# ระบบฝึกฝนการพัฒนาวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง

# VOCAL THAI TRAINING SYSTEM USING TONE RECOGNITION


ศันต์ อภิวัฒนดังการ

วิษณุวรรณ สุขไชยศรี


ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2553

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

# ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง

# VOCAL THAI TRAINING SYSTEM USING TONE RECOGNITION

T146218

โดย

กันต์ อภิวัฒนลังการ

KAN APIWATTANALANGKAN

วิชชุวรรณ สุขไชยศรี

WICHCHUWON SOOKCHAISRI

อาจารย์ที่ปรึกษา

ดร.สุภวรรณ อันนันหนับ

b.12841080
i.

หมายเลข 146218
ลงทะเบียน
วัน เดือน ปี 25 เม.ย. 2560

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2558

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง

# VOCAL THAI TRAINING SYSTEM USING TONE RECOGNITION

โดย


กันต์ อภิวัฒนลังการ

วิชชุวรรณ สุขไชยศรี

อาจารย์ที่ปรึกษา

ดร.สุภวรรณ อันนันหนับ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2558


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# VOCAL THAI TRAINING SYSTEM USING TONE RECOGNITION

KAN APIWATTANALANGKAN

WICHCHUWON SOOKCHAISRI

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

BACHELOR OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/2015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**COPYRIGHT 2016**

**FACULTY OF INFORMATION TECHNOLOGY**

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ใบรับรองปริญญานิพนธ์ ประจำปีการศึกษา 2558

## คณะเทคโนโลยีสารสนเทศ

## สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

**เรื่อง** ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง

**VOCAL THAI TRAINING SYSTEM USING TONE RECOGNITION**

**ผู้จัดทำ**

1. นายกันต์ อภิวัฒนลังการ รหัสนักศึกษา 55070006
2. นางสาววิชชุวรรณ สุขไชยศรี รหัสนักศึกษา 55070106

.....อาจารย์ที่ปรึกษา

(ดร.สุภวรรณ อันนันหนับ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

| | |
|---|---|
| **หัวข้อโครงงาน** | ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง |
| **นักศึกษา** | นายกันต์ อภิวัฒนลังการ รหัสนักศึกษา 55070006 |
| | นางสาววิชชุวรรณ สุขไชยศรี รหัสนักศึกษา 55070106 |
| **ปริญญา** | วิทยาศาสตรบัณฑิต |
| **สาขาวิชา** | เทคโนโลยีสารสนเทศ |
| **ปีการศึกษา** | 2558 |
| **อาจารย์ที่ปรึกษา** | ดร.สุภวรรณ อันนันหนับ |

## บทคัดย่อ

การออกเสียงวรรณยุกต์มีความสำคัญในการสื่อสารของภาษาไทย เพราะวรรณยุกต์ทำให้เกิดคำใหม่และมีความหมายที่แตกต่างกันออกไป ปัจจุบันชาวต่างชาติจำนวนไม่น้อยที่สนใจที่จะเรียนรู้ภาษาไทย ดังนั้นโครงการนี้จึงมีวัตถุประสงค์เพื่อช่วยในการฝึกฝนการออกเสียงวรรณยุกต์ไทยขึ้นโดยใช้เทคนิคการรู้จำเสียง ทางคณะผู้จัดทำได้ทำการเก็บข้อมูลเสียงเพื่อนำไปใช้เป็นข้อมูลฝึกสอนและสร้างแบบจำลองอ้างอิงของเสียงวรรณยุกต์ทั้ง 5 เสียง ด้วยวิธี Hidden Markov Model (HMM)

ภายในระบบจะมีฟังก์ชันหลักๆ 2 ฟังก์ชัน อย่างแรกคือ แบบทดสอบการฝึกออกเสียงวรรณยุกต์ให้ตรงกับแบบทดสอบที่เตรียมไว้ และอย่างที่สองคือประมวลผลเสียงที่รับเข้าไปว่าเป็นเสียงวรรณยุกต์ใด ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียงนี้ รองรับการทำงานบนคอมพิวเตอร์แพลตฟอร์ม โดยระบบนี้จะทำให้ผู้ใช้มีความสะดวกในการใช้งาน และจะสามารถทำให้ผู้ใช้สามารถเข้าใจและออกเสียงวรรณยุกต์ได้ถูกต้องและชัดเจนมากขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

| | |
|---|---|
| **Project Title** | Vocal Thai Training System using Tone Recognition |
| **Student** | Mr.Kan Apiwattanalangkan Student ID 55070006 |
| | Ms.Wichchuwon Sookchaisri Student ID 55070106 |
| **Degree** | Bachelor of Science |
| **Program** | Information Technology |
| **Academic Year** | 2015 |
| **Advisor** | Dr.Supawan Annanab |


## ABSTRACT

Phonetics are important part in Thai communication because intonation creates new words and creates different meaning. Nowadays, many foreigners are interested in learning Thai. The purpose of this project is to help them practice Thai tones pronunciation by recognizing techniques. Hidden Markov Model (HMM) is an algorithm that used to create training data and reference models of five tones.

There're two functions in the application. First, to practice pronunciation with the sample tests. Second, to differentiate each tone that the user is making. This application supports for computer platform and user friendly. The application helps users to improve and understand different Thai pronunciations.




เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยจากความช่วยเหลืออย่างดียิ่งจากหลาย ๆ ฝ่าย ซึ่งผู้จัดทำใคร่ขอขอบคุณทุก ๆ ท่านที่มีส่วนร่วมในการสนับสนุน ช่วยเหลือ และให้คำแนะนำในทุก ๆ ด้าน

ขอขอบพระคุณ ดร.สุภวรรณ อันนันหนับ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาเสียสละเวลา คอยสอบถามติดตามความคืบหน้า และให้คำแนะนำที่เป็นประโยชน์สำหรับการทำงานตลอดจนให้คำปรึกษาที่ดีเสมอมา จนทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.พิชญ สุพรรณกูล อาจารย์ประจำคณะวิศวกรรมศาสตร์ สาขาวิศวกรรมโทรคมนาคม ที่ช่วยสอน อธิบายเกี่ยวกับเนื้อหาทางวิศวะในหลายๆเรื่อง และให้คำปรึกษารวมถึงให้คำแนะนำที่ดีเสมอมา

ขอกราบขอบพระคุณบิดา มารดา ผู้ให้โอกาสและคอยให้กำลังใจเสมอ

ขอขอบคุณนายณัฐพล หอกกิ่ง ที่ช่วยอธิบาย ให้คำปรึกษาและคำแนะนำทำให้เข้าใจอย่างแจ่มแจ้ง

ขอขอบคุณนายอริยะ เหล่าวนิชชานนท์ นายศิรวิทย์ จันทร์ประภาพ นายธเนศ พุ่มไพจิตร นายวงศธร แซ่เล้า นางสาวฐาพิณภัค หิรัญไชยฤทธิ์ นายสรวุฒิ กิตติคีรีชัยกุล นายสารัช ท้วมชมธรรม นายธนกร มะลิขาว นายธีรวัจน์ นำศรีเจริญสุข และ นายณัฐวุฒิ มาลี ที่ช่วยให้คำปรึกษาและช่วยแก้บั๊กต่างๆ มาโดยตลอด

ขอขอบคุณเสียงจากเพื่อน ๆ รุ่น 10 นางสาวจันทรัสม์ อมฤตกุล นางสาววรรณิศา กำมะหยี่ นางสาวกีต้าร์ ไกรชิต นายพชร ภิญโญพรพาณิชย์ นายบวรวงศ์ สุขปราณี นายยุทธพิชัย ชาญนิตย์ นางสาววศินี ปุจฉาการ นางสาวสาวิตรี บุญสอน นางสาวจิดามาศ แสวงรุจิธรรม นายโสภณ จำเริญกิจ และนายอัฐวุฒิ วิริยะเกรียง ที่เสียสละเวลามาให้อัดเป็นเสียงต้นแบบในการสร้างแบบจำลองและใช้เสียงในการทดสอบแบบจำลอง ทำให้งานวิจัยนี้ผ่านไปได้ด้วยดี



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

หน้า


บทคัดย่อ ..... I

ABSTRACT ..... II

กิตติกรรมประกาศ ..... III

สารบัญ ..... IV

สารบัญตาราง ..... VIII

สารบัญรูป ..... IX

บทที่ 1 บทนำ ..... 1

1.1 ความเป็นมาและความสำคัญของปัญหา ..... 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา ..... 1

1.3 สมมติฐานของการศึกษา ..... 1

1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย ..... 2

1.5 การเปรียบเทียบระหว่างวิธีการที่นำเสนอกับวิธีการแบบพื้นฐาน ..... 2

1.6 ขอบเขตการวิจัย ..... 2

1.7 ขั้นตอนการศึกษา ..... 2

1.8 ประโยชน์ที่คาดว่าจะได้รับ ..... 3

บทที่ 2 วรรณกรรมที่เกี่ยวข้อง ..... 4

2.1 ทฤษฎีการสร้างเสียงพูด ..... 4

2.2 หน่วยเสียงสำคัญในภาษาไทย ..... 7

2.3 หน่วยเสียงสระ ..... 7

2.3.1 ลักษณะของเสียงสระ ..... 7

2.4 หน่วยเสียงพยัญชนะ ..... 8

2.4.1 ลักษณะของเสียงพยัญชนะ ..... 8

2.4.2 หน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย ..... 11

2.5 หน่วยเสียงวรรณยุกต์ ..... 11




เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ (ต่อ)

หน้า


2.5.1 ลักษณะของเสียงวรรณยุกต์ ..... 11
2.6 ลักษณะพยางค์ของไทย ..... 13
2.6.1 ลักษณะโครงสร้างของคำพยางค์เดียวต่อการผันของเสียงวรรณยุกต์ ..... 13
2.7 ระดับเสียงและคุณภาพของเสียง ..... 15
2.7.1 ระดับเสียง (Pitch) ..... 15
2.7.2 คุณภาพของเสียง (Timbre) ..... 15
2.8 สัญญาณเสียงระบบดิจิทัล ..... 16
2.8.1 การเปลี่ยนสัญญาณเสียงเป็นดิจิทัล (Digitizing) ..... 16
2.9 การหาค่าพิทช์ของสัญญาณเสียงพูด ..... 16
2.9.1 การวิเคราะห์ในโดเมนเวลา ..... 16
2.9.2 ทฤษฎีการประมาณค่าพิทช์โดยใช้ออโตคอร์รีเลชันฟังก์ชัน ..... 16
2.10 การหาค่าความถี่มูลฐานของสัญญาณเสียงพูด ..... 22
2.11 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง ..... 22
2.11.1 การปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง (Median Filtering) ..... 23
2.11.2 การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน ..... 24
2.12 SPEAK GOOD CHINESE PROGRAM ..... 27
2.13 PRAAT ..... 27
2.14 GTK+ ..... 28
2.15 PORTAUDIO ..... 28
2.16 XCODE ..... 28
2.17 MATLAB ..... 29
2.18 ฟังก์ชัน HMMTRAIN ของ MATLAB ..... 29
2.19 ฟังก์ชัน HMMVITERBI ของ MATLAB ..... 29
2.20 ADOBE PHOTOSHOP CC ..... 30




เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ (ต่อ)

หน้า

2.21 SWITCH AUDIO FILE CONVERTER ..... 30

บทที่ 3 วิธีดำเนินการวิจัย ..... 31
3.1 ศึกษาจากงานวิจัยเก่า ..... 31
3.2 แนวคิดในการพัฒนาแอปพลิเคชัน ..... 31
3.2.1 การสร้างแบบจำลอง ..... 31
3.2.2 การคัดเลือกผู้เป็นต้นแบบเสียง ..... 32
3.2.3 การเก็บข้อมูล ..... 32
3.2.4 เครื่องมือและอุปกรณ์ที่ใช้ในการวิจัย ..... 32
3.2.5 ข้อจำกัดในการดำเนินการวิจัย ..... 32
3.3 การวิเคราะห์ระบบที่ต้องการออกแบบ ..... 32
3.3.1 จุดประสงค์ของโครงงานนี้ ..... 32
3.3.2 รูปแบบของระบบ ..... 32
3.3.3 ความต้องการที่เป็นหน้าที่หลักของระบบ (Functional Requirement) ..... 33
3.3.4 ความต้องการที่ไม่ใช่หน้าที่หลักของระบบ (Non-Functional Requirement) ..... 33
3.4 การวิเคราะห์และออกแบบระบบ ..... 33
3.4.1 ยูสเคสไดอะแกรม ..... 33
3.5 USER STORY ..... 34
3.6 ACTIVITY DIAGRAM ..... 35
3.7 การดำเนินงาน ..... 37
3.7.1 การวิเคราะห์หาคาบพิทช์และเตรียมแบบจำลอง ..... 37
3.7.2 ขั้นตอนการทำแบบจำลอง ..... 37
3.7.3 การพัฒนาส่วนติดต่อกับผู้ใช้ ..... 38



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ (ต่อ)

หน้า


บทที่ 4 ผลการทดลอง ..... 39

4.1 การทดลองสร้างกราฟเสียงของแต่ละวรรณยุกต์ ..... 39

4.2 ผลการทดลองจากการสร้างกราฟเส้นเสียง ..... 40

4.2.1 การทดสอบเสียงวรรณยุกต์สามัญจากการออกเสียงคำว่า 'อา' ..... 40

4.2.2 การทดสอบเสียงวรรณยุกต์เอกจากการออกเสียงคำว่า 'อ่า' ..... 40

4.2.3 การทดสอบเสียงวรรณยุกต์โทจากการออกเสียงคำว่า 'อ้า' ..... 41

4.2.4 การทดสอบเสียงวรรณยุกต์ตรีจากการออกเสียงคำว่า 'อ๊า' ..... 41

4.2.5 การทดสอบเสียงวรรณยุกต์จัตวาจากการออกเสียงคำว่า 'อ๋า' ..... 42

4.3 คำที่ใช้ในการสร้างแบบจำลองและผลแบบจำลองอ้างอิง ..... 42

4.4 แบบจำลองการรู้จำเสียงของวรรณยุกต์ ..... 43

4.4.1 แบบจำลองการรู้จำเสียงวรรณยุกต์สามัญ (Middle Tone) ..... 43

4.4.2 แบบจำลองการรู้จำเสียงวรรณยุกต์เอก (Low Tone) ..... 44

4.4.3 แบบจำลองการรู้จำเสียงวรรณยุกต์โท (Falling Tone) ..... 45

4.4.4 แบบจำลองการรู้จำเสียงวรรณยุกต์ตรี (High Tone) ..... 46

4.4.5 แบบจำลองการรู้จำเสียงวรรณยุกต์จัตวา (Rising Tone) ..... 47

4.5 ผลจากการทดสอบแบบจำลองการรู้จำเสียงวรรณยุกต์ ..... 48

4.6 การออกแบบหน้าตาแอปพลิเคชัน ..... 50

บทที่ 5 สรุปผลการพัฒนา ..... 55

5.1 ผลในการพัฒนา ..... 55

5.2 ปัญหาและอุปสรรค ..... 55

บรรณานุกรม ..... 56

ภาคผนวก ก ..... 58




เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง

หน้า

ตารางที่

2.1 การออกเสียงสระเดี่ยว ..... 7
2.2 เสียงวรรณยุกต์ในภาษาไทย ..... 9
2.3 ลักษณะของคำพยางค์เดียวในภาษาไทย ..... 13
2.4 อักษรไตรยางค์ ..... 14
2.5 การจับคู่ในการผันเสียงวรรณยุกต์ ..... 15
3.1 User Story ของผู้ใช้งานแอปพลิเคชัน ..... 34
4.1 กลุ่มคำที่ใช้ในการสร้างแบบจำลองอ้างอิง ..... 42
4.2 ผลจากการรู้จำระดับเสียงวรรณยุกต์ จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงต้นแบบ..... 48



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญรูป

หน้า

รูปที่

2.1 ภาพตัดขวางแสดงอวัยวะในการทำให้เกิดเสียงพูด ..... 4
2.2 แผนภาพระดับการออกเสียงวรรณยุกต์ ..... 12
2.3 องค์ประกอบของพยางค์ในภาษาไทย ..... 13
2.4 รูปทรงในแต่ละคาบของสัญญาณเสียงพูด ..... 18
2.5 แสดงวิธีเซ็นเตอร์คลิปปิงของ Dobnoeski ..... 18
2.6 ตัวอย่างแสดงการคลิปสัญญาณเสียงพูด ..... 19
2.7 แสดงการออโตคอร์รีเลชันของสัญญาณเสียง ..... 21
2.8 ภาพแสดงการเปลี่ยนแปลงของความถี่มูลฐานในแต่ละเวลาของเสียงวรรณยุกต์ ..... 22
2.9 การจัดแบ่งความถี่มูลฐานออกเป็นชุดข้อมูล ..... 23
2.10 แสดงระดับความถี่มูลฐานที่แตกต่างกันระหว่าง (a) ผู้หญิง และ (b) ผู้ชาย ..... 25
2.11 ค่าความถี่มูลฐานตามทิศทางการเปลี่ยนแปลงต่อเวลาที่เพิ่มขึ้นออกเป็น 3 ระดับ ..... 26
2.12 หน้าโปรแกรม Speak Good Chinese ..... 27
2.13 ตัวอย่างภาพการวิเคราะห์เสียงของโปรแกรม Praat ..... 27
2.14 รูปภาพสัญลักษณ์ของซอฟต์แวร์ GTK+ ..... 28
2.15 รูปภาพสัญลักษณ์ของซอฟต์แวร์ PortAudio ..... 28
2.16 รูปภาพสัญลักษณ์ของซอฟต์แวร์ Xcode ..... 28
2.17 รูปภาพสัญลักษณ์ของซอฟต์แวร์ MATLAB ..... 29
2.18 รูปภาพของโปรแกรม Adobe Photoshop CC ..... 30
2.19 รูปภาพของโปรแกรม Switch Audio File Converter ..... 30
3.1 การทำแบบจำลองรู้จำเสียงวรรณยุกต์ ..... 31
3.2 ยูสเคสไดอะแกรม ..... 33
3.3 ลำดับการทำงานระบบฝึกฝนการออกเสียงวรรณยุกต์กับแบบทดสอบ ..... 35



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญรูป (ต่อ)

หน้า

รูปที่

3.4 ลำดับการทำงานการเปรียบเทียบการออกเสียงวรรณยุกต์กับเสียงที่รับเข้ามา ..... 36
4.1 แผนภูมิการทำงานของอัลกอริทึมการหาคลื่นเสียง ..... 39
4.2 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงสามัญ ..... 40
4.3 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงเอก ..... 40
4.4 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงโท ..... 41
4.5 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงตรี ..... 41
4.6 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงจัตวา ..... 42



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1
# บทนำ

## 1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากภาษาไทย มีวรรณยุกต์ที่ใช้ในการประสมคำเพื่อทำให้เกิดเป็นคำศัพท์ใหม่ขึ้นมา แต่ละคำที่มีวรรณยุกต์นั้นจะมีการออกเสียงที่ต่างกันและทำให้ความหมายแตกต่างกันออกไป การออกเสียงวรรณยุกต์จึงมีความสำคัญที่จะต้องออกเสียงให้ถูกต้องเพราะถ้าออกเสียงวรรณยุกต์ผิดเพี้ยนไปจะทำให้เกิดการสื่อความหมายที่ผิดตามไปด้วย ซึ่งในปัจจุบัน มีชาวต่างชาติสนใจและต้องการเรียนรู้ภาษาไทยเพิ่มมากขึ้น แต่ยังไม่ทราบถึงการออกเสียงของวรรณยุกต์ไทยว่ามีความถูกต้องและชัดเจนมากน้อยเพียงใด เพราะการออกเสียงวรรณยุกต์ในภาษาไทยเป็นเรื่องยากสำหรับชาวต่างชาติเพราะในภาษาอังกฤษไม่มีระบบวรรณยุกต์ และชาวต่างชาติบางคนก็ไม่มีผู้ที่จะคอยบอกได้ว่าถูกหรือผิด เป็นต้น

ด้วยเหตุนี้ทางคณะผู้จัดทำได้เล็งเห็นถึงความสำคัญของการออกเสียงวรรณยุกต์ไทย จึงอยากจะพัฒนาอีกหนึ่งช่องทางในการแก้ไขปัญหาการออกเสียงของวรรณยุกต์ไทยให้มีความถูกต้องและชัดเจนมากขึ้น เพื่อที่จะทำให้ผู้ใช้มีความมั่นใจและสามารถนำไปใช้ในชีวิตประจำวันได้อย่างถูกต้อง

## 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาและพัฒนาระบบโดยใช้เทคนิคการรู้จำเสียง (Voice Recognition)
2. เพื่อศึกษาและพัฒนาโปรแกรมบนระบบคอมพิวเตอร์
3. เพื่อออกแบบโปรแกรมที่ช่วยในการฝึกการออกเสียงให้ถูกต้องตามหลักวรรณยุกต์ไทย
4. เพื่อนำความรู้จากการศึกษามาประยุกต์ใช้ในการพัฒนาระบบ

## 1.3 สมมติฐานของการศึกษา

ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียงพัฒนาโดยใช้โปรแกรม MATLAB โดยทางผู้จัดทำหวังว่าตัวแอปพลิเคชันที่ทำขึ้นจะเป็นส่วนช่วยให้ชาวต่างชาติหรือผู้ที่สนใจสามารถออกเสียงวรรณยุกต์ภาษาไทยได้ถูกต้องตามหลักการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

- เทคนิคการรู้จำเสียง (Voice Recognition)
- เทคนิคการตรวจจับเสียง (Voice Detection)
- เทคนิคการควอนไตซ์พิทช์ (Pitch Quantization)
- Hidden Markov Modeling (HMM)
- ใช้โปรแกรม MATLAB

## 1.5 การเปรียบเทียบระหว่างวิธีการที่นำเสนอกับวิธีการแบบพื้นฐาน

แอปพลิเคชันที่มีอยู่เดิมไม่มีระบบให้ผู้ใช้สามารถทดสอบการออกเสียงได้โดยตรง มีเพียงให้ฟังเสียงและดูคำอ่านเท่านั้น โดยในแอปพลิเคชันที่จัดทำขึ้นในโครงงานนี้ ผู้ใช้จะสามารถทดสอบออกเสียงได้เลย และทราบได้ว่าสามารถออกเสียงได้ถูกต้องหรือไม่

## 1.6 ขอบเขตการวิจัย

1. ระบบสามารถตรวจสอบการออกเสียงวรรณยุกต์ได้
2. ระบบสามารถรองรับการใช้งานของผู้ใช้ได้ทีละ 1 คน
3. ระบบสามารถวิเคราะห์ระดับเสียงวรรณยุกต์ได้ครั้งละ 1 พยางค์
4. ระบบสามารถวิเคราะห์เสียงได้ในสภาพแวดล้อมที่เป็นห้องปิดและไม่มีเสียงรบกวน

## 1.7 ขั้นตอนการศึกษา

1. ศึกษางานวิจัยที่เกี่ยวข้องกับวิธีการออกเสียงที่ถูกต้องของวรรณยุกต์ไทย
2. ศึกษางานวิจัยที่เกี่ยวข้องกับเทคนิคการรู้จำเสียง (Voice Recognition)
3. ศึกษางานวิจัยที่เกี่ยวข้องกับเทคนิคควอนไตซ์พิทช์ (Pitch Quantization)
4. ศึกษางานวิจัยที่เกี่ยวข้องกับเทคนิค Hidden Markov Modeling
5. ศึกษาวิธีการรับไฟล์เสียงจากไมโครโฟน
6. ศึกษาวิธีการแปลงไฟล์จากเสียงให้เป็นค่ากราฟ
7. ศึกษาและวิเคราะห์ปัญหาเกี่ยวกับปัจจัยที่มีผลกระทบต่อระบบ
8. ศึกษาวิธีการพัฒนาระบบ
9. พัฒนาระบบตามที่ได้ออกแบบไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

10. ทำการทดสอบและประเมินผลตลอดระยะเวลาที่ทำการพัฒนาโปรแกรม

11. ทำการทดสอบระบบกับผู้ใช้งานจริง และนำผลตอบรับจากผู้ใช้มาทำการแก้ไขเพื่อให้ระบบมีประสิทธิภาพมากขึ้น

## 1.8 ประโยชน์ที่คาดว่าจะได้รับ

1. ชาวต่างชาติรวมถึงผู้ที่สนใจได้รับประโยชน์จากการใช้งานแอปพลิเคชัน สามารถฝึกฝนการออกเสียงวรรณยุกต์ได้

2. เป็นโปรแกรมช่วยสอนและฝึกออกเสียงภาษาไทยได้ด้วยตนเอง

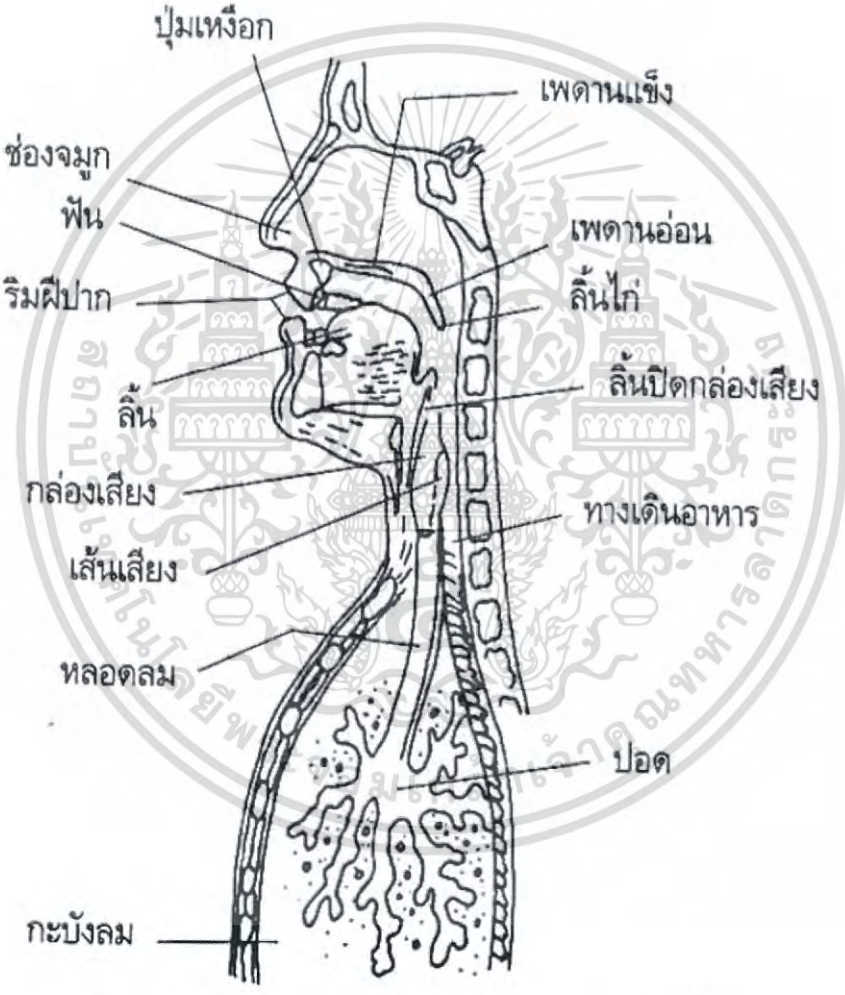
3. ช่วยเป็นแรงผลักดันให้ผู้คนมีความสนใจในภาษาไทย รักในภาษาไทย และช่วยกันอนุรักษ์ภาษาไทย ทำให้ภาษาไทยเป็นที่รู้จักแก่ชาวต่างชาติมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 2
# วรรณกรรมที่เกี่ยวข้อง

## 2.1 ทฤษฎีการสร้างเสียงพูด

เสียงที่ออกมาจากมนุษย์นั้นเริ่มจากการที่หายใจนำลมเข้าสู่ปอดแล้วใช้ลมนั้นหายใจออกผ่านอวัยวะต่าง ๆ จนเกิดเป็นเสียงพูด โดยปากจะเป็นตัวที่ควบคุมและทำให้เสียงเป็นไปในแบบที่ผู้พูดต้องการ




**รูปที่ 2.1** ภาพตัดขวางแสดงอวัยวะในการทำให้เกิดเสียงพูด [7]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**2.1.1 อวัยวะที่เกี่ยวข้องกับการออกเสียง**

1. ปอด ในปอดจะประกอบไปด้วยถุงลมเล็ก ๆ จำนวนมาก โดยปอดจะเป็นตัวระบายลมออกมาเพื่อใช้ในการปรับเปลี่ยนเป็นเสียงพูด

2. กะบังลม เป็นแผ่นกล้ามเนื้ออยู่ใต้ปอด เมื่อเราทำการหายใจเข้ากะบังลมจะลดต่ำลง ทำให้สามารถเก็บอากาศในช่องท้องได้และเมื่อยกสูงขึ้นจะเป็นการดันลมออกมา

3. หลอดลม เป็นเส้นทางผ่านของลมจากปอดมากล่องเสียง

4. เพดานแข็ง เรียกอีกอย่างว่า เพดานปาก เป็นส่วนที่อยู่เหนือลิ้นมีลักษณะเป็นกระดูกโค้ง

5. เพดานอ่อน เป็นส่วนที่ถัดเข้าไปจากเพดานแข็ง มีลักษณะนิ่มกว่าเพดานแข็ง เป็นกระดูกอ่อนที่สามารขยับขึ้นและลงได้

6. ลิ้น เป็นส่วนที่สามารถขยับและเคลื่อนไหวได้มาก แบ่งเป็น 3 ส่วน คือ

6.1 หน้าลิ้น เป็นส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานแข็ง

6.2 หลังลิ้น เป็นส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน

6.3 ปลายลิ้น เป็นส่วนของลิ้นที่อยู่ในช่วงปลาย สามารถขยับได้ง่าย

7. ริมฝีปาก เป็นส่วนที่ใช้ในการควบคุมการออกเสียงได้ในหลายลักษณะสามารถขยับเพื่อเปลี่ยนแปลงการออกเสียงได้ในหลายลักษณะ โดยจะเป็นการเปลี่ยนแปลงการออกจากปากของลม

8. ลิ้นไก่ เป็นก้อนเนื้อขนาดเล็กอยู่อยู่บริเวณเพดานอ่อน

9. ปุ่มเหงือก มีลักษณะเป็นพื้นผิวไม่เรียบอยู่หลังฟันบน

10. ฟัน เป็นอวัยวะที่มีผลต่อเสียงมาก เกี่ยวข้องกับการขยับของปากเพื่อให้เกิดเป็นเสียงที่ต้องการ และเกี่ยวข้องกับการลอดผ่านของลมในการออกเสียง

11. ช่องจมูก เป็นโพรงที่อยู่เหนือปาก ใช้สำหรับเป็นทางผ่านของลมและใช้ในการออกเสียงนาสิก ในเวลาพูดช่องนี้จะถูกปิดเพื่อให้ลมออกทางปาก

12. กล่องเสียง เป็นอวัยวะที่อยู่บริเวณลำคอ ใช้ในการป้องกันท่อลมและช่วยในการออกเสียง

13. สายเสียง เรียกอีกอย่างได้ว่า เส้นเสียง มีลักษณะเป็นเส้นกล้ามเนื้อ 2 แผ่น ระหว่างทั้งสองเส้นจะมีช่องว่างไว้สำหรับให้ลมผ่านเมื่อลมผ่านสายเสียงจะเกิดการสั่นทำให้เกิดเป็นเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

14. ลิ้นปิดกล่องเสียง เรียกอีกอย่างได้ว่า ฝาปิดกล่องเสียง เป็นแผ่นกระดูกอ่อน ทำหน้าที่ปิดทางเข้าออกของช่องระหว่างสายเสียง [4]

**2.1.2 ลักษณะต่าง ๆ ของเสียงพูด**

1. ความก้องของเสียง (Voice)

เสียงก้อง หรือ เสียงโฆษะ เป็นเสียงที่เกิดจากการตึงตัวของเส้นเสียงหรือเรียกว่าเส้นเสียงปิดในขณะที่เส้นเสียงปิดหากมีอากาศไหลผ่านจะทำให้เกิดการสั่นสะเทือนของเส้นเสียง สัญญาณเสียงจะมีลักษณะเป็นคาบ (Quasi-periodic) เรียกความถี่ในการปิด-เปิดของเส้นเสียงว่าความถี่มูลฐาน (Fundamental Frequency : $F_0$) ตัวอย่างพยัญชนะที่มีเสียงก้อง เช่น บ ใบไม้ ม ม้า เป็นต้น

เสียงไม่ก้อง หรือ เสียงอโฆษะ เป็นเสียงที่เกิดเมื่อเส้นเสียงมีลักษณะตึงหรือเรียกว่าเส้นเสียงเปิด เสียงจะมีลักษณะเป็นเสียงของสัญญาณรบกวน (Noise) ไม่เป็นลักษณะคาบ ตัวอย่างพยัญชนะที่มีเสียงไม่ก้อง เช่น ส เสือ ฟ ฟัน เป็นต้น

2. ความยาวของเส้นเสียง (Length)

คือ ความยาวของระยะเวลาที่ใช้ในการเปล่งเสียง โดยในภาษาไทยจะจำแนกได้จากชนิดของสระ ประกอบด้วยสระเสียงสั้น เช่น สระอะ สระอิ สระอุ สระโอะ เป็นต้น และสระเสียงยาว เช่น สระอา สระอี สระอู สระโอ เป็นต้น

3. ระดับเสียงสูง-ต่ำ (Pitch)

เสียงสูงต่ำขึ้นอยู่กับการสะบัดของเส้นเสียงโดยเกี่ยวข้องกับความถี่ที่เปล่งออกมา เสียงต่ำจะมีระดับเสียงต่ำ และเสียงสูงจะมีระดับเสียงสูง ในการพูดจะมีระดับของเสียงที่แตกต่างกันไปในหลายระดับ โดยในภาษาไทยจำแนกระดับเสียงสูง-ต่ำโดยใช้วรรณยุกต์เป็นตัวระบุ

4. ความดัง (Loudness)

เสียงที่ผู้พูดเปล่งออกมาจะมีระดับความดังตามปริมาณลมที่ใช้ในช่วงหนึ่ง ๆ

5. การลงน้ำหนัก (Stress)

เป็นการลงเสียงที่พยางค์ใดพยางค์หนึ่งซึ่งอาจมีระดับที่แตกต่างกับพยางค์ข้างเคียง โดยใช้ในหลายกรณี เช่น เพื่อความถูกต้องของคำพูด เพื่อการแสดงอารมณ์ เป็นต้น

6. ช่วงต่อของเสียง (Juncture)

เป็นช่วงต่อของพยางค์ในการเปล่งเสียง ถ้าคำที่เปล่งออกมาต่อกันแบบไม่มีระยะห่างจะเรียกว่าไม่เห็นร่องรอย (Close juncture) แต่ถ้ามีระยะห่างเป็นคนละพยางค์จะเรียกว่ามีช่วงต่อของเสียง (Open juncture) เราสามารถแยกแยะการแบ่งของคำพูดได้จากช่วงต่อของเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2 หน่วยเสียงสำคัญในภาษาไทย

"หน่วยเสียง" (Phoneme) ทำหน้าที่ให้ความหมายของคำนั้น ๆ เป็นหน่วยที่เล็กที่สุดของภาษา ในภาษาไทยมีหน่วยเสียงที่สำคัญอยู่ 3 ประเภทใหญ่ ๆ คือ หน่วยเสียงพยัญชนะ หน่วยเสียงสระ หน่วยเสียงวรรณยุกต์ โดยเมื่อนำสามหน่วยนี้มาประกอบกันจะเกิดเป็นคำขึ้นมานั่นเอง

เราจะพิจารณาเสียงพูดของมนุษย์ได้แบบกว้าง ๆ เป็น 2 ลักษณะ คือ

1. เสียงเรียง (Segmental Sound) เป็นเสียงที่มีลักษณะเด่นสามารถแยกจากเสียงอื่นได้ชัดเจน ได้แก่ เสียงสระและเสียงพยัญชนะ

2. เสียงซ้อน (Supra-segmental feature) เป็นเสียงที่ไม่สามารถแยกออกเสียงได้ตามลำพัง ได้แก่ เสียงวรรณยุกต์ และทำนองเสียง

## 2.3 หน่วยเสียงสระ

**2.3.1** ลักษณะของเสียงสระ

หน่วยเสียงสระมีลักษณะที่สำคัญคือ เสียงที่เปล่งออกมานั้นจะไม่ถูกลิ้นขวาง เวลาออกเสียงจะสะดวกและเปล่งเสียงได้นาน ถ้าลิ้นทำหน้าที่เพียงส่วนเดียวจะเป็น สระเดี่ยว แต่ถ้าทำหน้าที่หลายส่วนจะเป็น สระประสม

สำหรับภาษาไทยมีหน่วยเสียงสระทั้งหมด 24 หน่วยเสียง แยกออกเป็นสระเดี่ยว 18 หน่วยเสียง และสระประสม 6 หน่วยเสียง

สระเดี่ยว

สระเดี่ยวมีการเกิดแบ่งเป็น 2 รูปแบบใหญ่ ๆ คือ

1. พิจารณาว่าลมผ่านส่วนไหนของลิ้น ได้แก่ หน้า กลาง หลัง

2. พิจารณาว่าลิ้นอยู่ในระดับใดตอนที่ลมผ่าน ได้แก่ สูง กลาง ต่ำ

**ตารางที่ 2.1** การออกเสียงสระเดี่ยว

| สระ | หน้า | กลาง | หลัง |
|---|---|---|---|
| สูง | อิ อี | อึ อือ | อุ อู |
| กลาง | เอะ เอ | เออะ เออ | โอะ โอ |
| ต่ำ | แอะ แอ | อะ อา | เอาะ ออ |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้ยังสามารถแบ่งสระเดี่ยวได้เป็น 2 กลุ่มตามการออกเสียง คือ

- สระเดี่ยวเสียงสั้น 9 หน่วย คือ อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ
- สระเดี่ยวเสียงยาว 9 หน่วย คือ อา อี อือ อู เอ แอ โอ ออ เออ

สระประสม เป็นเสียงที่เกิดจากการที่ลมกระทบลิ้น 2 ส่วน คือ ส่วนบน และส่วนล่าง ในระหว่างที่ออกเสียงลิ้นจะมีลักษณะย้ายจากระดับสูงมาระดับต่ำ โดยเสียงช่วงหลังจะเป็นสระอะเสมอ

เสียงสระประสมที่เกิดจากลิ้นส่วนหน้า ได้แก่ เอียะ (อิ+อะ) และ เอีย (อี+อะ)

เสียงสระประสมที่เกิดจากลิ้นส่วนกลาง ได้แก่ เอือะ (อึ+อะ) และ เอือ (อือ+อะ)

เสียงสระประสมที่เกิดจากลิ้นส่วนหลัง ได้แก่ อัวะ (อัว+อะ) และ อัว (อู+อะ)

**2.3.2 หน้าที่ของหน่วยเสียงสระในภาษาไทย**

หน่วยเสียงสระทั้ง 24 รูปในภาษาไทย มีหน้าที่ประสมเกิดเป็นคำ ในภาษาไทยทุกคำต้องมีหน่วยเสียงสระประสมอยู่ด้วยเสมอ หน่วยเสียงสระที่ทำให้เกิดเป็นคำได้มากที่สุดมักเป็นหน่วยเสียงสระยาว

## 2.4 หน่วยเสียงพยัญชนะ

ในภาษาไทยพยัญชนะมีทั้งหมด 44 รูป 21 เสียง เสียงพยัญชนะเวลาออกเสียง ลมหายใจที่ออกมาจะถูกขวางโดยส่วนต่าง ๆ ของปาก และออกเสียงให้ยาวนานเหมือนเสียงสระไม่ได้

**2.4.1** ลักษณะของเสียงพยัญชนะ

หน่วยเสียงพยัญชนะ 21 หน่วยเสียงถูกแบ่งออกเป็นเสียงก้อง และเสียงไม่ก้อง ดังนี้

เสียงก้อง มี 9 หน่วย ได้แก่ /ง/ /ย/ /บ/ /ด/ /ม/ /น/ /ร/ /ล/ /ว/

เสียงไม่ก้อง มี 12 หน่วย ได้แก่ /ก/ /ค/ /จ/ /ช/ /ซ/ /ท/ /ต/ /ป/ /พ/ /ฟ/ /อ/ /ฮ/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 2.2** เสียงวรรณยุกต์ในภาษาไทย

| ลำดับที่ | อักษรไทยใช้แทนหน่วยเสียง | แทนสัญลักษณ์หน่วยเสียง |
|---|---|---|
| 1. | ก | ก |
| 2. | ข ฃ ค ฅ ฆ | ค |
| 3. | ง | ง |
| 4. | จ ( จร_*) | จ |
| 5. | ฉ ช ฌ | ช |
| 6. | ญ ย (หย_*) (หญ_*) | ย |
| 7. | ซ ศ ษ ส (ทร_*) | ซ |
| 8. | ฐ ฑ ฒ ถ ท ธ (ทร_*) | ท |
| 9. | บ | บ |
| 10. | ฎ ด (ฑ_*) | ด |
| 11. | ฏ ต | ต |
| 12. | ป | ป |
| 13. | ผ พ ภ | พ |
| 14. | ฝ ฟ | ฟ |
| 15. | ม (หม_*) | ม |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 2.2 (ต่อ)** เสียงวรรณยุกต์ในภาษาไทย

| | | |
|---|---|---|
| 16. | น ณ (หน_*) | น |
| 17. | ร | ร |
| 18. | ล ฬ (หล_*) | ล |
| 19. | ว (หว_*) | ว |
| 20. | อ | อ |
| 21. | ฮ ห | ฮ |
| อักษร 44 รูป | | 21 หน่วยเสียง |

หมายเหตุ (*) หมายถึง พยัญชนะที่ออกเสียงเป็นพยัญชนะต้นจะออกเสียงเหมือนพยัญชนะในลำดับนั้น

ลักษณะของเสียงพยัญชนะ แบ่งตามลักษณะของรูปเสียง [1] [4]

1. เสียงกัก หรือ เสียงหยุด (Stop)

เมื่อเสียงผ่านมาถึงช่องปากจะมีบางส่วนของปากกั้นเสียงนี้ไว้ในชั่วขณะหนึ่งแล้วเปิดออกเสียงจะถูกแผดออกมาในลักษณะระเบิด (Plosive sound) มี 9 หน่วยเสียง คือ /ป/ /พ/ /บ/ /ต/ /ท/ /ด/ /ก/ /ค/ /อ/

2. เสียงเสียดแทรก (Fricative)

เมื่อเสียงผ่านมาถึงช่องปากจะมีบางส่วนของปากกั้นเสียงนี้ไว้แต่กั้นไม่สนิท ยังมีเสียงแทรกสอดออกมา มี 3 หน่วยเสียงคือ /ซ/ /ฟ/ /ฮ/

3. เสียงกึ่งเสียดแทรก (Affricate)

เป็นเสียงที่มีลักษณะเหมือนเสียงกักแล้วตามด้วยเสียงเสียดแทรก มี 2 หน่วยเสียง ได้แก่ /จ/ และ /ช/

4. เสียงนาสิก (Nasal)

เป็นเสียงที่ผ่านขึ้นมาจากลำคอออกทางจมูกโดนในปากจะถูกปิดกั้นเอาไว้ มี 3 หน่วยเสียงได้แก่ /ม/ /น/ /ง/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. เสียงข้าง (Lateral)

เป็นเสียงที่ออกทางปากโดยผ่านทางข้าง ๆ ของลิ้น มีเสียงเดียวคือ /ล/

6. เสียงรัว (Trill)

เป็นเสียงที่ออกทางปากโดยจะมีการกระดกปลายลิ้น ได้แก่เสียง /ร/

7. เสียงครึ่งสระ (Semi-Vowel)

การออกเสียงชนิดนี้ตำแหน่งลิ้นตอนช่วงต้นและช่วงท้ายของเสียงจะแตกต่างกัน มี 2 หน่วยเสียงคือ /ว/ และ /ย/

**2.4.2 หน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย**

1. เป็นพยัญชนะต้นของพยางค์ คือสามารถประสมกับหน่วยเสียงสระเป็นพยางค์หนึ่งได้ โดยเกิดได้ 2 รูปแบบคือ แบบหน่วยเดี่ยว ที่ทุกตัวสามารถเกิดได้หมด และแบบสองหน่วยทำหน้าที่เป็นพยัญชนะควบ มีหน่วยแรกเป็น /ก/ /ค/ /ต/ /ป/ หรือ /พ/ และหน่วยเสียงที่สองเป็น /ร/ /ล/ หรือ /ว/

2. เป็นพยัญชนะสะกดของพยางค์ เกิดได้ 9 หน่วยเสียงคือ /ก/ (แม่กก) /บ/ (แม่กบ) /ด/ (แม่กด) /น/ (แม่กด) /ม/ (แม่กม) /ย/ (แม่เกย) /ว/ (แม่เกอว) /ง/ (แม่กง) และไม่มีตัวสะกด (แม่กา)

## 2.5 หน่วยเสียงวรรณยุกต์

เสียงวรรณยุกต์คือ ระดับเสียงสูงต่ำของคำในภาษาไทย ซึ่งความหมายจะขึ้นอยู่กับเสียงวรรณยุกต์นั้น ๆ ด้วย

ในภาษาไทยเสียงวรรณยุกต์เป็นเสียงที่สำคัญ เป็นเสมือนการสร้างคำใหม่ ๆ ให้มีจำนวนมากขึ้น ถ้าออกเสียงไม่ตรงวรรณยุกต์ก็จะทำให้ความหมายของคำนั้น ๆ เปลี่ยนไป

หน่วยเสียงวรรณยุกต์ในภาษาไทยมีทั้งหมด 5 เสียง 4 รูปวรรณยุกต์ โดยเสียงสามัญจะไม่มีรูปวรรณยุกต์ [1] [4]

**2.5.1 ลักษณะของเสียงวรรณยุกต์**

แบ่งเป็น 2 กลุ่มใหญ่ ๆ คือ

1. กลุ่มวรรณยุกต์ระดับ (Level tone) มี 3 หน่วยเสียง คือ

1.1 หน่วยเสียงวรรณยุกต์ระดับต่ำ (Low tone) แทนด้วยสัญลักษณ์ / \ /

คือ เสียงวรรณยุกต์เอก พบในภาษาไทยทุกแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 หน่วยเสียงวรรณยุกต์ระดับกลาง (Mid tone) ไม่มีสัญลักษณ์

คือ เสียงวรรณยุกต์สามัญ วรรณยุกต์ชนิดนี้จะไม่พบในพยางค์ที่มีตัวสะกดเป็นพยัญชนะกัก

1.3 หน่วยเสียงวรรณยุกต์ระดับสูง (High tone) แทนด้วยสัญลักษณ์ / ╱ /

คือ เสียงวรรณยุกต์ตรี วรรณยุกต์ชนิดนี้จะไม่พบในพยางค์ที่ประสมด้วยสระเสียงยาวและมีตัวสะกดเป็นเสียงกัก

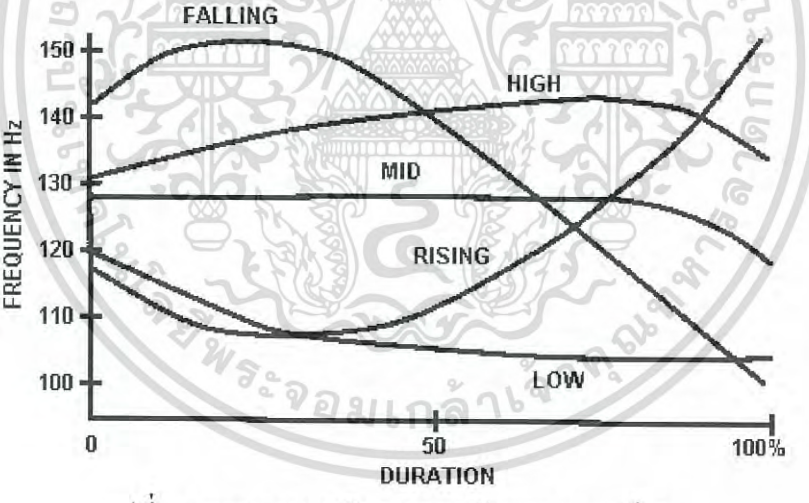
2. กลุ่มวรรณยุกต์เปลี่ยนระดับ (Contour tone) มี 2 หน่วยเสียง คือ

2.1 หน่วยเสียงวรรณยุกต์เปลี่ยนตก (Falling tone) แทนด้วยสัญลักษณ์ / Λ /

คือ เสียงวรรณยุกต์โท วรรณยุกต์ชนิดนี้จะไม่พบในพยางค์ที่ประสมด้วยสระเสียงสั้นและมีตัวสะกดเป็นเสียงกัก

2.2 หน่วยเสียงวรรณยุกต์เปลี่ยนขึ้น (Rising tone) แทนด้วยสัญลักษณ์ / V /

คือ เสียงวรรณยุกต์จัตวา วรรณยุกต์ชนิดนี้จะไม่พบในพยางค์ที่มีพยัญชนะสะกดเป็นเสียงกัก




รูปที่ 2.2 แผนภาพระดับการออกเสียงวรรณยุกต์ [5]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.6 ลักษณะพยางค์ของไทย

### 2.6.1 คำจำกัดความของพยางค์และคำในภาษาไทย

กาญจนา นาคสกุล (2520:104) ได้ให้ความหมายของพยางค์ในภาษาไทยไว้ว่า "พยางค์หมายถึง จำนวนเสียงที่ดังเด่นซึ่งปรากฏในกลุ่มเสียงที่เรียงกันเป็นคำพูด ส่วนเสียงอื่น ๆ ที่อยู่ข้างเคียงก็จะประกอบกันเข้าเป็นส่วนของพยางค์" โดยเสียงสระจะเป็นเสียงที่เด่นดังที่สุดและมีลักษณะเป็นเสียงก้อง ดังนั้นจำนวนสระจึงมักเป็นจำนวนพยางค์ด้วยเช่นกัน [4]

คำในภาษาไทยส่วนใหญ่เป็นคำพยางค์เดียว ซึ่งเป็นคำพื้นฐาน (Base words) จัดอยู่ในกลุ่มตระกูลคำโดด หรือพยางค์เดียว (Monosyllabic language) หน่วยเสียงที่ประกอบกันจะมีอย่างน้อย 3 หน่วยเสียงคือ หน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ แต่มีไม่เกิน 5 หน่วยเสียง ในกรณีที่เพิ่มเติมด้วย หน่วยเสียงควบกล้ำและหน่วยเสียงพยัญชนะสะกด [1] [4]

| | | วรรณยุกต์ | |
|---|---|---|---|
| พยัญชนะต้น | (ควบ) | สระ | (พยัญชนะสะกด) |

**รูปที่ 2.3** องค์ประกอบของพยางค์ในภาษาไทย [4]

### 2.6.1 ลักษณะโครงสร้างของคำพยางค์เดียวต่อการผันของเสียงวรรณยุกต์

ลักษณะโครงสร้างของคำพยางค์เดียวในไทยมี 5 รูปแบบ โดยจะมีผลต่อการผันเสียงวรรณยุกต์ ดังแสดงในตารางที่ 2.3

**ตารางที่ 2.3** ลักษณะของคำพยางค์เดียวในภาษาไทย

| เสียงวรรณยุกต์ / โครงสร้างพยางค์ | สามัญ | เอก | โท | ตรี | จัตวา |
|---|---|---|---|---|---|
| 1. พ (พ) ส ส $^{0-4}$ | + | + | + | + | + |
| 2. พ (พ) ส น $^{0-4}$ | + | + | + | + | + |
| 3. พ (พ) ส ส น $^{0-4}$ | + | + | + | + | + |
| 4. พ (พ) ส ก $^{1,3}$ | - | + | - | + | - |
| 5. พ (พ) ส ส ก $^{1,2}$ | - | + | + | - | - |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมายเหตุ + หมายถึง โครงสร้างพยางค์สามารถผันระดับเสียงวรรณยุกต์นั้นได้

- หมายถึง โครงสร้างพยางค์ไม่สามารถผันระดับเสียงวรรณยุกต์นั้นได้

กำหนดให้ พ แทนหน่วยเสียงพยัญชนะต้น 1 หน่วย

พพ แทนหน่วยเสียงพยัญชนะต้น 2 หน่วยควบกัน หรือพยัญชนะต้นควบ โดยหน่อยเสียงที่ 2 คือ /ร/ /ล/ หรือ /ว/

ส แทนหน่วยเสียงสระเดี่ยวสั้น

สส แทนหน่วยเสียงสระเดี่ยวยาว และหน่วยเสียงสระประสม

น แทนหน่วยเสียงพยัญชนะสะกดที่เป็นพยัญชนะนาสิกและครึ่งสระ

ก แทนหน่วยเสียงสะกดที่เป็นพยัญชนะกัก

0 แทนหน่วยเสียงวรรณยุกต์สามัญ

1 แทนหน่วยเสียงวรรณยุกต์เอก

2 แทนหน่วยเสียงวรรณยุกต์โท

3 แทนหน่วยเสียงวรรณยุกต์ตรี

4 แทนหน่วยเสียงวรรณยุกต์จัตวา

เพื่อความเข้าใจได้ง่ายในภาษาไทยจึงจำแนกอักษรในภาษาไทยอักษรเป็นอักษรไตรยางค์ ดังแสดงไว้ในตารางที่ 2.4

**ตารางที่ 2.4** อักษรไตรยางค์

| อักษรไตรยางค์ | | รูปวรรณยุกต์ | | | |
|---|---|---|---|---|---|
| | | เอก | โท | ตรี | จัตวา |
| อักษรสูง | ข ฃ ฉ ฐ ฑ ผ ฝ ศ ษ ส ห | + | + | - | - |
| อักษรกลาง | ก จ ด ต ฎ ฏ บ ป อ | + | + | + | + |
| อักษร ต่ำ- เดี่ยว | ม น ง ณ ย ญ ร ล ฬ ว | + | + | - | - |
| ต่ำ-คู่ | ค ฅ ฆ ช ฌ ซ ฑ ท ธ ฒ พ ภ ฟ ฮ | + | + | - | - |

อักษรสูง 11 ตัว ผันวรรณยุกต์ได้ 3 เสียง เช่น ผา ผ่า ผ้า

อักษรกลาง 9 ตัว ผันวรรณยุกต์ได้ 5 เสียง เช่น ปา ป่า ป้า ป๊า ป๋า

อักษรต่ำ 24 ตัว ผันวรรณยุกต์ได้ 3 เสียง เช่น คา ค่า ค้า

อักษรต่ำรูปวรรณยุกต์เอกจะออกเสียงเป็นเสียงวรรณยุกต์โท และรูปวรรณยุกต์โทจะออกเสียงเป็นวรรณยุกต์ตรี โดยจำแนกออกเป็นอักษรต่ำคู่และอักษรต่ำเดี่ยว โดยอักษรต่ำคู่จะนำมาผัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ร่วมกับคำที่เป็นอักษรสูงทำให้เมื่อผันรวมกันจะได้ครบ 5 เสียง ตัวอย่างเช่น คา ข่า ค่า/ข้า ค้า ขา เป็นต้น โดยสามารถจับคู่กันได้ 7 คู่ ดังราตารางที่ 2.5

**ตารางที่ 2.5** การจับคู่ในการผันเสียงวรรณยุกต์

| คู่ที่ | อักษรสูง | อักษรต่ำคู่ |
|---|---|---|
| 1 | ข ฃ | ค ฅ ฆ |
| 2 | ฉ | ช ฌ |
| 3 | ถ ฐ | ท ฑ ธ ฒ |
| 4 | ผ | พ ภ |
| 5 | ฝ | ฟ |
| 6 | ศ ษ ส | ซ |
| 7 | ห | ฮ |
| | 11 ตัว | 14 ตัว |

ในส่วนของอักษรต่ำเดี่ยวจะมีการใช้ ห หีบ มาช่วยในการผันทำให้สามารถผันได้ครับทั้ง 5 เสียง ตัวอย่างเช่น นี หนี่ หนี้ นี้ หนี เป็นต้น [1] [4]

## 2.7 ระดับเสียงและคุณภาพของเสียง

### 2.7.1 ระดับเสียง (Pitch)

การได้ยินเสียง แบ่งออกเป็น 2 ลักษณะ คือ เสียงระดับความดังมากหรือน้อยเกิดจากพลังงานของแหล่งกำเนิดเสียง และเสียงทุ้มหรือแหลมเกิดจากความถี่สูงหรือต่ำ ซึ่งเรียกว่า ระดับเสียง (Pitch) มนุษย์จะได้ยินเสียงที่ระดับความถี่ 20 – 20,000 Hz โดยความถี่ที่มีระดับต่ำกว่า 20 Hz จะเรียกว่า คลื่นใต้เสียง (Infra Sound) และความถี่เสียงที่มีระดับสูงกว่า 20,000 Hz จะเรียกว่า คลื่นเหนือเสียง (Ultra Sound) และในด้านของแหล่งกำเนิดเสียงก็จะให้ช่วงความถี่ที่ต่างกัน เช่น

เสียงมนุษย์เพศชายจะมีระดับความถี่ประมาณ 80 – 160 Hz

เสียงมนุษย์เพศหญิงจะมีระดับความถี่ประมาณ 160 – 400 Hz

### 2.7.2 คุณภาพของเสียง (Timbre)

ลักษณะของเสียงจะมีความเฉพาะตัวแตกต่างกันออกไป ขึ้นอยู่กับแหล่งกำเนิดเสียง คุณภาพของเสียงจะเป็นตัวช่วยให้เราแยกประเภทของแหล่งกำเนิดเสียงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.8 สัญญาณเสียงระบบดิจิทัล

ในการเปลี่ยนคลื่นเสียงเป็นคลื่นไฟฟ้าในระบบดิจิทัลนั้น จะทำการรับเสียงทางไมโครโฟน ระดับของความต่างศักย์ไฟฟ้าจะเป็นตัวกำหนดความดังของสัญญาณเสียงมีหน่วยเป็นโวลต์ (Volts) โดยเสียงที่ได้จะเป็นคลื่นความถี่ของสัญญาณไฟฟ้า (Waves) มีหน่วยเป็น เฮิร์ซ (Hz)

### 2.8.1 การเปลี่ยนสัญญาณเสียงเป็นดิจิทัล (Digitizing)

เครื่องคอมพิวเตอร์เป็นตัวสำคัญในการเปลี่ยนสัญญาณแอนะล็อก ที่ทำการเปลี่ยนจากสัญญาณแอนะล็อกเป็นดิจิทัลเพราะว่าสัญญาณเสียงในระบบแอนะล็อกใช้สัญญาณไฟฟ้าทำให้มีความผิดพลาดในการคำนวณจึงได้มีการพัฒนาการเข้ารหัสโดยใช้เลขฐานสองขึ้น

เทคนิคการเปลี่ยนสัญญาณเสียงแอนะล็อกเป็นดิจิทัลทำโดยการแบ่งเสียงออกเป็นส่วน ๆ เรียกว่า Sampling เพื่อกำหนดปริมาณของหน่วยวัดดิจิทัลในแต่ละส่วนได้

ความถี่ของสัญญาณเสียงที่ใช้ในการแบ่งคลื่นเสียง เรียกว่า Sampling Frequency โดยมีความถี่ประมาณ 2 เท่าของความถี่ที่ใช้งาน ในด้านความดังของเสียงนั้นแบ่งตามระดับมาตรฐานของเสียงในระบบดิจิทัล คือ 8 bit = 256 ระดับ และ 16 bit = 65,536 ระดับ ในกรณีที่ระดับของเสียงมีระดับสูงก็จะทำให้การเปลี่ยนกลับจากสัญญาณแอนะล็อกเป็นสัญญาณดิจิทัลมีความเหมือนเดิมมากตามไปด้วย

## 2.9 การหาค่าพิทช์ของสัญญาณเสียงพูด

### 2.9.1 การวิเคราะห์ในโดเมนเวลา

เสียงที่เปล่งออกมาในช่วงเวลาหนึ่งนั้นขึ้นอยู่กับลักษณะรูปทรงของช่องทางเดินเสียง (Vocal tract) และลักษณะการสั่นของเส้นเสียง (Vocal cord) เสียงพูดของมนุษย์จึงมีลักษณะที่เป็นสัญญาณที่มีคาบเวลาชั่วขณะ (Quisi-periodic) ในการวิเคราะห์จึงต้องแบ่งเสียงพูดออกเป็นช่วง ๆ (Frame) เพื่อให้มีลักษณะที่คงที่ [3]

### 2.9.2 ทฤษฎีการประมาณค่าพิทช์โดยใช้ออโตคอร์รีเลชันฟังก์ชัน

ออโตคอร์รีเลชัน (Autocorrelation) ใช้ตรวจหาคาบพิทช์ของสัญญาณเสียงพูด ทำหน้าที่ในการแสดงยอดกราฟหลัก (Prominent peak) ทำให้เราสามารถหาระยะคาบของแต่ละช่วงเสียงได้ [3]

#### 2.9.2.1 การจัดแบ่งวิเคราะห์สัญญาณออกเป็นช่วงสั้นๆ

กำหนดให้สัญญาณแทนด้วย s(m) ซึ่งออโตคอร์รีเลชันฟังก์ชันของ Discrete-time deterministic signal โดยทั่วไปเขียนได้เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง


$$\phi(\tau) = \sum_{m=-\infty}^{\infty} s(m)s(m+\tau) \quad (2.1)$$

ถ้าสัญญาณ s(m) มีคาบที่แน่นอนด้วยระยะ p จะได้ว่า

$$s(m) = s(m + p) \quad ; \text{สำหรับทุก } m$$

ดังนั้นสามารถเขียนได้ว่า

$$\phi(\tau) = \phi(\tau+p) \quad (2.2)$$

นั่นคือ ออโตคอร์รีเลชันมีความเป็นคาบด้วยระยะคาบเดียวกัน เป็นตัวบ่งชี้ความเป็นคาบของสัญญาณ โดยมีคุณสมบัติที่สำคัญ คือ

เป็นฟังก์ชันคู่ โดย $\phi(\tau) = \phi(-\tau)$ และ

มีค่ามากที่สุดที่ $\tau=0$ นั่นคือ $|\phi(\tau)| \leq \phi(0)$ ; สำหรับทุกค่า $\tau$

พิจารณาจากสมการที่ *2.2* ควบคู่ไปกับคุณสมบัติทั้งสองประการ จะพบความเป็นคาบโดยมีแซมเปิลของออโตคอร์รีเลชันที่มีค่ามากที่สุดที่ 0, ±p, ±2p, ... โดยไม่ต้องคำนึงถึงเวลาเริ่มต้น (Time origin) ของสัญญาณ ในส่วนของคาบของสัญญาณจะหาได้จากตำแหน่งแรกที่มีค่ามากที่สุดในออโตคอร์รีเลชันฟังก์ชัน และตำแหน่งถัดไปจากยอด*กราฟ* [3]

สัญญาณเสียงพูดจะต้องทำการแบ่งเป็นช่วงสั้น ๆ เพื่อหาข้อมูลที่ต้องการ โดยใช้ Short-time autocorrelation function ซึ่งนิยามได้ดังนี้

$$R(\tau) = \sum_{m=0}^{N-1-\tau} s(m)s(m+\tau) \quad (2.3)$$

เมื่อ N คือ จำนวนตัวอย่าง (Sample) ต่อ frame

$\tau$ คือ จุดเลื่อนไปในการคำนวนออโตคอร์รีเลชัน

ในการเลือกค่าของจำนวนตัวอย่าง (N) ในแต่ละเฟรม สัญญาณจะต้องมีความเป็นคาบที่สมบูรณ์ (Complete period) อย่างน้อย 2-3 คาบ โดยความยาวของคลื่นสัญญาณจะมีผลต่อการคำนวณ $R_n(\tau)$ เนื่องจากค่าของ $R_n(\tau)$ จะลดลงเรื่อย ๆ เมื่อ $\tau$ มีค่าเพิ่มขึ้น

**2.9.2.2 การกำจัดผลของโครงสร้างฟอร์แมนท์ด้วยวิธีเซ็นเตอร์คลิปปิง**

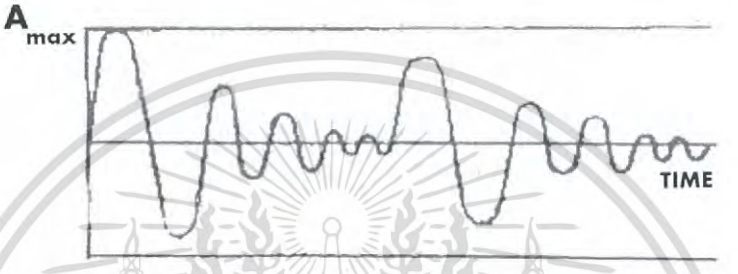
จากรูปภาพที่ 2.4 คลื่นสัญญาณตอบสนองทางความถี่ที่เกิดในช่องทางเดินที่มียอดกราฟจำนวนมาก หากนำไปทำออโตคอร์รีเลชันอาจทำให้เกิดความผิดพลาดไปจากคาบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

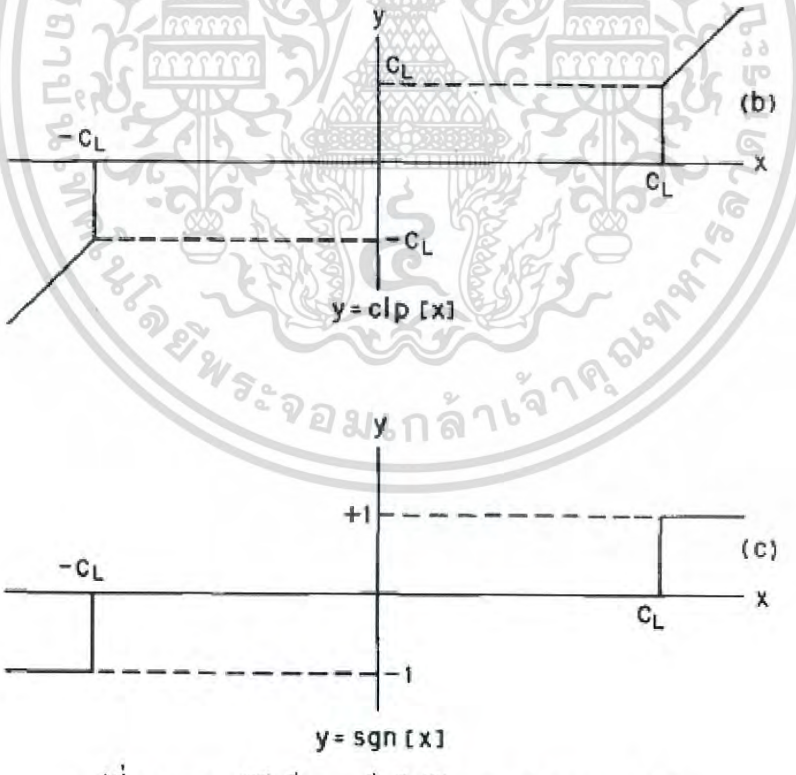
146218

จริง เนื่องจากผลตอบสนองความถี่ในช่องทางเดินเสียงจะมีขนาดใหญ่กว่ายอดกราฟที่เกิดจากความเป็นคาบของแหล่งกำเนิดเสียง (Vocal excitation)

เพื่อหลีกเลี่ยงจึงได้มีการคิดค้นวิธีการที่จะสามารถจัดการกับปัญหานี้ได้ คือทำให้คาบของสัญญาณมีความเด่นชัด โดยการขจัดลักษณะของสัญญาณที่จะทำให้เกิดความไขว้เขว (Distracting) เทคนิคนี้เรียกว่า การทำสเปกตรัมราบเรียบ (Spectrum flatteners) โดยวิธีเซ็นเตอร์คลิปปิงก็เป็นหนึ่งในนั้น สามารถนำมาใช้คำนวณได้จากสัญญาณโดยตรง ซึ่งถูกใช้การแปลงสัญญาณแบบไม่เป็นเชิงเส้นที่พัฒนามาจาก Sondhi ในปี 1968

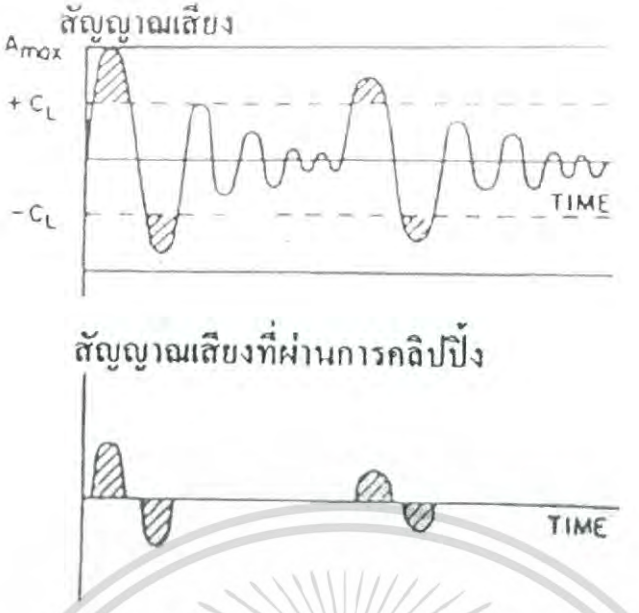



รูปที่ 2.4 รูปทรงในแต่ละคาบของสัญญาณเสียงพูด [3]




รูปที่ 2.5 แสดงวิธีเซ็นเตอร์คลิปปิงของ Dobnoeski [12]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 2.6** ตัวอย่างแสดงการคลิปสัญญาณเสียงพูด [3]

สัญญาณเสียงจะถูกนำมาหาค่าแอมพลิจูดสูงสุด $A_{max}$ เพื่อกำหนดระดับในการทำคลิปปิง (Clipping level : $C_L$) ค่าของสัญญาณที่มีค่าอยู่ต่ำกว่าระดับจะถูกกำหนดให้เป็นศูนย์ โดยจะต้องระมัดระวังไม่ให้ระดับการคลิปที่ตั้งไว้สูงเกินไปจนระดับสัญญาณหาย [3] สามารถคำนวณหาระดับคลิปปิงได้จากสมการต่อไปนี้

$$C_L=(S\%)\times \min(K_1,K_2) \tag{2.4}$$

โดยที่ $C_L$ = Clipping Level

$K_1$ = Absolute Amplitude Peak ของ 100 Samples แรกของ Frame

$K_2$ = Absolute Amplitude Peak ของ 100 Samples ท้ายของ Frame

S% = เปอร์เซ็นต์ที่กำหนดขึ้น (อยู่ภายในช่วง 20-80%)

ค่าของสัญญาณที่ผ่านการคลิปจะเป็นดังนี้

เป็นค่า +1 เมื่อ $A_{s(m)} > C_L$

เป็นค่า -1 เมื่อ $A_{s(m)} < C_L$

และเป็นค่า 0 เมื่อ $-C_L \le A_{s(m)} \le C_L$

โดยฟังก์ชันนี้เรียกว่า เซ็นเตอร์คลิปปิงแบบ 3 ระดับ (3-level center clipping) [3]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**2.9.2.3 การทำนอร์มัลไลซ์ออโตคอร์รีเลชัน**

จากรูปที่ 2.7 (a) เป็นช่วงแรกของสัญญาณเสียงที่ยังไม่คงที่ (Nonstationary) เมื่อนำมาคำนวณอาจเกิดข้อผิดพลาดขึ้นได้ดังรูปที่ 2.7 (b) จะเห็นได้ว่าผลที่ได้นั้นมีค่าขนาดของยอดคลื่นผิดพลาดไปจากตำแหน่งความเป็นจริง เพื่อให้การคำนวณหาระยะคาบมีความถูกต้องมากขึ้นนั้นจึงทำการนอร์มัลไลซ์ ดังรูป 2.7 (c) คือผลของการทำออโตคอร์รีเลชันด้วยการทำนอร์มัลไลซ์สำหรับสัญญาณต้นเสียงที่ยังไม่คงที่จากสมการที่ความคลาดเคลื่อนของพิทช์

$$E_n = \sum_n \left[ s(m) - s'(m) \right]^2 \tag{2.5a}$$

โดย n นั้นคือช่วงของสัญญาณที่ใช้คำนวณ เพื่อให้ได้ค่า $E_n$ ต่ำที่สุดจึงต้องมีเงื่อนไขว่า $\partial E(\tau,\beta)/\partial\beta = 0$ และจากสมการหาระยะยอดของคลื่นเสียง

$$E(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \left[ s(n) - \beta s(n+\tau) \right]^2 \tag{2.5b}$$

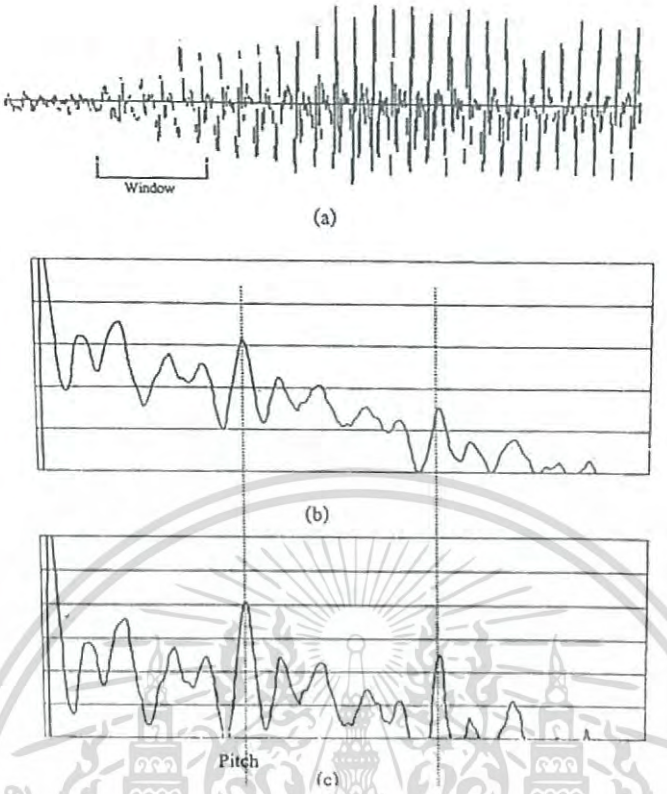
เมื่อค่า $\beta$ คือสัมประสิทธิ์ของอัตราขยายพิทช์ที่ทำการควบคุมการเปลี่ยนแปลงของระดับสัญญาณโดยหาได้จากสมการ (2.6)

$$\beta = \frac{\sum_{n=0}^{N-1} s(n)s(n+\tau)}{\sum_{n=0}^{N-1} s^2(n+\tau)} \tag{2.6}$$

แทนค่าของ $\beta$ เข้าในสมการที่ (2.5)

$$E(\tau,\beta) = \sum_{n=0}^{N-1} n^2(n) - \frac{\left[ \sum_{n=0}^{N-1} s(n)s(n+\tau) \right]^2}{\sum_{n=0}^{N-1} s^2(n+\tau)} \tag{2.7}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




รูปที่ 2.7 แสดงการออโตคอร์รีเลชันของสัญญาณเสียง [3]

(a) สัญญาณเสียงพูด

(b) สัญญาณที่ผ่านออโตคอร์รีเลชันที่ยังไม่ได้ทำการนอร์มัลไลซ์

(c) สัญญาณที่ผ่านออโตคอร์รีเลชันที่ยังทำการนอร์มัลไลซ์แล้ว

จากสมการ (2.7) จะมีส่วนของออโตคอร์รีเลชันกำลังสองตามสมการนี้

$$R_n^2(\tau)=\frac{\left[\sum_{n=0}^{N-1} s(n)s(n+\tau)\right]^2}{\sum_{n=0}^{N-1} s^2(n+\tau)} \tag{2.8}$$

$$R_n(\tau)=\frac{\sum_{n=0}^{N-1} s(n)s(n+\tau)}{\sqrt{\sum_{n=0}^{N-1} s^2(n+\tau)}} \tag{2.9}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.10 การหาค่าความถี่มูลฐานของสัญญาณเสียงพูด

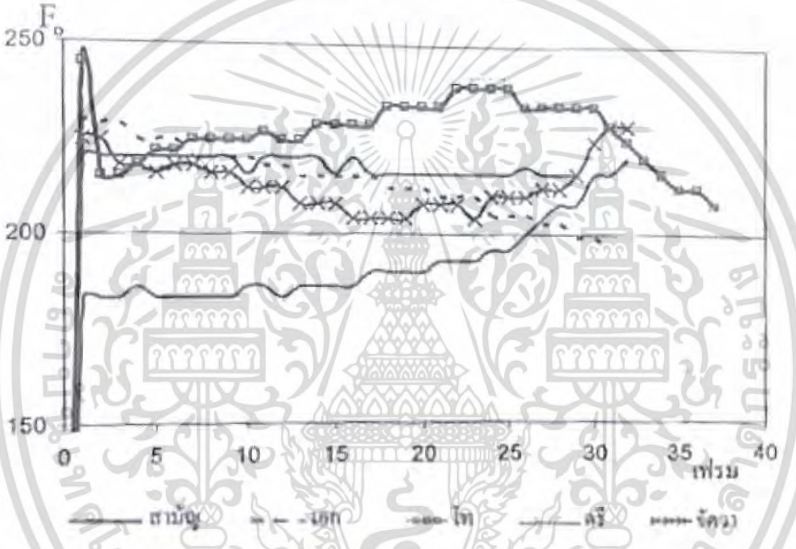
จากค่าคาบเวลาพิทช์ที่ได้นี้สามารถนำมาหาค่าความถี่มูลฐาน $F_0$ ได้จากความสัมพันธ์ที่ (2.10) และผลจากค่าความถี่มูลฐานจะแสดงให้เห็นถึงลักษณะพิเศษของเส้นเสียงของวรรณยุกต์ไทยทั้ง 5 เสียง ดังแสดงในรูปที่ 2.8

$$F_0 = \frac{F_s}{P} \tag{2.10}$$

เมื่อ $F_0$ = ความถี่มูลฐาน (Hz)

$F_s$ = ความถี่ที่ใช้ในการสุ่มสัญญาณ

P = คาบเวลาพิทช์




**รูปที่ 2.8** ภาพแสดงการเปลี่ยนแปลงของความถี่มูลฐานในแต่ละเวลาของเสียงวรรณยุกต์ [3]

## 2.11 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง

การจดจำเสียงพูด เป็นลักษณะหนึ่งของการจดจำรูปแบบ (Pattern recognition) คือการเปรียบเทียบระหว่างแบบทดสอบกับแบบอ้างอิงซึ่งเป็นรูปแบบที่เก็บไว้ล่วงหน้า

ขั้นตอนในการจดจำเสียงพูด แบ่งออกเป็น 2 ขั้นตอน ดังนี้

1. ขั้นตอนการเรียนรู้ (Learning) เป็นขั้นตอนการสร้างแบบอ้างอิงในการจดจำเสียงพูด โดยขั้นตอนนี้จะทำการวิเคราะห์เสียงพูดก่อน โดยจะทำการดึงลักษณะของพารามิเตอร์ที่ต้องการออกมา ซึ่งคือการเปลี่ยนแปลงของค่าความถี่มูลฐานของเสียง จากนั้นทำการจัดกลุ่มพารามิเตอร์โดยใช้การควอนไตซ์ข้อมูลเพื่อนำไปทำการสร้างแบบจำลองอ้างอิงในการรู้จำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ขั้นตอนการจดจำ (Recognition) เป็นการทดสอบการจดจำระหว่างแบบอ้างอิงกับแบบทดสอบ โดยจะทำการเปรียบเทียบพารามิเตอร์ของแบบอ้างอิงกับแบบทดสอบทั้งหมด ซึ่งแบบอ้างอิงที่จะเลือกคือแบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด

สัญญาณเสียงพูดที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้ว ก่อนที่จะถูกนำมาเป็นแบบอ้างอิงหรือใช้เป็นแบบทดสอบ จะต้องผ่านกระบวนการในการเตรียมข้อมูลเสียก่อน เพื่อที่จะขจัดข้อจำกัดอันเนื่องมาจากความถี่มูลฐานที่แตกต่างกันระหว่างการออกเสียงของเพศชายและเพศหญิง เพื่อทำให้แบบจำลองอ้างอิงที่สร้างขึ้นสามารถใช้งานร่วมกันได้ ไม่ว่าผู้ออกเสียงจะเป็นชายหรือหญิง ซึ่งกระบวนการเตรียมข้อมูลมี 2 ขั้นตอน คือ ขั้นตอนการปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง และขั้นตอนการควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน

**2.11.1 การปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง (Median Filtering)**

เนื่องจากเสียงที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้ว อาจมีความไม่ต่อเนื่องของลำดับความถี่ที่เกิดขึ้น ซึ่งเกิดจากความไม่ต่อเนื่องของสัญญาณเสียงในช่วงต้นของการออกเสียง และจากการปัดเศษในการคำนวณ ดังนั้น ขั้นตอนแรกของการเตรียมข้อมูลคือ นำสัญญาณเสียงที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้วมาผ่านตัวกรองค่ากลาง เพื่อปรับปรุงให้ข้อมูลมีความต่อเนื่องเพิ่มมากขึ้น [13] โดยลำดับของความถี่มูลฐานซึ่งเป็นข้อมูลอินพุตจะอยู่ในรูปของข้อมูล 1 มิติ ขนาด [1×T] เมื่อ T คือจำนวนเฟรมของสัญญาณเสียง [4]

**2.11.1.1 ขั้นตอนการกรองค่ากลาง**

ทำการจัดค่าความถี่มูลฐานออกเป็นชุดข้อมูล โดยแต่ละชุดข้อมูลจะประกอบด้วยค่าความถี่ 3 ค่า กำหนดให้มีการเลื่อนของชุดข้อมูล แสดงได้ดังรูป 2.9

**รูปที่ 2.9** การจัดแบ่งความถี่มูลฐานออกเป็นชุดข้อมูล [13]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ $F_1$ = ค่าความถี่มูลฐานของเฟรมที่ 1

$F_2$ = ค่าความถี่มูลฐานของเฟรมที่ 2

$F_T$ = ค่าความถี่มูลฐานของเฟรมสุดท้าย

จากนั้นนำค่าความถี่ทั้ง 3 ค่าในแต่ละชุดข้อมูลมาจัดเรียงใหม่ตามความสัมพันธ์ดังนี้

$$a \leq b \leq c \tag{2.11}$$

a = ความถี่ $F_0$ ที่มีค่าน้อยที่สุดของแต่ละชุดข้อมูล

b = ความถี่ $F_0$ ที่มีค่าอยู่ระหว่างกลาง

c = ความถี่ $F_0$ ที่มีค่ามากที่สุดของแต่ละชุดข้อมูล

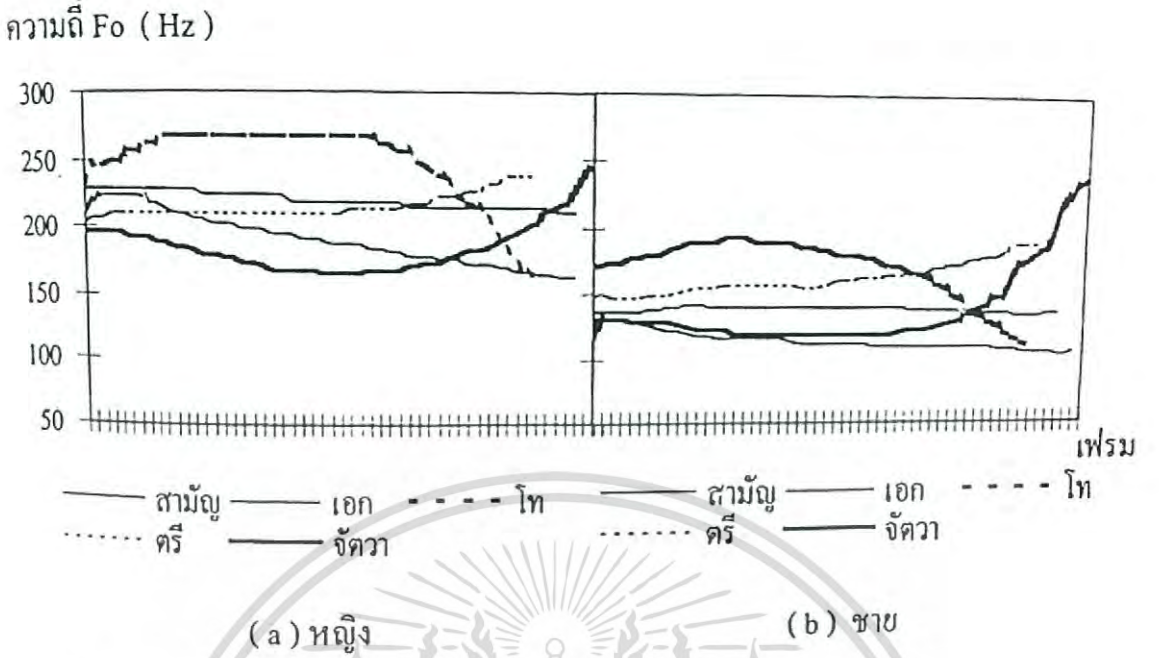
จากนั้นนำความถี่ค่ากลาง (b) ที่ได้จากชุดข้อมูลแต่ละชุดมาจัดเรียงลำดับ จะได้ความถี่มูลฐานชุดใหม่ที่ผ่านกระบวนการกรองค่ากลางแล้ว [4]

**2.11.2 การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน**

ขั้นตอนสุดท้ายของการเตรียมข้อมูลคือ การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน จากข้อเท็จจริงที่ว่าระดับความถี่มูลฐานของเสียงผู้ชายและผู้หญิงมีความแตกต่างกัน

โดยเฉลี่ยแล้ว ในผู้ชายความถี่มูลฐานจะมีค่าอยู่ในช่วง 80-160 Hz และผู้หญิงจะมีความถี่มูลฐานอยู่ในช่วง 160-400 Hz [4]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 2.10** แสดงระดับความถี่มูลฐานที่แตกต่างกันระหว่าง (a) ผู้หญิง และ (b) ผู้ชาย [14]

จากรูปจะเห็นได้ว่าลักษณะการเปลี่ยนแปลงของค่าความถี่มูลฐานในแต่ละระดับเสียงวรรณยุกต์จะมีรูปแบบการเปลี่ยนแปลงที่มีลักษณะเฉพาะอยู่แล้ว ไม่ได้ขึ้นอยู่กับเพศ [4]

ดังนั้นในการศึกษาครั้งนี้จึงเอาลักษณะเด่นนี้มาใช้ในการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ โดยทำการจัดกลุ่มค่าความถี่มูลฐานออกเป็น 3 ระดับตามแนวทางการเปลี่ยนแปลงของความถี่ ($\Delta F$) ที่เพิ่มขึ้นหรือลดลงเมื่อเวลาเปลี่ยนไป

จากสมการคือ

$$\Delta F_t = F_{t+1} - F_t \qquad (2.12)$$

เมื่อ $t = 1,2, \ldots, (T-1)$ โดย T คือ จำนวนเฟรม

$F_t$ = ความถี่ $F_0$ ที่เวลา t

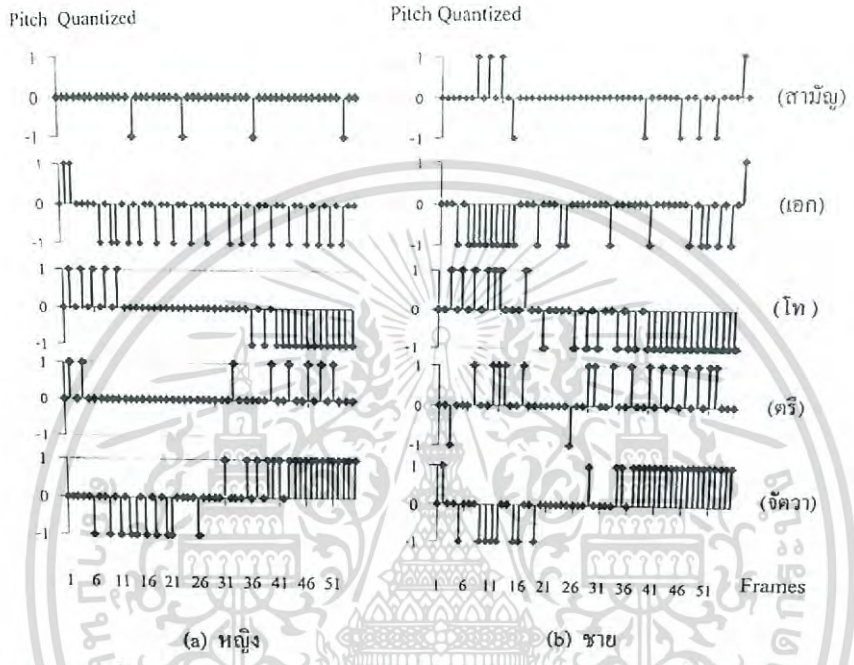
$F_{t+1}$ = ความถี่ $F_0$ ที่เวลา t+1

จากนั้น ทำการควอนไตซ์ $\Delta F$ โดยทำการแบ่งออกเป็น 3 ระดับ ตามทิศทางการเปลี่ยนแปลงของความถี่มูลฐาน กำหนดให้

$$V_t = \begin{cases} 1 \; ; \Delta F_t > 0 \\ 0 \; ; \Delta F_t = 0 \\ -1 \; ; \Delta F_t < 0 \end{cases} \qquad (2.13)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้น $V_t = \{-1, 0, 1\}$ จะถูกนำไปใช้เป็นข้อมูลในการฝึกสอน (Training) เพื่อนำไปใช้ในการสร้างแบบจำลองอ้างอิงของเสียงวรรณยุกต์ต่อไป ซึ่งจะเห็นได้ว่าการควอนไตซ์ความถี่ออกเป็น 3 ระดับนี้ นอกจากช่วยขจัดข้อจำกัดของความถี่มูลฐานที่แตกต่างระหว่างผู้ชายและผู้หญิงแล้ว ยังสามารถช่วยลดเนื้อที่ของหน่วยความจำที่ใช้ในการจัดเก็บข้อมูลได้ และทำให้การคำนวณทำได้เร็วขึ้น เมื่อเทียบกับการใช้ช่วงความถี่มูลฐานทั้งหมดมาสร้างแบบจำลอง [4]




**รูปที่ 2.11** ค่าความถี่มูลฐานตามทิศทางการเปลี่ยนแปลงต่อเวลาที่เพิ่มขึ้นออกเป็น 3 ระดับ [4]

เมื่อนำค่าความถี่มูลฐาน มาทำการจัดระดับค่าการเปลี่ยนแปลงความถี่ออกเป็น 3 ระดับ จากการพิจารณาแนวทางการเปลี่ยนแปลงในระดับเสียงวรรณยุกต์ทั้ง 5 ระดับ เราจะพบว่า

1. เสียงสามัญ ตลอดทั้งเสียง ความถี่มูลฐานมีการเปลี่ยนแปลงลดลงเล็กน้อย
2. เสียงเอก ความถี่มูลฐานของเสียงจะมีค่าลดลงอย่างต่อเนื่อง
3. เสียงโท ความถี่มูลฐานมีค่าเพิ่มขึ้นในช่วงแรกและลดลงต่อเนื่องในช่วงท้ายของเสียง
4. เสียงตรี ความถี่มูลฐานมีแนวโน้มเพิ่มขึ้น
5. เสียงจัตวา ความถี่มูลฐานมีค่าลดลงในช่วงแรกและเพิ่มขึ้นต่อเนื่องในช่วงสุดท้าย

โดยข้อมูลที่ได้จากการควอนไตซ์ออกเป็น 3 ระดับนี้ จะถูกใช้เป็นข้อมูลฝึกสอนหรือข้อมูลทดสอบ ของการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ทั้ง 5 ระดับด้วย Hidden Markov Model [4] [14]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

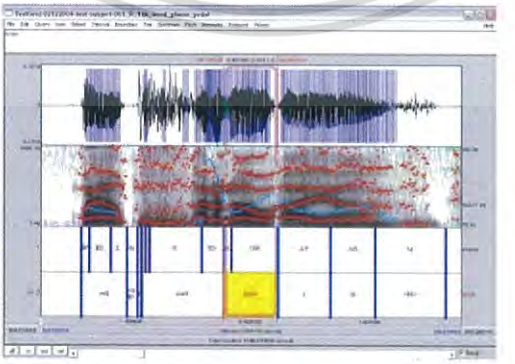
## 2.12 Speak Good Chinese Program

คือโปรแกรมที่พัฒนาโดย Lifeng Liu และ Guangqin Chen โดยเป็นโปรแกรมที่ช่วยในการฝึกออกเสียงคำศัพท์ภาษาจีน (แมนดาริน) ให้ถูกต้อง การทำงานของโปรแกรมจะมีคำตัวอย่างในการออกเสียงที่และมีการแสดง phonetic ที่ช่วยในการออกเสียงวรรณยุกต์ให้เข้าใจขึ้น มีกราฟแสดงการเปรียบเทียบของเสียงที่ถูกต้องกับเสียงของผู้พูดด้วย ซึ่งมีวิธีใช้คือ ผู้ใช้เปิดโปรแกรมขึ้นมาเลือกคำที่ต้องการจะฝึก เมื่อพบคำที่ต้องการก็ทำการบันทึกเสียง จะมีกราฟแสดงและมีผลลัพธ์บอกว่าผู้ใช้ได้ออกเสียงได้ถูกต้องหรือไม่ โดยผู้ใช้สามารถกดฟังเสียงตัวอย่างรวมถึงเสียงของตัวเองที่อัดไปแล้วก็ได้

**รูปที่ 2.12** หน้าโปรแกรม Speak Good Chinese [11]

## 2.13 Praat

คือโปรแกรมที่พัฒนาขึ้นโดย Paul Boersma และ David Weenink จากภาควิชาสัทศาสตร์ มหาวิทยาลัยอัมสเตอร์ดัม ประเทศเนเธอร์แลนด์ โดยโปรแกรมนี้เป็นโปรแกรมที่เกี่ยวข้องกับการวิเคราะห์คลื่นเสียง มีเครื่องมือที่จำเป็นและน่าสนใจในการอ่านค่าและแสดงผลของคลื่นเสียงในรูปแบบต่าง ๆ



**รูปที่ 2.13** ตัวอย่างภาพการวิเคราะห์เสียงของโปรแกรม Praat [10]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.14 GTK+

คือซอฟต์แวร์เสรีพัฒนาโดยมูลนิธิกโนม (GNOME Foundation) เป็นเครื่องมือสำหรับช่วยในการพัฒนาส่วนติดต่อกับผู้ใช้ (Graphical User Interface) ในภาษาซี

**รูปที่ 2.14** รูปภาพสัญลักษณ์ของซอฟต์แวร์ GTK+ [18]

## 2.15 PortAudio

คือซอฟต์แวร์ที่รวบรวม audio I/O library ที่สามารถพัฒนาด้วยภาษาซีได้ โดยสามารถใช้งานได้หลายแพลตฟอร์ม โดยมี API สำหรับการบันทึกเสียง เล่นเสียงให้ใช้ เป็นต้น

**รูปที่ 2.15** รูปภาพสัญลักษณ์ของซอฟต์แวร์ PortAudio

## 2.16 Xcode

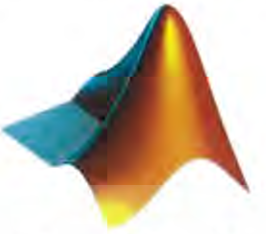
คือ เครื่องมือสำหรับนักพัฒนาบนแพลตฟอร์มของ OS X หรือ iOS รองรับได้หลายภาษา มีเครื่องมือที่ใช้งานง่ายและสะดวก

**รูปที่ 2.16** รูปภาพสัญลักษณ์ของซอฟต์แวร์ Xcode

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.17 MATLAB

คือซอฟต์แวร์ที่ใช้สำหรับการเขียนโปรแกรมมีความสามารถครอบคลุม มีอัลกอริทึมที่มีประโยชน์และใช้งานได้หลากหลาย ผลิตโดยบริษัทแมตเวิรกส์ เริ่มพัฒนาโดย Dr. Cleve Molor ด้วยภาษาฟอร์แทรน และถูกพัฒนาภายใต้โครงการ LINPACK และ EIPACK



รูปที่ 2.17 รูปภาพสัญลักษณ์ของซอฟต์แวร์ MATLAB [16]

## 2.18 ฟังก์ชัน hmmtrain ของ MATLAB

เป็นฟังก์ชันสำหรับการสร้างแบบจำลองโดยจะทำการวิเคราะห์ค่าจาก Transition probability และ Emission Probability ด้วยวิธีการ Hidden Makov Model โดยการใช้ Baum-Welch algorithm มีส่วนที่เป็น input ได้แก่ seq คือ เสียงที่ถูกนำมาแปลงให้อยู่ในรูปแบบ 3 ระดับแล้ว TRGUESS คือ Matrix ของความน่าจะเป็นในการย้ายข้ามมีขนาดเป็น N x N และ EMITGUESS คือ Matrix ของความน่าจะเป็นในการมีอยู่ของค่านั้น ๆ ในตำแหน่งนั้น ๆ มีขนาดเป็น N x จำนวนรูปแบบของ seq โดยค่า output ที่ได้ออกมานั้นจะมีสองค่า ได้แก่ ESTTR คือ Matrix ของ Transition probability ที่ได้จากการคำนวณและ ESTEMIT คือ Matrix ของ Emission probability ที่ได้จากการคำนวณ [17]

## 2.19 ฟังก์ชัน hmmviterbi ของ MATLAB

เป็นฟังก์ชันสำหรับตรวจสอบค่าความใกล้เคียงของ seq กับ Transition probyies และ Emission Probability โดยมี input 3 ค่าได้แก่ seq คือ เสียงที่ถูกนำมาแปลงให้อยู่ในรูปแบบ 3 ระดับแล้วโดยจะเป็นเสียงที่ใช้ในการเทียบทดสอบ TRANS คือ Matrix Model ของความน่าจะเป็นในการย้ายข้ามมีขนาดเป็น N x N และ EMITS คือ Matrix Model ของความน่าจะเป็นในการมีอยู่ของค่านั้น ๆ ในจุดตำแหน่งนั้น ๆ มีขนาดเป็น N x จำนวนรูปแบบของ seq โดย output ที่ได้จากฟังก์ชันจะเป็นค่าของความน่าจะเป็นของความใกล้เคียงระหว่าง seq และ model ซึ่งถ้าค่ายิ่งมากแปลว่ายิ่งใกล้เคียง [16]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

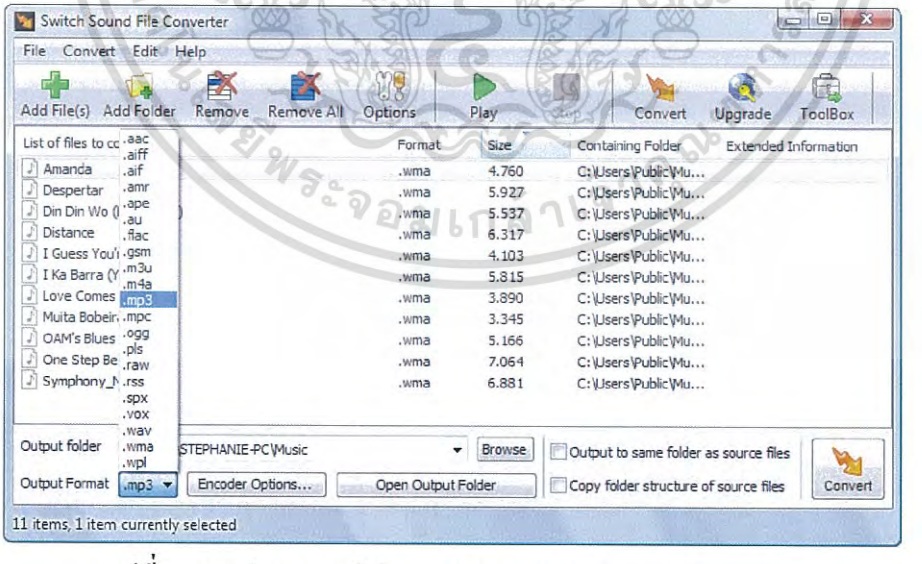
## 2.20 Adobe Photoshop CC

เป็นโปรแกรมประยุกต์ที่ใช้ในการแก้ไขตกแต่งรูปภาพแบบแรสเตอร์ ผลิตโดยบริษัทอะโดบีซิสเต็มส์ โดยรุ่น CC หรือ Creative Cloud เป็นรุ่นล่าสุดในปัจจุบัน

รูปที่ 2.18 รูปภาพของโปรแกรม Adobe Photoshop CC

## 2.21 Switch Audio File Converter

คือโปรแกรมที่ใช้สำหรับการแปลงค่าไฟล์เสียง และสามารถตั้งค่าต่าง ๆ ให้กับไฟล์เสียงได้ โปรแกรมนี้สามารถแปลงไฟล์เสียงได้ทีละหลายไฟล์ พัฒนาขึ้นโดย NCH Software




รูปที่ 2.19 รูปภาพของโปรแกรม Switch Audio File Converter

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 3
# วิธีดำเนินการวิจัย

## 3.1 ศึกษาจากงานวิจัยเก่า

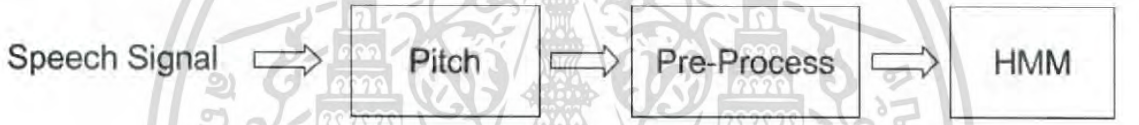
ในการทำแบบจำลองการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทย ได้แบ่งออกเป็น 3 ขั้นตอน ดังนี้

1. การหาค่าพิทช์ ซึ่งทำโดยวิธี Autocorrelation ที่มีการคลิปสัญญาณ
2. การเตรียมข้อมูล เพื่อใช้เป็นแบบจำลอง ซึ่งจะมีอยู่ 2 ขั้นตอน คือ

2.1 การปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองมัธยฐาน (Median Filtering)

2.2 การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน

3. การสร้างแบบจำลองด้วย Hidden Markov Modeling




รูปที่ 3.1 การทำแบบจำลองรู้จำเสียงวรรณยุกต์ [1]

ซึ่งเมื่อทำครบทุกขั้นตอนแล้ว เราจะได้แบบจำลองในการรู้จำเสียงออกมา และสามารถนำไปใช้ในการพัฒนาระบบต่อไปได้ [1]

## 3.2 แนวคิดในการพัฒนาแอปพลิเคชัน

จากการที่ได้ศึกษาแบบจำลองการรู้จักเสียงวรรณยุกต์สำหรับภาษาไทยมานั้น จึงคิดว่ามีความเหมาะสมที่จะสามารถนำไปประยุกต์เพื่อพัฒนาเป็นระบบเพื่อช่วยในการฝึกฝนการผันวรรณยุกต์ภาษาไทยต่อไปได้ โดยระบบจะมีคำมาให้ในการฝึกฝน ซึ่งจะรับเสียงของผู้ใช้มาเทียบกับแบบจำลองของเสียง และจะมีผลออกมาว่าที่ผู้ใช้ได้ทำการฝึกฝนกับระบบนั้น ผู้ใช้ออกเสียงถูกหรือผิด และควรแก้ไขอย่างไร ภายในระบบจะมีเสียงตัวอย่างให้ฟัง เพื่อทำความเข้าใจ และพูดตามได้ โดยจะมีวิธีการดังต่อไปนี้

### 3.2.1 การสร้างแบบจำลอง

ใช้ต้นแบบเสียงทั้งหมด 8 คน ประกอบด้วยผู้ชาย 4 คน และผู้หญิง 4 คน โดยจะให้พูดตามคำที่ได้ทำการเลือกไว้สำหรับการสร้างเป็นแบบจำลอง ซึ่งในการวิจัยนี้ จะใช้หน่วยเสียงพยัญชนะต้น พยัญชนะสะกด หน่วยเสียงสระ แตกต่างกันไป เพื่อให้ได้แบบจำลองที่มีความหลากหลายมากขึ้น และให้ครบกับเสียงวรรณยุกต์ทั้ง 5 เสียง [2]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**3.2.2 การคัดเลือกผู้เป็นต้นแบบเสียง**

ในการวิจัยนี้ จะใช้ผู้บอกภาษาที่เป็นคนไทย เป็นเพศชายจำนวน 4 คน และเพศหญิง จำนวน 4 คน โดยภาษาที่ใช้จะเป็นภาษาไทยของคนภาคกลาง

**3.2.3 การเก็บข้อมูล**

ในการเก็บข้อมูลของการวิจัยนี้ จะใช้คำพูดภาษาไทยให้ผู้ที่เป็นต้นแบบเสียงอ่านออกเสียงคำที่ได้กำหนดไว้จำนวน 10 ชุด แล้วทำการบันทึก ซึ่งข้อมูลเสียง 1 ตัวอย่าง จะถูกแทนด้วยข้อมูล 8 บิท โดยใช้ความถี่ในการ Sampling เท่ากับ 11.025 kHz รูปแบบไฟล์ข้อมูลเป็น .wav ซึ่งจะนำไปใช้ในการวิเคราะห์ต่อไป [2]

**3.2.4 เครื่องมือและอุปกรณ์ที่ใช้ในการวิจัย**

- คอมพิวเตอร์พกพา MacBook
- โปรแกรมที่ใช้ในการทำวิจัยใช้ MATLAB เวอร์ชั่น R2015a (8.5.0.197613)
- Xcode ใช้ในการเขียนภาษาซี
- GTK+ สำหรับพัฒนา GUI
- PortAudio สำหรับ addio I/O library
- โปรแกรม Adobe Photoshop CC
- โปรแกรม Switch Audio File Converter

**3.2.5 ข้อจำกัดในการดำเนินการวิจัย**

ต้นแบบของเสียง สามารถพูดและออกเสียงภาษาไทยของคนภาคกลางได้อย่างชัดเจน

## 3.3 การวิเคราะห์ระบบที่ต้องการออกแบบ

**3.3.1 จุดประสงค์ของโครงงานนี้**

ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง มีจุดประสงค์ที่จะช่วยในการฝึกผันและออกเสียงวรรณยุกต์ไทยให้แก่ชาวต่างชาติและผู้ที่ต้องการฝึกออกเสียงวรรณยุกต์ไทยได้ด้วยตนเอง และสามารถช่วยตรวจสอบได้ว่าถูกต้องหรือไม่

**3.3.2 รูปแบบของระบบ**

ระบบจะพัฒนาด้วยโปรแกรม MATLAB ซึ่งใช้ในส่วนของการทำแบบจำลองอ้างอิงเสียงวรรณยุกต์ และพัฒนาขึ้นเป็นแอปพลิเคชันด้วยภาษาซี ร่วมกับ GTK+ และ PortAudio

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**3.3.3 ความต้องการที่เป็นหน้าที่หลักของระบบ (Functional Requirement)**

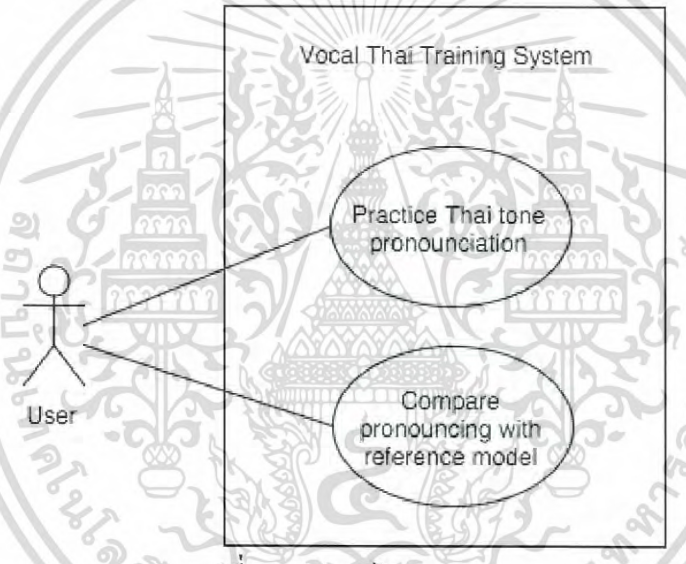
- ระบบสามารถเปรียบเทียบเสียงที่รับเข้ามากับแบบจำลองได้
- ระบบสามารถรู้จำเสียงวรรณยุกต์ได้
- ระบบสามารถแยกเสียงได้ตามเสียงวรรณยุกต์
- ระบบสามารถแสดงผลจากการฝึกฝนได้ว่าถูกหรือผิด

**3.3.4 ความต้องการที่ไม่ใช่หน้าที่หลักของระบบ (Non-Functional Requirement)**

- ระบบมีความสวยงาม

## 3.4 การวิเคราะห์และออกแบบระบบ

**3.4.1 ยูสเคสไดอะแกรม**




รูปที่ 3.2 ยูสเคสไดอะแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

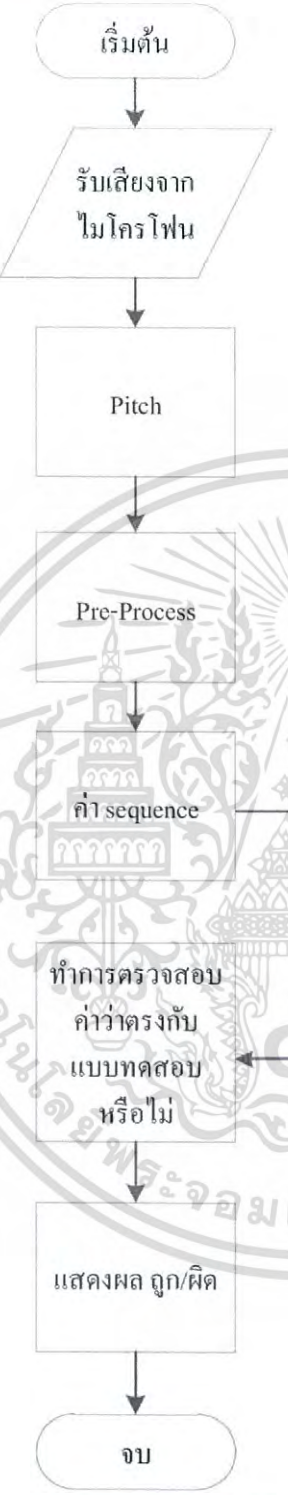
## 3.5 User Story

ตารางที่ 3.1 User Story ของผู้ใช้งานแอปพลิเคชัน

| | User Story |
|---|---|
| User 1 | As a user, I want to practice Thai tone pronunciation. |
| User 2 | As a user, I want to know what tone is, from my pronunciation. |
| User 3 | As a user, I want to know the different pronunciation of five tones. |
| User 4 | As a user, I want to improve my pronunciation. |
| User 5 | As a user, I want to pronounce Thai words confidently. |

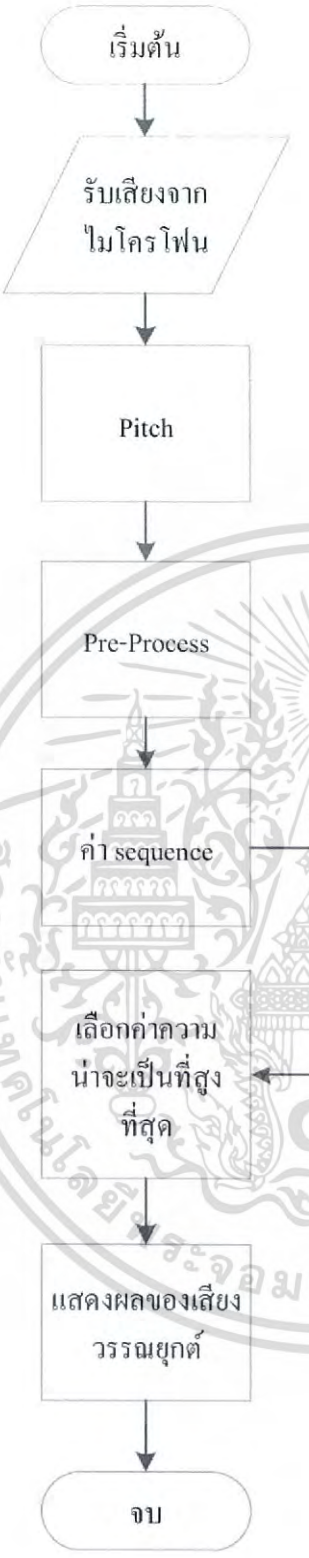
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.6 Activity Diagram




**รูปที่ 3.3** ลำดับการทำงานระบบฝึกฝนการออกเสียงวรรณยุกต์กับแบบทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 3.4** ลำดับการทำงานการเปรียบเทียบการออกเสียงวรรณยุกต์กับเสียงที่รับเข้ามา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 3.7 การดำเนินงาน

### 3.7.1 การวิเคราะห์หาคาบพิทช์และเตรียมแบบจำลอง

วิธีการทำแบบจำลองใช้โปรแกรม MATLAB ในการพัฒนา โดยเริ่มจากการเก็บเสียงจากกลุ่มตัวอย่างมาทั้งหมด 8 คน เป็นชาย 4 คน และหญิง 4 คน นำข้อมูลที่ได้มาแปลงเป็นไฟล์ .wav ขนาด มีความละเอียด 8 บิต Sampling rate อยู่ที่ 11.025 KHz มี Audio Channels เป็นแบบ mono เสียงจะถูกนำมาผ่านขั้นตอนการเตรียมข้อมูลทำแบบจำลอง (Pre-Process) ตามทฤษฎีโดยสามารถดูได้ในบทที่ 2

### 3.7.2 ขั้นตอนการทำแบบจำลอง

หลังจากได้ Sequence ของเสียงทั้ง 400 เสียงมาแล้วจะนำค่าที่ได้ไปทำการเรียนรู้เพื่อทำแบบจำลองโดยใช้โปรแกรม MATLAB ในการทำขั้นตอนนี้ โดยมีฟังก์ชันที่ใช้คือ hmmtrain มี Structure ของฟังก์ชันเป็น [ESTTR,ESTEMIT] = hmmtrain(seq,TRGUESS,EMITGUESS) โดยเป็นฟังก์ชันสำหรับการสร้างแบบจำลองโดยจะทำการวิเคราะห์ค่าจาก Transition probability และ Emission Probability ด้วยวิธีการ Hidden Makov Model โดยการใช้ Baum-Welch algorithm มีส่วนที่เป็น input ได้แก่ seq คือ เสียงที่ถูกนำมาแปลงให้อยู่ในรูปแบบ 3 ระดับแล้ว TRGUESS คือ Matrix ของความน่าจะเป็นในการย้ายข้ามมีขนาดเป็น N x N และ EMITGUESS คือ Matrix ของความน่าจะเป็นในการมีอยู่ของค่านั้น ๆ ในตำแหน่งนั้น ๆ มีขนาดเป็น N x จำนวนรูปแบบของ seq โดยค่า output ที่ได้ออกมานั้นจะมีสองค่า ได้แก่ ESTTR คือ Matrix ของ Transition probability ที่ได้จากการคำนวณและ ESTEMIT คือ Matrix ของ Emission probability ที่ได้จากการคำนวณ

ค่า Transition probability และ Emission probability เริ่มแรกที่ใช้จะเป็นตัวกำหนด State และการย้ายข้าม State ในงานวิจัยนี้ได้ใช้ State ทั้งหมด 7 States และมีการย้ายข้าม State ได้สูงสุด 2 States โดยเกิดจากการทดลองหลาย ๆ ครั้งแล้วพบว่าผลที่ได้จากจำนวนนี้มีความน่าพอใจในระดับหนึ่ง ค่า Transition probability และ Emission probability เริ่มแรกที่ใช้จะได้จากการสุ่ม โดยจะนำ Sequence ของเสียงที่ได้ทำการเก็บมาของแต่ละเสียงวรรณยุกต์เข้าไปทำการเรียนรู้ในการสร้างแบบจำลอง โดยจะมีการใช้คำสั่งเพิ่มเติมของฟังก์ชัน hmmtrain คือ 'Pseudoemissions',PSEUDOE และ 'Pseudotransitions',PSEUDOTR โดยทั้งสองคำสั่งนี้จะเป็นการเรียกค่า Transition Probability และ Emission Probability เพื่อหลีกเลี่ยงปัญหาเกี่ยวกับ zero probability ที่อาจเกิดขึ้นได้ในขั้นตอนการทำ viterbi หรือการเทียบเสียงกับแบบจำลอง ค่าที่ได้ใช้เป็น Pseudotransition และ Pseudomission ในการทดลองนี้คือค่า Transition probability และ ค่า Emission Probability เริ่มแรก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั่นเอง เสียงที่ใส่เข้าไปเพื่อใช้ในการสร้างแบบจำลองของแต่ละวรรณยุกต์จะมีขนาด 80 เสียง โดยผลที่ได้จากการเข้าฟังก์ชัน hmmtrain จะเป็นแบบจำลองเสียงที่ได้ โดยมี 2 Matrix ต่อแบบจำลองเสียงวรรณยุกต์เสียง 1 เสียง ในงานวิจัยนี้ Transition Probability ที่ได้มีขนาด 7x7 และ Emission Probability มีขนาด 7x3

**3.7.3 การพัฒนาส่วนติดต่อกับผู้ใช้**

ในส่วนติดต่อกับผู้ใช้ได้พัฒนาขึ้นด้วยภาษาซี เขียนบนโปรแกรม Xcode มีซอฟต์แวร์ช่วย คือ GTK+ เป็น Tools ในการสร้าง GUI (Graphical User Interface) เพื่อติดต่อกับผู้ใช้ ทำให้สามารถใช้งานได้ง่ายขึ้น โดยภาพที่ใช้ในการทำได้ทำการสร้างขึ้นจากโปรแกรม Adobe Photoshop CC ในด้านของเสียงได้ใช้ซอฟต์แวร์ PortAudio เข้ามาพัฒนาเกี่ยวกับ I/O Audio ซอฟต์แวร์นี้จะรวบรวม Library เกี่ยวกับเครื่องมือทางด้านเสียง กล่าวโดยสรุปอย่างง่ายคือสร้างภาพต่าง ๆ ด้วยโปรแกรม Adobe Photoshop CC แล้วนำภาพไปทำการเขียนเป็นแอปพลิเคชันด้วยโปรแกรม Xcode ร่วมกับ GTK+ และใช้ PortAudio ในด้านการรับเสียงจากผู้ใช้

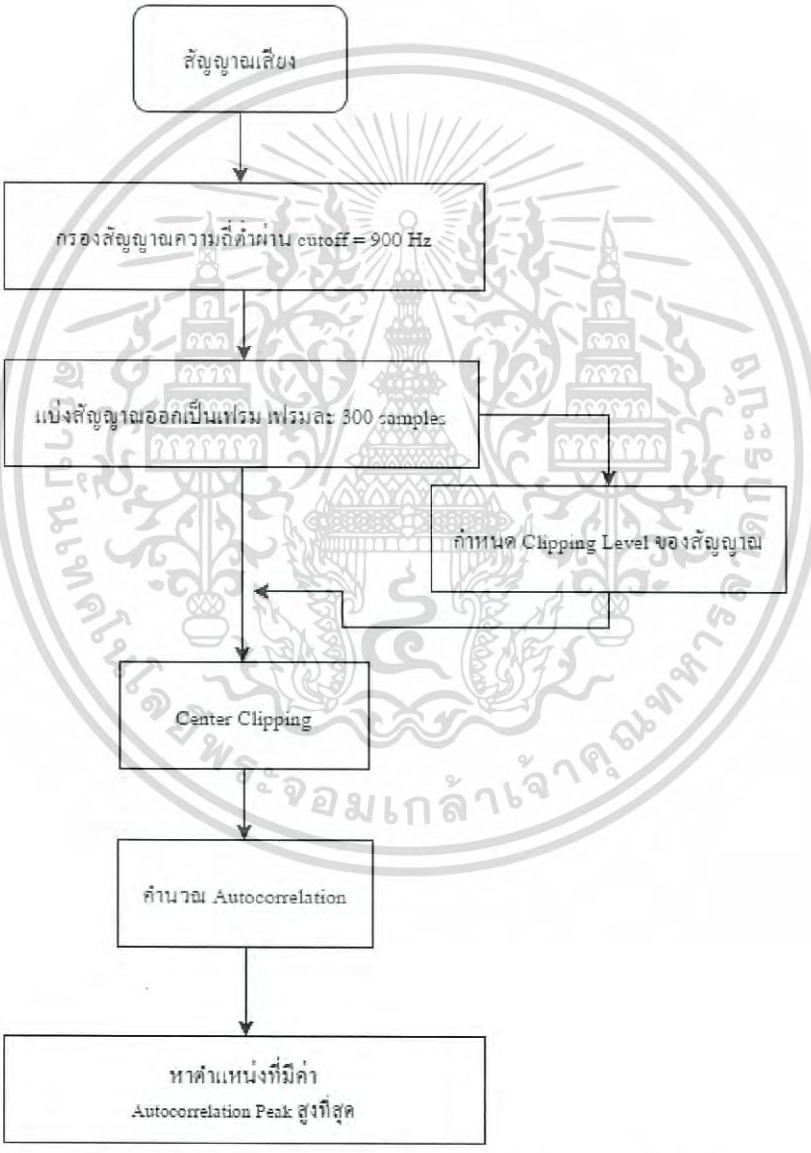
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 4

# ผลการทดลอง

## 4.1 การทดลองสร้างกราฟเสียงของแต่ละวรรณยุกต์

จากการทดลองในโปรแกรม MATLAB โดยการเขียนโปรแกรมอ้างอิงตามงานวิจัยที่ได้ศึกษามาและจากการศึกษาการเขียนโปรแกรมภาษา MATLAB ทำให้ได้อัลกอริทึมดังนี้

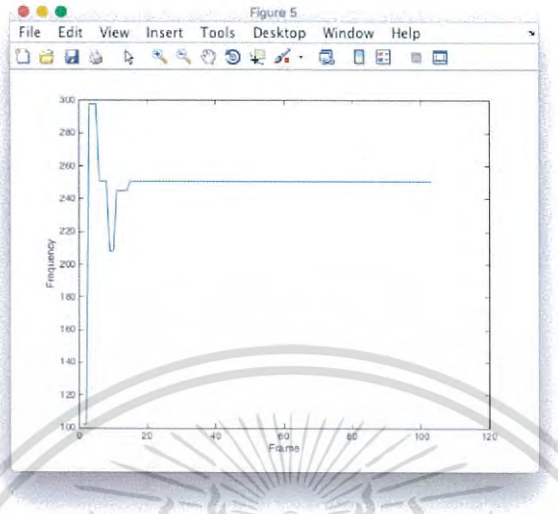



**รูปที่ 4.1** แผนภูมิการทำงานของอัลกอริทึมการหาคลื่นเสียง [1]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

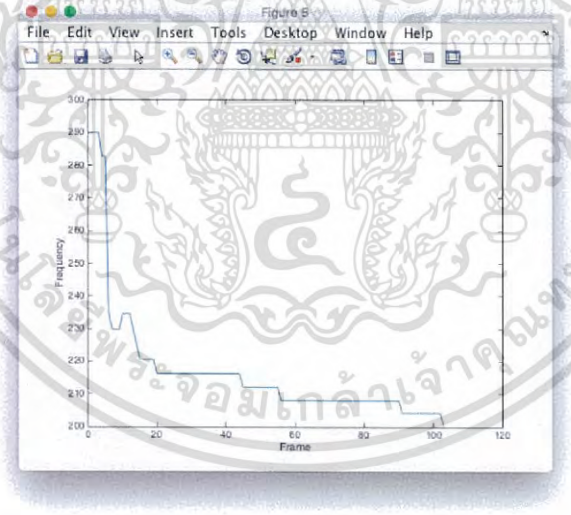
## 4.2 ผลการทดลองจากการสร้างกราฟเส้นเสียง

### 4.2.1 การทดสอบเสียงวรรณยุกต์สามัญจากการออกเสียงคำว่า 'อา'




รูปที่ 4.2 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงสามัญ

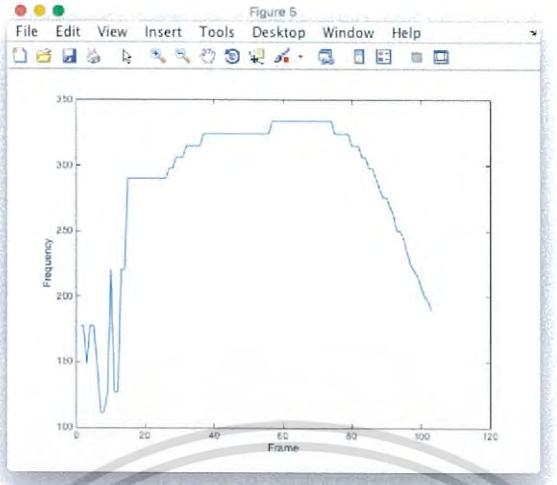
### 4.2.2 การทดสอบเสียงวรรณยุกต์เอกจากการออกเสียงคำว่า 'อ่า'




รูปที่ 4.3 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงเอก

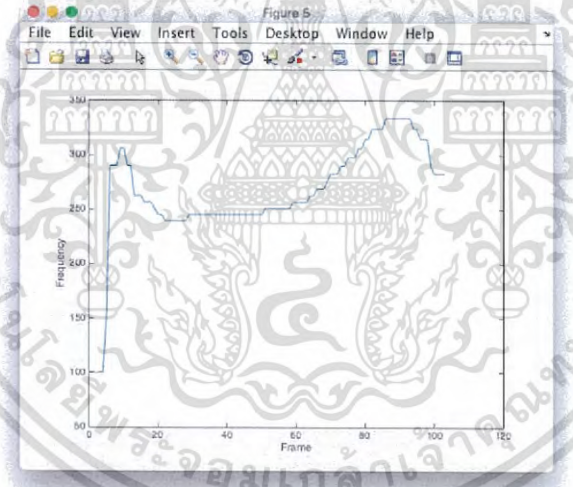
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.2.3 การทดสอบเสียงวรรณยุกต์โทจากการออกเสียงคำว่า ‘อ้า’




รูปที่ 4.4 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงโท

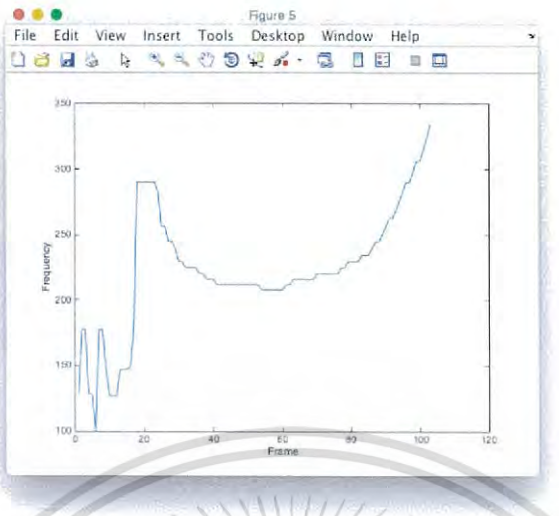
### 4.2.4 การทดสอบเสียงวรรณยุกต์ตรีจากการออกเสียงคำว่า ‘อ๊า’




รูปที่ 4.5 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงตรี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.2.5 การทดสอบเสียงวรรณยุกต์จัตวาจากการออกเสียงคำว่า 'อ๋า'




รูปที่ 4.6 ผลการทดสอบจากการทดลองเสียงวรรณยุกต์เสียงจัตวา

## 4.3 คำที่ใช้ในการสร้างแบบจำลองและผลแบบจำลองอ้างอิง

ตารางที่ 4.1 กลุ่มคำที่ใช้ในการสร้างแบบจำลองอ้างอิง

| คำที่ | ระดับเสียงวรรณยุกต์ | | | | |
|---|---|---|---|---|---|
| | สามัญ | เอก | โท | ตรี | จัตวา |
| 1 | ตอน | ถ่อ | เนิบ | เย้ย | ไถ |
| 2 | งา | สื่อ | ตุ้ย | ง้อ | แหวน |
| 3 | เอน | จอก | เทือก | แม้ | สาม |
| 4 | ดาว | เก่า | ลาบ | ท้อง | ผิว |
| 5 | ทอ | จุ่ม | อึ้ง | โม้ | หนอ |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 4.1 (ต่อ)** กลุ่มคำที่ใช้ในการสร้างแบบจำลองอ้างอิง

| | | | | | |
|---|---|---|---|---|---|
| 6 | แบ | ส่ง | ล่อ | เว้น | เขย |
| 7 | คาย | เกิด | ยืด | มื้อ | หยี |
| 8 | นาน | แก่ | เท่า | ยุ้ง | เหงา |
| 9 | ดู | จ่า | เลิก | แท้ | ขุย |
| 10 | แนว | ตืด | โง่ | ม้า | โขง |

## 4.4 แบบจำลองการรู้จำเสียงของวรรณยุกต์

### 4.4.1 แบบจำลองการรู้จำเสียงวรรณยุกต์สามัญ (Middle Tone)

$Trans_{7x7}$ =

$$\begin{bmatrix} 0.0009 & 0.4639 & 0.5352 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.9012 & 0.0001 & 0.0988 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0182 & 0.1278 & 0.8540 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.9068 & 0.0057 & 0.0876 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.1570 & 0.8430 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9719 & 0.0281 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

$Emis^{T}_{3x7}$ =

$$\begin{bmatrix} 0.0043 & 0.0000 & 0.0000 & 0.0734 & 0.6911 & 0.0859 & 0.4137 \\ 0.9905 & 1.0000 & 1.0000 & 0.8670 & 0.0036 & 0.8604 & 0.5294 \\ 0.0052 & 0.0000 & 0.0000 & 0.0596 & 0.3053 & 0.0537 & 0.0569 \end{bmatrix}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.4.2 แบบจำลองการรู้จำเสียงวรรณยุกต์เอก (Low Tone)

$Trans_{7x7}$ =

$$\begin{bmatrix} 0.4800 & 0.5078 & 0.0122 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.8871 & 0.1129 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.9743 & 0.0257 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0438 & 0.9562 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9997 & 0.0003 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0114 & 0.9886 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

$Emis^{T}_{3x7}$ =

$$\begin{bmatrix} 0.0000 & 0.4062 & 0.3048 & 0.0000 & 0.2472 & 0.0000 & 0.2697 \\ 0.9220 & 0.5892 & 0.6952 & 0.9617 & 0.0000 & 0.9713 & 0.5277 \\ 0.0780 & 0.0046 & 0.0000 & 0.0383 & 0.7528 & 0.0287 & 0.2026 \end{bmatrix}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.4.3 แบบจำลองการรู้จำเสียงวรรณยุกต์โท (Falling Tone)

$Trans_{7x7}$ =

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.9015 | 0.0985 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.8851 | 0.0000 | 0.1149 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.8125 | 0.1808 | 0.0067 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.8652 | 0.1348 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.9096 | 0.0904 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.7950 | 0.2050 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |

$Emis^{T}_{3x7}$ =

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.0468 | 0.0673 | 0.5291 | 0.2071 | 0.5187 | 0.8465 | 0.3127 |
| 0.8000 | 0.9188 | 0.4703 | 0.7929 | 0.4432 | 0.1535 | 0.4262 |
| 0.1532 | 0.0139 | 0.0006 | 0.0000 | 0.0381 | 0.4262 | 0.2612 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.4.4 แบบจำลองการรู้จำเสียงวรรณยุกต์ตรี (High Tone)

$Trans_{7x7}$ =

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.7505 | 0.2485 | 0.0010 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.9116 | 0.0000 | 0.0884 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.1786 | 0.5214 | 0.3000 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.9235 | 0.0765 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.8536 | 0.1464 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.9430 | 0.0570 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |

$Emis^{T}_{3x7}$ =

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.2482 | 0.1043 | 0.0507 | 0.0980 | 0.0040 | 0.0002 | 0.4580 |
| 0.7515 | 0.6556 | 0.9305 | 0.9020 | 0.8513 | 0.5784 | 0.4648 |
| 0.0003 | 0.2401 | 0.0188 | 0.0000 | 0.1447 | 0.4214 | 0.0772 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.4.5 แบบจำลองการรู้จำเสียงวรรณยุกต์จัตวา (Rising Tone)

$Trans_{7x7}$ =

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.8867 | 0.1133 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.9308 | 0.0692 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.9010 | 0.0990 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.9485 | 0.0515 | 0.0000 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.9288 | 0.0712 | 0.0000 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0605 | 0.9395 |
| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |

$Emis^{T}_{3x7}$ =

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.2497 | 0.3630 | 0.0480 | 0.0450 | 0.0000 | 0.0000 | 0.3339 |
| 0.6498 | 0.6369 | 0.9277 | 0.5384 | 0.1989 | 0.9974 | 0.2928 |
| 0.1005 | 0.0001 | 0.0243 | 0.4166 | 0.8011 | 0.0026 | 0.0704 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.5 ผลจากการทดสอบแบบจำลองการรู้จำเสียงวรรณยุกต์

**ตารางที่ 4.2** ผลการรู้จำระดับเสียงวรรณยุกต์จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงต้นแบบ

| คนที่ | การรู้จำระดับเสียงถูกต้อง | | | | |
|---|---|---|---|---|---|
| | สามัญ | เอก | โท | ตรี | จัตวา |
| M1 | 40.00% | 70.00% | 60.00% | 70.00% | 100.00% |
| M2 | 40.00% | 40.00% | 50.00% | 10.00% | 80.00% |
| M3 | 80.00% | 50.00% | 70.00% | 80.00% | 70.00% |
| M4 | 50.00% | 70.00% | 10.00% | 20.00% | 50.00% |
| F1 | 100.00% | 30.00% | 50.00% | 70.00% | 90.00% |
| F2 | 70.00% | 80.00% | 60.00% | 60.00% | 100.00% |
| F3 | 70.00% | 40.00% | 60.00% | 50.00% | 80.00% |
| F4 | 80.00% | 100.00% | 30.00% | 100.00% | 90.00% |
| เฉลี่ย | 66.25% | 58.75% | 48.75% | 57.50% | 82.50% |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

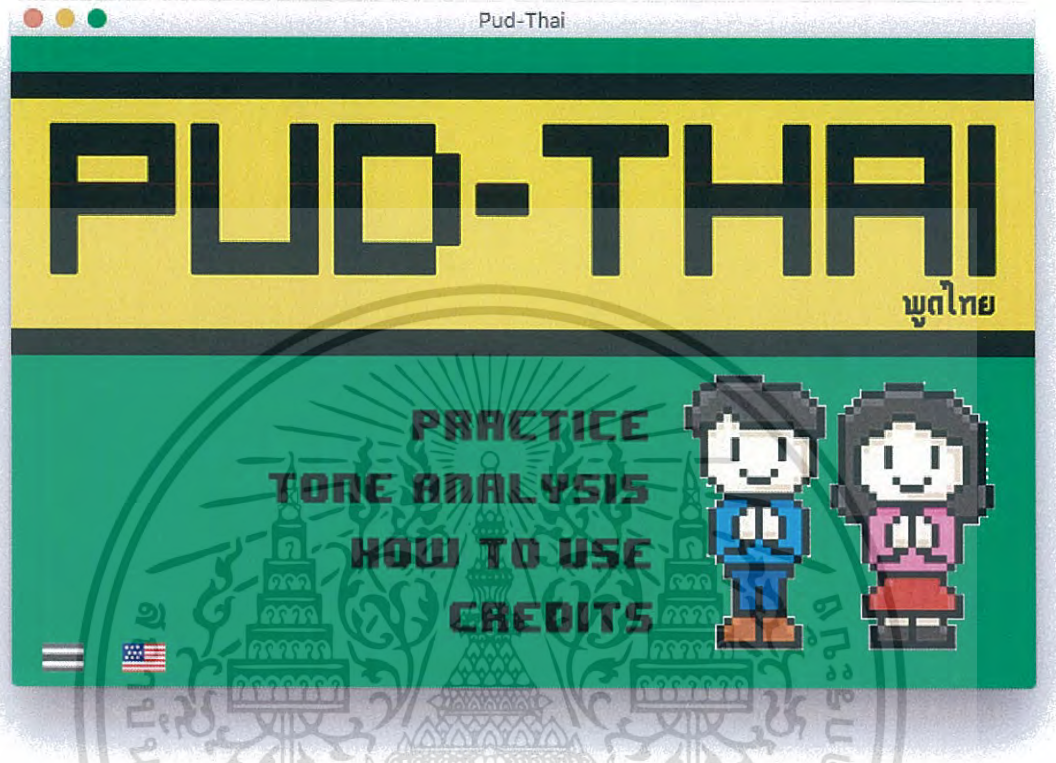
**ตารางที่ 4.2** ผลการรู้จำระดับเสียงวรรณยุกต์จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงกลุ่มใหม่

| คนที่ | การรู้จำระดับเสียงถูกต้อง | | | | |
|---|---|---|---|---|---|
| | สามัญ | เอก | โท | ตรี | จัตวา |
| M1 | 100.00% | 60.00% | 60.00% | 30.00% | 100.00% |
| M2 | 40.00% | 30.00% | 50.00% | 50.00% | 100.00% |
| F1 | 80.00% | 40.00% | 30.00% | 90.00% | 100.00% |
| F2 | 90.00% | 70.00% | 80.00% | 70.00% | 100.00% |
| เฉลี่ย | 77.50% | 50.00% | 55.00% | 60.00% | 100.00% |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.6 การออกแบบหน้าตาแอปพลิเคชัน

- หน้าแรกของแอปพลิเคชัน




รูปที่ 4.7 หน้าแรกของแอปพลิเคชัน

- การทำงานการทดสอบเสียงที่รับเข้ามากับแบบทดสอบที่มีให้จะอยู่ใน Menu Practice จะมีเสียงทั้งหมด 25 เสียง ที่เป็นแบบฝึกหัด โดยจะแบ่งเป็นเสียงวรรณยุกต์ละ 5 เสียงโดยจะมีคำกำหนดมาให้ และมีเสียงให้ฟังว่าอ่านอย่างไร จะให้ผู้ใช้ทำการพูดผ่านไมโครโฟน หลังจากนั้นระบบจะทำการประมวลผล และแสดงผลว่าที่ออกเสียงไปถูกต้องหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- กรณีที่ออกเสียงถูกต้อง ระบบจะแสดงผลว่าถูก

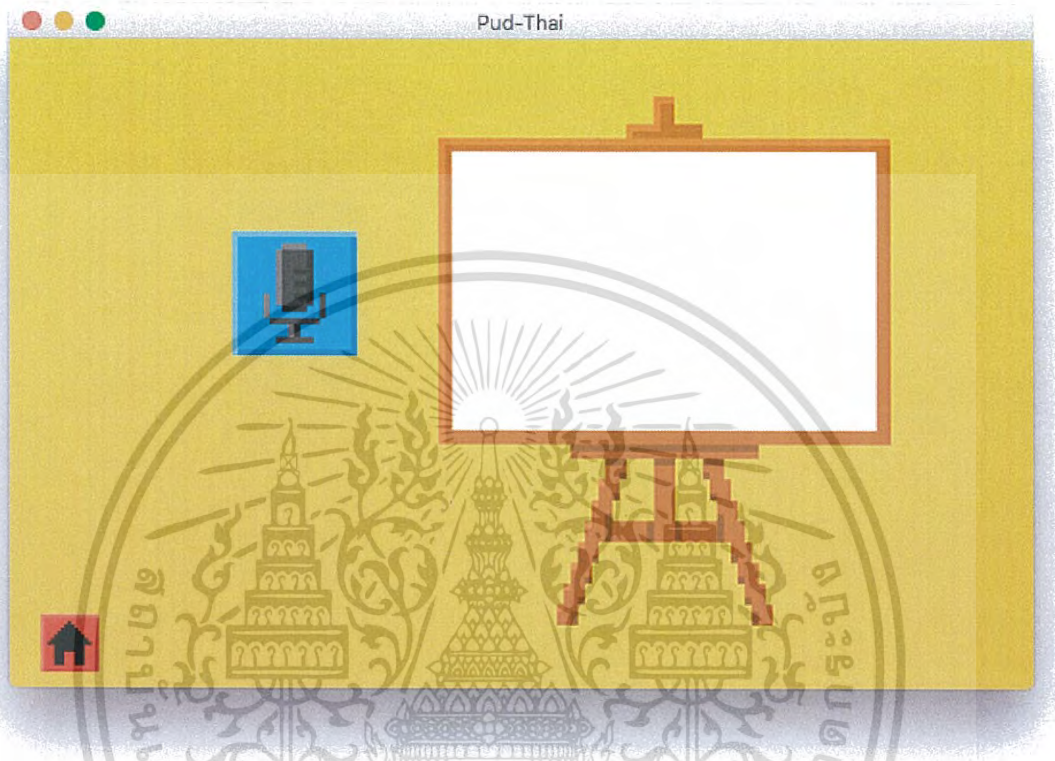
รูปที่ 4.8 ผลการทดสอบในกรณีที่ออกเสียงถูก

- กรณีที่ออกเสียงผิด ระบบจะแสดงผลว่าผิด

รูปที่ 4.9 ผลการทดสอบในกรณีที่ออกเสียงผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- การทำงานการเปรียบเทียบเสียงที่รับเข้ามากับแบบจำลองอ้างอิง จะอยู่ใน Menu Tone Analysis โดยจะให้ผู้ใช้พูดเสียงผ่านไมโครโฟน หลังจากนั้นระบบจะทำการแสดงผลเสียงของวรรณยุกต์ที่ผู้ใช้ได้พูดเข้าไป




**รูปที่ 4.10** หน้า Tone Analysis

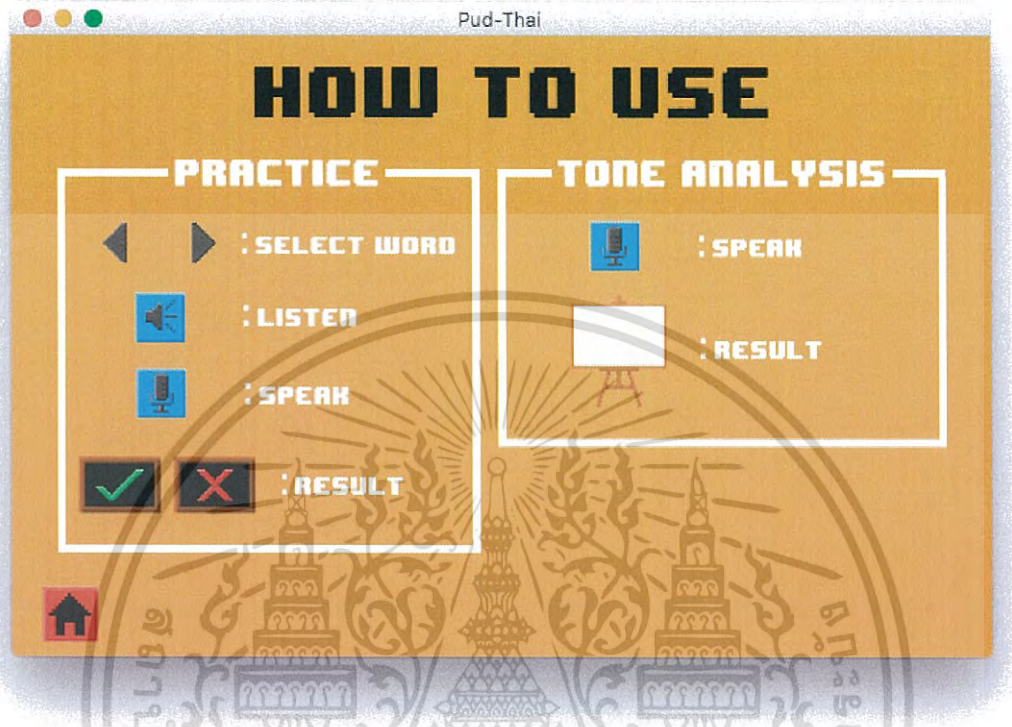
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ระบบแสดงผลเสียงของวรรณยุกต์ หลังจากที่ผู้ใช้ได้ทำการออกเสียง

รูปที่ 4.11 ผลลัพท์จากหน้า Tone Analysis

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- หน้า How to use อธิบายวิธีการใช้งานแอปพลิเคชัน




รูปที่ 4.12 หน้า How To Use

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 5
# สรุปผลการพัฒนา

## 5.1 ผลในการพัฒนา

ทุกวันนี้มีชาวต่างชาติจำนวนมากที่สนใจภาษาไทย และเรียนภาษาไทยมากขึ้น ซึ่งชาวต่างชาติบางคนสามารถออกเสียงภาษาไทยได้ อ่านภาษาไทยได้ แต่เสียงวรรณยุกต์อาจจะยังผิดอยู่ ในภาษาไทยนั้น วรรณยุกต์เป็นสิ่งที่สำคัญมาก เพราะวรรณยุกต์จะทำให้คำมีความหมายเปลี่ยนไปได้ และจะทำให้การสื่อสารผิดพลาดได้ คณะผู้จัดทำจึงได้คิดแอปพลิเคชันขึ้นมา เพื่อช่วยในการแก้ปัญหาเหล่านี้ การพัฒนาแอปพลิเคชัน ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง ผู้พัฒนาได้ศึกษาขั้นตอนและวิธีการสร้างแบบจำลองจากทฤษฎี Hidden Markov Model เพื่อนำมาใช้ในการทำเป็นแบบจำลองอ้างอิงกับเสียงวรรณยุกต์ที่ใช้ในการทดสอบภายในแอปพลิเคชั่น ซึ่งภายในแอปพลิเคชั่นได้มีฟังก์ชันการทำงานหลัก ๆ ของแอปพลิเคชัน 2 ฟังก์ชัน คือ การฝึกออกเสียงวรรณยุกต์จากแบบทดสอบที่มีให้ภายในแอปพลิเคชันโดยจะมีเสียงตัวอย่างให้ฟังก่อนออกเสียง เพื่อทำการทดสอบ และอีกฟังก์ชันคือการออกเสียงวรรณยุกต์ เพื่อดูว่าเป็นเสียงวรรณยุกต์ใด ซึ่งฟังก์ชันหลักนี้ จะสามารถเป็นส่วนที่ช่วยให้ผู้ใช้สามารถฝึกฝนการออกเสียงวรรณยุกต์ไทยได้ ให้ชัดเจนมากขึ้นได้ และเข้าใจถึงความแตกต่างในการออกเสียงของแต่ละเสียงวรรณยุกต์ในภาษาไทยได้ เพื่อที่จะช่วยให้ผู้ใช้ที่สนใจสามารถออกเสียงวรรณยุกต์ไทยได้อย่างมีความมั่นใจและมีความชัดเจนมากขึ้น และจะช่วยให้ภาษาไทยมีการเผยแพร่และเป็นที่รู้จักมากขึ้น ผู้ใช้สามารถใช้โปรแกรมได้หลังจากที่มีเวลาว่าง ๆ จากกิจกรรมต่าง ๆ ในแต่ละวันได้ ระบบจะช่วยให้ผู้ใช้สามารถฝึกฝนการออกเสียงวรรณยุกต์ไทยได้ง่าย และสะดวกมากขึ้น

## 5.2 ปัญหาและอุปสรรค

- แอปพลิเคชันนี้ ใช้ได้สำหรับคอมพิวเตอร์เท่านั้น
- แอปพลิเคชันนี้ ใช้ได้ในที่ ๆ เป็นห้องปิด ไม่มีเสียงรบกวน
- โปรแกรมที่ใช้ในการพัฒนาไม่เคยศึกษามาก่อน จึงต้องเรียนรู้นาน
- แบบจำลองอ้างอิงที่สร้างขึ้น ผลการรู้จำที่ผิดพลาดนั้น สาเหตุหนึ่งอาจเกิดจากการออกเสียงที่ไม่ชัดเจน หรือมีการเอื้อนเสียงในส่วนท้ายของพยางค์มากไป จึงทำให้เกิดการรู้จำที่ผิดพลาดขึ้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บรรณานุกรม

[1] จิตรลดา จารุมิศร์. **"การออกแบบ แบบจำลองในการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทยโดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling."** วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542

[2] ภักดี จันทร์เกษ. **"การรู้จำเสียงตัวเลขภาษาไทย โดยการวิเคราะห์เสียงสระและเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤตและควอนไตซ์พิทช์."** วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2550

[3] คำ ขันทะวีวอน. **"แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling."** วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2546

[4] กฤษกร สุนทรมนทกานติ. **"การหาแบบจำลองที่เหมาะสมการรู้จำเสียงวรรณยุกต์ภาษาไทย โดยใช้วิธีการออโตคอร์รีเลชั่นแบบเซ็นเตอร์คลิปปิ้ง สำหรับการหาค่าพิทช์ในสภาวะที่มีสัญญาณรบกวนเกาส์เซียนขาว."** วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2548

[5] ร้อยกรอง สุขเกิด, นางสาวทิพปภา พิสิษฐ์กุล, กมลพรรณ ทวิชศรี. **"โปรแกรมฝึกออกเสียงวรรณยุกต์ภาษาไทย."** โรงเรียนมหิดลวิทยานุสรณ์. 2550

[6] ธันวา ศรีประโมง. **"การวิเคราะห์เสียงพูดภาษาไทยในแกนความถี่ฮาร์โมนิก."** วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2537

[7] ชรินทร์ เนืองศรี. **"เสียงในภาษาไทย."** [Online]. เข้าถึงได้จาก: http://slideplayer.in.th/slide/1999453/. 2557

[8] วิกิพีเดีย สารานุกรมเสรี. มูลนิธิวิกิมีเดีย. อัปเดต 22 พฤศจิกายน 2557, 1.41 UTC. สารานุกรมออนไลน์. เข้าถึงได้จาก http://th.wikipedia.org/wiki/ฝากล่องเสียง. อินเทอร์เน็ต. เข้าถึงเมื่อ 1 พฤศจิกายน 2558.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บรรณานุกรม (ต่อ)

[9] วิกิพีเดีย สารานุกรมเสรี. มูลนิธิวิกิมีเดีย. อัปเดต 8 มีนาคม 2556, 1.58 UTC. สารานุกรมออนไลน์. เข้าถึงได้จาก http://th.wikipedia.org/wiki/กล่องเสียง. อินเทอร์เน็ต. เข้าถึงเมื่อ 1 พฤศจิกายน 2558.

[10] Fady Biadsy. **"Praat."** [Online]. Available: http://slideplayer.com/slide/5227944. 2009

[11] Lifeng Liu, Guangqin Chen. **"Speak Good Chinese."** [Online]
Available : http://www.speakgoodchinese.org. 2015

[12] LAWRENCE R. RABINER. **"On the Use of Autocorrelation Analysis for Pitch Detection"** IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL.ASSP-25, No. 1, FEBRUARY 1977.

[13] A.Rosenfeld, A.C.Kak. **"Digital picture Processing."** Orlando Florida : Academic Press Inc. 1982.

[14] W. Thomas. **"Voice and Speech Processing."** New York : McGraw-Hill, Inc. 1987.

[15] วิกิพีเดีย สารานุกรมเสรี. มูลนิธิวิกิมีเดีย. อัปเดต 23 กุมภาพันธ์ 2559, 1.58 UTC. สารานุกรมออนไลน์. เข้าถึงได้จาก https://th.wikipedia.org/wiki/อะโดบี_โฟโตชอป. อินเทอร์เน็ต. เข้าถึงเมื่อ 10 มีนาคม 2559.

[16] The MathWorks, Inc. "hmmviterbi." [Online]
Available : http://www.mathworks.com/help/stats/hmmviterbi.html 1998.

[17] The MathWorks, Inc. "hmmtrain." [Online]
Available : http://www.mathworks.com/help/stats/hmmtrain.html?refresh=true 1998.

[18] The GTK+ Team. "The GTK+ Project" [Online]
Available : http://www.gtk.org 2016

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ภาคผนวก ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ก.1 Use case description

**ตารางที่ ก.1** รายละเอียดการทำงานของยูสเคส Practice Thai tone pronunciation

| | | |
|---|---|---|
| Use Case Name : | Practice Thai tone pronunciation | ID : 1 |
| Triggering Name : | ผู้ใช้ต้องการฝึกฝนการออกเสียงวรรณยุกต์ | |
| Brief Description : | เป็นการฝึกฝนการออกเสียงของวรรณยุกต์โดยออกเสียงตามแบบทดสอบที่มีให้จำนวน 25 ชุด โดยแบ่งเป็น 5 เสียงวรรณยุกต์ อย่างละ 5 คำ | |
| Actors : | End user | |
| Related Use Cases : | - | |
| Stakeholders : | End user | |
| Pre-Condition : | - | |
| Post Condition : | - | |
| Minimal Guarantee : | - | |
| Success Guarantee : | สามารถทดสอบและฝึกฝนการออกเสียงได้ | |
| Flow of events : | 1. ผู้ใช้กดเลือกเมนู Practice<br>2. ผู้ใช้กดฟังเสียงตัวอย่างจากคำที่แสดงบนแอปพลิเคชัน<br>3. ผู้ใช้อ่านคำที่แสดงบนแอปพลิเคชัน<br>4. ระบบจะแสดงผลว่าผู้ใช้ออกเสียงได้ถูกหรือผิด | |
| Sub Flows : | - | |
| Extension : | - | |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.1 (ต่อ)** รายละเอียดการทำงานของยูสเคส Practice Thai tone pronunciation

| | |
|---|---|
| Alternative / Exceptional Flows : | - |

**ตารางที่ ก.2** Compare pronouncing with reference model

| | | |
|---|---|---|
| Use Case Name : | Compare pronouncing with reference model | ID : 2 |
| Triggering Name : | ผู้ใช้ต้องการทดสอบการออกเสียงวรรณยุกต์กับแบบอ้างอิง | |
| Brief Description : | เป็นการเปรียบเทียบการออกเสียงวรรณยุกต์ เพื่อดูว่าเป็นเสียงวรรณยุกต์ที่พูดออกมานั้น เป็นเสียงวรรณยุกต์ใด | |
| Actors : | End user | |
| Related Use Cases : | - | |
| Stakeholders : | End user | |
| Pre-Condition : | - | |
| Post Condition : | - | |
| Minimal Guarantee : | - | |
| Success Guarantee : | แสดงผลลัพธ์ของเสียงวรรณยุกต์นั้นได้ | |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ ก.2 (ต่อ)** Compare pronouncing with reference model

| | |
|---|---|
| Flow of events : | 1. ผู้ใช้พูดเสียงผ่านไมโครโฟน<br>2. ระบบจะทำการแสดงผลเสียงวรรณยุกต์ที่ผู้ใช้พูด |
| Sub Flows : | - |
| Extension : | - |
| Alternative / Exceptional Flows : | - |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทย โดยเทคนิคการรู้จำเสียง

**กันต์ อภิวัฒนลังการ[1] และ วิชชุวรรณ สุขไชยศรี[2]**

*คณะเทคโนโลยีสารสนเทศ*

*สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ*

*Emails: thisistakan@gmail.com, may.wichchuwon@gmail.com*

**บทคัดย่อ**

การออกเสียงวรรณยุกต์มีความสำคัญในการสื่อสารของภาษาไทย เพราะวรรณยุกต์ทำให้เกิดคำใหม่และมีความหมายที่แตกต่างกันออกไป ปัจจุบันชาวต่างชาติจำนวนไม่น้อยที่สนใจที่จะเรียนรู้ภาษาไทย ดังนั้นโครงการนี้จึงมีวัตถุประสงค์เพื่อช่วยในการฝึกฝนการออกเสียงวรรณยุกต์ไทยขึ้นโดยใช้เทคนิคการรู้จำเสียง ทางคณะผู้จัดทำได้ทำการเก็บข้อมูลเสียงเพื่อนำไปใช้เป็นข้อมูลฝึกสอนและสร้างแบบจำลองอ้างอิงของเสียงวรรณยุกต์ทั้ง 5 เสียง ด้วยวิธี Hidden Markov Model (HMM) ภายในระบบจะมีฟังก์ชันหลักๆ 2 ฟังก์ชัน อย่างแรกคือ แบบทดสอบการฝึกออกเสียงวรรณยุกต์ให้ตรงกับแบบทดสอบที่เตรียมไว้ และอย่างที่สองคือประมวลผลเสียงที่รับเข้าไปว่าเป็นเสียงวรรณยุกต์ใด ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียงนี้ รองรับการทำงานบนคอมพิวเตอร์แพลตฟอร์ม โดยระบบนี้จะทำให้ผู้ใช้มีความสะดวกในการใช้งานและจะสามารถทำให้ผู้ใช้สามารถเข้าใจและออกเสียงวรรณยุกต์ได้ถูกต้องและชัดเจนมากขึ้น

**คำสำคัญ** – การรู้จำเสียงวรรณยุกต์; วรรณยุกต์ไทย; ฮิดเดนมาร์คอฟโมเดล

## 1. บทนำ

เนื่องจากภาษาไทย มีวรรณยุกต์ที่ใช้ในการประสมคำเพื่อทำให้เกิดเป็นคำศัพท์ใหม่ขึ้นมา แต่ละคำที่มีวรรณยุกต์นั้นจะมีการออกเสียงที่ต่างกันและทำให้ความหมายแตกต่างกันออกไป การออกเสียงวรรณยุกต์จึงมีความสำคัญที่จะต้องออกเสียงให้ถูกต้องเพราะถ้าออกเสียงวรรณยุกต์ผิดเพี้ยนไปจะทำให้เกิดการสื่อความหมายที่ผิดตามไปด้วย ซึ่งในปัจจุบัน มีชาวต่างชาติสนใจและต้องการเรียนรู้ภาษาไทยเพิ่มมากขึ้น แต่ยังไม่ทราบถึงการออกเสียงของวรรณยุกต์ไทยว่ามีความถูกต้องและชัดเจนมากน้อยเพียงใด เพราะการออกเสียงวรรณยุกต์ในภาษาไทยเป็นเรื่องยากสำหรับชาวต่างชาติเพราะในภาษาอังกฤษไม่มีระบบวรรณยุกต์และชาวต่างชาติบางคนก็ไม่มีผู้ที่จะคอยบอกได้ว่าถูกหรือผิด เป็นต้น ด้วยเหตุนี้ทางคณะผู้จัดทำได้เล็งเห็นถึงความสำคัญของการออกเสียงวรรณยุกต์ไทย จึงอยากจะพัฒนาอีกหนึ่งช่องทางในการแก้ไขปัญหาการออกเสียงของวรรณยุกต์ไทยให้มีความถูกต้องและชัดเจนมากขึ้น เพื่อที่จะทำให้ผู้ใช้มีความมั่นใจและสามารถนำไปใช้ในชีวิตประจำวันได้อย่างถูกต้อง

## 2. การหาค่าพิทช์ของสัญญาณเสียงพูด

### 2.1 การวิเคราะห์ในโดเมนเวลา

เสียงที่เปล่งออกมาในช่วงเวลาหนึ่งนั้นขึ้นอยู่กับลักษณะรูปทรงของช่องทางเดินเสียง (Vocal tract) และลักษณะการสั่นของเส้นเสียง (Vocal cord) เสียงพูดของมนุษย์จึงมีลักษณะที่เป็นสัญญาณที่มีคาบเวลาชั่วขณะ (Quisi-periodic) ในการวิเคราะห์จึงต้องแบ่งเสียงพูดออกเป็นช่วงๆ (Frame) เพื่อให้มีลักษณะที่คงที่ [3]

### 2.2 ทฤษฎีการประมาณค่าพิทช์โดยใช้ออโตคอร์รีเลชัน

ฟังก์ชันออโตคอร์รีเลชัน (Autocorrelation) ใช้ตรวจหาคาบพิทช์ของสัญญาณเสียงพูด ทำหน้าที่ในการแสดงยอด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

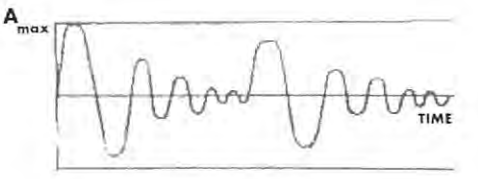
กราฟหลัก (Prominent peak) ทำให้เราสามารถหาระยะคาบของแต่ละช่วงเสียงได้

*2.2.1 การจัดแบ่งวิเคราะห์สัญญาณออกเป็นช่วงสั้นๆ*

กำหนดให้สัญญาณแทนด้วย s(m) ซึ่งออโตคอร์รีเลชันฟังก์ชันของ Discrete-time deterministic signal นั่นคือ ออโตคอร์รีเลชันมีความเป็นคาบด้วยระยะคาบเดียวกัน เป็นตัวบ่งชี้ความเป็นคาบของสัญญาณ พิจารณาจากคุณสมบัติที่สำคัญทั้งสองประการ จะพบความเป็นคาบโดยมีแซมเปิลของออโตคอร์รีเลชันที่มีค่ามากที่สุดที่ 0, ±p, ±2p, ... โดยไม่ต้องคำนึงถึงเวลาเริ่มต้น (Time origin) ของสัญญาณ ในส่วนของคาบของสัญญาณจะหาได้จากตำแหน่งแรกที่มีค่ามากที่สุดในออโตคอร์รีเลชันฟังก์ชัน และตำแหน่งถัดไปจากยอดกราฟ [3] สัญญาณเสียงพูดจะต้องทำการแบ่งเป็นช่วงสั้น ๆ เพื่อหาข้อมูลที่ต้องการ โดยใช้ Short-time autocorrelation function ในการเลือกค่าของจำนวนตัวอย่าง (N) ในแต่ละเฟรม สัญญาณจะต้องมีความเป็นคาบที่สมบูรณ์ (Complete period) อย่างน้อย 2-3 คาบ โดยความยาวของคลื่นสัญญาณจะมีผลต่อการคำนวณ

*2.2.2 การกำจัดผลของโครงสร้างฟอร์แมนท์ด้วยวิธีเซ็นเตอร์คลิปปิง*

จากรูปที่ 1 คลื่นสัญญาณตอบสนองทางความถี่ที่เกิดในช่องทางเดินที่มียอดกราฟจำนวนมาก หากนำไปทำออโตคอร์รีเลชันอาจทำให้เกิดความผิดพลาดไปจากคาบจริง เนื่องจากผลตอบสนองความถี่ในช่องทางเดินเสียงจะมีขนาดใหญ่กว่ายอดกราฟที่เกิดจากความเป็นคาบของแหล่งกำเนิดเสียง (Vocal excitation) เพื่อหลีกเลี่ยงจึงได้มีการคิดค้นวิธีการที่จะสามารถจัดการกับปัญหานี้ได้ คือทำให้คาบของสัญญาณมีความเด่นชัด โดยการขจัดลักษณะของสัญญาณที่จะทำให้เกิดความไขว้เขว (Distracting) เทคนิคนี้เรียกว่าการทำสเปกตรัมราบเรียบ (Spectrum flatteners) โดยวิธีเซ็นเตอร์คลิปปิงก็เป็นหนึ่งในนั้น สามารถนำมาใช้คำนวณได้จากสัญญาณโดยตรงซึ่งถูกใช้การแปลงสัญญาณแบบไม่เป็นเชิงเส้นที่พัฒนามาจาก Sondhi ในปี 1968




**รูปที่ 1.** รูปทรงในแต่ละคาบของสัญญาณเสียงพูด [3]

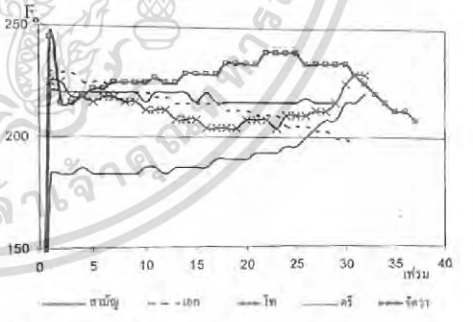
สัญญาณเสียงจะถูกนำมาหาค่าแอมพลิจูดสูงสุด Amax เพื่อกำหนดระดับในการทำคลิปปิง (Clipping level : CL) ค่าของสัญญาณที่มีค่าอยู่ต่ำกว่าระดับจะถูกกำหนดให้เป็นศูนย์ โดยจะต้องระมัดระวังไม่ให้ระดับการคลิปที่ตั้งไว้สูงเกินไปจนระดับสัญญาณหาย ค่าที่ถูกคลิปปิงจะเหลือ 3 ระดับ คือ -1, 0 ,1 โดยฟังก์ชันนี้เรียกว่า เซ็นเตอร์คลิปปิงแบบ 3 ระดับ (3-level center clipping) [3]

*2.2.3 การทำนอร์มัลไลซ์ออโตคอร์รีเลชัน*

การทำนอร์มัลไลซ์ออโตคอร์รีเลชันทำเพื่อให้การคำนวณหาระยะคาบมีความถูกต้องมากขึ้น

**2.3 การหาค่าความถี่มูลฐานของสัญญาณเสียงพูด**

จากค่าคาบเวลาพิทช์ที่ได้นี้สามารถนำมาหาค่าความถี่มูลฐาน $F_0$ และผลจากค่าความถี่มูลฐานจะแสดงให้เห็นถึงลักษณะพิเศษของเส้นเสียงของวรรณยุกต์ไทยทั้ง 5 เสียง ดังแสดงในรูปที่ 2.




**รูปที่ 2.** การเปลี่ยนแปลง $F_0$ ของเสียงวรรณยุกต์ [3]

**2.4 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง**

การจดจำเสียงพูด เป็นลักษณะหนึ่งของการจดจำรูปแบบ (Pattern recognition) คือการเปรียบเทียบระหว่างแบบทดสอบกับแบบอ้างอิงซึ่งเป็นรูปแบบที่เก็บไว้ล่วงหน้า โดยขั้นตอนในการจดจำเสียงพูด แบ่งออกเป็น 2 ขั้นตอน ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.ขั้นตอนการเรียนรู้ (Learning) เป็นขั้นตอนการสร้างแบบอ้างอิงในการจดจำเสียงพูด โดยขั้นตอนนี้จะทำการวิเคราะห์เสียงพูดก่อน โดยจะทำการดึงลักษณะของพารามิเตอร์ที่ต้องการออกมา ซึ่งคือการเปลี่ยนแปลงของค่าความถี่มูลฐานของเสียง จากนั้นทำการจัดกลุ่มพารามิเตอร์โดยใช้การควอนไตซ์ข้อมูลเพื่อนำไปทำการสร้างแบบจำลองอ้างอิงในการรู้จำ

2. ขั้นตอนการจดจำ (Recognition) เป็นการทดสอบการจดจำระหว่างแบบอ้างอิงกับแบบทดสอบ โดยจะทำการเปรียบเทียบพารามิเตอร์ของแบบอ้างอิงกับแบบทดสอบทั้งหมด ซึ่งแบบอ้างอิงที่จะเลือกคือแบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด

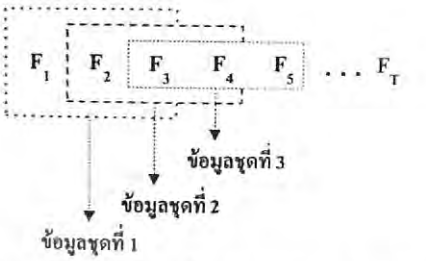
สัญญาณเสียงพูดที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้ว ก่อนที่จะถูกนำมาเป็นแบบอ้างอิงหรือใช้เป็นแบบทดสอบ จะต้องผ่านกระบวนการในการเตรียมข้อมูลเสียก่อน เพื่อที่จะขจัดข้อจำกัดอันเนื่องมาจากความถี่มูลฐานที่แตกต่างกันระหว่างการออกเสียงของเพศชายและเพศหญิง เพื่อทำให้แบบจำลองอ้างอิงที่สร้างขึ้นสามารถใช้งานร่วมกันได้ ไม่ว่าผู้ออกเสียงจะเป็นชายหรือหญิง ซึ่งกระบวนการเตรียมข้อมูลมี 2 ขั้นตอน คือ ขั้นตอนการปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง และขั้นตอนการควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน

*2.4.1 การปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง (Median Filtering)*

เนื่องจากเสียงที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้วอาจมีความไม่ต่อเนื่องของลำดับความถี่ที่เกิดขึ้น ซึ่งเกิดจากความไม่ต่อเนื่องของสัญญาณเสียงในช่วงต้นของการออกเสียงและจากการปิดเศษในการคำนวณ ดังนั้น ขั้นตอนแรกของการเตรียมข้อมูลคือ นำสัญญาณเสียงที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้วมาผ่านตัวกรองค่ากลาง เพื่อปรับปรุงให้ข้อมูลมีความต่อเนื่องเพิ่มมากขึ้น [13] โดยลำดับของความถี่มูลฐานซึ่งเป็นข้อมูลอินพุตจะอยู่ในรูปของข้อมูล 1 มิติ ขนาด [1×T] เมื่อ T คือจำนวนเฟรมของสัญญาณเสียง [4]

*2.4.2 ขั้นตอนการกรองค่ากลาง*

ทำการจัดค่าความถี่มูลฐานออกเป็นชุดข้อมูล โดยแต่ละชุดข้อมูลจะประกอบด้วยค่าความถี่ 3 ค่า กำหนดให้มีการเลื่อนของชุดข้อมูล แสดงได้ดังรูป 3




**รูปที่ 3.** การจัดแบ่งความถี่มูลฐานออกเป็นชุดข้อมูล [13]

เมื่อ $F_1$ = ค่าความถี่มูลฐานของเฟรมที่ 1, $F_2$ = ค่าความถี่มูลฐานของเฟรมที่ 2, $F_T$ = ค่าความถี่มูลฐานของเฟรมสุดท้าย จากนั้นนำค่าความถี่ทั้ง 3 ค่าในแต่ละชุดข้อมูลมาจัดเรียงใหม่ตามความสัมพันธ์ a = ความถี่ $F_0$ ที่มีค่าน้อยที่สุดของแต่ละชุดข้อมูล, b = ความถี่ $F_0$ ที่มีค่าอยู่ระหว่างกลาง, c = ความถี่ $F_0$ ที่มีค่ามากที่สุดของแต่ละชุดข้อมูล จากนั้นนำความถี่ค่ากลาง (b) ที่ได้จากชุดข้อมูลแต่ละชุดมาจัดเรียงลำดับ จะได้ความถี่มูลฐานชุดใหม่ที่ผ่านกระบวนการกรองค่ากลางแล้ว [4]

**2.5 การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน**

ขั้นตอนสุดท้ายของการเตรียมข้อมูลคือ การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน จากข้อเท็จจริงที่ว่าระดับความถี่มูลฐานของเสียงผู้ชายและผู้หญิงมีความแตกต่างกันลักษณะการเปลี่ยนแปลงของค่าความถี่มูลฐานในแต่ละระดับเสียงวรรณยุกต์จะมีรูปแบบการเปลี่ยนแปลงที่มีลักษณะเฉพาะอยู่แล้ว ไม่ได้ขึ้นอยู่กับเพศ [4]

ดังนั้นในการศึกษาครั้งนี้จึงเอาลักษณะเด่นนี้มาใช้ในการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ โดยทำการจัดกลุ่มค่าความถี่มูลฐานออกเป็น 3 ระดับตามแนวทางการเปลี่ยนแปลงของความถี่ ($\Delta F$) ที่เพิ่มขึ้นหรือลดลงเมื่อเวลาเปลี่ยนไป จากนั้น ทำการควอนไตซ์ $\Delta F$ โดยทำการแบ่งออกเป็น 3 ระดับ ตามทิศทางการเปลี่ยนแปลงของความถี่มูลฐาน เป็นค่า -1,0,1 จากนั้น จะถูกนำไปใช้เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลในการฝึกสอน (Training) เพื่อนำไปใช้ในการสร้างแบบจำลองอ้างอิงของเสียงวรรณยุกต์ต่อไป ซึ่งจะเห็นได้ว่าการควอนไตซ์ความถี่ออกเป็น 3 ระดับนี้ นอกจากช่วยขจัดข้อจำกัดของความถี่มูลฐานที่แตกต่างระหว่างผู้ชายและผู้หญิงแล้ว ยังสามารถช่วยลดเนื้อที่ของหน่วยความจำที่ใช้ในการจัดเก็บข้อมูลได้ และทำให้การคำนวณทำได้เร็วขึ้น เมื่อเทียบกับการใช้ช่วงความถี่มูลฐานทั้งหมดมาสร้างแบบจำลอง [4]

เมื่อนำค่าความถี่มูลฐาน มาทำการจัดระดับค่าการเปลี่ยนแปลงความถี่ออกเป็น 3 ระดับ จากการพิจารณาแนวทางการเปลี่ยนแปลงในระดับเสียงวรรณยุกต์ทั้ง 5 ระดับ เราจะพบว่า 1.เสียงสามัญตลอดทั้งเสียง ความถี่มูลฐานมีการเปลี่ยนแปลงลดลงเล็กน้อย 2.เสียงเอก ความถี่มูลฐานของเสียงจะมีค่าลดลงอย่างต่อเนื่อง 3.เสียงโท ความถี่มูลฐานมีค่าเพิ่มขึ้นในช่วงแรกและลดลงต่อเนื่องในช่วงท้ายของเสียง 4. เสียงตรี ความถี่มูลฐานมีแนวโน้มเพิ่มขึ้น 5. เสียงจัตวา ความถี่มูลฐานมีค่าลดลงในช่วงแรกและเพิ่มขึ้นต่อเนื่องในช่วงสุดท้าย

โดยข้อมูลที่ได้จากการควอนไตซ์ออกเป็น 3 ระดับนี้ จะถูกใช้เป็นข้อมูลฝึกสอนหรือข้อมูลทดสอบ ของการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ทั้ง 5 ระดับด้วย Hidden Markov Model [4] [14]

## 3. วิธีการดำเนินการวิจัย

### 3.1 การเก็บข้อมูล

ในการวิจัยนี้ จะใช้ผู้บอกภาษาที่เป็นคนไทย เป็นเพศชาย จำนวน 4 คน และเพศหญิงจำนวน 4 คน โดยภาษาที่ใช้จะเป็นภาษาไทยของคนภาคกลาง ในการเก็บข้อมูลของการวิจัยนี้ จะใช้คำพูดภาษาไทยให้ผู้ที่เป็นต้นแบบเสียงอ่านออกเสียง คำที่ได้กำหนดไว้จำนวน 10 ชุด แล้วทำการบันทึก ซึ่งข้อมูลเสียง 1 ตัวอย่าง จะถูกแทนด้วยข้อมูล 8 บิท โดยใช้ความถี่ในการ Sampling เท่ากับ 11.025 kHz รูปแบบไฟล์ข้อมูลเป็น .wav ซึ่งจะนำไปใช้ในการวิเคราะห์ต่อไป [2]

### 3.2 การวิเคราะห์หาคาบพิทช์และเตรียมแบบจำลอง

วิธีการทำแบบจำลองใช้โปรแกรม MATLAB ในการพัฒนา โดยเริ่มจากการเก็บเสียงจากกลุ่มตัวอย่างมาทั้งหมด 8 คน เป็นชาย 4 คน และหญิง 4 คน นำข้อมูลที่ได้มาแปลงเป็นไฟล์ .wav ขนาด มีความละเอียด 8 บิต Sampling rate อยู่ที่ 11.025 KHz มี Audio Channels เป็นแบบ mono เสียงจะถูกนำมาผ่านขั้นตอนการเตรียมข้อมูลทำแบบจำลอง (Pre-Process)

### 3.3 ขั้นตอนการทำแบบจำลอง

หลังจากได้ Sequence ของเสียงทั้ง 400 เสียงมาแล้วจะนำค่าที่ได้ไปทำการเรียนรู้เพื่อทำแบบจำลองโดยใช้โปรแกรม MATLAB ในการทำขั้นตอนนี้ โดยมีฟังก์ชันที่ใช้คือ hmmtrain โดยเป็นฟังก์ชันสำหรับการสร้างแบบจำลองโดยจะทำการวิเคราะห์ค่าจาก Transition probability และ Emission Probability ด้วยวิธีการ Hidden Makov Model โดยการใช้ Baum-Welch algorithm มีส่วนที่เป็น input ได้แก่ seq คือ เสียงที่ถูกนำมาแปลงให้อยู่ในรูปแบบ 3 ระดับแล้ว TRGUESS คือ Matrix ของความน่าจะเป็นในการย้ายข้ามมีขนาดเป็น N x N และ EMITGUESS คือ Matrix ของความน่าจะเป็นในการมีอยู่ของค่านั้น ๆ ในตำแหน่งนั้น ๆ มีขนาดเป็น N x จำนวนรูปแบบของ seq โดยค่า output ที่ได้ออกมานั้นจะมีสองค่า ได้แก่ ESTTR คือ Matrix ของ Transition probability ที่ได้จากการคำนวณ และ ESTEMIT คือ Matrix ของ Emission probability ที่ได้จากการคำนวณ

ค่า Transition probability และ Emission probability เริ่มแรกที่ใช้จะเป็นตัวกำหนด State และการย้ายข้าม State ในงานวิจัยนี้ได้ใช้ State ทั้งหมด 7 States และมีการย้ายข้าม State ได้สูงสุด 2 States โดยเกิดจากการทดลองหลาย ๆ ครั้งแล้วพบว่าผลที่ได้จากจำนวนนี้มีความน่าพอใจในระดับหนึ่ง ค่า Transition probability และ Emission probability เริ่มแรกที่ใช้จะได้จากการสุ่ม โดยจะนำ Sequence ของเสียงที่ได้ทำการเก็บมาของแต่ละเสียงวรรณยุกต์เข้าไปทำการเรียนรู้ในการสร้างแบบจำลอง โดยจะมีการใช้คำสั่งเพิ่มเติมของฟังก์ชัน hmmtrain คือ 'Pseudoemissions',PSEUDOE และ 'Pseudotransitions',PSEUDOTR โดยทั้งสองคำสั่งนี้จะเป็นการเรียกค่า Transition Probability และ Emission Probability เพื่อหลีกเลี่ยงปัญหาเกี่ยวกับ zero probability ที่อาจเกิดขึ้นได้ในขั้นตอนการทำ viterbi หรือการเทียบเสียงกับแบบจำลอง ค่าที่ได้ใช้เป็น Pseudotransition และ Pseudomission ในการทดลองนี้คือค่า Transition probability และ ค่า Emission

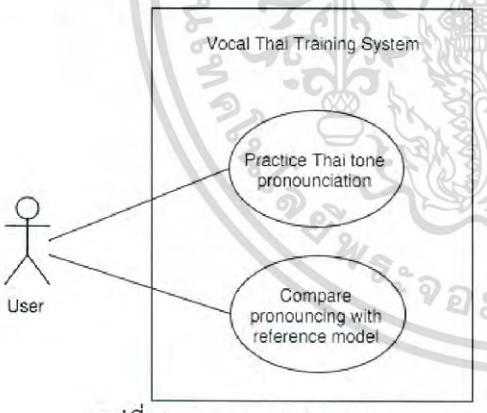
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Probability เริ่มแรกนั่นเอง เสียงที่ใส่เข้าไปเพื่อใช้ในการสร้างแบบจำลองของแต่ละวรรณยุกต์จะมีขนาด 80 เสียง โดยผลที่ได้จากการเข้าฟังก์ชัน hmmtrain จะเป็นแบบจำลองเสียงที่ได้ โดยมี 2 Matrix ต่อแบบจำลองเสียงวรรณยุกต์เสียง 1 เสียง ในงานวิจัยนี้ Transition Probability ที่ได้มีขนาด 7x7 และ Emission Probability มีขนาด 7x3

**3.4 การพัฒนาส่วนติดต่อกับผู้ใช้**

ในส่วนติดต่อกับผู้ใช้ได้พัฒนาขึ้นด้วยภาษาซี เขียนบนโปรแกรม Xcode มีซอฟต์แวร์ช่วยคือ GTK+ เป็น Tools ในการสร้าง GUI (Graphical User Interface) เพื่อติดต่อกับผู้ใช้ ทำให้สามารถใช้งานได้ง่ายขึ้น โดยภาพที่ใช้ในการทำได้ทำการสร้างขึ้นจากโปรแกรม Adobe Photoshop CC ในด้านของเสียงได้ใช้ซอฟต์แวร์ PortAudio เข้ามาพัฒนาเกี่ยวกับ I/O Audio ซอฟต์แวร์นี้จะรวบรวม Library เกี่ยวกับเครื่องมือทางด้านเสียง กล่าวโดยสรุปอย่างง่ายคือ สร้างภาพต่าง ๆ ด้วยโปรแกรม Adobe Photoshop CC แล้วนำภาพไปทำการเขียนเป็นแอปพลิเคชันด้วยโปรแกรม Xcode ร่วมกับ GTK+ และใช้ PortAudio ในด้านการรับเสียงจากผู้ใช้

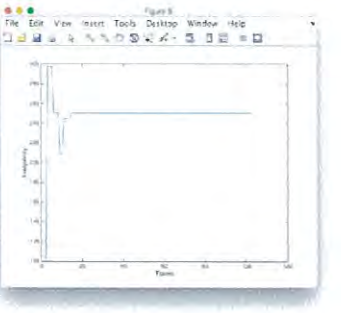
**3.5 Use case Diagram**



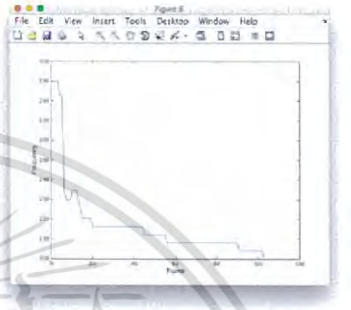

รูปที่ 4. Use case diagram

## 4. ผลการทดลอง

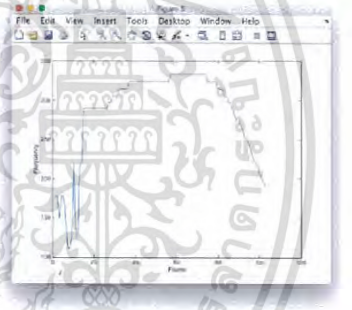
4.1 ผลการทดลองจากการสร้างกราฟเส้นเสียง



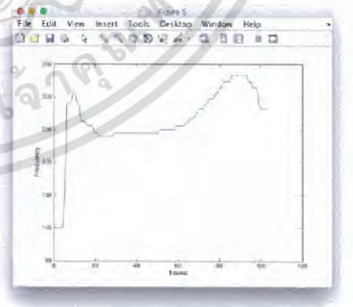
รูปที่ 5. กราฟวรรณยุกต์เสียงสามัญ



รูปที่ 6. กราฟวรรณยุกต์เสียงเอก

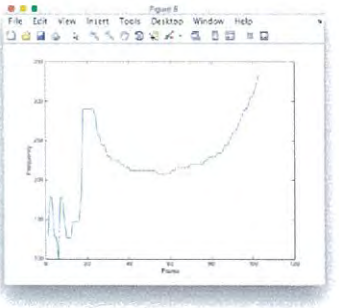


รูปที่ 7. กราฟวรรณยุกต์เสียงโท



รูปที่ 8. กราฟวรรณยุกต์เสียงตรี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



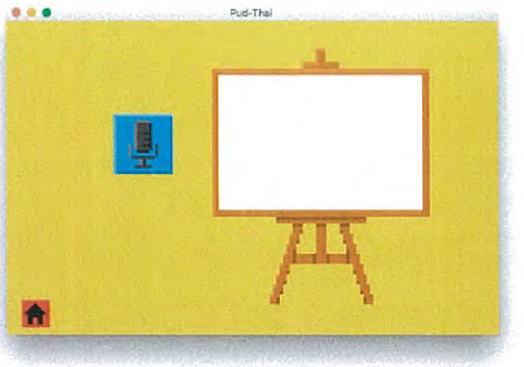
รูปที่ 9. กราฟวรรณยุกต์เสียงจัตวา

4.2 การออกแบบหน้าตาแอปพลิเคชัน

รูปที่ 10. หน้าหลักของโปรแกรม

รูปที่ 11. ผลการทดสอบในกรณีที่ออกเสียงถูก

รูปที่ 12. ผลการทดสอบในกรณีที่ออกเสียงผิด




รูปที่ 13. หน้า Tone Analysis

รูปที่ 14. ผลลัพธ์หน้า Tone Analysis

## 3. สรุปผลการทดลอง

ทุกวันนี้มีชาวต่างชาติจำนวนมากที่สนใจภาษาไทย และเรียนภาษาไทยมากขึ้น ซึ่งชาวต่างชาติบางคนสามารถออกเสียงภาษาไทยได้ อ่านภาษาไทยได้ แต่เสียงวรรณยุกต์อาจจะยังผิดอยู่ ในภาษาไทยนั้น วรรณยุกต์เป็นสิ่งที่สำคัญมาก เพราะวรรณยุกต์จะทำให้คำมีความหมายเปลี่ยนไปได้ และจะทำให้การสื่อสารผิดพลาดได้ คณะผู้จัดทำจึงได้คิดแอปพลิเคชันขึ้นมา เพื่อช่วยในการแก้ปัญหาเหล่านี้ การพัฒนาแอปพลิเคชัน ระบบฝึกฝนการผันวรรณยุกต์ภาษาไทยโดยเทคนิคการรู้จำเสียง ผู้พัฒนาได้ศึกษาขั้นตอนและวิธีการสร้างแบบจำลองจากทฤษฎี Hidden Markov Model เพื่อนำมาใช้ในการทำเป็นแบบจำลองอ้างอิงกับเสียงวรรณยุกต์ที่ใช้ในการทดสอบ ภายในแอปพลิเคชั่น ซึ่งภายในแอปพลิเคชั่นได้มีฟังก์ชันการทำงานหลัก ๆ ของแอปพลิเคชัน 2 ฟังก์ชัน คือ การฝึกออกเสียงวรรณยุกต์จากแบบทดสอบที่มีให้ภายในแอปพลิเคชันโดยจะมีเสียงตัวอย่างให้ฟังก่อนออกเสียง เพื่อทำการทดสอบ และอีก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฟังก์ชันคือการออกเสียงวรรณยุกต์ เพื่อดูว่าเป็นเสียงวรรณยุกต์ใด ซึ่งฟังก์ชันหลักนี้ จะสามารถเป็นส่วนที่ช่วยให้ผู้ใช้สามารถฝึกฝนการออกเสียงวรรณยุกต์ไทยได้ ให้ชัดเจนมากขึ้นได้ และเข้าใจถึงความแตกต่างในการออกเสียงของแต่ละเสียงวรรณยุกต์ในภาษาไทยได้ เพื่อที่จะช่วยให้ผู้ใช้ที่สนใจสามารถออกเสียงวรรณยุกต์ไทยได้อย่างมีความมั่นใจและมีความชัดเจนมากขึ้น และจะช่วยให้ภาษาไทยมีการเผยแพร่และเป็นที่รู้จักมากขึ้น ผู้ใช้สามารถใช้โปรแกรมได้หลังจากที่มีเวลาว่าง ๆ จากกิจกรรมต่าง ๆ ในแต่ละวันได้ ระบบจะช่วยให้ผู้ใช้สามารถฝึกฝนการออกเสียงวรรณยุกต์ไทยได้ง่าย และสะดวกมากขึ้น

## เอกสารอ้างอิง


[1] จิตรลดา จารุมิศร์. "การออกแบบ แบบจำลองในการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทยโดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling." วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542

[2] ภักดี จันทร์เกษ. "การรู้จำเสียงตัวเลขภาษาไทย โดยการวิเคราะห์เสียงสระและเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤตและควอนไตซ์พิทช์." วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2550

[3] คำ ขันทะวีวอน. "แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling." วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2546

[4] กฤษกร สุนทรมนทกานติ. "การหาแบบจำลองที่เหมาะสมการรู้จำเสียงวรรณยุกต์ภาษาไทย โดยใช้วิธีการออโตคอร์รีเลชั่นแบบเซ็นเตอร์คลิปปิ้ง สำหรับการหาค่าพิทช์ในสภาวะที่มีสัญญาณรบกวนเกาส์เซียนขาว." วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2548

[5] ร้อยกรอง สุขเกิด, นางสาวทิพปภา พิสิษฐ์กุล, กมลพรรณ ทวิชศรี. "โปรแกรมฝึกออกเสียงวรรณยุกต์ภาษาไทย." โรงเรียนมหิดลวิทยานุสรณ์. 2550

[6] ธันวา ศรีประโมง. "การวิเคราะห์เสียงพูดภาษาไทยในแกนความถี่ฮาร์โมนิก." วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2537

[12] LAWRENCE R. RABINER. "On the Use of Autocorrelation Analysis for Pitch Detection" IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL.ASSP-25, No. 1, FEBRUARY 1977.

[13] A.Rosenfeld, A.C.Kak. "Digital picture Processing." Orlando Florida : Academic Press Inc. 1982.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้