมี 135 คำ อีกทั้งเมื่อพิจารณาโอกาสสังเกตเห็นสายข้อมูลด้วยแล้วยังยิ่งน้อยใหญ่โดยปกติ ในการพิจารณาค่าความน่าจะเป็นจึงพิจารณาความน่าจะเป็นในสเกลลอการิทึมที่มีฐานธรรมชาติ แต่ในทางกลับกันความน่าจะเป็นในการจำแนกเสียงวรรณยุกต์ไม่ได้มีขนาดเล็กเท่า ดังนั้นตัวถ่วงน้ำหนัก w จึงจำเป็นและจากการทดลองพบว่า w ที่มีค่า 52 ให้ผลในการรู้จำเสียงพูดมากที่สุด ซึ่งเป็นค่าที่สมดุลกันระหว่างความน่าจะเป็นของการเกิดคำ และความน่าจะเป็นในการเกิดเสียงวรรณยุกต์ ซึ่งเมื่อเพิ่มค่าน้ำหนักมากกว่านี้จะเห็นได้ว่าค่าความน่าจะเป็นลดลงเนื่องจากน้ำหนักของความน่าจะเป็นในการเกิดเสียงวรรณยุกต์มากเกินไปจนทำให้ ความน่าจะเป็นของการเกิดคำนั้นมีความสำคัญน้อยลงส่งผลให้ประสิทธิภาพลดต่ำลง ดังนั้นในการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์จำเป็นต้องคำนึงถึงความเหมาะสมของค่าน้ำหนัก w ด้วย

เมื่อพิจารณาการสะกดคำที่ไม่มีส่วนขยายที่มีเสียงของพยัญชนะและสระเป็นเสียงเดียวกัน ในการสะกดคำภาษาไทยจะได้คู่ของการสะกดที่มีเสียงของพยัญชนะและสระเป็นเสียงเดียวกันแต่แตกต่างกันในเสียงวรรณยุกต์ ดังแสดงในตารางที่ 5.3

การสะกดคำในภาษาไทยจะมีฐานเสียงวรรณยุกต์ 2 เสียงได้แก่เสียงสามัญ และเสียงจัตวา เมื่อพิจารณาความสับสนในการรู้จำเสียงระหว่างคำที่มีเสียงพยัญชนะ และสระเหมือนกัน แต่ต่างเสียงวรรณยุกต์ โดยวัดความสับสนของคู่คำนั้นเป็นร้อยละความสับสนในการรู้จำเสียงพูด จากคำที่เป็นเสียงสามัญไปเป็นเสียงจัตวา และคำที่มีเสียงจัตวาเป็นเสียงสามัญ จะได้ผลลัพธ์ ดังที่แสดงในตาราง 5.3 โดยพบว่าการนำคะแนนความน่าจะเป็นของเสียงวรรณยุกต์มาช่วยในการรู้จำเสียงพูดการสะกดคำพยัญชนะที่ไม่มีส่วนขยายพบว่า ระบบที่เสนอสามารถจัดการคู่ของความสับสนที่เกิดขึ้นได้อย่างมีประสิทธิภาพมากกว่าระบบอ้างอิงอย่างเห็นได้ชัด

ตารางที่ 5.3 ความสับสนของคู่คำที่มีเสียงพยัญชนะและสระเดียวกันของระบบอ้างอิงและระบบที่เสนอ

คู่คำที่เสียงพยัญชนะและสระเดียวกัน		ความสับสนในการรู้จำเสียงของคู่คำ (%)	
สามัญ	จัตวา	ระบบอ้างอิง	ระบบที่นำเสนอ
ช, ฉ	จ	3.3	0.0
ช	ช, ศ, ส,	5.0	0.0
พ, ภ	ผ	10.0	0.0
ท, ฒ, ท, ฑ	ฐ, ถ	11.7	0.0
ค, ต, ษ	ช, ฌ	24.0	2.0
ฮ	ห	25.0	5.0
ฟ	ฝ	25.0	15.0

นอกจากนี้จากการพิจารณาคำที่มีการรู้จำที่ผิดพลาดมากกว่าร้อยละ 40 ของทั้งสองระบบ พบว่าเป็นไปตามตารางที่ 5.4 ซึ่งจากตารางสิ่งที่เห็นได้เด่นชัดอย่างแรกคือคำที่มีความผิดพลาดในการรู้จำเกินร้อยละ 40 ลดลงไปเป็นอย่างมากในระบบที่นำเสนอ และจะเห็นว่าในการรู้จำตัวสะกดภาษาไทยที่ไม่มีส่วนขยายสำหรับระบบที่นำเสนอจะได้ความแม่นยำที่สูงขึ้นอย่างเห็นได้ชัดเห็นได้จากความผิดพลาดในการรู้จำการสะกดในส่วนที่ไม่คำขยายลดลงเป็นอย่างมากจนบางคำ ออกจากรายการที่มีความผิดพลาดมากกว่าร้อยละ 40 แต่ก็เป็นที่น่าสังเกตว่าความผิดพลาดในการรู้จำคำภาษาไทยที่มีส่วนขยายมีค่าเพิ่มขึ้นเล็กน้อย ซึ่งเมื่อพิจารณาข้อผิดพลาดตรงนี้จะพบว่า

คำว่า “๓ นางมณโฑ” จะผิดพลาดไปเป็น “๓ มณโฑ” คำว่า “๓ ทหาร” จะผิดพลาดไปเป็น “ฐฐาน” และ คำว่า “๓ ศาลา” จะผิดพลาดไปเป็น “๓ ฝา” จะพบว่าเสียงวรรณยุกต์ไม่มีส่วนช่วยให้ความถูกต้องมากขึ้น เนื่องจากคำที่ผิดไปเป็นคำที่มีเสียงวรรณยุกต์เดียวกัน นอกจากนี้สังเกตได้ว่าความผิดพลาดจะเป็นคำที่มีเสียงใกล้เคียงกันแต่ความยาวพยางค์แตกต่างกัน เป็นที่น่าสังเกตว่าระบบจะเลือกคำที่มีความยาวสั้นกว่า

ตารางที่ 5.4 คำที่มีการรู้จำที่ผิดพลาดมากกว่าร้อยละ 40 ของระบบอ้างอิงและระบบที่เสนอ

ระบบอ้างอิง		ระบบที่เสนอ	
คำ	ความผิดพลาด	คำ	ความผิดพลาด
ผ	80%	๓ นางมณโฑ	70%
สระ อี	70%	ผ	60%
ฐ, ถ	60%	อ	40%
ค, ต, ฌ	50%	๓ ทหาร	40%
๓ นางมณโฑ	50%	๓ ศาลา	40%
ข, ฃ	40%		
ช	40%		
ฟ	40%		
อ	40%		
สระ ใอ ไม้ม้วน	40%		
อา	40%		
เอก	40%		

ดังนั้นในการแก้ปัญหาในลักษณะที่จำนวนพยางค์มีผลต่อการรู้จำเสียงการสะกดคำโดดภาษาไทยจำเป็นต้องพิจารณาถึงแบบจำลองที่แสดงถึงความยาวของช่วงระยะเวลา (Duration model) เนื่องจากถ้ามีแบบจำลองดังกล่าวจะช่วยควบคุมจำนวนพยางค์ที่มีมากเกินไป หรือมีน้อยเกินไปได้

บทที่ 6

สรุปผลการวิจัย และข้อเสนอแนะ

สรุปผลการวิจัย

วิทยานิพนธ์นี้ทำการศึกษาเกี่ยวกับการจำแนกเสียงวรรณยุกต์ภาษาไทย โดยได้เสนอการทดลองเกี่ยวกับการรับรู้เสียงวรรณยุกต์ของมนุษย์ การจำแนกเสียงวรรณยุกต์ และการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูด จากการทดลองแต่ละการทดลองสามารถสรุปผลได้ดังต่อไปนี้

การทดลองการรับรู้เสียงวรรณยุกต์ของมนุษย์

การทดลองการรับรู้เสียงวรรณยุกต์ของมนุษย์ เริ่มต้นจากการเตรียมสิ่งเร้าในรูปแบบต่างๆ เพื่อใช้ในการทดสอบความสามารถในการจำแนกเสียงวรรณยุกต์ ซึ่งในการศึกษานี้จะให้ความสำคัญกับความถี่มูลฐาน พลังงาน และลักษณะทางสเปกตรัมของสัญญาณเสียงพูด ในการทดลองได้ทำการศึกษาสิ่งเร้าตามจำนวนพยางค์ซึ่งแบ่งออกเป็น 3 ชนิดได้แก่ สิ่งเร้าในรูปแบบหนึ่งพยางค์ สิ่งเร้าในรูปแบบสองพยางค์ และสิ่งเร้าในรูปแบบสามพยางค์ จากการทดลองพบว่าสิ่งเร้าในรูปแบบหนึ่งพยางค์มีความถูกต้องมากที่สุดและความถูกต้องจะลดหลั่นลงมาตามจำนวนพยางค์ที่เพิ่มขึ้น ซึ่งกล่าวได้อีกนัยหนึ่งว่าความสามารถของมนุษย์ในการจำแนกเสียงวรรณยุกต์จะลดลงเมื่อเสียงพูดมีความต่อเนื่องมากยิ่งขึ้น จากการทดลองยังพบว่าเฉพาะค่าความถี่มูลฐานให้ความสามารถในการรับรู้เสียงวรรณยุกต์ที่มีความถูกต้องในระดับหนึ่ง แต่ก็ยังไม่สามารถเทียบเท่าเสียงต้นแบบ ดังนั้นในการทดลองแรกจึงได้ศึกษาเกี่ยวกับพลังงานของสัญญาณเสียงพูดทำได้โดยการนำรูปร่างของพลังงานของคำต้นแบบมาประยุกต์ในเสียงความถี่มูลฐานที่สังเคราะห์ขึ้น ผลที่ได้พบว่าในระดับหนึ่งพยางค์มีความถูกต้องเพิ่มมากขึ้นเป็นอย่างมาก แต่ในระดับสองและสามพยางค์นั้นผลที่ได้มีค่าในช่วงเดิม ดังนั้นในการทดลองแรกจึงได้ข้อสรุปว่าพลังงานมีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ในกรณีที่เป็นคำโดด แต่ในกรณีที่เป็นเสียงพูดต่อเนื่องพลังงานไม่มีส่วนช่วยในการรับรู้เสียงวรรณยุกต์ ในการทดลองที่สองจะเป็นการศึกษาลักษณะทางสเปกตรัมของสัญญาณเสียงที่มีผลต่อการรับรู้เสียงวรรณยุกต์ โดยในการทดลองได้ใช้วงจรกรองความถี่ต่ำกรอง

ความถี่ในช่วงต่างๆ 3 ช่วง ได้แก่ ช่วงสัญญาณที่ปราศจากความถี่ฟอร์แมนต์ที่หนึ่ง, ช่วงสัญญาณที่ปราศจากความถี่ฟอร์แมนต์ที่สอง และช่วงสัญญาณที่ปราศจากความถี่ฟอร์แมนต์ที่สาม พบว่า ช่วงสัญญาณที่ปราศจากความถี่ฟอร์แมนต์ที่หนึ่งของทั้งสามรูปแบบพยางค์จะให้การรับรู้เสียงวรรณยุกต์มีค่าเช่นเดียวกับการรับรู้เสียงวรรณยุกต์โดยความถี่มูลฐาน แต่เมื่อพิจารณาช่วงสัญญาณที่ปราศจากความถี่ฟอร์แมนต์ที่สองและช่วงสัญญาณที่ปราศจากความถี่ฟอร์แมนต์ที่สามของทั้งสามรูปแบบพยางค์พบว่าความสามารถในการจำแนกเสียงวรรณยุกต์เพิ่มขึ้นอย่างมีนัยสำคัญ ดังนั้นในการทดลองที่สองจึงสรุปว่าสเปกตรัมในช่วงความถี่ฟอร์แมนต์ที่หนึ่งเป็นต้นไปมีผลต่อการรับรู้เสียงวรรณยุกต์ภาษาไทย

การจำแนกเสียงวรรณยุกต์

การทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยแบ่งออกเป็น 2 ส่วน ได้แก่ การทดลองจำแนกเสียงวรรณยุกต์ด้วยความถี่มูลฐาน และการทดลองการจำแนกเสียงวรรณยุกต์ด้วยลักษณะเด่นเพิ่มเติมอื่น ซึ่งการทดลองทั้งสองทดลองบนฐานข้อมูลเสียงคำโดดและฐานข้อมูลเสียงพูดต่อเนื่อง โดยในการทดลองการจำแนกเสียงวรรณยุกต์ด้วยความถี่มูลฐาน ได้ทำการทดลองโดยใช้ระบบอ้างอิงที่ใช้โครงข่ายประสาทเทียม [6, 7] และระบบอ้างอิงที่ใช้แบบจำลองฮิดเดนมาร์คอฟ [8] ในการจำแนกเสียงวรรณยุกต์ ซึ่งจากการทดลองพบว่าระบบอ้างอิงที่ดีที่สุดคือระบบอ้างอิงที่ใช้โครงข่ายประสาทเทียมที่มีการเปลี่ยนแปลงความถี่เป็นความถี่อัตราฮาร์โมนีที่ใช้วิธีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐานและใช้ลักษณะเด่นชุด 5F_5dF โดยความถูกต้องในการจำแนกเสียงวรรณยุกต์ในฐานข้อมูลเสียงคำโดดและฐานข้อมูลเสียงพูดต่อเนื่องเท่ากับ 93.48% และ 67.05% ตามลำดับ และสำหรับระบบที่นำเสนอโดยใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟีลด์มีความถูกต้องในการรู้จำมากที่สุดเมื่อใช้ชุดลักษณะเด่น F_dF_aF และใช้การเปลี่ยนแปลงความถี่เป็นความถี่อัตราฮาร์โมนีที่มีวิธีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน ซึ่งความถูกต้องในการจำแนกเสียงวรรณยุกต์ในฐานข้อมูลเสียงคำโดดและฐานข้อมูลเสียงพูดต่อเนื่องเท่ากับ 94.24% และ 71.01% ตามลำดับ ซึ่งพบว่าได้ความถูกต้องกว่าระบบอ้างอิง สำหรับในการทดลองส่วนที่สองจะเป็นการทดลองการจำแนกเสียงวรรณยุกต์ด้วยลักษณะเด่นเพิ่มเติมอื่น โดยในการทดลองนี้ได้ใช้ระบบที่นำเสนอที่ดีที่สุดในการทดลองก่อนหน้านี้เป็นระบบอ้างอิง และทำการทดลองโดยเพิ่ม

ลักษณะเด่นเพิ่มเติมอื่น ซึ่งการเพิ่มลักษณะเด่นของทั้งการทดลองบนฐานข้อมูลคำโดดและฐานข้อมูลเสียงพูดต่อเนื่องได้พิจารณาเพิ่มเติมลักษณะเด่นตามการทดลองการรับรู้เสียงวรรณยุกต์ของมนุษย์ โดยในการทดลองบนฐานข้อมูลเสียงคำโดดได้พิจารณาเพิ่มค่าพลังงานเป็นลักษณะเด่นเพิ่มเติม ส่วนในการทดลองบนฐานข้อมูลเสียงพูดต่อเนื่องได้ทดลองเพิ่มค่าสเปกตรัมของสัญญาณเสียงพูดซึ่งได้เลือกใช้ค่า 3 อย่างที่จะใช้ในศึกษาได้แก่ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น, ค่าการประมาณพันธะแบบอิงการรับฟังของมนุษย์ และค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมล จากการทดลองบนฐานข้อมูลเสียงคำโดดพบว่าจากระบบที่นำเสนอได้ความถูกต้องที่ 94.24% เมื่อทำการเพิ่มลักษณะเด่นพลังงานได้ความถูกต้องในการจำแนกเสียงวรรณยุกต์เพิ่มขึ้นเป็น 95.53% ส่วนการทดลองจำแนกเสียงวรรณยุกต์บนฐานข้อมูลเสียงพูดต่อเนื่องพบว่าจากเดิมระบบที่นำเสนอได้ความถูกต้องที่ 71.01% แต่เมื่อเพิ่มลักษณะเด่นทางสเปกตรัมของสัญญาณพบว่าความถูกต้องในการจำแนกมีความถูกต้องเพิ่มขึ้นโดยที่ได้ความถูกต้องมากที่สุดเมื่อเลือกใช้ค่าการประมาณพันธะแบบอิงการรับฟังของมนุษย์ซึ่งได้ความถูกต้องที่ 75.04% ในขณะที่เมื่อเลือกใช้ค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่เมลจะได้อัตราความถูกต้องที่ 74.91% และจะได้อัตราความถูกต้องเพิ่มขึ้นน้อยที่สุดเมื่อใช้ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นซึ่งจะได้ค่าความถูกต้องที่ 72.08% ดังนั้นในการทดลองทั้งสองสามารถสรุปได้ว่าการเพิ่มลักษณะเด่นเพิ่มเติมที่เหมาะสมสามารถที่จะเพิ่มประสิทธิภาพและความถูกต้องในการจำแนกเสียงวรรณยุกต์ภาษาไทยได้ซึ่งสอดคล้องกับการทดลองการรับรู้เสียงวรรณยุกต์ภาษาไทย

การประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูด

การประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูดภาษาไทยเป็นการทดลองเพื่อนำระบบจำแนกเสียงวรรณยุกต์มาใช้ในการเพิ่มประสิทธิภาพของระบบรู้จำเสียงพูด ซึ่งทำได้โดยการใช้วิธีคิดคะแนนความน่าจะเป็นใหม่โดยนำความน่าจะเป็นในการเกิดเสียงวรรณยุกต์นั้นไปพิจารณาด้วย ในการทดลองการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูดภาษาไทยจะดำเนินการทดลองบนการรู้จำเสียงพูดภาษาไทยในโดเมนการสะกดคำโดดภาษาไทย เนื่องจากการสะกดคำโดดในภาษาไทยมีคำหลายคู่ที่เสียงพยัญชนะ และเสียงสระเหมือนกัน แต่ต่างกันที่เสียงวรรณยุกต์ซึ่งทำให้ระบบรู้จำเสียงพูดภาษาไทยโดยทั่วไปเกิดความสับสนในคำ

เหล่านี้ การทดลองการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูดภาษาไทยจะเริ่มจากการสร้างระบบการรู้จำเสียงพูดที่เหมาะสมกับการสะกดคำ โดยในการสร้างระบบรู้จำเสียงพูดได้เลือกใช้แบบจำลองฮิดเดนมาร์คอฟ ซึ่งใช้ลักษณะเด่นเป็นค่าสัมประสิทธิ์เซปสตรัมในย่านความถี่ 39 มิติ มาสร้างระบบรู้จำเสียงพูดในระดับคำ ซึ่งจำเป็นต้องพิจารณาค่าพารามิเตอร์ให้เหมาะสม โดยค่าพารามิเตอร์ที่สนใจได้แก่ จำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ และจำนวนเกาส์เซียนมิกเจอร์ จากการทดลองพบว่าจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ และจำนวนเกาส์เซียนมิกเจอร์ที่เหมาะสมที่สุดคือ 7 และ 16 ตามลำดับ ซึ่งด้วยค่าดังกล่าวให้ความแม่นยำในการรู้จำเสียงพูด 84.15% ในลำดับถัดมาจะเป็นการสร้างระบบการจำแนกเสียงวรรณยุกต์ภาษาไทย โดยได้เลือกใช้ระบบจำแนกเสียงวรรณยุกต์สำหรับเสียงพูดต่อเนื่อง เนื่องจากในการสะกดคำจะพบคำที่มีมากกว่าหนึ่งพยางค์ ดังนั้นจึงเลือกใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ และใช้การเปลี่ยนสเกลความถี่เป็นความถี่อัตราอาร์บีทีมีวิธีการทำให้เป็นบรรทัดฐานด้วยค่ามาตรฐาน อีกทั้งเพิ่มชุดลักษณะเด่นด้วยค่าการประมาณพันธะแบบอิงการรับฟังของมนุษย์ ทำให้การจำแนกเสียงวรรณยุกต์มีความถูกต้องเท่ากับ 87.79% สำหรับระบบที่นำเสนอจะนำคำตอบที่ดีที่สุดจำนวน 20 คำตอบมาคำนวณคะแนนใหม่ โดยพิจารณาจากระบบรู้จำเสียงพูดและระบบการจำแนกเสียงวรรณยุกต์ภาษาไทยรวมถึงการปรับค่าน้ำหนักของเสียงวรรณยุกต์พบว่าที่ค่าน้ำหนัก 52 ให้ความถูกต้องในการรู้จำเสียงพูดมากที่สุดเป็น 87.93% และเมื่อพิจารณาผลการทดลองจะเห็นได้ว่าในคู่ที่มีความสับสนของเสียงวรรณยุกต์มีความผิดพลาดลดลงอย่างเด่นชัด ดังนั้นในการทดลองจึงสรุปได้ว่าการนำการจำแนกเสียงวรรณยุกต์ไปใช้ในการรู้จำเสียงพูดการสะกดคำโดดในภาษาไทยจะช่วงเพิ่มประสิทธิภาพในการรู้จำเสียงพูดอย่างมีนัยสำคัญ และลดข้อผิดพลาดที่เกิดจากคำที่มีเสียงพยัญชนะ และเสียงสระเดียวกันแต่ต่างกันที่เสียงวรรณยุกต์ได้อย่างชัดเจน

ในวิทยานิพนธ์นี้ได้ศึกษาลักษณะเด่นของเสียงที่ผลต่อการรับรู้เสียงพูดของมนุษย์ และได้นำลักษณะเด่นเดียวกันนั้นมาเพิ่มประสิทธิภาพในการจำแนกเสียงวรรณยุกต์ทั้งในเสียงพูดที่เป็นคำโดด และเสียงพูดต่อเนื่อง นอกจากนี้ในการทดลองการจำแนกเสียงวรรณยุกต์ยังศึกษาและประยุกต์ใช้แบบจำลองฮิดเดนคอนดิชันนัลแรนดอมฟิลด์ ซึ่งเป็นแบบจำลองใหม่ที่ยังไม่เคย

นำมาใช้ในปัญหาในการจำแนกเสียงวรรณยุกต์ทำให้ประสิทธิภาพในการรู้จำเสียงวรรณยุกต์เพิ่มขึ้น อีกทั้งในวิทยานิพนธ์นี้ได้ทำการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูดภาษาไทยซึ่งทำให้การรู้จำเสียงพูดมีประสิทธิภาพสูงขึ้นอีกด้วย

ข้อเสนอแนะ

จากการทดลองการจำแนกเสียงวรรณยุกต์ภาษาไทยสำหรับฐานข้อมูลเสียงคำโดดพบว่าได้ประสิทธิภาพในระดับสูง แต่อย่างไรก็ตามเมื่อทดลองบนฐานข้อมูลเสียงพูดต่อเนื่องแล้วประสิทธิภาพลดลงอย่างเห็นได้ชัด ปัจจัยหนึ่งมาจากผลกระทบของเสียงวรรณยุกต์ของคำข้างเคียง ซึ่งส่งผลให้รูปร่างของความถี่มูลฐานของคำที่สนใจเปลี่ยนแปลงไป ดังนั้นในการทดลองในอนาคตจึงควรพิจารณาผลของคำข้างเคียงที่มีผลต่อการรู้จำเสียงวรรณยุกต์ อีกทั้งยังควรพิจารณาคำถึงปัจจัยอื่นที่ส่งผลกระทบต่อความถี่มูลฐานในระดับที่สูงกว่าพยางค์ เช่นการเน้นหนักเบา รวมไปถึงท่วงทำนองในการพูดของประโยค

ในการทดลองการประยุกต์ใช้การจำแนกเสียงวรรณยุกต์ในการรู้จำเสียงพูดภาษาไทยพบว่าข้อผิดพลาดที่พบหลังจากนำการจำแนกเสียงวรรณยุกต์ไปประยุกต์ใช้คือ คำที่มีเสียงพยัญชนะและเสียงสระรวมถึงเสียงวรรณยุกต์คล้ายกัน แต่ต่างกันในจำนวนพยางค์ จะไม่สามารถรู้จำเสียงของคำเหล่านั้นได้มีประสิทธิภาพ ดังนั้นการทดลองในอนาคตควรพิจารณาแบบจำลองระยะเวลาของคำด้วยเพื่อลดข้อผิดพลาดที่เกิดขึ้น หรืออาจจะทดลองโดยใช้การรู้จำเสียงพูดด้วยระดับหน่วยเสียงย่อยแทนที่จะใช้ระบบรู้จำเสียงพูดในระดับคำ

นอกจากนี้ในการนำการจำแนกเสียงวรรณยุกต์ไปใช้ในการรู้จำเสียงพูดภาษาไทยในสถานการณ์จริงจำเป็นต้องเพิ่มประสิทธิภาพของภาพของระบบให้เร็วกว่านี้เนื่องจากในการทดลองได้ใช้เครื่องมือหลายชนิดในการทดลองทำให้เกิดปัญหาคอขวดในการส่งข้อมูลระหว่างเครื่องมือแต่ละชนิดทำให้ประสิทธิภาพโดยรวมของระบบมีความล่าช้าเกิดขึ้น ดังนั้นเพื่อหลีกเลี่ยงปัญหาดังกล่าวในการนำไปใช้จริง จึงจำเป็นต้องทำการพัฒนาระบบให้มีความสมบูรณ์มากกว่านี้

รายการอ้างอิง

- [1] Apple Inc. Siri [Online].2012.Available from: <http://www.apple.com/ios/siri/> [2012, February 7]
- [2] NUANCE Inc. Dragon Dictation [Online].2012.Available from: <http://www.nuance.com/for-business/by-product/dragon-dictation-iphone> [2012,February 7]
- [3] NECTEC. TVIS Traffic Talk [Online].2012.Available from: http://www.nectec.or.th/index.php?option=com_content&view=article&id=454 [2012,February 7]
- [4] Luksaneeyanawin, S. Intonation in Thai, in D. Hirst and A. D. Cristo (Eds.), Intonation Systems A Survey of Twenty Language, pp. 376–394. Cambridge University Press, 1998.
- [5] Thubthong, N. Kijisirikul, B. and Pusittrakul, A. A method for isolated Thai tone recognition using combination of neural networks. Computational Intelligence (2002) : 313-335.
- [6] Thubthong, N. and Kijisirikul, B. An empirical study for constructing Thai tone models. In Proc. the 5th Symposium on Natural Language Processing and Oriental COCOSDA Workshop (2002) : 179–186.
- [7] Tan, L. Karnjanadecha, M. and Khaorapapong, T. A study of tone classification for continuous Thai speech recognition. In 8th International Conference on Spoken Language Processing. (2004) : 3033-3036.
- [8] Tungthangthum, A. Tone recognition for Thai. In Proc. IEEE Asia-Pacific Conf. Circuits and System (1998) : 157–160.
- [9] Xing, Z. Pei, J. and Keogh, E. A brief survey on sequence classification. In ACM SIGKDD Explorations, Volume 12, Issue 1 (June 2010) : 40-48.
- [10] Gunawardana, A. Mahajan, M. Acero, A. Platt, J. C. Hidden conditional random fields for phone classification, In INTERSPEECH-2005 (2005) : 1117-1120.

- [11] Sung, Y. Jurafsky, D. Hidden Conditional Random Fields for phone recognition In Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop (2009) : 107-112.
- [12] Kasuriya, S., Sornlertlamvanich, V., Cotsomrong, P., Kanokphara, S., Thatphithakkul, N. Thai speech corpus for speech recognition, In International Conference on Speech Databases and Assessments (Oriental-COCOSDA) (2003) : 54-61.
- [13] กาญจนา นาคสกุล. ระบบเสียงภาษาไทย. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2541.
- [14] Kertkeidkachorn, N. Vorapatratom, S. Tangruamsub, S. Punyabukkana, P. Suchato, A. Contribution of Spectral Shapes to Tone Perception In 13th Annual Conference of the International Speech Communication Association. (2012).
- [15] Maneenoi, E., Ahkuputra, V., Luksaneeyanawin, S. and Jitapunkul, S. Acoustic modeling of onset-rhyme for Thai continuous speech recognition. In Proceeding of the 9 Australian International Conference on Speech Science & Technology (2002) 462–467.
- [16] Leelaphattarakij, P. Acoustic-phonetic approaches to improving segment-based speech recognition. Dissertation Ph.D, Faculty of Department of Computer Engineering Faculty of Engineering Chulalongkorn University, 2009.
- [17] วิทยา ไรจน์กิตติเจริญ. พารามิเตอร์ทางเสียงสำหรับการจำแนกลักษณะการเปล่งเสียงในเสียงพูดต่อเนื่องภาษาไทย วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2554.
- [18] Owens F. J., Signal processing of speech, First Published, The MACMILLAN PRESS LTD, 1993

- [19] วิทิต ลิ้มปิยะภิมรณย์ การออกแบบวงจรรวม EPD และ DTW ของระบบรู้จำเสียงแบบขึ้นกับผู้พูด สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, 2546.
- [20] ธีราพร เรือนจำบุญ แนวทางการพัฒนา Speech Recognition กับ ภาษาไทย Trends in Thai Speech Recognition Development, สารสารร่วมพฤษ ปีที่ 28 ฉบับที่ 1 (ตุลาคม 2552 - มกราคม 2553) : 24-52
- [21] Hermansky H. Perceptual linear predictive (PLP) analysis for speech, In the Journal of The Acoustical Society of America 87, (1990) : 1738 -1752
- [22] Gillick, L. and CoxSome, S. J. Some Statistical issues in the comparison of speech recognition algorithms. In proceedings of the 1989 IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP 1989) (1989) : 532-535.
- [23] Maleerat, S. Supot, N. and Choochart H. Tone Classification for Isolated Thai Words using Multi-Layer Perceptron. In Proceedings of the World Congress on Engineering and Computer Science (2009) : 1322-1325.
- [24] Nuttakorn Thubthong. A STUDY OF VARIOUS LINGUISTIC EFFECTS ON TONE RECOGNITION IN THAI CONTINUOUS SPEECH. Ph.D Dissertation, Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, 2001.
- [25] Lee, T. Ching, P. C. Chan, L. W. Cheng, Y. H.; and Mark, B. Tone recognition of isolated Cantonese syllables. IEEE Transactions on Speech Audio Processing 3 (1995) : 204–209.
- [26] Jian, F. H. L. Classification of Taiwanese tones based on pitch and energy movement. In Proc. Int. Conf. Spoken Language Processing, Vol. 2, (1998) : 329–332.
- [27] Tian, Y., Zhou, J.-L., Chu, M., Chang, E. Tone Recognition with Fractionized Models and Outlined Features. In Proc. ICASSP 2004, (2004) : 105-108.
- [28] Nguyen, H. Q. Nocera, P. Castelli, E. and Van Loan T. Tone recognition of Vietnamese continuous speech using hidden markov model. In

- Communications and Electronics, 2008. ICCE 2008. Second International (2008) : 235–239.
- [29] Nguyen, H. Q. Nocera, P. Castelli, E. and Van Loan T. Using tone information for Vietnamese continuous speech recognition. In Research, Innovation and Vision for the Future (RIVF 2008) (2008) : 103-106.
- [30] Wei, H. Wang, X. Wu, H. Luo, D. and Wu X. Exploiting prosodic and lexical features for tone modeling in a conditional random field framework In Acoustics, Speech and Signal Processing(ICASSP) (2008) : 4549-4552.
- [31] Lafferty, J. McCallum, A. and Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, (2001) : 282–289.
- [32] Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H. and Charnvivit, P., Perception of tone and vowel quality in Thai, In ICSLP-2002, (2002) : 753–756.
- [33] Svantesson, J. and House, D., Tone production, tone perception and Kammu tonogenesis, In Phonology 23, (2006) : 309–333.
- [34] Zsiga, E. and Nitisaroj, R., Tone features, tone perception, and peak alignment in Thai, In Language and Speech 50, (2007) : 343-383.
- [35] Li, Y. and Lee, T., Perceptual Equivalence of Approximated Cantonese Tone Contours, In proceedings of Interspeech 2007 (2007) : 2677-2680.
- [36] Lee, T., Lau, W. Wong, Y. W. and Ching P.C. Using tone Information In Cantonese Continuous Speech Recognition ACM Transactions on Asian Language Information Processing (2002) : 83-102.
- [37] Boersma, P. and Weenink, D., Praat 5. 3.11, A system for doing phonetics by computer, 2011.
- [38] Rabiner, L. R. Sambur, M. R. Schmidt, C.E. Applications of Nonlinear Smoothing Algorithm to Speech Processing, In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-23, No. 6, (1975) : 552-557.

- [39] Punyayodhin, S., Misra, D., Yadav, R. and Narang, V., A Study of F1 Correlation with F0 in a Tone Language: Case of Thai, In Asian Language Processing (IALP), (2010) : 330-333.
- [40] Steinberg Media Technologies GmbH. Cubase [Online].2012.Available from: <http://www.steinberg.net/en/products/cubase/start.html> [2012, February 7]
- [41] NECTEC. ISPEECH [Online].2012.Available from:<http://www.nectec.or.th/2008/r-d/hlt.html> [2012, February 7]
- [42] Wutiw WATCHAI, C. Thangthai, K. and Sertsi,P. Thai ASR development for network-based speech translation. In Proc. of O-COCOSDA, (2012).
- [43] Speech, Music and Hearing part of School of Computer Science and Communication. WaveSurfer [Online].2012.Available from:<http://www.speech.kth.se/wavesurfer/> [2012, February 7]
- [44] Boersma P. ACCURATE SHORT-TERM ANALYSIS OF THE FUNDAMENTAL FREQUENCY AND THE HARMONICS-TO-NOISE RATIO OF A SAMPLED SOUND In Proceedings 17 Institute of Phonetic Sciences, University of Amsterdam, (1993) : 97-110.
- [45] Nissen, S. et al. Fast Artificial Neural Network (FANN) Library [Online].2012.Available from: <http://leenissen.dk/fann/wp/> [2012, February 7]
- [46] Young, S., Evermann, G., Galse, M., Kershaw, D., Moore, G. Hidden Markov model toolkit – speech recognition toolkit. [Online]. 2012. Available from: <http://htk.eng.cam.ac.uk> [2012, February 7]
- [47] Morency, L.P. et al. Hidden-state Conditional Random Field (HCRF) Library. [Online]. 2012. Available from: <http://sourceforge.net/projects/hcrf/> [2012, February 7]
- [48] Theodoridis, S. Koutroumbas, K. Pattern Recognition, Academic Press, 2006

- [49] Kasuriya, S. Sornlertlamvanich, V. Cotsomrong, P. Jitsuhiro, T. Kikui, G. and Sagisaka, Y. NECTEC-ATR Thai Speech Corpus, In Proc. of Oriental-COCOSDA-2003 (2003): 105-111
- [50] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, In Proceedings of the IEEE, vol. 77, (1989) : 257-286.

ภาคผนวก

ภาคผนวก ก

ตารางเทียบหน่วยเสียงไทยที่ใช้ในวิทยานิพนธ์

Thai Phoneme

พยัญชนะต้น			
เดี่ยว	ตัวอย่าง	ผสม	ตัวอย่าง
p	ปาก	pr	ประสาน
t	เต็น, กฏี	ph r	พธาน
c	จะ	tr	เตรียม
k	ก่อน	kr	กราบ
z	อาน	kh r	คร่า
ph	พบ, ภัย, ผ่าน	pl	ปลา
th	ทั้ง, ชง, เตา, ลาน, มด, โท	phl	พลาด
ch	ชอบ, เจอ	thr	จันทรา
kh	คน, เขิน, ช้า	kl	กลอ
b	บอก	khl	เคลื่อน
d	ด้าน, ชญา	kw	กวาง
m	ไม้	kh w	ขวา
n	นาน, เกร	เสียงทับศัพท์	
ng	เงิน	br	เบรน
l	เลน, กีฬา	bl	บล
r	รอ, ภัย	fr	ฟราย
f	ฝน, ฟัน	fl	เฟลม
s	สาย, ศิลา, รักษา, ซ่อน	dr	ดราคอน
h	โหน, เฮฮา	17 หน่วย	
w	วา		
j	ย่อน, หญิง		
21 หน่วย			

สระ			
เดี่ยว	ตัวอย่าง	ผสม	ตัวอย่าง
a	อะ	ia	เอียะ
aa	อา	iaa	เอีย
i	อี	va	เอือะ
ii	เอี	vva	เอือ
v	อือ	ua	อัวะ
vv	อือ	uua	อิว
u	อุ	6 หน่วย	
uu	อุ		
e	เอะ		
ee	เอ		
x	แอะ		
xx	แอ		
o	โอะ		
oo	โอ		
@	เออะ		
@@	ออ		
q	เออะ		
qq	เออ		
18 หน่วย			

ตัวสะกด	
เดี่ยว	ตัวอย่าง
p^	พบ
t^	เทรีด
k^	ปาก
n^	หาร
m^	ลม
ng^	ฟาง
j^	ยาย
w^	กาวิ
เสียงทับศัพท์	
f^	กราฟ
l^	แอล
s^	เอส
ch^	คลัช
12 หน่วย	

ภาคผนวก ข

คำที่เลือกใช้สำหรับการสะกดคำภาษาไทย

ลำดับ	คำ	รูปหน่วยเสียง
1	ก	k-@@-0
2	ข	kh-@@-4
3	ฃ	kh-@@-4
4	ค	kh-@@-0
5	ค	kh-@@-0
6	ฌ	kh-@@-0
7	ง	ng-@@-0
8	จ	c-@@-0
9	ฉ	ch-@@-4
10	ช	ch-@@-0
11	ซ	s-@@-0
12	ฌ	ch-@@-0
13	ญ	j-@@-0
14	ฎ	d-@@-0
15	ฏ	t-@@-0
16	ฐ	th-@@-4
17	ฑ	th-@@-0
18	ฒ	th-@@-0
19	ณ	n-@@-0
20	ด	d-@@-0
21	ต	t-@@-0
22	ถ	th-@@-4
23	ท	th-@@-0
24	ธ	th-@@-0
25	น	n-@@-0
26	บ	b-@@-0
27	ป	p-@@-0
28	ผ	ph-@@-4

ลำดับ	คำ	รูปหน่วยเสียง
29	ฝ	f-@@-4
30	พ	ph-@@-0
31	ฟ	f-@@-0
32	ภ	ph-@@-0
33	ม	m-@@-0
34	ย	j-@@-0
35	ร	r-@@-0
36	ล	l-@@-0
37	ว	w-@@-0
38	ศ	s-@@-4
39	ษ	s-@@-4
40	ส	s-@@-4
41	ห	h-@@-4
42	ฬ	l-@@-0
43	อ	z-@@-0
44	ฮ	h-@@-0
45	ก ไก่	k-@@-0 k-a-j ⁻¹
46	ข ไข่	kh-@@-4 kh-a-j ⁻¹
47	ช ชวด	kh-@@-4 kh-uua-t ⁻¹
48	ค ควาย	kh-@@-0 khw-aa-j ⁻⁰
49	ต คน	kh-@@-0 kh-oo-n ⁻⁰
50	ฌ ระวัง	kh-@@-0 r-a-3 kh-a-ng ⁻⁰
51	ง งู	ng-@@-0 ng-uu-0
52	จ จาน	c-@@-0 c-aa-n ⁻⁰
53	ฉ ฉิ่ง	ch-@@-4 ch-i-ng ⁻¹
54	ช ช้าง	ch-@@-0 ch-aa-ng ⁻³
55	ซ โซ่	s-@@-0 s-oo-2
56	ฌ เฉลอ	ch-@@-0 ch-cq-0

ลำดับ	คำ	รูปหน่วยเสียง
57	ฉ กะเฉอ	ch-@@-0 k-a-1 ch-qq-0
58	ญ หญิง	j-@@-0 j-i-ng^-4
59	ฎ ฎา	d-@@-0 ch-a-3 d-aa-0
60	ฏ ปฏัก	t-@@-0 p-a-1 t-a-k^-1
61	ฐ ฐาน	th-@@-4 th-aa-n^-4
62	ฑ มณโฑ	th-@@-0 m-o-n^-0 th-oo-0
63	ฑ นางมณโฑ	th-@@-0 n-aa-ng^-0 m-o-n^-0 th-oo-0
64	ฒ ผู้เฒ่า	th-@@-0 ph-uu-2 th-a-w^-2
65	ณ เณร	n-@@-0 n-ee-n^-0
66	ด เด็ก	d-@@-0 d-e-k^-1
67	ต เต่า	t-@@-0 t-a-w^-1
68	ถ ถุง	th-@@-4 th-u-ng^-4
69	ท ทหาร	th-@@-0 th-a-3 h-aa-n^-4
70	ธ ธง	th-@@-0 th-o-ng^-0
71	น หนู	n-@@-0 n-uu-4
72	บ ใบไม้	b-@@-0 b-a-j^-0 m-aa-j^-3
73	ป ปลา	p-@@-0 pl-aa-0
74	ผ ผึ้ง	ph-@@-4 ph-v-ng^-2
75	ฝ ฝา	f-@@-4 f-aa-4
76	พ พาน	ph-@@-0 ph-aa-n^-0
77	ฟ ฟัน	f-@@-0 f-a-n^-0
78	ภ สำเภา	ph-@@-0 s-a-m^-4 ph-a-w^-0
79	ม ม้า	m-@@-0 m-aa-3
80	ย ยักษ์	j-@@-0 j-a-k^-3
81	ร เรือ	r-@@-0 r-vva-0
82	ล ลิง	l-@@-0 l-i-ng^-0
83	ว แหวน	w-@@-0 w-xx-n^-4
84	ศ ศาลา	s-@@-4 s-aa-4 l-aa-0

ลำดับ	คำ	รูปหน่วยเสียง
85	ษ ฤษี	s-@@-4 r-vv-0 s-ii-4
86	ส เสือ	s-@@-4 s-vva-4
87	ห หีบ	h-@@-4 h-ii-p^-1
88	ฬ จุฬา	l-@@-0 c-u-1 l-aa-0
89	จ ฉ่าง	z-@@-0 z-aa-ng^-1
90	ฮ นกฮูก	h-@@-0 n-oo-k^-3 h-uu-k^-2
91	สระ อะ	s-a-1 r-a-1 z-a-1
92	สระ อา	s-a-1 r-a-1 z-aa-0
93	สระ อิ	s-a-1 r-a-1 z-i-1
94	สระ อี	s-a-1 r-a-1 z-ii-0
95	สระ อึ	s-a-1 r-a-1 z-v-0
96	สระ อือ	s-a-1 r-a-1 z-vv-0
97	สระ อุ	s-a-1 r-a-1 z-u-1
98	สระ อู	s-a-1 r-a-1 z-uu-0
99	สระ เอ	s-a-1 r-a-1 z-ee-0
100	สระ แอ	s-a-1 r-a-1 z-xx-0
101	สระ โอ	s-a-1 r-a-1 z-oo-0
102	สระ อัม	s-a-1 r-a-1 z-a-m^-0
103	สระ ไอ ไม้ม้วน	s-a-1 r-a-1 z-a-j^-0 m-a-j^-4 m-uua-n^-3
104	สระ ไอ	s-a-1 r-a-1 z-a-j^-0
105	สระ ไอ ไม่มลาย	s-a-1 r-a-1 z-a-j^-0 m-a-j^-4 m-a-0 l-a-j^-0
106	อะ	z-a-1
107	อา	z-aa-0
108	อิ	z-i-1
109	อี	z-ii-0
110	อึ	z-v-0
111	อือ	z-vv-0
112	อุ	z-u-1

ลำดับ	คำ	รูปหน่วยเสียง
113	คู	z-uu-0
114	เอ	z-ee-0
115	แอ	z-xx-0
116	โอ	z-oo-0
117	อ๋า	z-a-m [^] -0
118	ไอ	z-a-j [^] -0
119	ฤ	r-v-0
120	ฤา	r-vv-0
121	ไม้เอก	m-aa-j [^] -3 z-ee-k [^] -1
122	ไม้โท	m-aa-j [^] -3 th-oo-0
123	ไม้ตรี	m-aa-j [^] -3 tr-ii-0
124	ไม้จัตวา	m-aa-j [^] -3 c-a-t [^] -1 t-a-1 w-aa-0
125	เอก	z-ee-k [^] -1
126	โท	th-oo-0
127	ตรี	tr-ii-0
128	จัตวา	c-a-t [^] -1 t-a-1 w-aa-0
129	ไม้ไต่คู้	m-aa-j [^] -3 t-a-j [^] -1 kh-uu-3
130	การ์ตัน	k-aa-0 r-a-n [^] -0
131	ไปยาลน้อย	p-a-j [^] -0 j-aa-n [^] -0 n-@@-j [^] -3
132	ไปยาลใหญ่	p-a-j [^] -0 j-aa-n [^] -0 j-@@-j [^] -1
133	ไม้ยมก	m-a-j [^] -3 j-a-3 m-o-k [^] -3
134	พินทุ	p-i-n [^] -0 t-u-3
135	นิคหิต	n-i-3 kh-a-3 h-i-t [^] -1

ประวัติผู้เขียนวิทยานิพนธ์

นายณัฐวุฒิ เกิดเกียรติขจร เกิดวันที่ 22 พฤษภาคม พ.ศ.2532 ที่จังหวัดกรุงเทพมหานคร เข้าศึกษาระดับประถมศึกษาโรงเรียนพันธุวัฒนา จังหวัดกรุงเทพมหานคร สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นและมัธยมศึกษาตอนปลายที่ โรงเรียนเทพศิรินทร์ จังหวัดกรุงเทพมหานคร จากนั้นในปี 2550 ได้เข้าศึกษาระดับปริญญาบัณฑิต ในหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จังหวัดกรุงเทพมหานคร และในปี 2554 ได้เข้าศึกษาต่อระดับปริญญาโท สาขาวิชาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จังหวัดกรุงเทพมหานคร