แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling

TONE RECOGNITION MODEL FOR LAO LANGUAGE USING PITCH QUANTIZATION AND HIDDEN MARKOV MODELING TECHNIQUES

ค่ำ ขันทะวีวอน

KHAM KHANTHAVIVONE

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมโทรคมนาคม

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2546

ISBN 974-324-793-9

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling

TONE RECOGNITION MODEL FOR LAO LANGUAGE USING PITCH QUANTIZATION AND HIDDEN MARKOV MODELING TECHNIQUES

T048899

คำ ขันทะวีวอน

KHAM KHANTHAVIVONE

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมโทรคมนาคม

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เลขหมู่.....
เลขทะเบียน 48899
วัน, เดือน, ปี 12 มิ.ย. 2547

พ.ศ.2546

ISBN 974-324-793-9

.b.....
.i.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TONE RECOGNITION MODEL FOR LAO LANGUAGE USING PITCH QUANTIZATION AND HIDDEN MARKOV MODELING TECHNIQUES

KHAM KHANKHAVIVONE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN TELECOMMUNICATION ENGINEERING

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2003

ISBN 974-324-793-9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2003

SCHOOL OF GRADUATE SYUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	**แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิคการควอนไตซ์พิทช์ และ Hidden Markov Modeling**
นักศึกษา	**นายคำ ขันทะวีวอน**
รหัสประจำตัว	44611217
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมโทรคมนาคม
พ.ศ.	2546
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.ไกรสิน ส่งวัฒนา

บทคัดย่อ

วิทยานิพนธ์นี้ได้นำเสนอวิธีการการสร้างแบบจำลอง ระดับเสียงวรรณยุกต์สำหรับภาษาลาว (เฉพาะเสียงพูดคนลาวที่เป็น เสียงเวียงจันทน์) ซึ่งเสียงพูดในภาษาลาวมี 5 ระดับด้วยกัน เสียงจากการบันทึกแต่ละคำได้ถูกแบ่งให้เป็นส่วนย่อยๆแล้วนำ แต่ละส่วนมาคำนวณหาคาบเวลาพิทช์โดยใช้วิธีออโตโครีเลชั่น ตามด้วยทำการแปลงคาบเวลาพิทช์มาเป็นค่าความถี่มูลฐาน ค่าความถี่มูลฐานที่ได้ถูกนำเข้ากระบวนการควอนไตซ์ ตามการเบี่ยงเบนออกเป็น 3 ระดับตามทิศทางการเพิ่มขึ้น คงที่หรือลดลง ของค่าความถี่มูลฐาน ซึ่งจากการจัดแบ่งข้อมูลนี้จะทำให้ช่วงความถี่เสียงที่แตกต่างกันของชายและหญิงไม่มีผลต่อการสร้างแบบจำลองการรู้จำ ทำให้แบบจำลองที่สร้างขึ้นนี้สามารถใช้ร่วมกันได้ทั้งชายและหญิง โดยข้อมูลจากกระบวนการควอนไตซ์จะถูกนำเข้าฝึกสอนสำหรับการสร้างแบบจำลองการรู้จำของหน่วยเสียงวรรณยุกต์ทั้ง 5 ระดับด้วยวิธี Hidden Markov Model (HMM)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Tone Recognition Model For Lao Language Using Pitch Quantization and Hidden Markov Modeling Techniques
Student	Mr. Kham Khanthavivone
Student ID.	44611217
Degree	Master of Engineering
Programme	Telecommunication Engineering
Year	2003
Thesis Advisor	Assoc. Prof. Dr. Kraisin Songwattana

ABSTRACT

This Thesis presents the designing of tones level recognition modeling for spoken Laotian Language (local Vientiane pronunciation). The spoken word is divided into frames. Then the auto-correlation method using center clipping is applied to each frame of speech to determine the pitch period and it fundamental frequency. The sequence of fundamental frequency is improved to make the connecting of data more smoother by using median filtered. The observation sequence of pitch levels are preprocessed to find the pitch differences and the sequence of pitch differences are then grouped into three quantized levels, which indicates a possibility of gender independent tone recognition. The quantization of three levels has the properties of frequency independent. The resultant sequence is used as bases for training a Hidden Markov Model and recognition of 5 tones.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ผู้จัดทำใคร่ขอขอบพระคุณ รศ.ดร. ไกรสิน ส่งวัฒนา อาจารย์ผู้ควบคุมวิทยานิพนธ์ ที่ได้กรุณาเสียสละเวลา ให้คำปรึกษา และข้อแนะนำที่เป็นประโยชน์มา ตลอดถึงห้องทดลองและการทำงาน ทำให้การทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี

ขอขอบพระคุณอาจารย์ในภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกท่าน ที่มีส่วนร่วมสนับสนุน ช่วยเหลือและแนะนำในทุกๆด้านตลอดมา

ขอขอบคุณ ภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ให้โอกาส สถานที่ศึกษาเล่าเรียนตลอดมา

ขอขอบอคุณมหาวิทยาลัยแห่งชาติลาวที่ได้ให้การสนับสนุนและโอกาสตลอดมา

ขอขอบคุณองค์การ AUN / SEED-Net ที่เป็นผู้ให้การสนับสนุนทุนการศึกษานี้

ขอขอบคุณ พี่ๆ และ เพื่อนๆ ห้อง T-310 ที่ช่วยให้คำปรึกษาและกำลังใจมาโดยตลอด

สุดท้าย วิทยานิพนธ์ฉบับนี้คงไม่สำเร็จลงได้ ถ้าขาดเจ้าของเสียงทุกท่านที่ให้ความร่วมมือในการเก็บข้อมูลเพื่อทำการวิจัยเป็นอย่างดี ผู้วิจัยขอขอบคุณไว้ ณ ที่นี่ด้วย

คำ ขันทะวีวอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VIII

สารบัญภาพ IX

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ในการทำวิทยานิพนธ์ 2

1.3 ข้อกำหนดในการทำวิทยานิพนธ์ 2

1.4 โครงประกอบของวิทยานิพนธ์ 3

บทที่ 2 ภาษาและความหมายสำคัญของภาษาลาว 4

2.1 การกำเนิดและการขยายตัวของภาษา 4

2.1.1 ภาษาคืออะไร 4

2.1.2 ภาษามีการกำเนิดและขยายตัวได้อย่างใด 4

2.2 ความหมายสำคัญของภาษา 5

2.2.1 ภาษาคือสื่อสัมพันธ์ระที่หนึ่งของมนุษย์ 5

2.2.2 ความสัมพันธ์ของภาษาพูดและภาษาเขียน 5

2.3 การกำเนิดและการขยายตัวของภาษาลาว 5

บทที่ 3 หลักของภาษาลาว 7

3.1 เสียงในภาษาคืออะไร 7

3.1.1 ความก้อง หรือ ไม่ก้องของเสียง 7

3.1.2 ระดับเสียงสูง-ต่ำ (Pitch) 7

3.1.3 ความดัง (Loudness) 7

3.1.4 การลงน้ำหนัก (Stress) 8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.1.5 ช่วงต่อของเสียง (Juncture) 8

3.2 หน่วยเสียงสำคัญในภาษาลาว 8

3.2.1 หน่วยเสียงสระ 8

3.2.2 หน่วยเสียงพยัญชนะ 8

3.2.3 หน่วยเสียงวรรณยุกต์ 8

3.3 ตัวอักษรลาว 9

3.3.1 พยัญชนะ 9

3.3.1.1 พยัญชนะเดิม 9

3.3.1.2 พยัญชนะผสม 10

3.3.1.3 พยัญชนะควบ 10

3.3.1.4 อักษรหลักและตัวสะกด 10

3.3.2 สระ 11

3.3.3 วรรณยุกต์ 11

3.4 การประสมอักษร 11

3.4.1 แม่กะกา 11

3.4.2 แม่สะกดทั้ง 8 12

3.4.2.1 แม่สะกดทั้ง 8 แบบธรรมดา 12

3.4.2.2 แม่สะกดทั้ง 8 ใส่วรรณยุกต์ 13

3.5 วิธีผันเสียง 14

บทที่ 4 การหาค่าพิทช์ของสัญญาณเสียงพูด 15

4.1 กล่าวนำ 15

4.1.1 ขั้นตอนการเรียนรู้ 15

4.1.2 ขั้นตอนการจดจำ 16

4.2 การหาค่าพิทช์ของสัญญาณเสียงพูด 16

4.2.1 การวิเคราะห์ในโดเมนเวลา 16

4.2.2 ทฤษฎีการประมาณค่าพิทช์โดยใช้ออโตคอร์รีเลชั่นฟังก์ชั่น 16

4.2.2.1 การจัดแบ่งการวิเคราะห์สัญญาณออกเป็นช่วงสั้นๆ 16

4.2.2.2 การกำจัดผลของโครงสร้างฟอร์แมนท์ด้วยวิธีเซนเตอร์คลิปปิ้ง .18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2.2.3 การทำนอร์แมลไลซ์ออโตคอร์รีเลชั่น 20

4.3 การหาค่าความถี่มูลฐานของสัญญาณเสียงพูด..... 22

4.4 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง..... 24

4.4.1 การปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง 24

4.4.2 การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน 26

4.5 การสร้างแบบจำลองการรู้จำด้วยวิธี Hidden Markov Model 28

4.5.1 ส่วนประกอบของแบบจำลองมาร์คอฟ 28

4.5.2 คุณสมบัติการย้ายสเตทของ HMM 29

4.5.3 ปัญหาพื้นฐานของแบบจำลอง มาร์คอฟ 29

4.5.4 การคำนวณเพื่อแก้ปัญหาของ HMM 30

4.5.4.1 การแก้ปัญหาที่ 1 30

4.5.4.1.1 กระบวนการไปข้างหน้า(Forward Procedure) 31

4.5.4.1.2 กระบวนการย้อนกลับ(Backward Procedure) 32

4.5.4.2 การแก้ปัญหาที่ 2 33

4.5.4.3 การแก้ปัญหาที่ 3 34

4.5.5 การปรับปรุงค่าพารามิเตอร์ของ HMM 37

4.5.5.1 การสเกลลิง (Scaling) 37

4.5.1.2 ลำดับของค่าปรากฏหลายลำดับ 41

4.5.6 ขบวนการสร้างแบบจำลองอ้างอิง 42

4.6 สรุป 44

บทที่ 5 การทดลอง และผลการทดลอง 45

5.1 การกำหนดขอบเขตของพยางค์ หรือคำ 45

5.2 ผลการวิเคราะห์แบบจำลองเสียงวรรณยุกต์ 46

5.2.1 ผลการหาค่าความถี่มูลฐานและการควอนไตซ์พิทช์ 46

5.2.1.1 ผลการหาค่าความถี่มูลฐาน 46

5.2.1.2 การกรองค่ากลาง 48

5.2.2 การควอนไตซ์พิทช์ 49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

5.2.3 การทดสอบหารูปแบบ HMM ที่เหมาะสม 51

5.2.4 การสร้างแบบจำลองอ้างอิงเพื่อใช้ในการรู้จำระดับเสียงวรรณยุกต์ 54

5.3 การทดสอบแบบจำลองอ้างอิง 61

5.3.1 การทดสอบแบบจำลองอ้างอิงโดยใช้กลุ่มคำต้นแบบ 61

5.3.2 การทดสอบแบบจำลองอ้างอิงโดยใช้กลุ่มคำใหม่ 62

บทที่ 6 สรุปผล และข้อเสนอแนะ 65

6.1 การทดลอง 65

6.1.1 ขบวนการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ 65

6.1.2 การสร้างแบบจำลองอ้างอิง 66

6.1.3 ผลการทดสอบแบบจำลองอ้างอิงที่สร้างขึ้น 66

6.2 ข้อสังเกต ปัญหาที่พบในการทดลอง และข้อแสนอแนะ 66

เอกสารอ้างอิง 68

ภาคผนวก 70

ประวัติผู้เขียน 77

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

3.1 แสดงการสะกดอักษรของแม่สะกดในภาษาลาวแบบธรรมดา 12

3.2 แสดงแม่สะกดในภาษาลาวแบบใส่วรรณยุกต์กับอักษรกลาง 13

3.3 แสดงการสะกดอักษรในภาษาลาวแบบใส่วรรณยุกต์กับอักษรสูงและอักษรต่ำ 13

3.4 แสดงตัวอย่างการผันเสียงอักษรไปตามวรรณยุกต์..... 14

5.1 กลุ่มคำที่ 1 ใช้ในการทดสอบหาแบบจำลอง HMM ที่เหมาะสม 53

5.2 กลุ่มคำที่ 2 ใช้ในการสร้างแบบจำลองอ้างอิง 54

5.3 ผลการทดสอบแบบจำลองอ้างอิง โดยใช้กลุ่มคำและเสียงต้นแบบ จำนวน 500 เสียง 62

5.4 กลุ่มคำที่ 3 ใช้ในการทดสอบแบบจำลองอ้างอิง 62

5.5 ผลการรู้จำระดับเสียงวรรณยุกต์ จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงต้นแบบ 63

5.6 ผลการรู้จำระดับเสียงวรรณยุกต์ จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงกลุ่มใหม่ 63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

ภาพที่ หน้า

1.1 ส่วนประกอบของระบบการรู้จำภาษาลาวโดยวิธีแยกจำลักษณะตามหน่วยเสียง 2

4.1 ขั้นตอนในการวิเคราะห์ 15

4.2 แสดงการหาค่าพิทช์จากการคำนวณออโตโครีเลชั่นฟังก์ชั่นของสัญญาณเสียงพูด 18

4.3 แสดงรูปทรงในแต่ละคาบของสัญญาณเสียงพูด 19

4.4 แสดงวิธีเซนเตอร์คลิปปิ้งของ Dobnoeski 19

4.5 ตัวอย่างแสดงการคลิปสัญญาณเสียงพูด 19

4.6 แสดงการออโตคอร์รีเลชั่นของสัญญาณเสียง 21

4.7 โฟลว์ชาร์ตของขบวนการหาค่าพิทช์ เทียบกับผลที่ได้รับ 23

4.8 แสดงการเปียนแปลงของความถี่มูลฐานในแต่ละเวลาของวรรณยุกต์ 24

4.9 การจัดแบ่งความถี่มูลฐานออกเป็นชุดข้อมูล 25

4.10 แสดงตัวอย่างความถี่มูลฐานที่ผ่านวิธีการกรองค่ากลาง 25

4.11 แสดงการจัดแบ่งค่าความถี่มูลฐานออกเป็น 3 ระดับตามการเปลี่ยนแปลงต่อเวลาเพิ่มขึ้น 27

4.12 แบบจำลองLeft-Right Model ของ HMM 29

4.13 กระบวนการไปข้างหน้า 32

4.14 กระบวนการย้อนกลับ 33

4.15 ลำดับการคำนวณเกิดค่าปรากฏร่วมซึ่งอยู่ที่สเตท i ที่เวลา t และอยู่ที่ สเตท j ในเวลา t+1 ... 36

4.16 โฟลว์ชาร์ต การคำนวณหาค่าพารามิเตอร์ของแบบจำลองอ้างอิง 43

4.17 บล็อคไดอะแกรม ของการรู้จำระดับเสียงวรรณยุกต์ด้วยแบบจำลองมาร์คอฟ 44

5.1 ตัวอย่างสัญญาณเสียงพูดของคำว่า "ซอง" จากผู้ออกเสียงเพศชาย 45

5.2 แสดงตัวอย่างการการเปลี่ยนแปลงค่าความถี่มูลฐานของทางเดินเสียงวรรณยุกต์ 47

5.3 แสดงตัวอย่างการกองค่ากลางความถี่มูลฐานของทางเดินเสียงวรรณยุกต์ 49

5.4 แสดงตัวอย่างการการเปลี่ยนแปลงค่าของพิทช์จากการทำควอนไตซ์ 51

5.5 Left-Right Model N สเตท 52

5.6 เปอร์เซนต์ความถูกต้องเมื่อมีการเปลี่ยนแปลงสเตท และการย้ายข้ามสเตทของ HMM 53

5.7 แบบจำลอง HMM ที่เหมาะสมกับอัลกอริธึมที่พัฒนาขึ้น 13 สเตท 54

5.8 แสดงผลของโปรแกรมในการทดสอบการรู้จำ 61

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้คอมพิวเตอร์ได้เข้ามามีบทบาทต่อชีวิตมนุษย์เป็นอย่างมาก โดยได้ถูกนำใช้ร่วมกับการทำงานชีวิตประจำวันในหลายๆอย่าง ซึ่งโดยปกติการทำงานร่วมกับคอมพิวเตอร์จะทำได้โดยการสื่อสารสั่งงานผ่านทางคีย์บอร์ดและเมาส์ แต่ก็ยังไม่สะดวกพอเท่ากับการสื่อสารด้วยเสียงพูดที่ใช้กับการสั่งงานทั่วไปและจะสะดวกมากถ้าหากคอมพิวเตอร์สามารถเข้าใจภาษาพูดของมนุษย์ได้ ซึ่งจะทำให้สามารถสั่งงานด้วยเสียงพูดทำให้การติดต่อสื่อสารระหว่างเครื่องคอมพิวเตอร์กับมนุษย์เป็นเรื่องง่ายและสะดวกมากยิ่งขึ้น

การรู้จำภาษาสื่อสารของมนุษย์ในเครื่องคอมพิวเตอร์ โดยส่วนมากแล้วจะมีความต่าง กัน อันเนื่องมาจากภาษาของแต่ละเขต แต่ละชาติมีความต่างกันทั้งภาษาเขียนและภาษาพูด ภาษาลาวก็เป็นภาษาหนึ่งที่ในจำนวนภาษาต่างๆของโลก ถูกกำหนดขึ้นโดยคนชาติลาวมาเนิ่นนานแล้ว อันทำให้คนลาวมีการสื่อสารทางด้านภาษาระหว่างคนในชาติมีลักษณะเฉพาะ โดยคำพูดที่ใช้ในการสื่อสารกันก็มีสำเนียงเป็นลักษณะเฉพาะตามแต่ละเขต แต่ละภาคในประเทศลาว ซึ่งทุกภาคก็ล้วนแล้วแต่มีหลัก การของการใช้ภาษาเป็นอย่างเดียวกัน เสียงพูดของคนลาวโดยทั่วไปจะออกเสียงไปตามวรรณยุกต์ (Tonal language) เพื่อสื่อสารความหมายให้ได้ครบถ้วนและถูกต้อง

ดังนั้นเพื่อให้การรู้จำเสียงพูดของภาษาลาวให้ได้อย่างถูกต้อง จึ่งจำเป็นต้องมีการเรียนรู้จำการออกเสียงของวรรณยุกต์ด้วย ซึ่งเหมือนกันกับเสียงภาษาไทย[1] เพราะการออกเสียงวรรณยุกต์จะเป็นตัวบอกความหมายของคำพูดในการสื่อสารให้ได้อย่างครบถ้วน

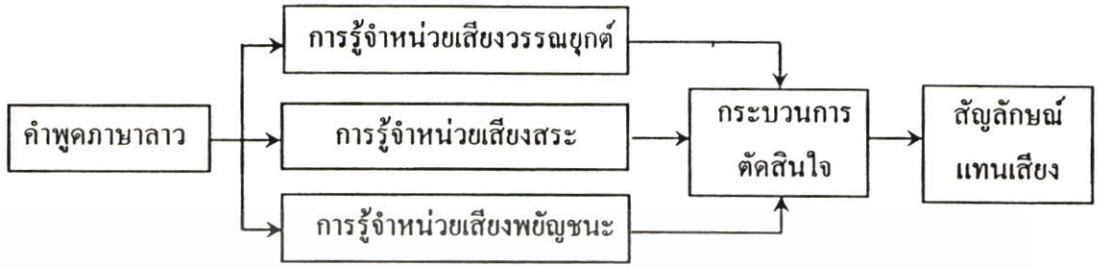
เนื่องจากภาษาพูดของมนุษย์มีความซับซ้อนมาก โดยเฉพาะภาษาลาวที่มีการพูด การออกเสียงจะแตกต่างไปตามผู้พูดแต่ละบุคคลและประโยคในการพูด ดังนั้นเพื่อให้การรู้จำเสียงพูดได้ดีจึ่งมีการแบ่งวิธีตามกระบวนการดังนี้

1.1.1 พิจารณาทั้งหน่วยภาษาที่เปล่งเสียงออกมาทั้งชุด มีทั้งระบบการรู้จำคำเดี่ยว [2]-[3] (Isolated word Recognition) และระบบรู้จำคำพูดต่อเนื่อง (Continuous word Recognition) ซึ่งเป็นวิธีจดจำเป็นกลุ่มคำ ทำให้หลีกเลี่ยงผลกระทบอันเนื่องจากฐานของเสียงภายในคำพูดหรือกลุ่มคำ แต่มีความสามารถรู้จำได้ในจำนวนคำที่จำกัด เนื่องจากแบบจำลองอ้างอิงต้องใช้เนื้อที่จำนวนมากในการจัดเก็บกลุ่มคำของแบบจำลองอ้างอิง ทำให้ต้องใช้เวลามากในการคำนวณเพื่อเปรียบเทียบกับจำนวนของแบบจำลองอ้างอิงที่มีอยู่

1.1.2 พิจารณาโดยการแยกแยะรายละเอียดของหน่วยเสียง (Phonetic Recognition)[4]-[5] วิธีนี้จะพิจารณาลักษณะของหน่วยเสียงที่มีขนาดเล็กลงไป เช่น หน่วยเสียงพยัญชนะ หน่วยเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สระและหน่วยเสียงวรรณยุกต์ ดังแสดงในรูปที่ 1.1 โดยจะใช้หน่วยเสียงย่อยเหล่านี้เป็นหลักในการรู้จำเสียงพูด ซึ่งวิธีนี้เหมาะสำหรับการพัฒนาไปสู่ระบบการรู้จำคำจำนวนมาก



รูปที่ 1.1 ส่วนประกอบของระบบการรู้จำภาษาลาวโดยวิธีแยกจำลักษณะตามหน่วยเสียง

ด้วยเหตุนี้ เพื่อพัฒนาไปสู่ระบบการรู้จำเสียงพูดภาษาลาว ซึ่งมีคำจำนวนมาก วิทยานิพนธ์นี้จึงทำการศึกษาพัฒนาระบบการรู้จำแบบแยกแยะหน่วยเสียง โดยมุ่งหวังที่จะสร้างระบบการรู้จำหน่วยเสียงวรรณยุกต์ ซึ่งเป็นภาคส่วนสำคัญต่อการออกเสียงพูดให้มีความหมายต่างกัน ดังนั้นจะเห็นว่าหน่วยเสียงวรรณยุกต์จึงมีหน้าที่ๆจะทำให้เกิดคำใช้งานในภาษามากขึ้น

1.2 วัตถุประสงค์ในการทำวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้ เป็นการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ของภาษาลาวด้วยคำพยางค์เดียว (Monosyllabic) โดยใช้การหาค่าคาบเวลาพิทช์ในโดเมนของเวลาแต่ระคำพูดแล้วนำตัวอย่างข้อมูลเสียงเหล่านั้นมาสร้างแบบจำลองการรู้จำด้วยฮิดเดนมาร์คอฟโมเดล โดยมีวัตถุประสงค์ดังนี้

1.2.1 ศึกษาวิธีการที่จะทำให้เครื่องคอมพิวเตอร์สามารถรู้จำเสียงพูดได้

1.2.2 เพื่อศึกษารูปแบบการเปลี่ยนแปลงเสียงในแต่ละ วรรณยุกต์ของพยางค์

1.2.3 เพื่อศึกษาและออกแบบระบบการรู้จำหน่วยเสียงวรรณยุกต์ในภาษาลาว โดยมุ่งเน้นให้แบบจำลองการรู้จำที่สร้างขึ้นนี้สามารถรู้จำเสียงพูดกับบุคคลต่างๆ โดยแบบจำลองที่พัฒนาขึ้นนี้สามารถใช้ร่วมกันได้ทั้งผู้ออกเสียงชายและหญิง

1.2.4 เพื่อเป็นองค์ประกอบหนึ่งในการพัฒนาไปสู่ระบบการรู้จำคำพูดภาษาลาวทั้งภาษา

1.3 ข้อกำหนดในการทำวิทยานิพนธ์

1.3.1 งานวิจัยนี้มุ่งศึกษาพัฒนาอัลกอริธึมบนเครื่องคอมพิวเตอร์ส่วนบุคคล โดยได้นำใช้โปรแกรมภาษา Visual C++ 6.0 โดยมีอุปกรณ์เพิ่มเติมได้แก่ ไมโครโฟนและลำโพง

1.3.2 ข้อมูลเสียงที่ใช้ในวิทยานิพนธ์ฉบับนี้ เป็นเสียงสำเนียงภาษาเวียงจันทน์ภาคกลาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3.3 ในการทดลองได้แบ่งกลุ่มของคำทดสอบออกเป็น 3 กลุ่มจากผู้ออกเสียง ได้แก่

กลุ่มที่ 1 ประกอบด้วยเสียงพูดที่ผันได้ระดับวรรณยุกต์ทั้ง 5 เสียง 50 คำ จำนวน 300 เสียงเพื่อใช้ในการทดสอบหาแบบจำลองของ HMM (Hidden Markov Model) ที่เหมาะสมกับอัลกอริธึมที่พัฒนาขึ้น

กลุ่มที่ 2 ประกอบด้วยคำพยางค์เดียวที่ผันเสียงครบได้ระดับวรรณยุกต์ครบทั้ง 5 ระดับ จำนวนทั้งสิ้น 100 คำ จำนวน 1000 เพื่อใช้เป็นคำต้นแบบในการสร้างแบบจำลองอ้างอิงการรู้จำระดับเสียงวรรณยุกต์

กลุ่มที่ 3 ประกอบด้วยคำพยางค์เดียวที่มีระดับเสียงวรรณยุกต์ครบทั้ง 5 ระดับที่แตก ต่างจากกลุ่มคำที่ 2 จำนวนทั้งสิ้น 50 คำ เพื่อใช้เป็นคำทดสอบแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ที่สร้างขึ้น

1.4 โครงประกอบของวิทยานิพนธ์

แบ่งออกเป็น 6 บทดังนี้

บทที่ 1 บทนำ

บทที่ 2 ที่มาและความหมายของภาษาโดยเฉพาะภาษาลาว ซึ่งได้กล่าวถึงการกำเนิดและการขยายตัวของภาษา

บทที่ 3 หลักของภาษาลาว ซึ่งได้กล่าวถึงหลักการนำใช้ภาษาลาวโดยรวมมีทั้งหลักในการเขียน และ การออกผันเสียงพูด

บทที่ 4 ทฤษฎีการวิเคราะห์และสร้างแบบจำลองเสียงวรรณยุกต์ในภาษาลาว

บทที่ 5 ขั้นตอนในการทดลอง

บทที่ 6 บทสรุปเกี่ยวกับการทดลองทั้งหมดที่ทำมา พร้อมทั้งข้อสังเกต ปัญหาที่พบในการทดลอง และข้อเสนอแนะสำหรับผู้ที่ทำการวิจัย และพัฒนาระบบการรู้จำเสียงพูดต่อไป

ภาคผนวก

ผลงานวิจัยที่ได้รับการตีพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ภาษาและความหมายสำคัญของภาษาลาว

2.1 การกำเนิดและการขยายตัวของภาษา

2.1.1 ภาษาคืออะไร ?

ภาษา คือการแสดงความในใจของตนออกมาให้ผู้อื่นได้รับรู้ อาจจะเป็นอาการ การให้สิ่งของหรือเป็นลายลักษณ์อักษรก็ได้ แต่อาการและสิ่งเหล่านั้นต้องเป็นสื่อสัมพันธ์ ซึ่งจะเป็นการช่วยให้คนอื่นรู้ความหมายและเข้าใจได้

ภาษา คือการแสดงความรู้สึกของตนเราออกให้คนอื่นได้เข้าใจได้ด้วยการพูดการวาดเขียนและเครื่องหมายสัญลักษณ์ต่างๆเพื่อเป็นสื่อสัมพันธ์ในการติดต่อกัน เมื่อมนุษย์เราแสดงคำพูดสื่อสารกับคนอื่นนั้นจะหมายถึงการแสดงความคิดออกด้วยการพูด ความหมายของคำพูดจะถูกแปลเป็นข้อมูลออกมา ซึ่งจะเป็นการสื่อถึงสิ่งของสถานที่และกาละเวลา

ภาษา คือรูปลักษณะต่างๆที่ใช้เป็นสื่อสารติดต่อให้เข้าใจซึ่งกันและกัน รวมมีภาษาใบ้ ภาษาตา ภาษาพูด ภาษาเขียนและเครื่องหมายต่างๆอันทำให้มนุษย์เราสามารถเข้าใจความหมายกันของการกระทำได้

2.1.2 ภาษามีการกำเนิดและขยายตัวได้อย่างใด ?

ภาษากำเนิดมาจากการเรียนเสียงของธรรมชาติ เสียงในธรรมชาติอาจจะเป็นเสียงสัตว์ร้องซึ่งมนุษย์เราจะเรียกเสียงสัตว์ตามที่ได้ยินอย่างเช่น เสียงร้องของ กบ เขียด กับแกะ วัว แพะ แมว และสัตว์ต่างๆ บางคำพูดมนุษย์เราก็เรียนตามเสียงธรรมชาติอย่างอื่น เช่น เสียงลมพัด น้ำตก ฟ้าร้อง เป็นต้น แต่เสียงที่มนุษย์นำมาใช้จะผิดต่างไปตามแต่ละภูมิประเทศ

ภาษาเกิดจากการอุทาน เสียงอุทานของมนุษย์เกิดจากการเปล่งเสียงออกโดยไม่รู้สึกตัวเมื่อมีความเจ็บปวด เสียใจ ตกใจหรือตื่นตันใจต่อสิ่งใดหนึ่งแล้วร้องขึ้นด้วยอารมณ์เองอย่างเช่น โอย! (ໂອຍ!) โอ! (ໂອ!) ปะโท! (ปะโທ!) อ้าว! (ອ້າວ!) ๆ

ภาษาเกิดจากเสียงเด็กน้อย เด็กน้อยที่เกิดขึ้นมาเมื่อถึงเวลาพูดจะต้องเปล่งเสียงออก ซึ่งมักจะร้องออกเสียงหาผู้ใกล้ชิดของตนเช่น แม่ พ่อ เป็นต้น เมื่อเด็กร้องออกเสียง แม่หรือพ่อ ก็หมายความว่าภาษาได้กำเนิดขึ้น โดยการกำหนดว่าเสียงนั้นมีความหมายที่เข้าใจกันได้

ภาษาที่เกิดขึ้นจากมนุษย์สร้างขึ้นมา เพื่อให้มีภาษามากขึ้นตามการขยายตัวของสังคมมนุษย์ จึงมีการสมมุติสิ่งต่างๆขึ้นมาตามรูปร่างลักษณะสีขนาด ๆลๆ ซึ่งทำให้เกิดภาษาที่ต่างกันออกไปตามความรู้สึกของมนุษย์ ดังตัวอย่างคำว่า ใหญ่ น้อย แดง ฟ้า ๆลๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 ความหมายสำคัญของภาษา

2.2.1 ภาษาคือสื่อสัมพันธ์ระดับที่หนึ่งของมนุษย์

ภาษาคือพาหนะสื่อสารพื้นฐานอย่างหนึ่งที่รับใช้ติดต่อในชีวิตของสังคม ภาษาหรือคำพูดเกิดขึ้นโดยอาใส่เหตุการณ์ต่างๆอย่างเช่น ท่าทาง สำเนียง อาการ เป็นต้น ภาษากำเนิดขึ้นด้วยความต้องการอันจำเป็นของมนุษย์ที่อยากแสดงความรู้สึกนึกคิดของตนให้คนอื่นได้รับรู้และเข้าใจนำ ซึ่งเริ่มจากใช้วาดชงการแสดงออกคือ ภาษาใบ้ ต่อมันได้มีการออกเสียงบอกโดยปะสานกับจังหวะท่าทางของอวัยวะต่างๆอย่างเช่น ยกมือ ตีคิ้ว หลับตา เป็นต้น ฉะนั้นจึงพูดได้ว่าถ้าไม่ มีภาษามนุษย์ก็ไม่อาจจะมีความเข้าใจต่อกันและกันได้

2.2.2 ความสัมพันธ์ของภาษาพูดและภาษาเขียน

ภาษาพูดและภาษาเขียนเป็นเครื่องมือแสดงความรุ่งเรืองในอดีต และเป็นหลักฐานที่สำคัญในการเรียนรู้ความเป็นมาของมนุษย์ ภาษาพูดกำเนิดก่อนภาษาเขียน โดยมีภาษาพูดเพื่อใช้ติดต่อสื่อสารความคิดและเพื่อหาความรู้ใหม่ๆ ภาษาเขียนเกิดขึ้นตามความต้องการในการบันทึกข้อความของภาษาพูดไว้เป็นหลักฐานอ้าวอิง

การที่ทั้งภาษาพูดและภาษาเขียนเป็นความสะดวกสำหรับช่วยสื่อความหมาย ความคิดความรู้สึกและความเข้าใจให้เกิดขึ้นในชีวิตประจำวันไม่ว่าจะเป็นการส่งข่าวสาร การศึกษาเล่าเรียนก็คือการประกอบอาชีพ ถ้าหากภาษาพูดและภาษาเขียนเป็นอย่างเดียวกันแล้วจะทำให้มีภาษากายเป็นเครื่องผูกพันความรู้สึกของมนุษย์ให้ดีกว่าสิ่งอื่นใด

2.3 การกำเนิดและการขยายตัวของภาษาลาว

ชาติลาวมีภาษาเป็นของตัวเองมาแต่ดั้งเดิมและใช้สืบต่อกันมาจนถึงทุกวันนี้ ภาษาลาวคือภาษาหนึ่งในภาษาต่างๆของโลก ที่กำเนิดขึ้นพร้อมกับการกำเนิดของชนชาติลาว โดยผ่านยุคสมัยต่างๆภาษาลาวได้มีการขยายตัวจากที่มีจำนวนถ้อยคำจำกัดอยู่ในวงแคบ แล้วก็มีการแผ่กว้างออกไป ภาษาลาวเดิมเป็นภาษาลาวคำเดี่ยวหรือคำที่มีพะยางเดียวอย่างเช่น พ่อ (พ่) แม่ (แม่) ดิน (ดิน) น้ำ (มาก) ลม (ลิน) ไฟ (ไฟ) มา (มา) เหล่านี้เป็นต้น ภาษาลาวมีลักษณะพิเศษคือ คำหนึ่งๆมีความหมายเฉพาะอย่างหนึ่งหรือบางคำที่ออกเสียงเหมือนกันแต่มีความหมายต่างกัน แต่ถ้าต้องการความหมายให้สื่อเข้าใจให้ได้ดีนั้นต้องนำคำที่ต้องการมาเรียงกันให้ถูกตำแหน่ง และหน้าที่เพื่อให้มีข้อความสัมพันธ์กันอย่างสมเหตุผล และทำให้ผู้รับฝังเข้าใจตามความประสงค์ของผู้พูดก็ถือเป็นอันว่าได้ประโยคที่สมบูรณ์แล้ว

ดังที่ได้กล่าวมาแล้ว ภาษาลาวเป็นภาษาคำเดี่ยวที่มีพยางค์เดียว แต่ต่อมาสังคมลาวได้ขยายตัวออกมีการพบปะกับชนชาติอื่น อีกอย่างหนึ่งคนในชาติก็มีความต้องการใช้ศัพท์ภาษามาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขึ้นเมื่อเห็นว่าภาษาเดิม ซึ่งมีแต่คำเดี่ยวนั้นไม่เพียงพอในการใช้พูดจาสนทนาก็จึงรู้จักดัดให้เป็นคำประสมหลายพยางค์เพื่อให้เพียงพอกับความต้องการ

เมื่อชนชาติลาวได้ติดต่อกับชาติต่างๆ ซึ่งนิยมพูดภาษาที่ใช้คำหลายพยางค์ จากนั้นก็นำเอาคำเหล่านั้นมาใช้สับปนกับภาษาของตน อันเป็นเหตุให้มีคำพูดของภาษาอื่นๆมาประสมกับภาษาลาว เมื่อพุทธศาสนาแผ่เข้ามาถึงประเทศลาวภาษาบาลีหรือสันสกฤตที่ใช้อยู่ในคำสอนของพระพุทธศาสนาจึงเข้ามามีบทบาทให้ภาษาลาวขยายตัวทางด้านคำศัพท์ อันเป็นเหตุให้ภาษาลาวมีคำพูดที่มีหลายพยางค์เพิ่มขึ้นอีกอย่างเช่น ประเทศ(ปะเทด) นคร(นะคอน) กรรมกร(กำมะกอน) เกษตรกรรม(กะสิกำ) หัตถกรรม(หัดถะกำ) อุตสาหกรรม(อุดสาหะกำ) วัฒนธรรม (วัดทะนะทำ) เหล่านี้เป็นต้น ดังนั้นหลักการเขียนภาษาบาลีจึงกายมาเป็นแม่แบบให้แก่การเขียนคำศัพท์ภาษาลาว

โดยอิงตามการค้นคว้าภาษาศาสตร์ภายในและต่างประเทศ[6] ได้พบว่าหนังสือลาวได้ประดิษฐ์แต่งขึ้นโดยเริ่มจากหนังสือขอมตั้งแต่นานมาแล้ว อีกจำพวกหนึ่งให้เหตุผลว่าอักษรลาวกำเนิดขึ้นตามแบบอย่างของอักษรสันสกฤตที่ใช้อยู่ในพุทธศาสนานิกายมหายาน ซึ่งขยายเข้ามาถึงชนเผ่าลาวไทในตอนแรก หลังจากนั้นพุทธศาสนานิกายมะหายานที่ใช้หนังสือบาลีแผ่เข้ามาถึงประเทศมอนหรือพม่า คนชาติลาวจึงได้หลักตัวหนังสือแบบมอน-พม่าที่ดัดแปลงคำจากอักษรคะลินมาเป็นตัวอักษรเพื่อใช้บันทึกคำสอนแห่งพุทธศาสนาเท่านั้น ส่วนตัวหนังสือดั้งเดิมของลาวนั้นเรียกว่าตัวลาวหรืออักษรลาวได้ใช้ทางราชการสำหรับบันทึกเหตุการณ์บ้านเมืองตลอดมาถึงทุกวันนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
หลักของภาษาลาว

3.1 เสียงในภาษาคืออะไร ?

เสียงในภาษาคือเสียงที่เป่งออกมาจากปาก ซึ่งเสียงที่เป่งออกมาคือลมที่ถูกบังคับให้ผ่านอวัยวะต่างๆเช่น เส้นเสียงซึ่งอยู่ในช่องของหลอดลมผ่านกรวยคอ แล้วลมนั้นจะผ่านต่อออกมาทางช่องปากหรือช่องจมูก อวัยวะในช่องปากก็จะดัดแปลงให้เป็นเสียงต่างๆตามที่ผู้พูดต้องการ โดยเสียงที่ใช้ในภาษาพูดนั้นจะมีลักษณะที่สำคัญบางประการร่วมกันคือ

3.1.1 ความก้อง หรือ ไม่ก้องของเสียง

เสียงก้อง (Voice) คือเสียงที่เกิดในขณะที่เส้นเสียงเกิดการตึงตัวหรือเรียกว่าเส้นเสียงปิดเมื่อมีแรงดันให้อากาศไหลผ่านกล่องเสียง ในขณะที่เส้นเสียงปิดจะเกิดการสั่นสะบัดของเส้นเสียงเป็นผลให้สัญญาณเสียงที่ได้ (speech waveform) มีลักษณะเป็นคาบ (quasi-periodic) ซึ่งสามารถเรียกความถี่ในการปิด-เปิดของเส้นเสียงนี้ว่า "ความถี่มูลฐาน" (Fundamental Frequency: F_0) ตัวอย่างของเสียงก้องได้แก่ เสียงสระต่างๆ และเสียงพยัญชนะเช่น บ (ɓ) ด(ɗ) ที่เกิดจากการเปล่งเสียงออกทางปาก หรือเสียงพยัญชนะ ม(ɱ) น(ɲ) ง(ŋ) ที่เกิดจากการเปล่งเสียงออกทางจมูก

เสียงไม่ก้อง (Unvoice) คือเสียงที่เกิดในขณะที่เส้นเสียงคลายจากการตึงหรือเรียกว่าเส้นเสียงเปิด เมื่อมีแรงดันให้อากาศไหลผ่านกล่องเสียงในขณะที่เส้นเสียงเปิด อากาศที่ไหลผ่านอย่างรวดเร็วจะเกิดการไหลวนและปั่นป่วนทำให้เกิดเสียงที่มีลักษณะเป็นเสียงของสัญญาณรบกวน

3.1.2 ระดับเสียงสูง-ต่ำ (Pitch)

เสียงพูดจะมีระดับ สูง หรือ ต่ำ อยู่ที่ความถี่ของเสียง (Fundamental frequency) ถ้าความถี่ต่ำเสียงก็จะต่ำ อวัยวะส่วนที่ทำให้เสียงมีระดับ สูง-ต่ำ คือเส้นเสียง ดังนั้นระดับเสียงสูง-ต่ำก็คืออัตราการสั่นสะบัดของเส้นเสียงที่ทำให้เกิดมีความถี่ระดับแตกต่าง ในภาษาลาวระดับเสียง สูง-ต่ำของคำพูดเราเรียกว่าวรรณยุกต์

3.1.3 ความดัง (Loudness)

ความดังขึ้นอยู่กับปริมาณของลม ที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่งๆ

3.1.4 การลงน้ำหนัก (Stress)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมายถึง การออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นที่อยู่ข้างเคียงเพื่อต้องการเรียกร้องความสนใจเป็นพิเศษหรือแสดงอารมณ์อย่างใดอย่างหนึ่ง

3.1.5 ช่วงต่อของเสียง (Juncture)

หมายถึงช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งอีกเสียงหนึ่ง ซึ่งเรียงกันมาเป็นลำดับเสียงที่ประกอบกันเข้าเป็นพยางค์จะมีช่วงต่อของเสียงแนบสนิทจนไม่เห็นร่องรอย (close juncture) แต่ถ้าเสียงปรากฏอยู่คนละพยางค์หรือคนละคำ จะมีช่วงต่อระยะห่างจนสังเกตได้ชัด (open juncture) ดังนั้นช่วงต่อของเสียงโดยเพาะช่วงห่างจะมีความสำคัญมากในการแบ่งคำในภาษา

3.2 หน่วยเสียงสำคัญในภาษาลาว

หน่วยเสียง(phoneme) เป็นหน่วยเล็กที่สุดของภาษา หน่วยดังกล่าวได้แก่เสียงสำคัญๆในภาษาใดภาษาหนึ่ง ซึ่งทำหน้าที่ให้ความหมายของคำที่ใช้ในภาษานั้นและทำให้ความหมายของคำนั้นๆให้มีความหมายแตกต่างจากคำอื่นๆ หน่วยเสียงสำคัญในภาษาลาวมี 3 ประเภทใหญ่ๆคือ เสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์

3.2.1 หน่วยเสียงสระ

ลักษณะสำคัญของเสียงสระก็คือ เป็นเสียงก้องที่เปล่งเสียงออกมาโดยให้ลมออกทางช่องปากโดยไม่ถูกลิ้นกักหรือขัดขวาง ดังนั้นเวลาเราออกเสียงสระจะออกเสียงได้สะดวกและออกเสียงได้นาน ทั้งนี้เพราะคุณสมบัติของเสียงสระมีความดังเด่นกว่าเสียงอื่นๆที่เรียงอยู่ข้างเสมอ อวัยวะที่เกี่ยวข้องกับการออกเสียงสระได้แก่ ลิ้น กับริมฝีปาก ถ้าลิ้นส่วนใดทำหน้าที่เพียงส่วนเดียว เสียงที่เกิดขึ้นก็จะมีเพียงเสียงเดียว เสียงสระเราเรียกว่า สระเดี่ยว แต่ถ้าลิ้นส่วนอื่นทำหน้าที่ร่วมด้วย เสียงสระนั้นเรียงว่า สระประสม

3.2.2 หน่วยเสียงพยัญชนะ

หน่วยเสียงพยัญชนะออกเสียงได้ไม่สะดวกเท่าหน่วยเสียงสระ เพราะเวลาออกเสียงลมหายใจที่พุ่งออกมาจากหลอดลมจะถูกขัดขวางตามส่วนต่างๆของปาก เสียงพยัญชนะจึงออกเสียงให้ยาวนานอย่างเสียงสระไม่ได้และเสียงพยัญชนะก็ไม่ใช่เสียงก้องทุกตัว

3.2.3 หน่วยเสียงวรรณยุกต์

เสียงวรรณยุกต์ คือระดับเสียงสูง-ต่ำ ของคำพูดในภาษาลาว เช่นเดียวกับภาษาไทย และภาษาอื่นๆที่เป็นภาษาคำโดด ซึ่งมีการกำหนดเสียงสูงต่ำไว้ตายตัวในคำแต่ละคำ ถ้าออกเสียงสูง-ต่ำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผิดไปความหมายย่อมผิดตามไปด้วย โดยภาษาลาวหน่วยเสียงวรรณยุกต์เป็นหน่วยเสียงสำคัญ ที่ทำให้คำพูดที่มีความหมายต่างกัน ดังนั้นอาจกล่าวได้ว่าหน้าที่ของหน่วยเสียงวรรณยุกต์ก็คือการทำให้เกิดคำพูดขึ้นใช้ในภาษามากขึ้น เพราะถ้าเราเปลี่ยนเสียงวรรณยุกต์ก็จะทำให้คำเกิดความหมายเพิ่มขึ้นใหม่เช่นคำว่าเสือ(ເສືອ) เสื่อ (ເສື່ອ) เสื้อ (ເສື້ອ) อา (ອາ) อ่า (ອ່າ) อ้า(ອ້າ) ฯลฯ

เสียง สูง-ต่ำ ในภาษาพูดเกิดจากการสั่นสะเทือนของเส้นเสียงในอัตราต่างๆกัน โดยเสียงที่เปล่งออกมาในขณะที่เส้นเสียงสั่นนั้นจะต้องเป็นเสียงก้อง ดังนั้นหน่วยเสียงวรรณยุกต์ในภาษาลาวจึงจัดเป็นหน่วยเสียงซ้อน โดยสัตย์อักษรที่ใช้เป็นรูปเครื่องหมายเขียนซ้อนข้างบนหน่วยเสียงสระซึ่งมีรูปวรรณยุกต์อยู่ 4 รูป แทนเสียงวรรณยุกต์ทั้งหมด 5 หน่วยเสียงโดยเสียงสามัญไม่มีรูปวรรณยุกต์ ลักษณะของเสียงวรรณยุกต์สามารถแบ่งออกตามลักษณะระดับเสียงสำเนียงการพูดได้ 5 หน่วยเสียง [7] คือ

3.2.3.1 เสียงสามัญ มีระดับหน่วยเสียง คงที่หรือกลาง (mid tone) แทนด้วยสัญลักษณ์ [—]

3.2.3.2 เสียงเอก มีระดับหน่วยเสียง สูง-ตก (high falling) แทนด้วยสัญลักษณ์ [\ ]

3.2.3.3 เสียงโท มีระดับหน่วยเสียง กลาง-ตก (mid-falling) แทนด้วยสัญลักษณ์ [∧]

3.2.3.4 เสียงตรี มีระดับหน่วยเสียง ต่ำ-ขึ้น (rising-tone) แทนด้วยสัญลักษณ์ [/] และ

5.2.3.5 เสียงจัตวา มีระดับหน่วยเสียง กลาง-ขึ้น (mid-rising) แทนด้วยสัญลักษณ์ [∨]

3.3 ตัวอักษรลาว

ภาษาลาวมีตัวอักษรใช้สำหรับเขียนและอ่านคือกับภาษาของชาติต่างๆ ตัวอักษรลาวแบ่งออก เป็นสามหมวดคือ พยัญชนะ สระและวรรณยุกต์ เพื่อจะเขียนคำพูดคำหนึ่งจะต้องใช้ตัวอักษรหลายส่วนมาประกอบเข้าด้วยกันอย่างเช่นคำว่า แก้ว(ແກ້ວ) ◌ແ x คือสระ ก คือพยัญชนะหลักหรือเรียกว่าพยัญชนะตัวออกเสียง ◌ວ คือพยัญชนะตัวสะกดและ ◌ ้ คือวรรณยุกต์

3.3.1 พยัญชนะ

3.3.1.1 พยัญชนะเดิม

ในจำนวนพยัญชนะทั้งหมด 32 ตัว มีพยัญชนะเดิม ซึ่งเป็นพยัญชนะตัวที่ยังไม่ได้ผสมกับตัวอื่นมี 26 ตัวด้วยกันคือ ก(ก) ข(ຂ) ค(ຄ) ง(ງ) จ(จ) ส(ส) ช(ຊ) ย(ย) ด(ด) ต(ต) ถ(ถ) ท(ท) น(น) บ(บ) ป(ป) ผ(ผ) ฝ(ฝ) พ(พ) ฟ(ฟ) ม(ม) ยย(ย) ล(ล) ว(ว) ห(ห) อ(อ) ฮ(ฮ) โดยที่พยัญชนะเดิมได้แบ่งออกเป็นสามหมู่อักษรคือ อักษรสูง อักษรกลางและอักษรต่ำ ซึ่งอักษรสูงคือ จำพวกอักษรที่ออกเสียงหักขึ้นสูง ซึ่งรวมมี 6 ตัวด้วยกันคือ ຂ ສ ຖ ผ ฝ ห อักษรกลางคือ จำพวกอักษรที่มีเสียงอยู่ในระดับกลาง (ไม่สูง ไม่ต่ำ) ซึ่งรวมมี 8 ตัวด้วยกันคือ ก จ ด ต บ ป ย อ และอักษรต่ำคือ จำพวกอักษรที่มีเสียงต่ำซึ่งรวมมี 12 ตัวคือ ค ง ช ย ท น พ ฟ ม ล ว ฮ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.1.2 พยัญชนะผสม

พยัญชนะผสม คือพยัญชนะสองตัวที่เอามาผสมกันให้เป็นตัวเดียว ซึ่งทั้งหมดมี 6 ตัว ด้วยกัน โดยแต่ละตัวจะตั้งต้นด้วยตัว ຫ ที่เป็นอักษรเสียงสูงแล้วเอาตัว ງ ຍ ນ ມ ລ ວ ที่เป็นอักษรเสียงต่ำมาผสมใส่เมื่อประสมใส่แล้วตัว ຫ ไม่ออกเสียง แต่ทำให้อักษรต่ำเหล่านั้นผันเสียงเป็นอักษรเสียงสูงดังนี้ ຫງ(ห+ง) ຫຍ(ห+ย) ຫນหรือໜ(ห+น) ຫມหรือໝ(ห+ม) ຫລหรือໜ(ห+ล) และ ຫວ(ห+ว)

3.3.1.3 พยัญชนะควบ

พยัญชนะควบ คือพยัญชนะที่นำเสียงไปควบกับเสียงพยัญชนะตัวอื่น โดยในภาษาลาวมีตัวเดียว คือตัว ວ กำหนดไว้หลังพยัญชนะหลัก เพื่อนำเสียงของมันควบเข้ากับพยัญชนะนั้นให้มีเสียงตัว ວ อยู่นำในเวลาอ่านออกเสียงควบกันหรือออกเสียงพร้อมกันทั้งสองตัวเช่น ກວ(แกว่ง) ຂວ(แขวง) ຄວ(ความ) ງວ(งูมงวา) ສວ(สว่าง) ฯลฯ

3.3.1.4 อักษรหลักและตัวสะกด

อักษรหลัก คือพยัญชนะตัวที่เอาเสียงของตัวเองเป็นหลักเสียงของคำพูดอย่างเช่น ไป ตัว ป คืออักษรหลัก โดยอักษรหลักทั้งหมด 32 ตัว ตามจำนวนพยัญชนะทั้งหมด ในนี้มีจำนวนหนึ่งเป็นอักษรหลักอย่างเดียวและอีกจำนวนหนึ่งเป็นทั้งอักษรหลักและอักษรสะกดไปพร้อม ซึ่งอักษรหลักอย่างเดียวมี 24 ตัวคือ ຂ ຄ ຈ ສ ຊ ฑ ฒ ท ป ผ ฝ พ ฟ ย ล ห อ ฮ ຫງ ຫຍ ໜ ໝ ໜ ຫວ และจำนวนที่เป็นอักษรหลักและเป็นทั้งตัวสะกดมี 8 ตัวคือ ก ง ย ด น บ ม ว โดยตัวสะกด คือพยัญชนะที่อยู่ท้ายของพยางค์ ซึ่งไม่มีสระผสมอยู่ด้วยตัวอย่างเช่น ภาก ภาๆ ภวม ภิม ภุม ภาย ติว พยัญชนะตัว ก ง บ ม ย แม่นตัวสะกด ซึ่งเป็นทั้งตัวสะกดเป็นและตัวสะกดตายดังนี้

ตัวสะกดเป็น คือตัวสะกดที่แปลเสียงไปตามวรรณยุกต์มี 5 ตัวคือ ง ย น ม ว ดังตัวอย่าง ภ้าง ภ้อย ภิ่น ภ้อม ภ้าว และตัวสะกดตาย คือตัวสะกดที่แปรเสียงไปตามวรรณยุกต์ไม่ได้ครบถ้วน คือ ก ด บ ตัวอย่างของศัพท์คือ ภัก ภับ ภัด เป็นต้น

3.3.2 สระ

สระคืออักษรหมวดหนึ่งที่ใช้ประกอบใส่พยัญชนะ ให้อ่านออกเสียงเป็นพยางค์หรือคำพูด สระรวมมีทั้งหมด 28 ตัว ซึ่งในนั้นรวมมีทั้งสระเดิมและสระผสม โดยที่สระเดิมมี 12 ตัวคือ ะ า ิ ี ึ ื ุ ู เ แ โ ำ และสระผสมมี 12 ตัวคือ เะ แะ เิ เี เือ เือ โะ เาะ ิอะ ิอ เีย เือย นอกจากนี้ยังมีสระพิเศษอีก 4 ตัวคือ ไ ใ เิา ฤ ที่ไม่มีสระอื่นเป็นครู่ ฉะนั้นจึงจัดเป็นสระพิเศษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.3 วรรณยุกต์

วรรณยุกต์ คืออักษรจำพวกหนึ่งที่ใช้ควบคุมเสียง เพื่อให้สำเนียงเสียงพูดต่างจากเสียงของสระและพยัญชนะเดิมดังตัวอย่างได้แก่คำว่า ม้าแล่น(ม้าแล่น) ถ้าไม่ใส่วรรณยุกต์จะอ่านออกเป็นมาแลน ซึ่งจะไม่มีความหมายอะไร

วรรณยุกต์ในภาษาลาวมี 4 เสียง แทนเครื่องหมาย 4 รูปคือ ไม้เอก ไม้โท ไม้ตรี และไม้จัตวา โดยกระบวนการเขียนคำพูดกับรูปเสียงวรรณยุกต์อาจจะไม่ตรงกันเสมอไป เช่นคำที่ใช้รูปวรรณยุกต์เอกอาจจะมีเสียงโท คำที่ใช้รูปเสียงโทร อาจจะเป็นเสียงตรี หรือคำที่ไม่มีรูปวรรณยุกต์กำกับอาจจะมีเสียงวรรณยุกต์ที่เป็น เอก โท ตรี และ จัตวาก็ได้ โดยรูปวรรณยุกต์ในภาษาลาว

- ◌่ เรียกว่าไม้เอก หรือเรียกว่าไม้ที่หนึ่ง
- ◌้ เรียกว่าไม้โท หรือเรียกว่าไม้ที่สอง
- ◌๊ เรียกว่าไม้ตรี หรือเรียกว่าไม้ที่สาม
- ◌๋ เรียกว่าไม้จัตวา หรือเรียกว่าไม้ที่สี่

โดยการเขียน ◌่ ◌้ ◌๊ และ ◌๋ ให้ไว้ข้างบนของอักษรหลักของคำพูดทุกๆครั้งอย่างเช่น ก่า ก้า ก๊า ก๋า

3.4 การผสมอักษร

การผสมอักษร คือการเอาพยัญชนะกับสระผสมเข้ากัน ซึ่งทำให้อ่านออกเสียงได้วิธีผสมอักษรมีอยู่ด้วยกันสองแบบคือ

3.4.1 แม่กะกา

แม่กะกา คือการเอาพยัญชนะกับสระผสมกัน ซึ่งรวมมีแม่กะกาธรรมดาและแม่กะกาใส่วรรณยุกต์ โดยแม่กะกาธรรมดา คือเพียงแต่เอาพยัญชนะผสมกับสระเท่านั้นอย่างเช่น ก ຂ ถ ผสมกับสระ ◌ะ เป็น กะ ຂะ ถะ และแม่กะกาใส่วรรณยุกต์ดังตัวอย่างคือ อ่า อ้า อ๊า อ๋า

3.4.2 แม่สะกดทั้ง 8

แม่สะกดทั้ง 8 คือแม่กะกาที่ผสมตัวสะกดเข้านำตัวสะกดในอักษรลาวทั้งหมดมี 8 ตัว จึงเรียกว่าแม่สะกดทั้งแปดมีดังนี้

ตัว ก	สะกดเรียกว่า	กัก
ตัว ง	สะกดเรียกว่า	กัง
ตัว ด	สะกดเรียกว่า	กัด
ตัว บ	สะกดเรียกว่า	กับ
ตัว ຍ	สะกดเรียกว่า	กัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัว ม สะกดเรียกว่า มัม (ใช้เขียนเป็น ภาำ)

ตัว ย สะกดเรียกว่า มัย (ใช้เขียนเป็น ไภ)

ตัว อ สะกดเรียกว่า มัอ (ใช้เขียนเป็น เภิา)

การผสมแม่สะกดทั้งแปดมีอยู่ด้วยกัน 2 วิธีคือแม่สะกดทั้ง 8 ธรรมดาและแม่สะกดทั้ง 8 ใส่วรรณยุกต์

3.4.2.1 แม่สะกดทั้ง 8 แบบธรรมดา

ตารางที่ 3.1 แสดงการสะกดอักษรของแม่สะกดในภาษาลาวแบบธรรมดา

สระ	แม่ มัภ ภ สะกด	แม่ มัา า สะกด	แม่ มัด ดสะกด	แม่ มัม มสะกด	แม่ มัย ย สะกด	แม่ ภาำ มสะกด	แม่ภาย ย สะกด	แม่ ภาอ อ สะกด
xะ	xัภ	xัา	xัด	xัม	xัย	xำ	ไx	xัอ
xา	xาภ	xาา	xาด	xาม	xาย	xาม	xาย	xาอ
xิ	xิภ	xิา	xิด	xิม	xิย	xิม	xิย	xิอ
xี	xีภ	xีา	xีด	xีม	xีย	xีม	xีย	xีอ
xึ	xึภ	xึา	xึด	xึม	xึย	xึม	xึย	xึอ
xื	xืภ	xืา	xืด	xืม	xืย	xืม	xืย	xือ
xุ	xุภ	xุา	xุด	xุม	xุย	xุม	xุย	xุอ
xู	xูภ	xูา	xูด	xูม	xูย	xูม	xูย	xูย
เxะ	เxัภ	เxัา	เxัด	เxัม	เxัย	เxัม	เxัย	เxัอ
เx	เxภ	เxา	เxด	เxม	เxย	เxม	เxย	เxอ
แxะ	แxัภ	แxัา	แxัด	แxัม	แxัย	แxัม	แxัย	แxัอ
แx	แxภ	แxา	แxด	แxม	แxย	แxม	แxย	แxอ
โxะ	xภ	xา	xด	xม	xย	xม	โxัย	โxัอ
โx	โxภ	โxา	โxด	โxม	โxย	โxม	โxย	โxอ
เxาะ	xอัภ	xอัา	xอัด	xอัม	xอัย	xอัม	xอัย	xัออ
xำ	xอภ	xอา	xอด	xอม	xอย	xอม	xอย	xออ
เxิ	เxิภ	เxิา	เxิด	เxิม	เxิย	เxิม	เxิย	เxิอ
เxิ	เxิภ	เxิา	เxิด	เxิม	เxิย	เxิม	เxิย	เxิอ
เxือ	เxือภ	เxือา	เxือด	เxือม	เxือย	เxือม	เxือย	เxืออ
เxือ	เxือภ	เxือภ	เxือด	เxือด	เxือย	เxือม	เxือย	เxืออ
เxัย	xัยภ	xัยา	xยด	xัยม	xัยย	xัยม	xัยย	xัยอ
เxย	xยภ	xยา	xยด	xยม	xยย	xยม	xยย	xยอ
xัอะ	xอัภ	xอัา	xอัด	xอัม	xอัย	xอัม	xอัย	xอัอ
xัอ	xอภ	xอา	xอด	xอม	xอย	xอม	xอย	xออ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.2.2 แม่สะกดทั้ง 8 ใส่วรรณยุกต์

คือการประสมพยัญชนะ สระและตัวสะกดเข้าด้วยกันแล้วใส่วรรณยุกต์เข้าไปดังตัวอย่างนี้

ตารางที่ 3.2 แสดงแม่สะกดในภาษาลาวแบบใส่วรรณยุกต์กับอักษรกลาง

แม่สะกดทั้ง 8	อักษรกลาง			
	ไม้เอก ◌่	ไม้โท ◌้	ไม้ตรี ◌๊	ไม้จัตวา ◌๋
แม่ กัก (ก สะกด)	ภาก	ภ้าก	ภ๊าก	ภ๋าก
แม่ กัง (ง สะกด)	ภ่าง	ภ้าง	ภ๊าง	ภ๋าง
แม่ กัด (ด สะกด)	ภาด	ภ้าด	ภ๊าด	ภ๋าด
แม่ กับ (บ สะกด)	ภ่าบ	ภ้าบ	ภ๊าบ	ภ๋าบ
แม่ กับ (บ สะกด)	ภาบ	ภ้าบ	ภ๊าบ	ภ๋าบ
แม่ กำ (ม สะกด)	ภ่าม	ภ้าม	ภ๊าม	ภ๋าม
แม่ กาย (ย สะกด)	ภ่าย	ภ้าย	ภ๊าย	ภ๋าย
แม่ กาว (ว สะกด)	ภ่าว	ภ้าว	ภ๊าว	ภ๋าว

ตารางที่ 3.3 แสดงการสะกดอักษรในภาษาลาวแบบใส่วรรณยุกต์กับอักษรสูงและอักษรต่ำ

แม่สะกดทั้ง 8	อักษรสูง		อักษรต่ำ	
	ไม้เอก ◌่	ไม้โท ◌้	ไม้เอก ◌่	ไม้โท ◌้
แม่ กัก (ก สะกด)	สาก	สาก	ถาก	ถาก
แม่ กัง (ง สะกด)	ส่าง	ส้าง	ถ่าง	ถ้าง
แม่ กัด (ด สะกด)	สาด	สาด	ถาด	ถาด
แม่ กับ (บ สะกด)	ส่าบ	ส้าบ	ถ่าบ	ถ้าบ
แม่ กับ (บ สะกด)	สาบ	สาบ	ถาบ	ถาบ
แม่ กำ (ม สะกด)	ส่าม	ส้าม	ถ่าม	ถ้าม
แม่ กาย (ย สะกด)	ส่าย	ส้าย	ถ่าย	ถ้าย
แม่ กาว (ว สะกด)	ส่าว	ส้าว	ถ่าว	ถ้าว

หมายเหตุ คำที่ขีดเส้นใต้ไว้ตามหมวดอักษรกลาง สูง และ ต่ำ นั้นคือคำสะกดตาย ซึ่งไม่สามารถผันแปรไปตามวรรณยุกต์ได้ครบถ้วน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 วิธีผันเสียง

วิธีผันเสียงคือการประกอบวรรณยุกต์เข้ากับคำพูดที่เขียนด้วยพยัญชนะและสระ เพื่อให้คำพูดนั้นเป็นไปตามสำเนียงคำพูด โดยการผันเสียงจะเป็นไปตามเสียงพยัญชนะ ซึ่งมีตัวอย่างดังนี้

ตารางที่ 3.4 แสดงตัวอย่างการผันเสียงอักษรไปตามวรรณยุกต์

อักษรกลาง	เสียงสามัญ [—]	เสียงเอก [\]	เสียงโท [∧]	เสียงตรี [/]	เสียงจัตวา [∨]
ก	กา	ก่า	ก้า	ก๊า	ก๋า
จ	จาง	จ่าง	จ้าง	จ๊าง	จ๋าง
ป	ปาม	ป่าม	ป้าม	ป๊าม	ป๋าม
ต	ตอง	ต่อง	ต้อง	ต๊อง	ต๋อง
บ	บา	บ่า	บ้า	บ๊า	บ๋า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ทฤษฎีการวิเคราะห์และสร้างแบบจำลองเสียงวรรณยุกต์ในภาษาลาว

4.1 กล่าวนำ

เสียงพูดมีหลายระดับที่แตกต่างกันตามการเปล่งเสียงออกมาของผู้พูด การออกเสียงที่มีระดับสูง-ต่ำในเสียงพูดนั้นเรียกว่าเสียงวรรณยุกต์ ซึ่งเกิดจากการสั่นสะบัดของเส้นเสียงเป็นจังหวะที่มีความเป็นคาบของสัญญาณ และความถี่ของเสียงจะสูงหรือต่ำนั้นสามารถสังเกตได้จากคาบหรือเรียกว่าพิทช์ ซึ่งความถี่ของการสั่นหากมีระยะความเป็นคาบมากจะทำให้ความถี่มูลฐานมีค่าต่ำและหากมีระยะความเป็นคาบน้อยก็จะทำให้ความถี่มูลฐานมีค่าสูง ซึ่งเป็นส่วนสำคัญในการศึกษาเสียงวรรณยุกต์เพื่อนำไปสร้างแบบจำลองจดจำเสียงพูด โดยรูปแบบการจดจำก็คือลักษณะหนึ่งของการจดจำรูปแบบ (Pattern Recognition) ที่เปรียบเทียบกันระหว่างแบบทดสอบ (Test Pattern) กับแบบอ้างอิง(Reference Pattern) ซึ่งเป็นรูปแบบที่ทราบและเก็บไว้ล่วงหน้าจากขั้นตอนการเรียนรู้และจดจำมาก่อน

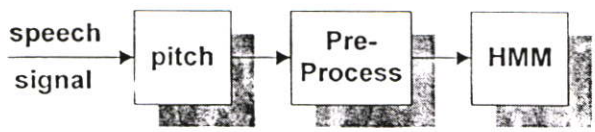
4.1.1 ขั้นตอนการเรียนรู้

เป็นการสร้างกลุ่มของแบบอ้างอิงในการจดจำเสียงพูด ขั้นตอนนี้จะทำการวิเคราะห์เสียงก่อน โดยดึงลักษณะของพารามิเตอร์ที่ต้องการออกมา ซึ่งในวิทยานิพนธ์นี้ได้ใช้แนวทางการเปลี่ยนแปลงของค่าความถี่มูลฐานของเสียง เพื่อทำการจัดกลุ่มพารามิเตอร์โดยใช้การควอนไตซ์ข้อมูลเพื่อนำไปสร้างแบบจำลองอ้างอิงในการรู้จำ ซึ่งมีขั้นตอนการวิเคราะห์ดังแสดงในรูปที่ 4.1 ได้แบ่งขั้นตอนในการวิเคราะห์และการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ออกเป็น 3 ขั้นตอน

4.1.1.1 การหาค่าพิทช์

4.1.1.2 การเตรียมข้อมูล เพื่อใช้เป็นข้อมูลฝึกสอนในการสร้างแบบจำลองอ้างอิง

4.1.1.3 การสร้างแบบจำลองอ้างอิงด้วยฮิดเดนมาร์คอฟโมเดล



รูปที่ 4.1 ขั้นตอนในการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 ขั้นตอนการจดจำ

การจดจำจะได้จากการทดสอบระหว่างแบบอ้างอิงกับแบบทดสอบ โดยจะทำการเปรียบเทียบพารามิเตอร์ของแบบทดสอบกับแบบอ้างอิงทั้งหมด ซึ่งแบบอ้างอิงที่เลือกจะต้องเป็นแบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด

4.2 การหาค่าพิทช์ของสัญญาณเสียงพูด

4.2.1 การวิเคราะห์ในโดเมนเวลา

สัญญาณเสียงพูดเป็นสัญญาณต่อเนื่องที่เปลี่ยนแปลงไปตามเวลา โดยเกิดในลักษณะแบบสุ่ม(random) แต่ก็ขึ้นกับการควบคุมเสียงของผู้พูดด้วยเพราะเสียงที่เปล่งออกมาในระยะเวลาหนึ่งนั้นจะขึ้นอยู่กับรูปทรงของช่องทางเดินเสียง(vocal tract) และลักษณะการสั่นของเส้นเสียง(vocal cord) เสียงพูดจึงเป็นสัญญาณที่มีคาบเวลาชั่วขณะ(quisi-periodic) คือมีความเป็นคาบคงที่ในเวลาอันสั้นและมีการเปลี่ยนแปลงในช่วงระหว่างเวลานั้น ดังนั้นในการวิเคราะห์จึงต้องทำการแบ่งเสียงพูดออกเป็นช่วงๆ(Frame) โดยมีช่วงเวลาอยู่ระหว่าง 10-30 มิลลิวินาที ในช่วงเวลาดังกล่าวถือว่าเสียงจะมีการเปลี่ยนแปลงคุณสมบัติน้อยมาก ดังนั้นในแต่ละเฟรมจึงมีคุณสมมติเป็นสัญญาณที่มีลักษณะคงที่ ซึ่งทำให้การวิเคราะห์ทำได้ง่ายขึ้น

4.2.2 ทฤษฎีการประมาณค่าพิทช์โดยใช้ออโตคอร์รีเลชั่นฟังก์ชั่น

ออโตคอร์รีเลชั่นฟังก์ชั่น [8] เป็นวิธีหนึ่งที่ได้ถูกนำมาใช้ตรวจหาคาบพิทช์ของสัญญาณเสียงพูด โดยออโตคอร์รีเลชั่นจะทำหน้าที่ในการแสดงยอดกราฟหลัก (prominent peak) ที่เป็นคาบในแต่ละส่วนของเสียง (section) ซึ่งจะทำให้เราสามารถหาระยะความเป็นคาบของสัญญาณตามแต่ละช่วงของเส้นเสียงได้

4.2.2.1 การจัดแบ่งการวิเคราะห์สัญญาณออกเป็นช่วงสั้นๆ

กำหนดให้สัญญาณเสียงแทนด้วย s(m) ซึ่งออโตคอร์รีเลชั่นฟังก์ชั่นของ discrete-time deterministic signal โดยทั่วไปเขียนได้เป็น

$$\phi(\tau) = \sum_{m=-\infty}^{\infty} s(m)\, s(m+\tau) \qquad (4.1)$$

ถ้าสัญญาณ s(m) มีความเป็นคาบที่แน่นอนด้วยระยะ P นั่นคือ

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ดังนั้นสามารถเขียนได้ว่า

$$\phi(\tau) = \phi(\tau + p) \qquad (4.2)$$

นั่นคือ ออโตคอร์รีเลชั่นฟังก์ชั่นก็มีความเป็นคาบด้วยระยะคาบเดียวกัน หรือในทางกลับกันก็คือความเป็นคาบในออโตคอร์รีเลชั่นฟังก์ชั่น เป็นตัวบ่งชี้ให้เห็นถึงความเป็นคาบในสัญญาณ โดยคุณสมบัติของออโตโครีเลชั่นฟังก์ชั่นที่สำคัญคือ

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

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

ถ้าพิจารณาสมการที่ 4.2 ควบคู่ไปกับคุณสมบัติในข้อ 1 และ 2 จะพบความเป็นคาบของสัญญาณ โดยแซมเปิ้ลของออโตโครีเลชั่นจะมีค่ามากที่สุดที่ $0, \pm p, \pm 2p, \ldots$ โดยไม่ต้องคำนึงถึงเวลาเริ่มต้น (time origin) ของสัญญาณ การคำนวณหาคาบของสัญญาณสามารถประมาณได้จากตำแหน่งแรกที่มีค่ามากที่สุดในออโตโครีเลชั่นฟังก์ชั่น และตำแหน่งถัดไปจากยอดกราฟ ซึ่งจากคุณสมบัติเหล่านี้ทำให้ออโตโครีเลชั่นฟังก์ชั่นเป็นหลักการพื้นฐาน ที่ใช้ประมาณค่าความเป็นคาบในสัญญาณได้

สำหรับสัญญาณที่มีลักษณะเปลี่ยนแปลงอยู่ตลอดเวลาเช่น สัญญาณเสียงพูดจะต้องทำการแบ่งสัญญาณออกเป็นช่วงสั้นๆเพื่อหาข้อมูล (Information) ที่ต้องการ โดย short-time auto-correlation function ซึ่งสามารถนิยามได้เป็น

$$R(\tau) = \sum_{m=0}^{N-1-\tau} s(m)\, s(m+\tau) \qquad (4.3)$$

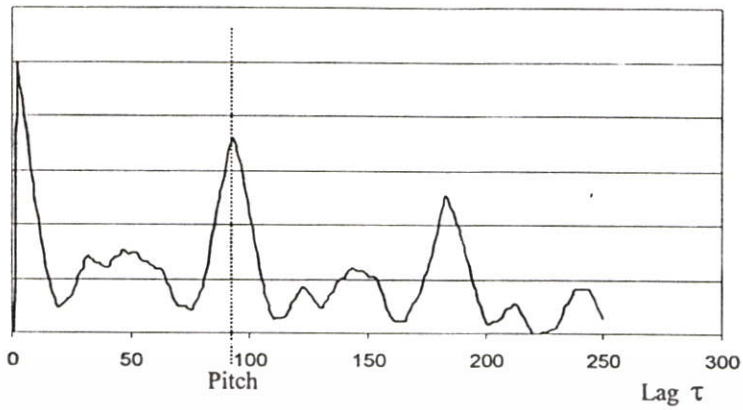
เมื่อ N คือ จำนวนตัวอย่างสัญญาณ (sample) ต่อเฟรม

τ คือ จุดเลื่อนไปในการคำนวณออโตคอร์รีเลชั่น

โดยในการเลือกค่าของจำนวนตัวอย่าง (N) ที่ใช้ในแต่ละเฟรม รูปคลื่นสัญญาณจะต้องมีความเป็นคาบที่สมบูรณ์ (complete period) อย่างน้อย 2-3 คาบ ซึ่งในความเป็นจริงแล้วความยาวของสัญญาณเสียงพูดมีผลต่อการคำนวณของ $R_n(\tau)$ เนื่องจากค่าของ $R_n(\tau)$ จะลดลงเรื่อยๆเมื่อ τ มีค่าเพิ่มมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

48893



รูปที่ 4.2 แสดงการหาค่าพิทช์จากการคำนวณออโตโครีเลชั่นฟังก์ชั่นของสัญญาณเสียงพูด

พิจารณารูป 4.2 แสดงตัวอย่างการคำนวณออโตโครีเลชั่นฟังก์ชั่นของสัญญาณเสียงพูดที่มีอัตราการแซมปลิ้งด้วยความถี่ 11.025 kHz โดยใช้สมการ 4.3 ด้วย N = 300 ในหนึ่งเฟรมและค่าการเลื่อนของเวลา (lag) เป็น $0 \le \tau \le 250$ จากรูปจะเห็นว่าจุดสูงสุด (peak) ยกเว้นที่ตำแหน่ง $\tau = 0$ เกิดที่ตำแหน่ง $\tau = 83$ นั่นคือสัญญาณมีคาบที่ระยะ 83 นาทีหรือมีค่าความถี่มูลฐานประมาณ 133 Hz (11.025 kHz / 83)

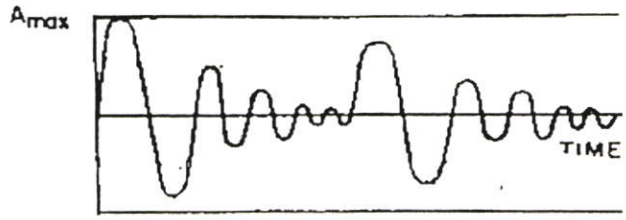
4.2.2.2 การกำจัดผลของโครงสร้างฟอร์แมนท์ด้วยวิธีเซนเตอร์คลิปปิ้ง

จากตัวอย่างรูป 4.3 จะเห็นว่าคลื่นของสัญญาณตอบสนองทางความถี่ที่เกิดในช่องทางเดินเสียงมียอดของกราฟจำนวนมาก ฉะนั้นถ้าหากนำไปหาตำแหน่งสูงสุดด้วยออโตโครีเลชั่นอาจจะผิดไปจากตำแหน่งคาบจริง เนื่องจากผลตอบสนองทางความถี่ในช่องทางเดินเสียงจะมีขนาดใหญ่กว่ายอดกราฟ ที่เกิดจากความเป็นคาบของแหล่งกำเนิดเสียง (vocal excitation) ซึ่งเหตุการณ์ลักษณะเช่นนี้จะทำให้การเลือกตำแหน่งยอดกราฟสูงที่สุดของออโตโครีเลชั่นฟังก์ชั่น เกิดการผิดพลาดด้วย

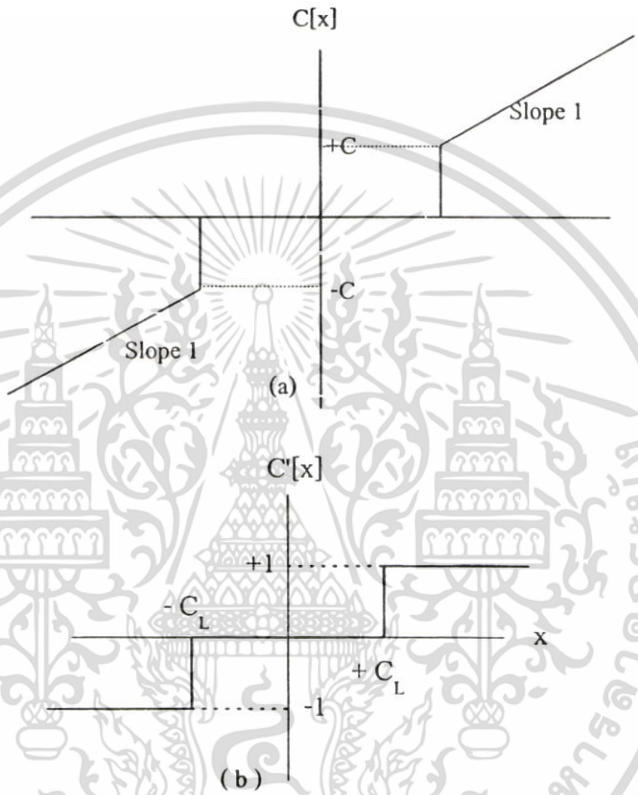
ดังนั้นเพื่อที่จะหลีกเลี่ยงปัญหานี้ จึงได้มีการเสนอกรรมวิธีเพื่อที่จะจัดการสัญญาณให้ความเป็นคาบของสัญญาณนี้เด่นชัดขึ้น โดยการขจัดลักษณะของสัญญาณที่จะทำให้เกิดความไข้วเขว (distracting) ออกไป เทคนิคนี้เรียกว่า การทำสเปกตรัมราบเรียบ (spectrum flatteners) ซึ่งมีอยู่หลายวิธี [9] เซนเตอร์คลิปปิ้งก็เป็นวิธีหนึ่งที่สะดวกและสามารถคำนวณได้จากสัญญาณโดยตรง ซึ่งถูกใช้การแปลงสัญญาณแบบไม่เป็นเชิงเส้นที่พัฒนามาจาก Sondhi ในปี 1968

จากวิธีดังกล่าวจะเห็นว่า วิธีเซนเตอร์กลิปปิ้งเป็นวิธีที่สะดวกในการทำให้สเปกตรัมราบเรียบ ซึ่งต่อมาได้มีการพัฒนาวิธีการนี้บนอุปกรณ์ดิจิตอล (Dobnoeski. 1975) โดยทำการปรับปรุงฟังก์ชั่นของเซนเตอร์คลิปปิ้งให้ง่ายต่อการคำนวณดังแสดงในรูป 4.4

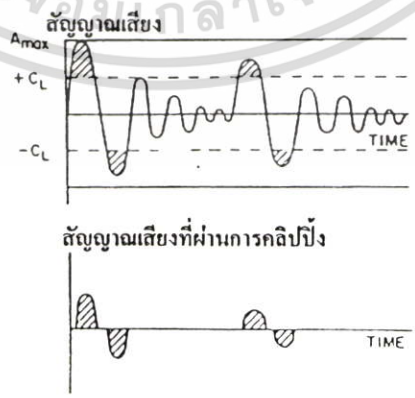
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 แสดงรูปทรงในแต่ละคาบของสัญญาณเสียงพูด



รูปที่ 4.4 (a , b) แสดงวิธีเซนเตอร์คลิปปิ้งของ Dobnoeski



รูปที่ 4.5 ตัวอย่างแสดงการคลิปสัญญาณเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลักการทำงานของเซนเตอร์คลิปปิ้ง คือสัญญาณเสียงจะถูกนำมาหาค่าแอมปลิจูดสูงสุด A_{max} เพื่อนำมากำหนดระดับในการคลิปสัญญาณ (clipping level : C_L) จากนั้นค่าของสัญญาณที่มีระดับต่ำกว่าระดับคลิปปิ้งจะถูกกำหนดให้มีค่าเป็นศูนย์ ส่วนสัญญาณที่มีระดับสูงกว่าระดับคลิปปิ้งจะยังคงไว้ด้วยระดับคลิปปิ้งดังรูป 4.5 สัญญาณยังคงความเป็นคาบของสัญญาณเดิมแต่ส่วนของสัญญาณที่เกิดฟอร์แมนท์ อันเนื่องมาจากการตอบสนองทางความถี่ภายในช่องทางเดินเสียงจะถูกกำจัดออกไป แต่ในการกำหนดระดับการคลิปสัญญาณจะต้องระมัดระวังว่าระดับที่กำหนดจะต้องไม่สูงเกินไปจนทำให้ข้อมูลสูญหาย [10] สัญญาณที่ผ่านการคลิปจะมีค่าเป็น +1 ถ้าขนาดของสัญญาณ $A_{S(m)} > C_L$, –1 ถ้า $A_{S(m)} < C_L$ และมีค่าเป็น 0 ถ้า $-C_L \le A_{s(m)} \le C_L$ ฟังก์ชั่นนี้เรียกว่า เซนเตอร์คลิปปิ้งแบบ 3 ระดับ (3-level center clipping) การกำหนดค่าในลักษณะนี้จะช่วยลดความซับซ้อนในการคำนวณออโตโครีเลชั่นฟังก์ชั่น เนื่องจากแต่ละพจน์ในสมการ (4.3) อยู่ในรูปของ $s(m)s(m+\tau)$ และค่า $s(m)$ จะมีค่าได้เพียง 3 ค่า คือ +1, 0, -1 เท่านั้น

4.2.2.3 การทำนอร์แมลไลซ์ออโตคอร์รีเลชั่น

รูปที่ 4.6 (a) แสดงรูปคลื่นสัญญาณเสียงมีช่วงแรกของของสัญญาณเสียงยังไม่คงที่ (non-stationary) เมื่อนำมาทำการคำนวณหาระยะคาบของพิทช์ก็อาจจะเกิดผิดพลาดจากตำแหน่งที่เลื่อนไปจากค่าจริง ดังรูปที่ 4.6(b) แสดงผลที่ได้จากการใช้ออโตคอร์รีเลชั่นฟังก์ชั่นในช่วงแรกของสัญญาณเสียง ซึ่งผลที่เกิดคือค่าขนาดของยอดคลื่นมีค่าผิดจากตำแหน่งเป็นจริง จึงทำให้การหาค่าความเป็นคาบอาจจะเกิดผิดพลาดได้ เพื่อให้การคำนวณหาระยะคาบมีความถูกต้องมากขึ้น จึงทำการทำนอร์แมลไลซ์[11] ดังแสดงในรูปที่ 4.6(c) คือผลของการทำออโตคอร์รีเลชั่น ด้วยการทำนอร์แมลไลซ์สำหรับสัญญาณต้นเสียงที่ยังไม่คงที่จากสมการที่ความคลาดเคลื่อนของพิทช์ (4.5a)

$$E_n = \sum_n \left[s(m) - s'(m) \right]^2 \qquad (4.5a)$$

โดย n คือช่วงของสัญญาณที่ใช้คำนวณ เพราะฉะนั้นเพื่อให้ได้ค่า E_n ต่ำที่สุดจะต้องมีเงื่อนว่า $\partial E(\tau,\beta)/\partial\beta = 0$ และจากสมการหาระยะยอดของคลื่นเสียง

$$E(\tau) = \frac{1}{N}\sum_{n=0}^{N-1}\left[s(n) - \beta s(n+\tau) \right]^2 \qquad (4.5b)$$

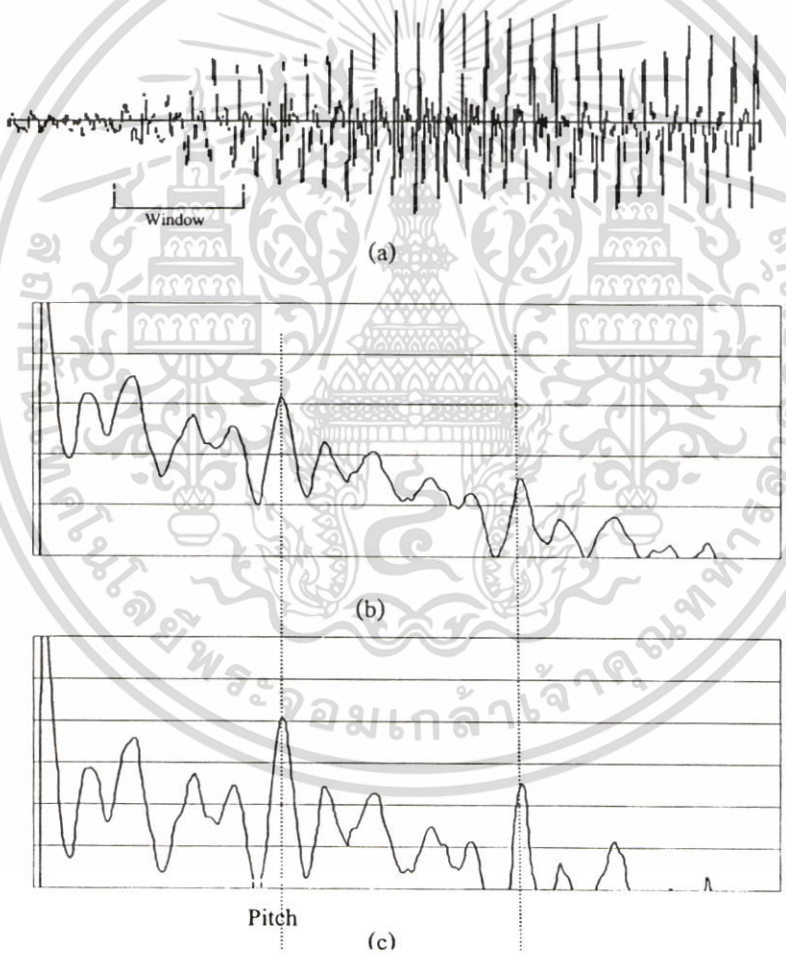
เมื่อ β คือค่าสัมประสิทธิ์ของอัตราขยายพิทช์ที่ทำการควบคุมการเปลี่ยนแปลงของระดับสัญญาณ หาได้จากสมการ(4.6)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\beta = \frac{\sum_{n=0}^{N-1} s(n)s(n+\tau)}{\sum_{n=0}^{N-1} s^2(n+\tau)} \tag{4.6}$$

แทนค่าของ β เข้าในสมการที่ (4.5)

$$E(\tau,\beta) = \sum_{n=0}^{N-1} s^2(n) - \frac{\left[\sum_{n=0}^{N-1} s(n)s(n+\tau)\right]^2}{\sum_{n=0}^{N-1} s^2(n+\tau)} \tag{4.7}$$



รูปที่ 4.6 แสดงการออโตคอร์รีเลชั่นของสัญญาณเสียง

(a). สัญญาณเสียงพูด

(b). สัญญาณที่ผ่านออโตคอร์รีเลชั่นที่ยังไม่ได้ทำการนอร์แมลไลซ์

(c). สัญญาณที่ผ่านออโตคอร์รีเลชั่นที่ยังทำการนอร์แมลไลซ์ แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการ (4.7) จะมีส่วนของออโตคอร์รีเลชั่นฟังก์ชั่นกำลังสองตามสมการนี้

$$R_n^2(\tau)=\frac{\left[\sum_{n=0}^{N-1}s(n)s(n+\tau)\right]^2}{\sum_{n=0}^{N-1}s^2(n+\tau)} \tag{4.8}$$

$$R_n(\tau)=\frac{\sum_{n=0}^{N-1}s(n)s(n+\tau)}{\sqrt{\sum_{n=0}^{N-1}s^2(n+\tau)}} \tag{4.9}$$

ดังที่ได้กล่าวทฤษฎีการวิเคราะห์มาทั้งหมดนั้น จะมีขบวนการหาค่าพิทช์มีตามขั้นตอนของโฟลว์ชาร์ตรูปที่ 4.7

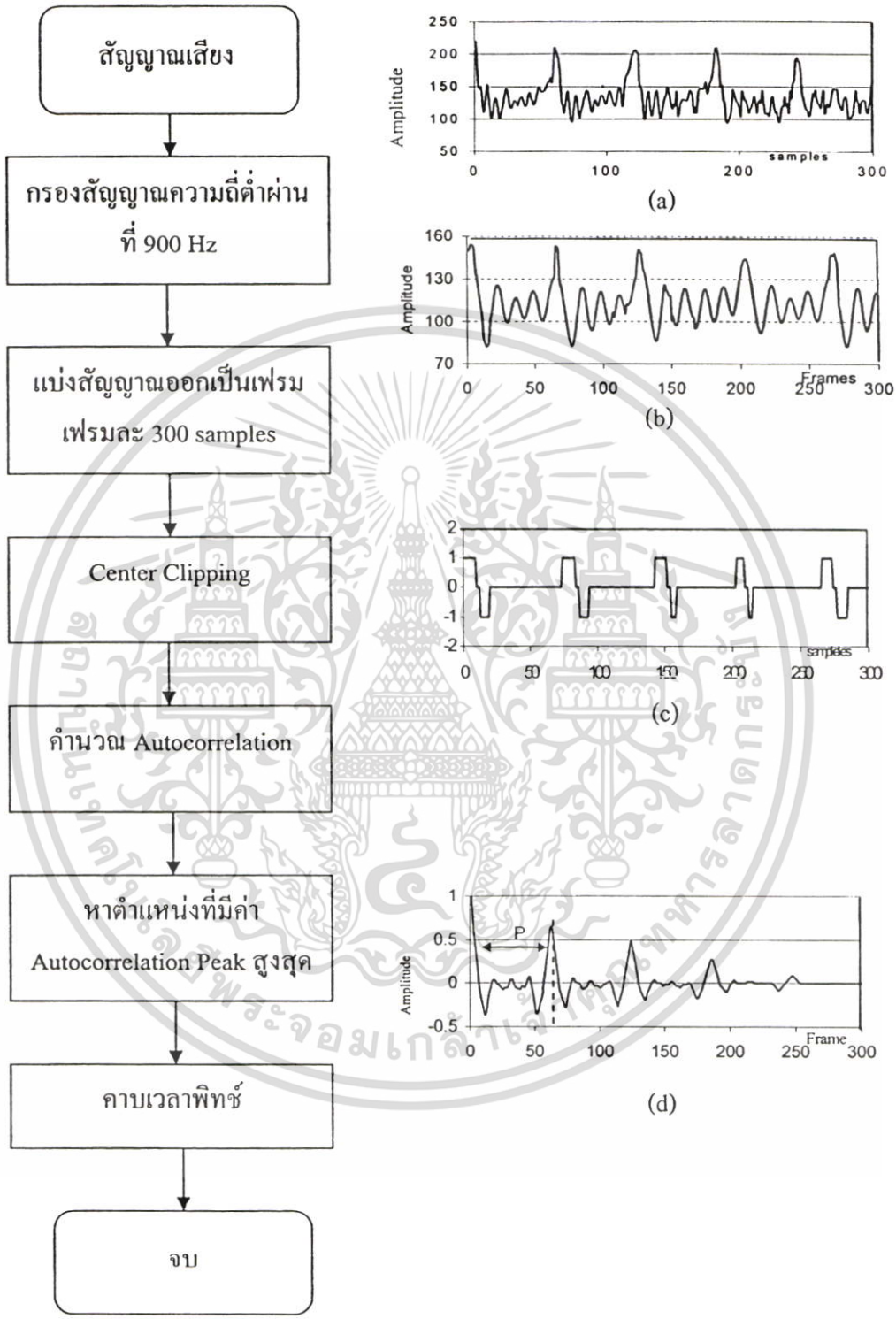
จากขบวนการตามโฟลว์ชาร์รูปที่ 4.7 จะได้ผลออกมาตามการทำงานคือรูปที่ 4.7(a) เป็นสัญญาณ ที่ได้จากการแซมปลิ้งที่ 11.025 kHz ซึ่งจะสังเกตุได้ว่าสัญญาณข้อมูลจะประกอบด้วยองค์ประกอบของความถี่จำนวนมาก อันเป็นผลมาจากการตอบสนองทางความถี่ภายในช่องทาง เดินเสียง ซึ่งความถี่เหล่านี้อาจมีผลทำให้การกำหนดตำแหน่งพิทช์ ในการคำนวณออโตโครีเลชั่นคลาดเคลื่อนไปจากตำแหน่งจริงได้ ดังนั้นเพื่อเป็นการกำจัดผลของความถี่เหล่านี้ออกไป จึงนำสัญญาณเสียงมาผ่านการกรองความถี่ต่ำผ่าน 900 Hz ซึ่งได้สัญญาณที่มีความราบเรียบมากขึ้นดังรูปที่ 4.7 (b) โดยเส้นประในรูปแสดงระดับในการคลิปสัญญาณ ซึ่งในวิทยานิพนธ์นี้ใช้ระดับการคลิปที่ 65 เปอร์เซนต์ [1] และสัญญาณที่ผ่านการคลิปแสดงได้ดังรูปที่ 4.7 (c) จากนั้นนำสัญญาณที่ผ่านการคลิปมาทำการคำนวณออโตโครีเลชั่นเพื่อหาพิทช์ จะได้สัญญาณที่มีลักษณะดังรูปที่ 4.7 (d) โดยระยะห่างระหว่าง R(0) กับจุดยอดที่สูงที่สุดถัดไปก็คือคาบพิทช์ ซึ่งจากรูปได้ตำแหน่งที่ 65 และสามารถหาค่าความถี่มูลฐานได้ตามสมการที่ (4.10) จะมีค่าเท่ากับ 170 Hz

4.3 การหาค่าความถี่มูลฐานของสัญญาณเสียงพูด

จากค่าคาบเวลาพิทช์ที่ได้นี้สามารถนำมาหาค่าความถี่มูลฐาน F_o ได้จากความสัมพันธ์ที่ (4.10) และผลจากค่าความถี่มูลฐานทำให้เรารู้ถึงคุณลักษณะพิเศษของเส้นเสียง ในแต่ละช่วงได้อย่างครบถ้วนดังแสดงในรูปที่ 4.8 คือตัวอย่างความถี่มูลฐานวรรณยุกต์ทั้ง 5 ที่มีค่าเปลี่ยนแปลงตามเวลาของเส้นเสียง

$$F_o=\frac{F_s}{P} \tag{4.10}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



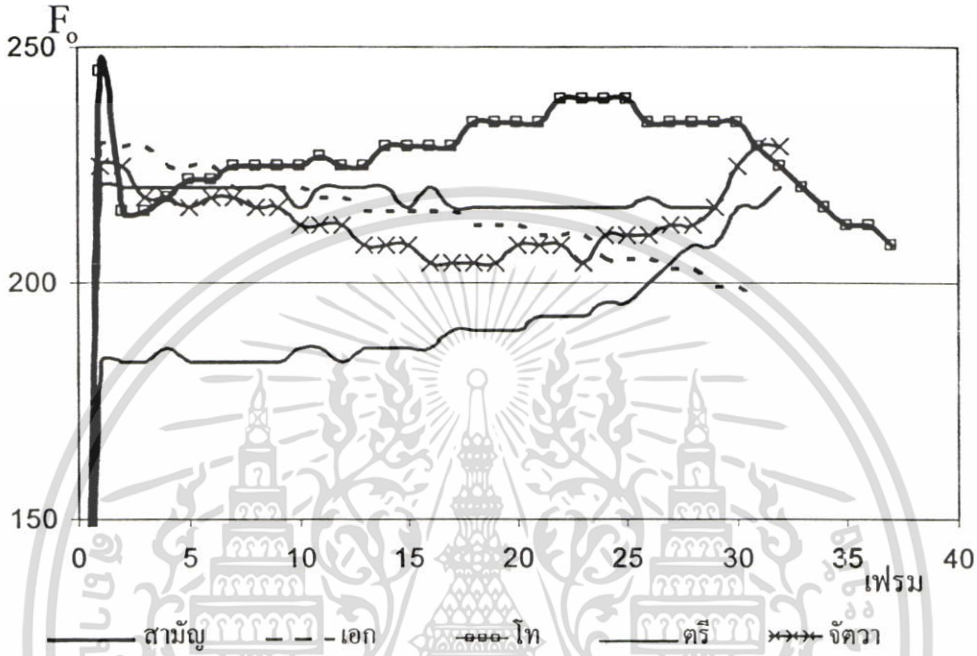
รูปที่ 4.7 โฟลว์ชาร์ตของขบวนการหาค่าพิทช์ เทียบกับผลที่ได้รับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ F_o = ความถี่มูลฐาน (Hz)

F_S = ความถี่ที่ใช้ในการสุ่มสัญญาณ

P = คาบเวลา พิทช์



รูปที่ 4.8 แสดงการเปลี่ยนแปลงองความถี่มูลฐานในแต่ละเวลาของเสียงวรรณยุกต์

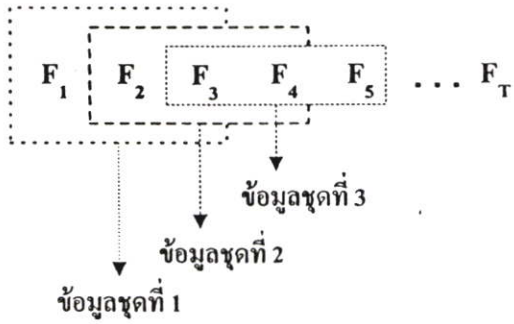
4.4 การเตรียมข้อมูลเพื่อสร้างแบบจำลอง

4.4.1 การปรับปรุงความต่อเนื่องของข้อมูลด้วยวิธีการกรองค่ากลาง (Median Filtering)

สัญญาณเสียงที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้ว อาจมีความไม่ต่อเนื่องของลำดับความถี่เกิดขึ้น เนื่องจากความไม่ต่อเนื่องของสัญญาณเสียงในช่วงต้นของการออกเสียงพูดและในการคำนวณแสดงในรูปที่ 4.8 ดังนั้นขั้นตอนแรกของการเตรียมข้อมูลก็คือการนำสัญญาณเสียงที่ผ่านขั้นตอนในการหาค่าความถี่มูลฐานแล้วมาผ่านตัวกรองค่ากลาง เพื่อปรับปรุงให้ข้อมูลมีความต่อเนื่องเพิ่มขึ้น [12] โดยลำดับของความถี่มูลฐานซึ่งเป็นข้อมูลอินพุทจะอยู่ในรูปของข้อมูล 1 มิติ ขนาด [1xT] เมื่อ T คือจำนวนเฟรมของสัญญาณเสียง

ขั้นตอนการทำงานของการกรองค่ากลาง คือทำการจัดเรียงค่าความถี่มูลฐานออกเป็นชุดๆ ข้อมูล โดยในแต่ละชุดข้อมูลประกอบด้วยค่าความถี่ 3 ค่า และกำหนดให้มีการเลื่อนของชุดข้อมูลแสดงได้ดังรูป 4.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.9 การจัดแบ่งความถี่มูลฐานออกเป็นชุดข้อมูล

เมื่อ F_1 = ค่าความถี่มูลฐานของเฟรมที่ 1

F_2 = ค่าความถี่มูลฐานของเฟรมที่ 2

F_T = ค่าความถี่มูลฐานของเฟรมสุดท้าย

จากนั้นนำค่าความถี่ทั้ง 3 ค่า ในแต่ละชุดข้อมูลมาจัดเรียงใหม่ตามความสัมพันธ์

$$a \leq b \leq c \tag{4.11}$$

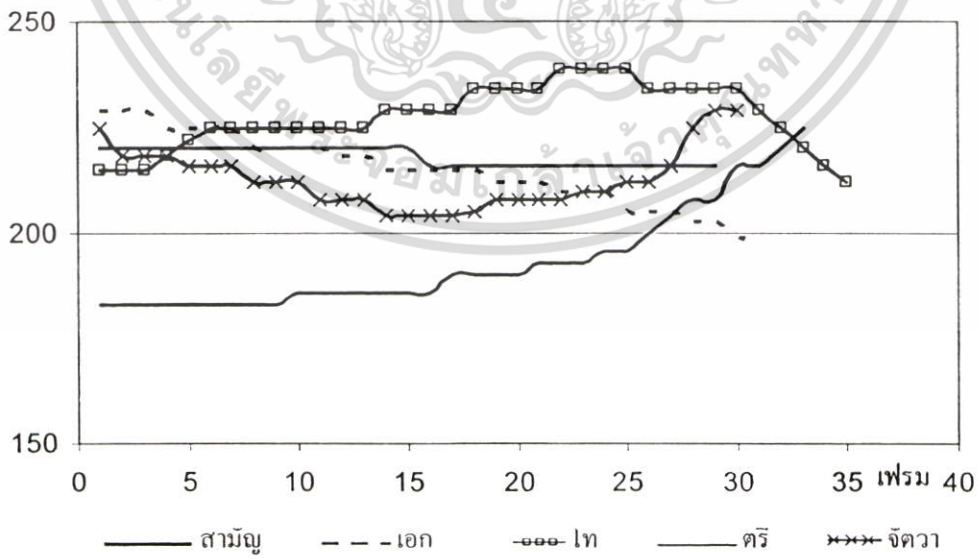
โดยที่

a = ความถี่ F_O ที่มีค่าน้อยที่สุดของแต่ละชุดข้อมูล

b = ความถี่ F_O ที่มีค่าอยู่ระหว่างกลาง

c = ความถี่ F_O ที่มีค่ามากที่สุดของแต่ละชุดข้อมูล

แล้วนำความถี่ค่ากลาง b จากสมการ 4.11 ที่ได้จากชุดข้อมูลแต่ละชุดมาจัดเรียงใหม่จะได้ความถี่มูลฐานชุดใหม่ที่ผ่านกระบวนการกรองค่ากลางแล้วผลแสดงในรูปที่ 4.10



รูปที่ 4.10 แสดงตัวอย่างความถี่มูลฐานที่ผ่านวิธีการกรองค่ากลาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนสุดท้ายของการเตรียมข้อมูลก็คือ การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน จากข้อเท็จจริงที่ว่าระดับความถี่มูลฐานของเสียงชายและหญิงมีความแตกต่างกัน ซึ่งโดยเฉลี่ยแล้วผู้ชายมีค่าความถี่มูลฐานจะมีค่าอยู่ในช่วง 80-160 Hz และ ผู้หญิงมีค่าความถี่มูลฐานในช่วง 160-400 Hz

4.4.2 การควอนไตซ์ทิศทางการเปลี่ยนแปลงของค่าความถี่มูลฐาน

จากรูป 4.10 จะสังเกตเห็นว่าลักษณะการเปลี่ยนแปลงของค่าความถี่มูลฐานในแต่ละระดับเสียงวรรณยุกต์จะมีรูปแบบการเปลี่ยนแปลงที่มีลักษณะเฉพาะ โดยไม่ขึ้นกับผู้ออกเสียงว่าเป็นเพศใด ดังนั้นในวิทยานิพนธ์นี้จึงได้ดึงเอาลักษณะเด่นนี้มาใช้ในการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ โดยทำการจัดกลุ่มค่าความถี่มูลฐานออกเป็น 3 ระดับตามแนวทางการเปลี่ยนแปลงของความถี่ (ΔF) ที่เพิ่มขึ้นหรือลดลงตามเวลา

โดย

$$\Delta F_t = F_{t+1} - F_t \tag{4.12}$$

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

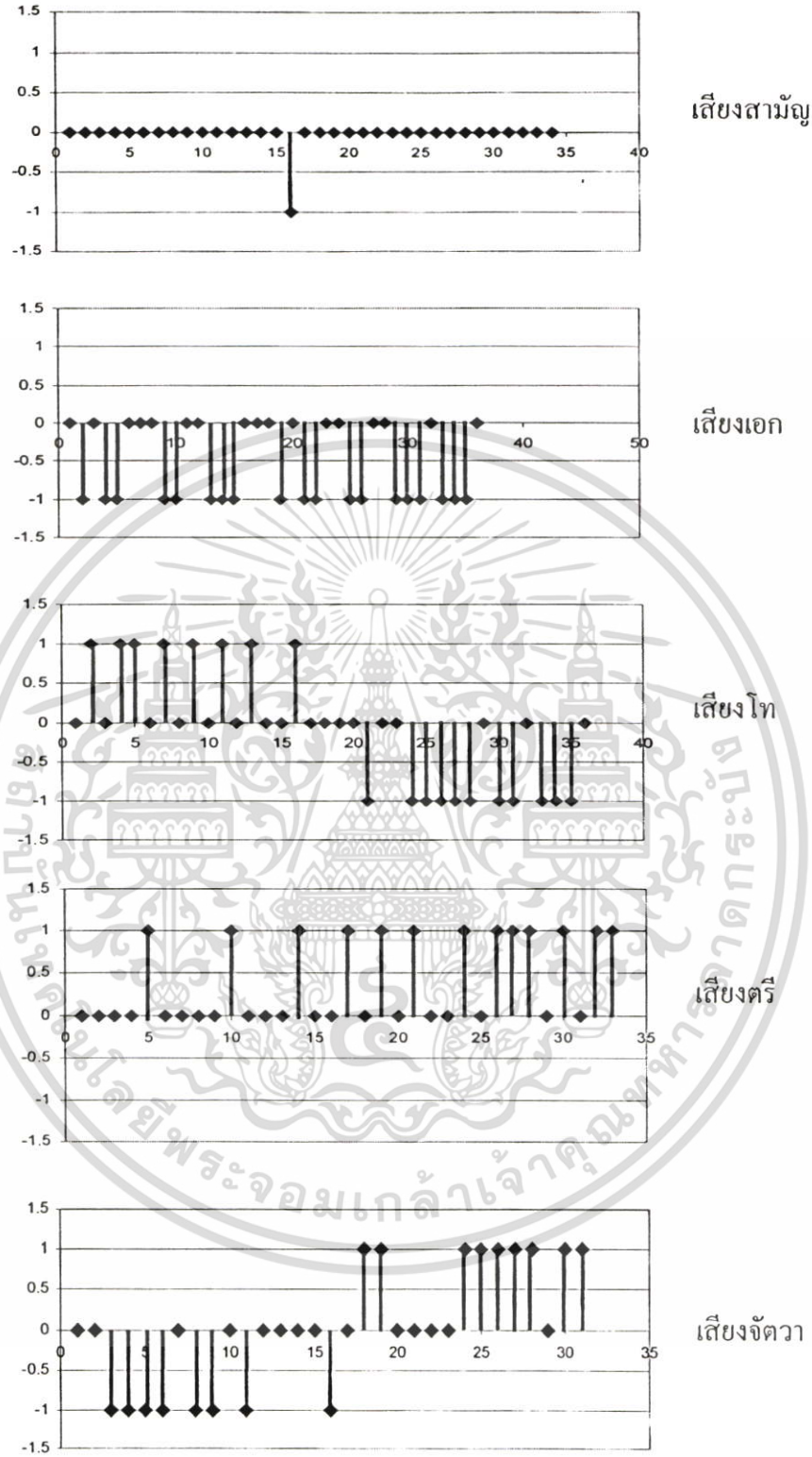
จากนั้นทำการควอนไตซ์ ΔF โดยแบ่งออกเป็น 3 ระดับตามทิศทางการเปลี่ยนแปลงของความถี่มูลฐาน โดยกำหนดให้

$$V_t = \begin{cases} 1 \; ; \; \Delta F_t > 0 \\ 0 \; ; \; \Delta F_t = 0 \\ -1 \; ; \; \Delta F_t < 0 \end{cases} \tag{4.13}$$

ค่า $V_t = \{-1, 0, 1\}$ จะถูกนำไปใช้เป็นข้อมูลฝึกสอน (training) เพื่อใช้ในการสร้างแบบจำลองอ้างอิงของเสียงวรรณยุกต์ต่อไป ซึ่งจะเห็นว่าการควอนไตซ์ความถี่ออกเป็น 3 ระดับนี้นอกจากจะกำจัดข้อจำกัดของความถี่มูลฐานที่แตกต่างกันระหว่างชายและหญิงแล้ว ยังช่วยลดเนื้อที่ของหน่วยความจำในการจัดเก็บข้อมูล และทำให้การคำนวณทำได้เร็วขึ้นเมื่อเทียบกับการใช้ช่วงความถี่มูลฐานทั้งหมดมาสร้างแบบจำลอง

เมื่อนำค่าความถี่มูลฐานในรูป 4.10 มาทำการจัดระดับค่าการเปลี่ยนแปลงความถี่ออกเป็น 3 ระดับจะแสดงได้ดังรูปที่ 4.11 และเมื่อพิจารณาตามแนวทางการเปลี่ยนแปลงค่าความถี่มูลฐานของเสียงค่าการควอนไตซ์จะมีลักษณะเฉพาะตามเสียงของวรรณยุกต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 แสดงการจัดแบ่งค่าความถี่มูลฐานของวรรณยุกต์ออกเป็น 3 ระดับตามทิศทางการเปลี่ยนแปลงตามเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 การสร้างแบบจำลองการรู้จำด้วยวิธี Hidden Markov Model

แบบจำลองมาร์คอฟเป็นแบบจำลองทางสถิติ ซึ่งพัฒนามาเพื่อแบ่งกลุ่มของอนุกรมทางเวลาหรือสัญญาณที่ไม่คงที่ นั่นคือใช้สำหรับจัดกลุ่มของสัญญาณที่ไม่รู้จัก (Unknown signal) ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่งแบบจำลองมาร์คอฟได้ถูกนำมาประยุกต์ใช้ในการรู้จำเสียงพูด [14] และเป็นวิธีการที่วิทยานิพนธ์นี้เลือกใช้

4.5.1 ส่วนประกอบของแบบจำลองมาร์คอฟ

พารามิเตอร์สำคัญที่เกี่ยวข้องในการสร้างแบบจำลองอ้างอิง ที่ต้องรู้จักได้แก่

4.5.1.1 T คือ ความยาวของลำดับข้อมูลที่ได้จากการควอนไตซ์ค่าความถี่มูลฐาน ซึ่งจะใช้เป็นข้อมูลอินพุทในส่วนของ HMM โดยต่อไปจะเรียกแทนว่า "ลำดับของค่าปรากฏ"(Observation sequence) ซึ่งมีขนาดความยาวของลำดับ เท่ากับจำนวนเฟรมทั้งหมดในเสียงแต่ละเสียง

4.5.1.2 N คือจำนวนสเตทในแบบจำลอง ถ้ากำหนดให้เซ็ตของสเตทเป็น $\{1, 2, ..., N\}$ จะสามารถแทนสเตทที่เปลี่ยนไปตามเวลา t ด้วยเซ็ตของ $Q = \{q_1, q_2, q_3 ....., q_N\}$

4.5.1.3 M คือจำนวนของค่าปรากฏที่สามารถเป็นไปได้ต่อหนึ่งสเตท แทนสัญลักษณ์ด้วย $V = \{v_1, v_2, v_3, ...v_M\}$ ซึ่งจากการจัดระดับของการเปลี่ยนแปลงของความถี่ (ΔF_i) ออกเป็น 3 ระดับ จะได้เซ็ตของค่าปรากฏที่สามารถเป็นไปได้ในแต่ละสเตทมีค่าเป็น $V = \{-1, 0, 1\}$

4.5.1.4 ค่าความน่าจะเป็นในการย้ายสเตท $A = \{a_{ij}\}$ โดย a_{ij} แทนการย้ายสเตทจาก i ไป j เมื่อ

$$a_{ij} = P[q_t = j | q_{t-1} = i] \quad ; 1 \le i, j \ge N \tag{4.14}$$

4.5.1.5 การกระจายความน่าจะเป็น ของค่าปรากฏที่สามารถเป็นไปได้ภายในสเตท B ซึ่ง $B = \{b_j(\tau)\}$

โดยที่ $b_j(\tau) = P[\ v_\tau$ ที่เวลา $t | q_j$ ที่เวลา $t\ ]; \ 1 \le \tau \le M$ (4.15)

ตามการกระจายสัญลักษณ์ในสเตท j เมื่อ $j = 1, 2, ..., N$

4.5.1.6 ค่าความน่าจะเป็นของการเป็นสเตทเริ่มต้น

$$\pi_i = P[\ q_i \text{ที่เวลา } t = 1\ ] \quad ; 1 \le i \le N \tag{4.16}$$

จะเห็นว่า Hidden Markov Model ต้องการพารามิเตอร์ของแบบจำลองคือ N, M และ กลุ่มของความน่าจะเป็น A, B, π ดังนั้นในการแสดงเซ็ตของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองอ้างอิงจะแทนด้วยสัญลักษณ์

$$\lambda = (A, B, \pi) \tag{4.17}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.2 คุณสมบัติการย้ายสเตทของ HMM

การสร้างแบบจำลองนี้ใช้ Left-Right Model ซึ่งเป็นแบบจำลองแบบที่เหมาะกับสัญญาณที่มีลักษณะเปลี่ยนแปลงตามเวลาอย่างต่อเนื่องเช่น เสียงพูด โดยมีคุณสมบัติของสัมประสิทธิ์ในการย้ายสเตทดังนี้

การย้ายสเตทจะย้ายจากซ้ายไปขวาหมายความว่า $a_{ij} = 0 , j < i$ แสดงว่าจะไม่มีการย้ายสเตทไปยังสเตทที่ต่ำกว่าสเตทปัจจุบันและนอกจากนี้ก็ยังมีความน่าจะเป็นของสเตทเริ่มต้น

$$\pi_i = \begin{cases} 0 & , i \neq 1 \\ 1 & , i = 1 \end{cases}$$

ลำดับของสเตทจะต้องเริ่มที่สเตทที่ 1 เสมอ และ Left-Right Model นี้มักมีกฎบังคับการย้ายสเตทเพื่อไม่ให้มีการเปลี่ยนแปลงดัชนีของสเตทมากนักกล่าวคือ

$$a_{ij} = 0 \quad , j > i + \Delta i$$

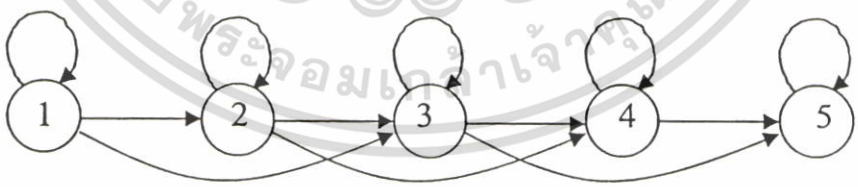
ดังรูปที่ 4.12 ค่าของ $\Delta i = 2$ คือจะไม่มีการย้ายข้ามสเตทไปเกิน 2 สเตท และมีเมตริกซ์ในการย้ายสเตทเป็น

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 \\ 0 & a_{22} & a_{23} & a_{24} & 0 \\ 0 & 0 & a_{33} & a_{34} & a_{35} \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

จะเห็นว่าสเตทสุดท้าย สัมประสิทธิ์การย้ายสเตทจะเป็น

$$a_{NN} = 1$$

$$a_{Ni} = 0 \quad , i < N$$



รูปที่ 4.12 แบบจำลอง Left-Right Model ของ HMM

4.5.3 ปัญหาพื้นฐานของแบบจำลอง มาร์คอฟ

ปัญหาของ HMM มี 3 ข้อ ซึ่งต้องใช้อัลกอริธึมวิธีต่างๆในการคำนวณเพื่อแก้ปัญหานี้

ปัญหาที่ 1 เมื่อมีลำดับของค่าปรากฏ $O = \{O_1 O_2 O_3 \ldots O_T\}$ และมีแบบจำลอง $\lambda = (A, B, \pi)$ จะคำนวณหาค่าความน่าจะเป็น $P(O|\lambda)$ ของลำดับค่าปรากฏนั้นได้อย่างไร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปัญหาที่ 2 เมื่อมีลำดับของค่าปรากฏ $O = \{O_1 O_2 O_3 \ldots O_T\}$และแบบจำลอง $\lambda = (A,B,\pi)$ จะคำนวณหาลำดับสเตท $q = \{ q_1 q_2 q_3 \ldots q_T\}$ ที่เหมาะสมกับลำดับค่าปรากฏนั้นได้อย่างไร

ปัญหาที่ 3 เราจะปรับพารามิเตอร์ของแบบจำลอง $\lambda = (A,B,\pi)$ เพื่อให้ได้ค่า $P(O|\lambda)$ สูงสุดได้อย่างไร

4.5.4 การคำนวณเพื่อแก้ปัญหาของ HMM

4.5.4.1 การแก้ปัญหาที่ 1

เป็นการคำนวณหาว่าแบบจำลอง λ ใดๆ มีโอกาสจะให้ค่าลำดับเป็นไปตามลำดับของค่าปรากฏนั้นด้วยค่าของความน่าจะเป็นมากหรือน้อยเท่าใด การแก้ปัญหาสามารถทำได้โดยระบุสเตทให้กับลำดับของค่าปรากฏซึ่งยาว T (โดยที่ค่าปรากฏหนึ่งตัวมีความเป็นไปได้ที่จะอยู่ในสเตทได้ N สเตท) ซึ่งสามารถเป็นไปได้ถึง N^T แบบ โดยให้สเตทต่างๆแทนด้วย

$$q = q_1\ q_2\ q_3 \cdots q_T \tag{4.18}$$

เมื่อ q_1 เป็นสเตทเริ่มต้นที่เวลา $t = 1$ ความน่าจะเป็นของลำดับของค่าปรากฏ O ที่กำหนดคือ

$$P(O|q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda) \tag{4.19}$$

ความน่าจะเป็นในการเกิดค่าปรากฏคือ

$$P(O|q,\lambda) = b_{q_1}(O_1)\cdot b_{q_2}(O_2)\cdot\ldots\cdot b_{q_T}(O_T) \tag{4.20}$$

และความน่าจะเป็นในการย้ายข้ามสเตท q จะเป็น

$$P(q|\lambda) = \pi_{q_1}\cdot a_{q_1q_2}\cdot a_{q_2q_3}\cdot a_{q_3q_4}\cdot\ldots\cdot a_{q_{T-1}q_T} \tag{4.21}$$

ดังนั้นเมื่อนำความน่าจะเป็นของการเกิดค่าปรากฏ O และความน่าจะเป็นในการย้ายสเตท q มารวมกัน ซึ่งนั่นก็คือความน่าจะเป็นที่ O และ q จะเกิดขึ้นพร้อมกันจะได้

$$\begin{aligned} P(O,q|\lambda) &= P(O|q,\lambda)P(q|\lambda) \\ &= \left(b_{q_1}O_1\cdot b_{q_2}O_2\cdot\ldots\cdot b_{q_T}O_T\right)\left(\pi_{q_1}\cdot a_{q_1q_2}\cdot a_{q_2q_3}\cdot\ldots\cdot a_{q_{T-1}q_T}\right) \end{aligned} \tag{4.22}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยความน่าจะเป็นของ O ได้มาจากผลรวมของความน่าจะเป็น O และ q เกิดขึ้นพร้อมกัน โดยคิดจากทุกสเตท q ที่จะเป็นไปได้ดังนี้

$$P(O|\lambda) = \sum_{allq} P(O|q,\lambda)P(q|\lambda)$$

$$= \left(\sum_{q_1 q_2 \cdots q_T} \pi_{q_1} \cdot b_{q_1}(O_1) a_{q_1 q_2} \cdot b_{q_2}(O_2) \cdot \ldots \cdot a_{q_{T-1} q_T} b_T O_T \right) \quad (4.23)$$

ที่เวลาเริ่มต้น (t = 1) เราจะอยู่ที่สเตท q_1 ด้วยค่าความน่าจะเป็น π_{q_1} และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏ O_1 ที่สเตทนี้ด้วย $b_{q_1}(O_1)$ ที่เวลา เพิ่มขึ้นจาก $t \longrightarrow t+1$ (t=2) เราแทน การย้ายสเตทจากสเตท q_1 ไปยัง q_2 ด้วยค่าความน่าจะเป็น $a_{q_1 q_2}$ และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏเป็น O_2 ด้วยค่าความน่าจะเป็น $b_{q_2}(O_2)$ จนกระทั่ง ที่เวลา T เราแทนการย้ายสเตทจากสเตท q_{T-1} ไปยัง q_T ด้วยค่าความน่าจะเป็น $a_{q_{T-1} q_T}$ และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏเป็น O_T ด้วยค่าความน่าจะเป็น $b_{q_T}(O_T)$

จะเห็นว่าสมการนี้มีการคำนวณที่ยุ่งยากเนื่องจากมีการคูณกันเป็นจำนวนมากในรูปของลำดับ $2T \times N^T$ ดังนั้นจึงมีการคิดหาวิธีมาช่วย ซึ่งแบ่งออกเป็น

4.5.4.1.1 กระบวนการไปข้างหน้า (Forward Procedure)

นิยาม ของ $\alpha_t(i)$ = Forward variable

$$\alpha_t(i) = P(O_1 O_2 \ldots O_T, q_t = i | \lambda) \quad (4.24)$$

คือความน่าจะเป็นของการเกิดลำดับของค่าปรากฏ $O_1 O_2 \ldots O_T$ และอยู่ที่สเตท q_i ณ เวลา t โดยมีแบบจำลองเป็น λ เราสามารถหา $\alpha_t(i)$ ได้ดังนี้

4.5.4.1.1.1 การเริ่มต้น (Initialization)

$$\alpha_1(i) = \pi_i b_i O_1 \quad ; \quad 1 \le i \le N \quad (4.25)$$

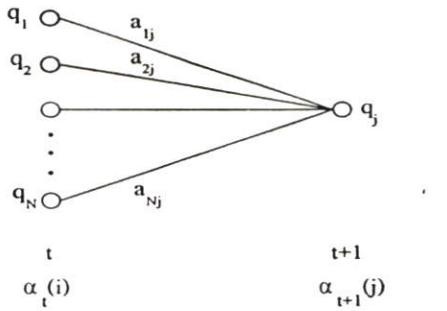
เริ่มด้วยการกำหนดความน่าจะเป็นไปข้างหน้าซึ่งเป็นความน่าจะเป็นร่วมของสเตท i และมีเหตุการณ์เริ่มต้นเป็น O_1

4.7.1.1.1.2 การเหนี่ยวนำ (Induction)

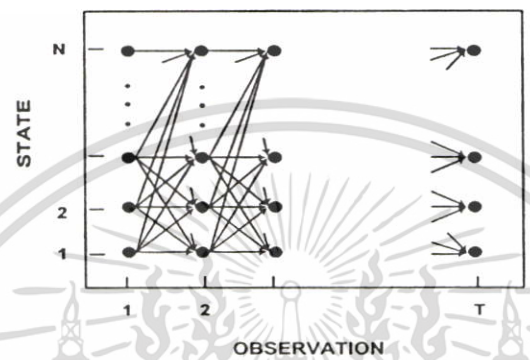
$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad ; \quad \begin{matrix} 1 \le i \le T-1 \\ 1 \le j \le N \end{matrix} \quad (4.26)$$

หมายความว่าสเตท j ที่เวลา t+1 สามารถมาได้จากสเตทก่อนหน้านี้ซึ่งเป็นไปได้ถึง N สเตท (สเตท i ณ เวลา t โดยที่) ดังรูป 4.13 (a)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(a)



(b)

รูปที่ 4.13 (a,b) กระบวนการไปข้างหน้า

จากรูป 4.13 (b) แสดงให้เห็นว่าการคำนวณค่าความน่าจะเป็นแบบไปข้างหน้า (Forward probability) มีโครงสร้างการคำนวณคล้ายๆกับลักษณะของโครงผลึก และเนื่องจากมีจำนวนสเตทเพียง N สเตท (แทนด้วยจำนวนโหนดในแต่ละช่วงเวลา t ใดๆในโครงผลึก) จำนวนลำดับสเตทจะถูกจัดเรียงลงในโหนดเหล่านี้ โดยในเวลา t = 1 จะทำการคำนวณค่าของ $\alpha_t(i)$ ในทุกๆสเตท $1 \le i \le N$ และที่เวลา t = 2, 3, ..., T จะทำการคำนวณค่าของ $\alpha_t(j)$ ในทุกๆสเตท, $1 \le j \le N$ โดยในแต่ละค่าจะทำการคำนวณมาจาก $\alpha_{t-1}(i)$ จำนวน N ค่าก่อนหน้านี้

4.5.1.1.3 การสิ้นสุด (Termination) คือการหา $P(O|\lambda)$ ได้จากผลรวมของ $\alpha_t(i)$ จากทุกๆสเตท

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \quad ; \quad 1 \le i \le N \tag{4.27}$$

4.5.1.1.2 กระบวนการย้อนกลับ (Backward Procedure)

นิยาม $\beta_t(i)$ = Backward variable

$$\beta_t(i) = P(O_{t+1}O_{t+2}O_{t+3}.....O_T \,|\, i_t = q_i, \lambda) \tag{4.28}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คือ ความน่าจะเป็นของลำดับค่าปรากฏส่วนหลังจากเวลา t+1 ไปจนจบโดยกำหนดว่าต้องอยู่ที่สเตท i ที่เวลา t และมีแบบจำลองเป็น เราจะคำนวณหาค่าได้ดังนี้

4.5.1.1.2.1 การเริ่มต้น (Initialization)

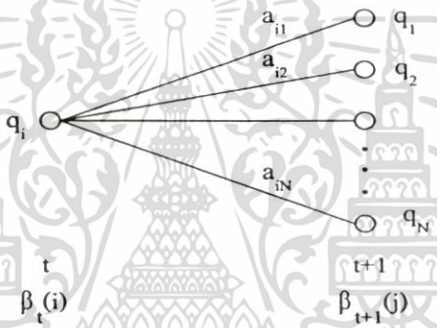
$$\beta_T(i) = 1 \quad ; \quad 1 \le i \le N \quad . \tag{4.29}$$

4.7.1.1.2.2 การเหนี่ยวนำ (Induction)

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \tag{4.30}$$

เมื่อ $t = T-1, T-2, \ldots, 1 \quad , 1 \le i \le N$

จากรูป 4.14 เพื่อที่จะให้ค่าปรากฏอยู่ที่สเตท i ณ เวลา t โดยคาดคะเนจากลำดับค่าปรากฏจากเวลา t+1 ซึ่งเราจะต้องพิจารณาจากสเตท j ที่เป็นไปได้ทั้งหมด โดยจะขึ้นอยู่กับค่า a_{ij} และ $b_j(O_{t+1})$



รูปที่ 4.14 กระบวนการย้อนกลับ

4.5.4.2 การแก้ปัญหาที่ 2

ใช้ วิเทอร์บีอัลกอริธึม (Viterbi Algorithm) เพื่อที่จะหาลำดับสเตทที่ดีที่สุด, $q = (q_1\ q_2\ q_3 \ldots q_T)$ ให้กับลำดับของค่าปรากฏ $O = \{O_1 O_2 O_3 \ldots O_T\}$ ที่มีอยู่ โดยนิยามให้

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P[q_1\ q_2 \ldots q_{t-1}, q_t = i, O_1\ O_2 \ldots O_t \mid \lambda] \tag{4.31}$$

เมื่อ $\delta_t(i)$ คือ ความน่าจะเป็นสูงสุด (hightest probability) ของเส้นทาง (path) ซึ่งจะหาได้จากค่าความน่าจะเป็นสูงสุด เมื่อเทียบกับสเตททุกสเตทในการให้ค่าปรากฏเป็นไปตามค่าปรากฏที่กำหนดให้ ที่ขณะเวลา t ใดๆ และจากการอาศัยคุณสมบัติของการเหนี่ยวนำจะได้

$$\delta_{t+1}(j) = \left[\max \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1}) \tag{4.32}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยกำหนดให้เป็นอาเรย์ที่เก็บตำแหน่งของสเตทที่ให้ค่าความน่าจะเป็นสูงสุด ที่คำนวณได้ในแต่ละเวลา t และแต่ละลำดับ j ซึ่งจะสามารถหาลำดับสเตทที่ดีที่สุดได้โดยใช้กระบวนการต่อไปนี้

4.5.4.21 การเริ่มต้น (Initialization)

$$\delta_1(i) = \pi_i b_i(O_1) \quad ; \qquad 1 \le i \le N \tag{4.33}$$

$$\psi_1(i) = 0 \tag{4.34}$$

4.5.4.2.2 การย้อนกลับ (Recursion)

$$\delta_t(j) = \left[\max_{1 \le i \le N} \delta_{t-1}(i) a_{ij} b_j(O_t) \right] \quad ; \quad \begin{matrix} 2 \le t \le T \\ 1 \le j \le N \end{matrix} \tag{4.35}$$

$$\psi_t(j) = \left[\arg \max_{1 \le i \le N} \delta_{t-1}(i) a_{ij} b_j \right] \quad ; \quad \begin{matrix} 2 \le t \le T \\ 1 \le j \le N \end{matrix} \tag{4.36}$$

4.5.4.2.3 การสิ้นสุด (Termination)

$$P^* = \max_{1 \le i \le N} [\delta_T(i)] \tag{4.37}$$

$$q_T = \arg \max_{1 \le i \le N} [\delta_T(i)] \tag{4.38}$$

4.5.4.2.4 เส้นทางเดินย้อนกลับ (Backtracking)

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad ; \; t = T\text{-}1, T\text{-}2, \ldots, 1 \tag{4.39}$$

4.5.4.3 การแก้ปัญหาที่ 3

จากที่กล่าวมาแล้วข้างต้นว่าแบบจำลองเสียงจะแทนด้วยค่าพารามิเตอร์ $\lambda = (A, B, \pi)$ ดังนั้นเมื่อมีลำดับของค่าปรากฏจำนวนหนึ่ง เพื่อที่จะนำมาสร้างแบบจำลองอ้างอิง จะต้องทำการคำนวณหาค่าพารามิเตอร์ ของแบบจำลองซึ่งจะอยู่ในรูปของค่าความน่าจะเป็น โดยวิธีที่เลือกใช้ก็คือวิธีของ บาม-เวลส์ (Baum-Welch method) หรือเรียกอีกชื่อหนึ่งว่า EM (Expectation-Maximization method) โดยมี

นิยาม 1. คือ

$$\gamma_t(i) = P(q_t = i | O, \lambda) \tag{4.40}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ $\gamma_t(i)$ คือ ค่าความน่าจะเป็นที่จะอยู่ที่สเตท i ที่ขณะเวลา t โดยให้ลำดับของค่าปรากฏด้วยโมเดล λ โดยที่กำหนดลำดับของค่าปรากฏให้ สามารถแสดงค่า $\gamma_t(i)$ ได้ดังนี้

$$\gamma_t(i) = P(q_t = i \mid O, \lambda)$$

$$= \frac{P(O, q_t = i|\lambda)}{P(O|\lambda)}$$

$$= \frac{P(O, q_t = i|\lambda)}{\sum_{i=1}^{N} P(O, q_t = i|\lambda)} \tag{4.41}$$

เนื่องจาก $P(O, q_t = i|\lambda)$ มีค่าเท่ากับ $\alpha_t(i)\beta_t(i)$ ดังนั้นสามารถเขียน $\gamma_t(i)$ ได้เป็น

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{4.42}$$

โดย $\alpha_t(i)$ เริ่มจาก $O_1, O_2, \ldots, O_t$ จนถึงสเตท i ที่เวลา t

โดย $\beta_t(i)$ เริ่มจาก $O_{t+1}, O_{t+2}, \ldots, O_T$ จนถึงสเตท $q_t = i$ ที่เวลา t

นิยาม 2.

$$\varepsilon_t(i,j) = P(q_t = i, q_{t+1} = j|O, \lambda) \tag{4.43}$$

เมื่อ $\varepsilon_t(i,j)$ คือความน่าจะเป็นที่จะอยู่ที่สเตท i ที่เวลา t และสเตท j ที่เวลา t+1 เมื่อกำหนดแบบจำลองและลำดับค่าปรากฏให้

จากรูปที่ 4.15 แสดง ลำดับการคำนวณการเกิดค่าปรากฏร่วม ซึ่งระบบจะอยู่ในสเตท i ที่เวลา t และอยู่ที่ สเตท j ที่เวลา t+1 โดย $\alpha_t(i)$ เริ่มจากเวลา t = 1 ที่ค่าปรากฏแรก จนถึงสเตท q_i ที่เวลา t และ $a_{ij}b_jO_{t+1}$ เป็นการเปลี่ยนสเตทที่เวลา t ไปเป็น q_j ที่เวลา t+1 และให้ค่าปรากฏเป็น O_{t+1}

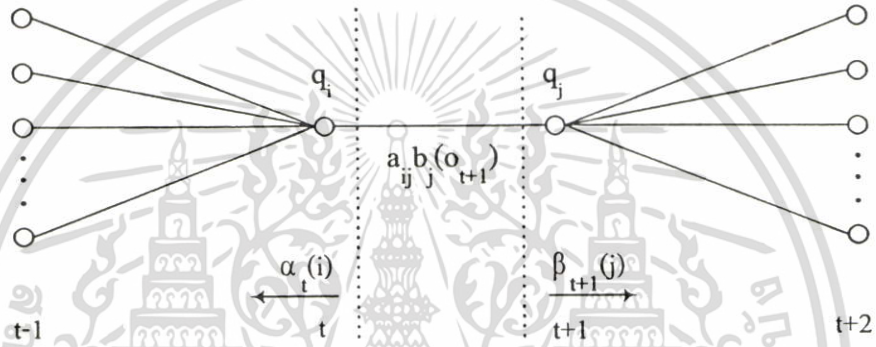
ซึ่งจากนิยามของตัวแปรไปข้างหน้า $\alpha_t(i)$ และตัวแปรย้อนกลับ $\beta_t(i)$ สามารถนำมาสัมพันธ์กับ $\varepsilon_t(i,j)$ ได้เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\varepsilon_t(i,j) = \frac{P(q_t = i, q_{t+1} = j, O|\lambda)}{P(O|\lambda)}$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} .$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)} \quad (4.44)$$



รูปที่ 4.15 ลำดับการคำนวณการเกิดค่าปรากฏร่วมซึ่งจะอยู่ ที่สเตท i ที่เวลา t และอยู่ที่ สเตท j ที่เวลา t+1

จากที่ได้นิยาม $\gamma_t(i)$ แล้วนำมาสัมพันธ์กับ $\varepsilon_t(i,j)$ ได้เป็น

$$\gamma_t(i) = \sum_{j=1}^{N} \varepsilon_t(i,j) \quad (4.45)$$

เมื่อ $\sum_{t=1}^{T-1} \gamma_t(i)$ = จำนวนของการย้ายสเตทจากสเตท i ในลำดับค่าปรากฏ O (4.46)

$\sum_{t=1}^{T-1} \varepsilon_t(i,j)$ = จำนวนของการย้ายสเตทจากสเตท i ไป j ในลำดับค่าปรากฏ O (4.47)

ดังนั้นสามารถคำนวณหาค่าของพารามิเตอร์ได้ดังนี้

π_i' = จำนวนครั้งในการอยู่ที่สเตท i ที่เวลา t = 1

$\pi_i' = \gamma_1(i)$; $1 \le i \le N$ (4.47)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$a'_{ij} = \frac{\text{จำนวนครั้งที่คาดไว้ของการย้ายสเตทจาก i ไป j}}{\text{จำนวนครั้งที่คาดว่าจะย้ายจากสเตท i}}$$

$$a'_{ij} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad . \tag{4.48}$$

$$b'_j(k) = \frac{\text{จำนวนครั้งที่คาดว่าจะอยู่ในสเตท j และเกิดค่าปรากฏเป็น } V_K}{\text{จำนวนครั้งที่คาดว่าจะอยู่ที่สเตท j}}$$

$$b'_j(k) = \frac{\sum_{t=1, O_t = V_K}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{4.49}$$

จากกระบวนการข้างต้นถ้าให้ $\lambda = (A,B,\pi)$ เป็นแบบจำลองปัจจุบัน และใช้ λ นี้คำนวณในด้านขวาของสมการที่(4.47-4.49)และให้แบบจำลองได้จากการคำนวณซ้ำเป็น $\lambda' = (A',B',\pi')$ เป็นแบบจำลองที่ได้จากด้านซ้ายของสมการที่(4.47-4.49) ซึ่งจะได้จุดวิกฤตของฟังก์ชั่นความน่าจะเป็นในกรณีที่ $\lambda' = \lambda$ หรือถ้า λ' มีความน่าจะเป็นมากกว่าแบบจำลอง λ [$P(O|\lambda') > P(O|\lambda)$] นั่นคือจะได้แบบจำลอง λ' ใหม่ที่น่าจะทำให้เกิดลำดับของค่าปรากฏ O ที่ดีกว่า

4.5.5 การปรับปรุงค่าพารามิเตอร์ของ HMM

4.5.5.1 การสเกลลิง (Scaling)

พิจารณาคำจำกัดความของ $\alpha_t(i)$ ในสมการที่ 4.24 จะเห็นว่า $\alpha_t(i)$ ประกอบไปด้วยผลรวมเทอมขนาดใหญ่ที่อยู่ในรูป

$$\prod_{s=1}^{t-1} a_{q_s q_{s+1}} \prod_{s=1}^{t} b_{q_s}(O_s)$$

เนื่องจากค่า a และ b เป็นค่าความน่าจะเป็น ซึ่งโดยทั่วไปแล้วมีค่าน้อยกว่า 1 ด้วยเหตุนี้เมื่อ t มากขึ้นค่าแต่ละเทอมของ $\alpha_t(i)$ จะเข้าสู่ศูนย์ ทำให้ช่วงไดนามิก (Dinamic Range) ของการคำนวณ $\alpha_t(i)$ มีค่าสูงเกินขอบเขตการทำงานของเครื่องคำนวณทำให้ค่าที่ได้ไม่ถูกต้อง ซึ่งเราสามารถแก้ไขปัญหานี้ได้โดยใช้กระบวนการสเกลลิง (Scaling Procedure)

การสเกลลิ่งทำได้โดยการคูณ $\alpha_t(i)$ ด้วยสัมประสิทธ์การสเกลลิง ซึ่งไม่ขึ้นกับ i (นั่นคือขึ้นอยู่กับค่าของเวลา t เท่านั้น) เพื่อให้ $\alpha_t(i)$ ที่ผ่านการสเกลลิงแล้วมีค่าอยู่ภายในช่วง Dynamic

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Range ของเครื่องคำนวณในทุกๆค่าเวลาภายใต้ $1 \le t \le T$ และในทำนองเดียวกันจะต้องทำการคำนวณค่าสัมประสิทธิ์การสเกลลิงของค่า $\beta_t(i)$ ด้วย ซึ่งในขั้นตอนสุดท้ายของการคำนวณ ค่าสัมประสิทธ์ของการสเกลลิงจะตัดกันหมดไป

เพื่อให้เข้าใจการทำงานของกระบวนการสเกลลิงดีขึ้น เราจะพิจารณาสมการของการย้ายสเตท (a_{ij}) ที่อยู่ในเทอมของตัวแปรไปข้างหน้าและตัวแปรย้อนกลับ

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{t=1}^{T}\sum_{j=1}^{N} \alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\beta_{t+1}(j)} \qquad (4.50)$$

พิจารณาสัญลักษณ์ในการคำนวณ $\alpha_t(i)$ เมื่อกำหนดให้

$\alpha_t(i)$ แทน α ที่ยังไม่ผ่านการสเกล

$\hat{\alpha}_t(i)$ แทน α ที่สเกลแล้ว

$\hat{\bar{\alpha}}_t(i)$ แทน α แทนเวอร์ชั่นของ α ก่อนการสเกล

เมื่อเวลาเริ่มต้น $t = 1$

คำนวณ $\alpha_t(i)$ ตามสมการที่ 4.25 และกำหนดให้ $\hat{\bar{\alpha}}_1(i) = \alpha_1(i)$

เมื่อ $$c_1 = \frac{1}{\sum_{i=1}^{N} \alpha_1(i)}$$

และ $\hat{\alpha}_1(i) = c_1 \alpha_1(i)$ เมื่อเวลา $2 \le t \le T$

เริ่มแรกทำการคำนวนหา $\hat{\bar{\alpha}}_t(i)$ ตามสมการการเหนี่ยวนำ สมการที่ 4.26 โดยใช้เทอมของค่าที่ผ่านการสเกลแล้ว $\hat{\alpha}_t(i)$ จะได้ดังนี้

$$\hat{\bar{\alpha}}_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j)\, a_{ji}\, b_i(O_t) \qquad (4.51a)$$

เมื่อกำหนดค่าสัมประสิทธ์การสเกลลิง ; c_t เป็น

$$c_t = \frac{1}{\sum_{i=1}^{N} \hat{\bar{\alpha}}_t(i)} \qquad (4.51b)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อให้

$$\hat{\alpha}_t(i) = c_t \bar{\hat{\alpha}}_t(i) \tag{4.51c}$$

จากสมการที่ 4.51 a-c สามารถเขียนสมการได้เป็น

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \hat{\alpha}_{t-1}(j)\, a_{ji}\, b_i(O_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\alpha}_{t-1}(j)\, a_{ji}\, b_i(O_t)} \tag{4.52}$$

และโดยการเหนี่ยวนำสามารถเขียน $\hat{\alpha}_{t-1}(j)$ ได้เป็น

$$\hat{\alpha}_{t-1}(j) = \left(\prod_{T=1}^{t-1} c_T\right) \alpha_{t-1}(j) \tag{4.53}$$

ดังนั้นสามารถเขียน $\hat{\alpha}_t(i)$ ได้เป็น

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \alpha_{t-1}(j)\left(\prod_{T=1}^{t-1} c_T\right) a_{ji}\, b_i(O_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{t-1}(j)\left(\prod_{T=1}^{t-1} c_T\right) a_{ji}\, b_i(O_t)} = \frac{\alpha_t(i)}{\sum_{i=1}^{N} \alpha_t(i)} \tag{4.54}$$

นั่นคือการสเกลลิงจะทำได้โดยนำ $\alpha_t(i)$ แต่ละค่า มาหารด้วยผลรวมของ $\alpha_t(i)$ ทุกสเตท จากนั้นทำการคำนวณลักษณะเดียวกันนี้กับเทอมของตัวแปรย้อนกลับ $\beta_t(i)$ โดยใช้สเกลเฟคเตอร์เดียวกันในรูปของ

$$\hat{\beta}_t(i) = c_t \beta_t(i) \tag{4.55}$$

พิจารณาสมการ 4.50 ในเทอมของตัวแปรที่ผ่านการสเกล จะได้เป็น

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i)\, a_{ij}\, b_j(O_{t+1})\hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N} \hat{\alpha}_t(i)\, a_{ij}\, b_j(O_{t+1})\hat{\beta}_{t+1}(j)} \tag{4.56}$$

โดยแต่ละ $\hat{\alpha}_t(i)$, $\hat{\beta}_{t+1}(j)$ สามารถเขียนให้อยู่ในรูปของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\hat{\alpha}_t(i) = \left[\prod_{s=1}^{T} c_s\right] \alpha_t(i) = C_t \alpha_t(i) \tag{4.57}$$

$$\hat{\beta}_{t+1}(j) = \left[\prod_{s=t+1}^{T} c_s\right] \beta_{t+1}(j) = D_{t+1} \beta_{t+1}(j) \tag{4.58}$$

ดังนั้นสมการ 4.56 สามารถเขียนได้เป็น

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} C_t \alpha_t(i)\, a_{ij}\, b_j(O_{t+1}) D_{t+1} \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} C_t \alpha_t(i)\, a_{ij}\, b_j(O_{t+1}) D_{t+1} \beta_{t+1}(j)} \tag{4.59}$$

โดยเทอม C_tD_{t+1} สามารถเขียนให้อยู่ในรูปของ

$$C_t D_{t+1} = \prod_{s=1}^{t} c_s \prod_{s=t+1}^{T} c_s = \prod_{s=1}^{T} c_s = C_T \tag{4.60}$$

จะเห็นว่าเทอม C_tD_{t+1} เป็นค่าที่ไม่ขึ้นกับเวลา ดังนั้นสามารถตัดออกจากทั้งเศษและส่วนของสมการ 4.59 ได้ ซึ่งจะทำให้ได้สูตรของการคำนวณซ้ำ กระบวนการสเกลลิงดังกล่าวนี้จะถูกนำไปใช้กับสัมประสิทธ์ π และ β การสเกลลิงนี้จะทำให้การคำนวณค่า $P(O|\lambda)$ เปลี่ยนไปเราจะไม่สามารถหาได้จากการรวมเทอมของ $\hat{\alpha}_T(i)$ เนื่องจากเป็นค่าที่ถูกสเกลแล้ว แต่เราสามารถคำนวณได้จากคุณสมบัติ

$$\prod_{t=1}^{T} c_t \prod_{i=1}^{N} \alpha_T(i) = C_T \sum_{i=1}^{N} \alpha_T(i) = 1 \tag{4.61}$$

ดังนั้นจะได้

$$\prod_{t=1}^{T} c_t \cdot P(O|\lambda) = 1 \tag{4.62}$$

หรือ

$$P(O|\lambda) = \frac{1}{\prod_{t=1}^{T} c_t} \tag{4.63}$$

หรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\log [\, P(O|\lambda) \,] \;=\; -\sum_{t=1}^{T} \log C_t \qquad (4.64)$$

นั่นคือ การคำนวณค่า P จะอยู่ในรูป log ของ Pเพื่อไม่ให้เกินช่วงไดนามิก (Dinamic Range) ของเครื่องคำนวณ

4.5.1.2 ลำดับของค่าปรากฏหลายลำดับ (Multiple Observation Sequences)

ในการสร้างแบบจำลองด้วย Left-Right Model จำเป็นจะต้องใช้จำนวนลำดับของเหตุการณ์หลายๆลำดับเข้ามาเทรนเพื่อให้การประมาณค่าพารามิเตอร์ ของแบบจำลองที่ได้มีความน่าเชื่อถือที่สุด ถ้ากำหนดให้ k แทนเซ็ตของลำดับค่าปรากฏ ดังนี้

$$O \;=\; [\, O^{(1)}, O^{(2)}, \ldots, O^{(k)} \,] \qquad (4.65)$$

เมื่อ $O^{(k)} = (O_1^{(k)}\ O_2^{(k)} \ldots O_{T_k}^{(k)})$ คือ ลำดับค่าปรากฏอันดับที่ k โดยสมมติให้แต่ละอันดับของค่าปรากฏเป็นอิสระต่อกัน โดยมีจุดประสงค์เพื่อที่จะปรับค่าพารามิเตอร์ของแบบจำลอง λ ให้มีค่ามากที่สุด

$$P(O|\lambda) \;=\; \prod_{k=1}^{K} P(O^{(k)}|\lambda) \qquad (4.65)$$

$$= \prod_{k=1}^{K} P_k \qquad (4.66)$$

ดังนั้นจะได้สมการของการคำนวณซ้ำที่ใช้ในการปรับค่า $\bar{a}_{ij}$ และ $\bar{b}_j(l)$ เป็น

$$\bar{a}_{ij} \;=\; \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\, a_{ij}\, b_j(O_{t+1}^{(k)})\beta_{t+1}^k(j)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\beta_t^k(i)} \qquad (4.67)$$

และ

$$\bar{b}_j(l) \;=\; \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1,\, O_t = v_l}^{T_k-1} \alpha_t^k(i)\, \beta_t^k(i)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\beta_t^k(i)} \qquad (4.68)$$

ส่วนค่า π_i ไม่ต้องมีการคำนวณซ้ำเนื่องจาก $\pi_1 = 1, \pi_i = 0, i \neq 1$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการของการสเกลลิงสมการที่ 4.67-4.68 เราสามารถเขียนสมการที่อยู่ในเทอมของตัวแปรที่สเกลแล้วได้เป็น

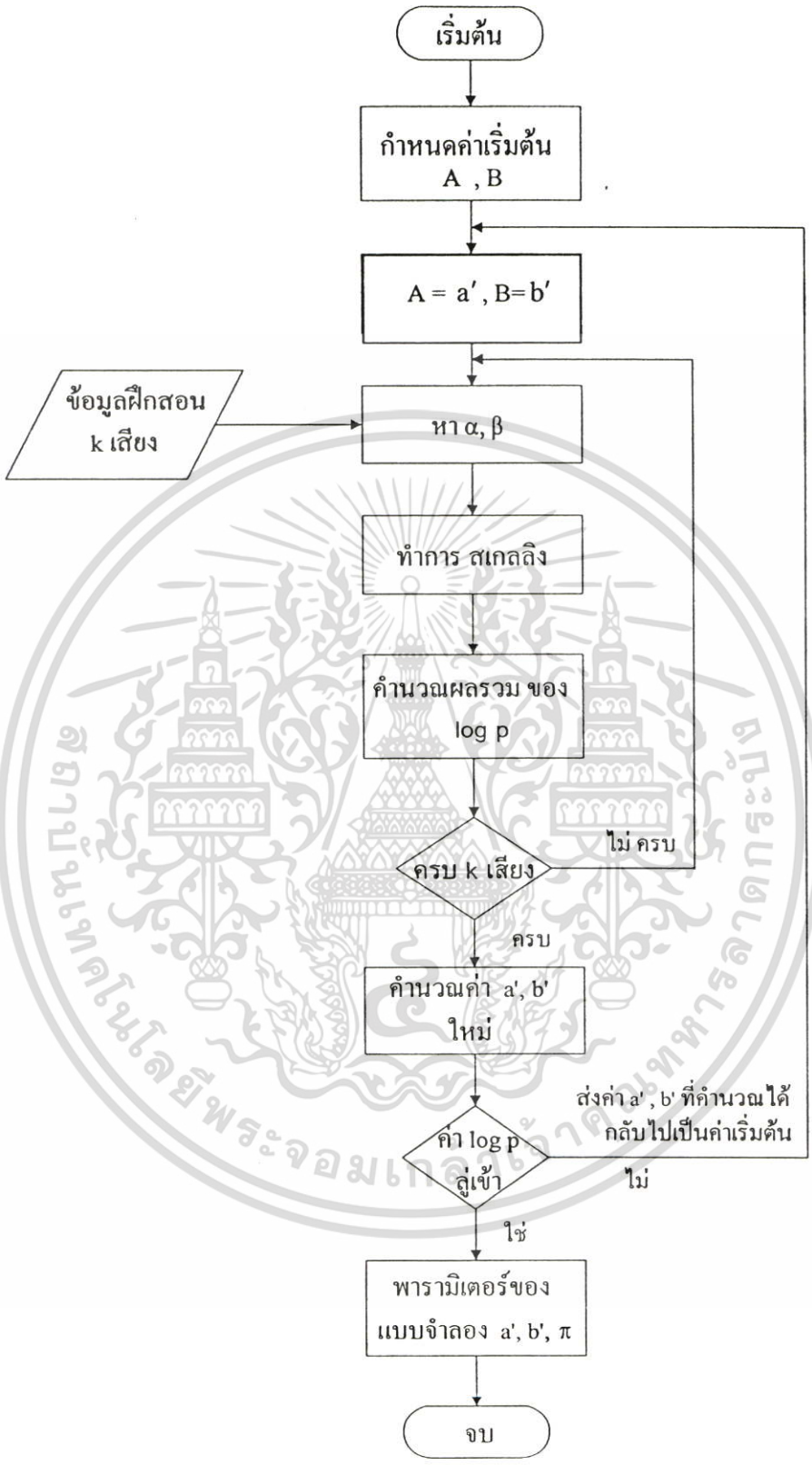
$$\bar{a}_{ij} = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^k(i)\, a_{ij}\, b_j(O_{t+1}^{(k)}) \hat{\beta}_{t+1}^k(j)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^k(i) \hat{\beta}_t^k(i)} \quad (4.69)$$

$$\bar{b}_j(k) = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1, O_t=v_l}^{T_k-1} \hat{\alpha}_t^k(i)\, \hat{\beta}_t^k(i)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^k(i) \hat{\beta}_t^k(i)} \quad (4.70)$$

4.5.6 ขบวนการสร้างแบบจำลองอ้างอิง

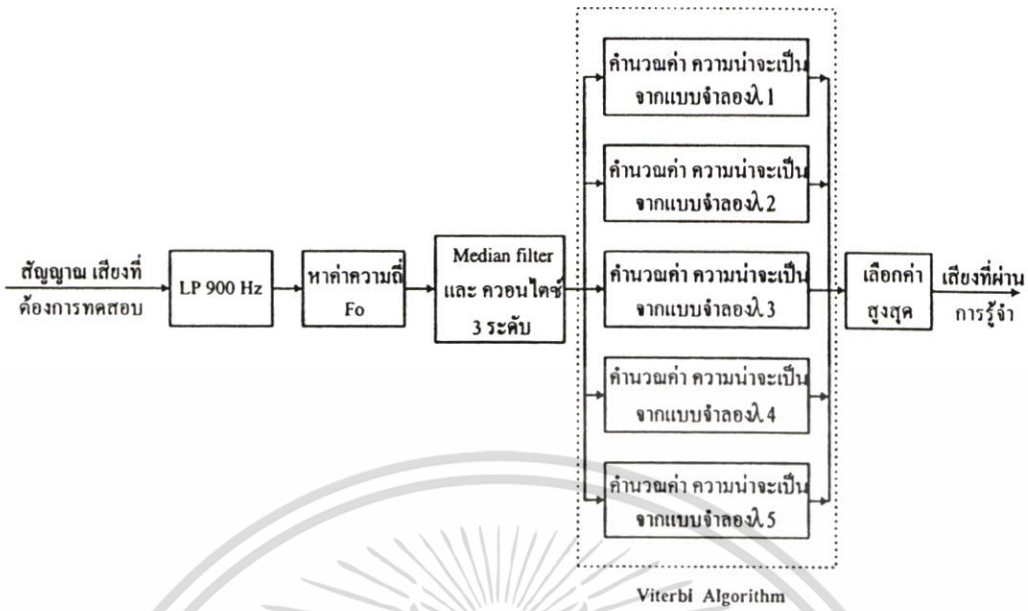
การสร้างแบบจำลองอ้างอิงเสียงวรรณยุกต์ทั้ง 5 ระดับเสียง สิ่งที่จะต้องมีก็คือกลุ่มเสียงต้นแบบหรือลำดับของค่าปรากฏทั้ง 5 กลุ่ม เพื่อใช้เป็นข้อมูลฝึกสอน (Training data) จากรูปที่ 4.16 แสดงขั้นตอนของการคำนวณการสร้างแบบจำลองอ้างอิง โดยในขั้นแรกจะต้องทำการกำหนดค่า A, B เริ่มต้น จากนั้น ทำการคำนวณหาค่า α, β โดยใช้การแก้ปัญหาที่ 1 ของ HMM แล้วทำการสเกลลิง เพื่อไม่ให้ค่าที่ได้จากการคำนวณมีค่าเกินช่วงไดนามิกของเครื่องคำนวณ (Dynamic Range) จากนั้นนำค่าสัมประสิทธิ์ของการสเกลลิงมาคำนวณหาค่า ความน่าจะเป็น $P(O|\lambda)$ ซึ่งจะอยู่ในรูปของค่า log P และเนื่องจากในการสร้างแบบจำลองอ้างอิง จำเป็นจะต้องใช้ข้อมูลฝึกสอนจำนวนมาก เพื่อให้แบบจำลองอ้างอิงที่สร้างขึ้น ครอบคลุมความแปรปรวนของลักษณะเสียงให้ได้มากที่สุด ดังนั้นจึงจะต้องมีการคำนวณซ้ำเกิดขึ้น ตามจำนวนของเสียงที่นำมาฝึกสอน จากนั้นทำการหาค่าผลรวมของค่าความน่าจะเป็น (ผลรวมของ log P จากจำนวนเสียงทั้งหมด) เพื่อมาใช้ในการคำนวณหาค่าพารามิเตอร์ของแบบจำลอง $\lambda=(A',B',\pi)$ โดยใช้การแก้ปัญหาที่ 3 ของ HMM จากนั้นทำการคำนวณซ้ำจนกว่าค่าผลรวมของ log P ที่ได้ในแต่ละรอบมีค่าลู่เข้า หรือไม่เปลี่ยนแปลง พารามิเตอร์ของแบบจำลอง $\lambda'=(A',B',\pi)$ ค่าสุดท้ายจะเป็นแบบจำลองที่น่าจะทำให้เกิดลำดับของค่าปรากฏ O ที่ดีกว่า โดยรายละเอียดของขั้นตอนทฤษฎีการวิเคราะห์ต่างๆดังได้กล่าวมาแล้วก่อนหน้านี้ ซึ่งเมื่อหาค่าพารามิเตอร์ของแต่ละระดับเสียงได้แล้วจะนำมาทำการสร้างแบบจำลองเสียงวรรณยุกต์ทั้ง 5 รูปแบบแสดงได้ดังรูปที่ 4.17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.16 โฟลว์ชาร์ต การคำนวณหาค่าพารามิเตอร์ของแบบจำลองอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.17 บล็อคไดอะแกรมของการรู้จำระดับเสียงวรรณยุกต์ด้วยแบบจำลองมาร์คอฟ

4.6 สรุป

เนื้อหาในบทนี้กล่าวถึงทฤษฎีการวิเคราะห์หาคาบพิทช์เพื่อนำมาหาค่าความถี่มูลฐาน ซึ่งเป็นตัวบ่งบอกถึงระดับเสียงสูง-ต่ำ ของคำในภาษา โดยค่าความถี่มูลฐานนี้หามาได้จากขบวนการประมาณคาบพิทช์ด้วยวิธีนอร์แมลไลซ์ออโตโครีเลชั่น จากนั้นค่าคาบพิทช์จะถูกแปลงให้อยู่ในรูปของค่าความถี่มูลฐาน โดยรูปแบบการเปลี่ยนแปลงของค่าความถี่มูลฐานเมื่อเทียบกับเวลาจะเป็นตัวบ่งบอกถึงระดับเสียงวรรณยุกต์ที่แตกต่างกันของคำหรือพยางค์ในภาษาลาว ซึ่งลำดับของค่าความถี่มูลฐานที่แตกต่างกันนี้จะถูกนำไปเข้าสู่กระบวนการเตรียมข้อมูล เพื่อใช้เป็นข้อมูลฝึกสอนในการสร้างแบบจำลองอ้างอิงของการรู้จำระดับเสียงวรรณยุกต์ด้วยวิธีของฮิดเดนมาร์คอฟ ซึ่งด้กล่าวถึงการคำนวณหาพารามิเตอร์ของแบบจำลองอ้างอิงในการรู้จำเพื่อสร้างให้เป็นแบบจำเสียงพูด โดยค่าพารามิเตอร์ผลของแบบจำลองก็คือค่าที่ได้จากการทดลองจะกล่าวถึงในบทต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

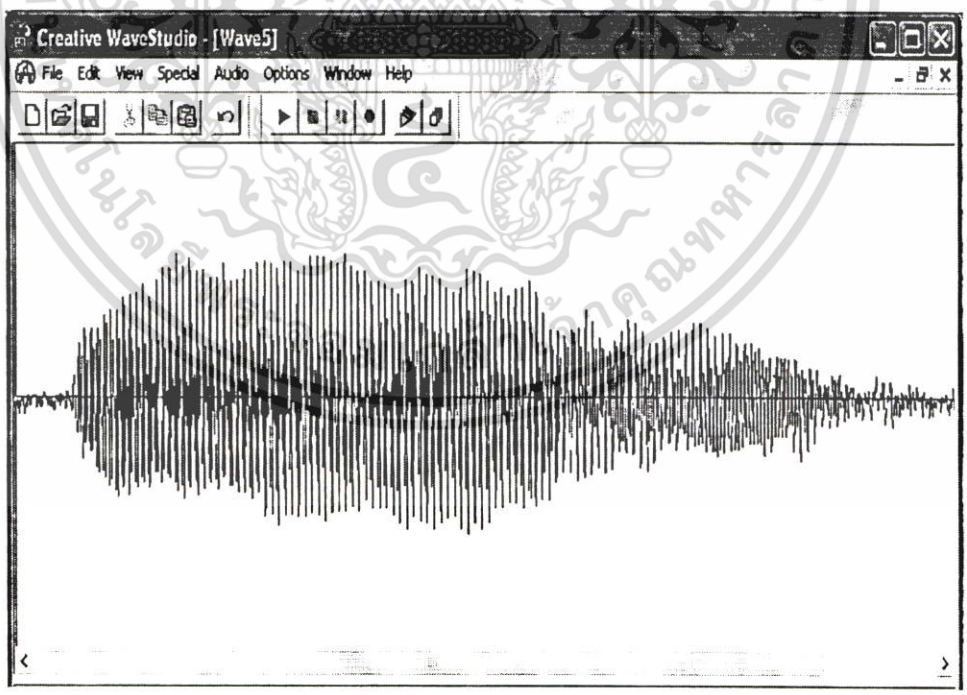
บทที่ 5

การทดลองและผลการทดลอง

5.1 เสียงของพยางค์ หรือคำที่ใช้ในการทดลอง

สัญญาณเสียงพูดที่ใช้ในวิทยานิพนธ์นี้ เป็นคำพยางค์เดียวที่ได้จากการเก็บตัวอย่างข้อมูลเสียง โดยใช้เครื่องคอมพิวเตอร์ส่วนบุคคล ซึ่งข้อมูลจะถูกเก็บอยู่ในรูปของไฟล์ '.wav' ข้อมูล 1 ตัวอย่างของเสียงจะถูกแทนด้วยข้อมูลขนาด 8 บิท โดยใช้ความถี่ในการแซมปลิ้งเท่ากับ 11.025 kHz และไฟล์ข้อมูล '.wav' นี้จะถูกใช้เป็นข้อมูลอินพุทสำหรับการคำนวณของโปรแกรมที่เขียนขึ้น โดยในวิทยานิพนธ์นี้เลือกใช้โปรแกรม Visual C++ ในการทดลองและพัฒนาวิธีการต่างที่ใช้ในการรู้จำเสียงพูด

จากที่กล่าวมาแล้วว่า หน่วยเสียงวรรณยุกต์เป็นหน่วยเสียงซ้อนวางตัวอยู่ในเสียงก้อง ซึ่งคุณสมบัติของเสียงก้องที่สังเกตได้เด่นชัดในโดเมนของเวลาที่มีความเป็นคาบที่ชัดเจน โดยการกำหนดขอบเขตของคำพูดเพื่อวิเคราะห์หาหน่วยเสียงวรรณยุกต์ จะกำหนดขอบเขตภายในช่วงที่มีความเป็นคาบทั้งหมดของคำหรือพยางค์นั้นๆ ดังตัวอย่างสัญญาณเสียงพูดแสดงได้ดังรูปที่ 5.1



รูปที่ 5.1 ตัวอย่างสัญญาณเสียงพูดของคำว่า "ชอง" จากผู้ออกเสียงเพศชาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

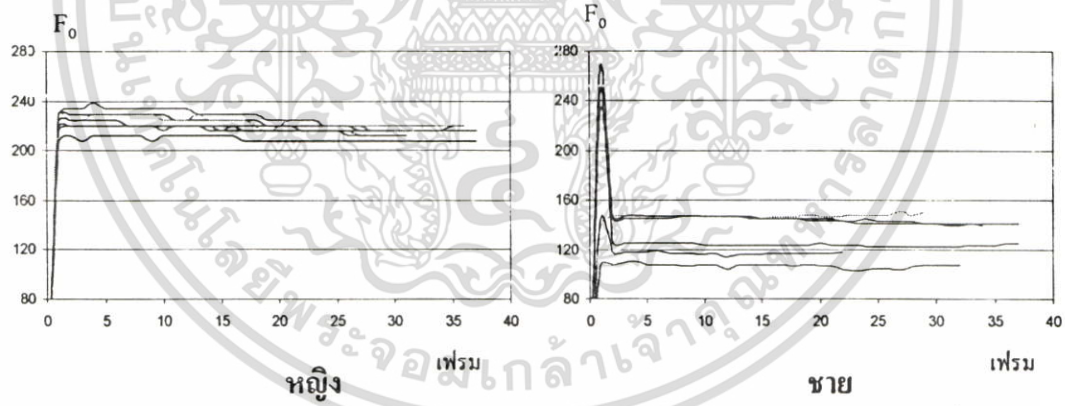
5.2 ผลการวิเคราะห์แบบจำลองเสียงวรรณยุกต์

การวิเคราะห์ในการสร้างแบบจำลองเสียงมีผลการทดลองตามขั้นตอนดังนี้

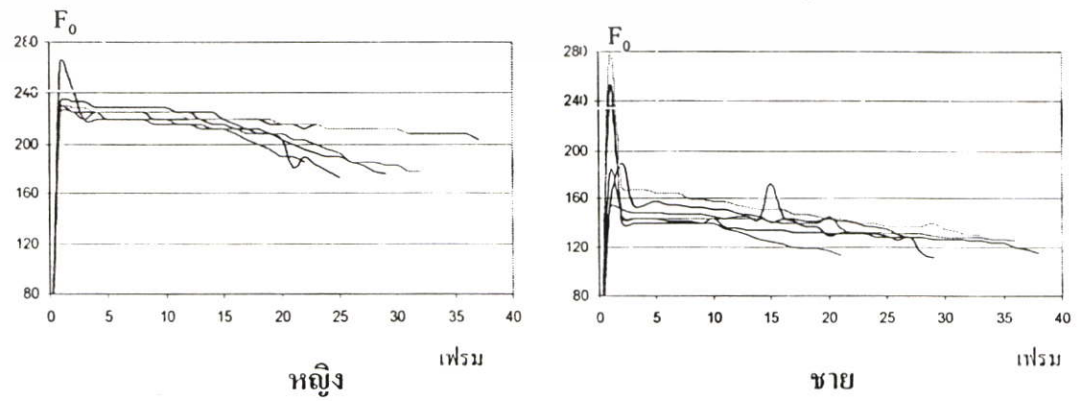
5.2.1 ผลการหาค่าความถี่มูลฐานและการควอนไตซ์พิทช์

5.2.1.1 ผลการหาค่าความถี่มูลฐาน

การหาค่าพิทช์ทำได้โดยใช้วิธีออโตโครีเลชั่นตามทฤษฎีการวิเคราะห์บทที่ 4 ดังที่ได้กล่าวมาแล้ว โดยอัลกอริธึมในการหาค่าพิทช์มีขั้นตอนตามโฟลว์ชาร์ตรูปที่ 4.7 ซึ่งผลที่ได้จากค่าคาบพิทช์จะถูกแปลงให้อยู่ในรูปของค่าความถี่มูลฐาน โดยรูปแบบการเปลี่ยนแปลงของค่าความถี่มูลฐานเมื่อเทียบกับเวลาตามสมการที่ (4.10) ดังตัวอย่างจากการทดลองเสียงหญิง 6 คน และชาย 6 คนดังรูปที่ 5.2 จะเห็นว่าผลตอบสนองความค่าความถี่มูลฐานทางเดินเสียงวรรณยุกต์แต่ละเสียงมีค่าการเปลี่ยนแปลงตามลักษณะเฉพาะของวรรณยุกต์ โดยไม่ขึ้นกับเสียงของผู้ออกเสียงชายและหญิง รูปที่ 5.2 (a) คือเสียงสามัญ จะมีคาบหรือความถี่มูลฐานเทียบกับเวลามีค่าคงที่ รูปที่ 5.2 (b) คือเสียงเอกที่มีค่าความถี่มูลฐานลดลงแต่เริ่มต้นหาท้ายสุดของเส้นเสียง รูปที่ 5.2 (c) คือเสียงโทรมีค่าความถี่มูลฐานเพิ่มขึ้นจากช่วงแรกหาช่วงกลางจากนั้นมีค่าลดลงถึงสุดท้ายของเส้นเสียง รูปที่ 5.2 (d) คือเสียงตรีมีค่าความถี่มูลฐานเพิ่มขึ้นจากช่วงแรกถึงสุดท้ายของเส้นเสียงและรูปที่ 5.2(e) คือเสียงจัตวามีค่าความถี่มูลฐานลดลงจากช่วงแรกหาช่วงกลางจากนั้นจะมีค่าเพิ่มขึ้นถึงสุดท้ายของเส้นเสียง

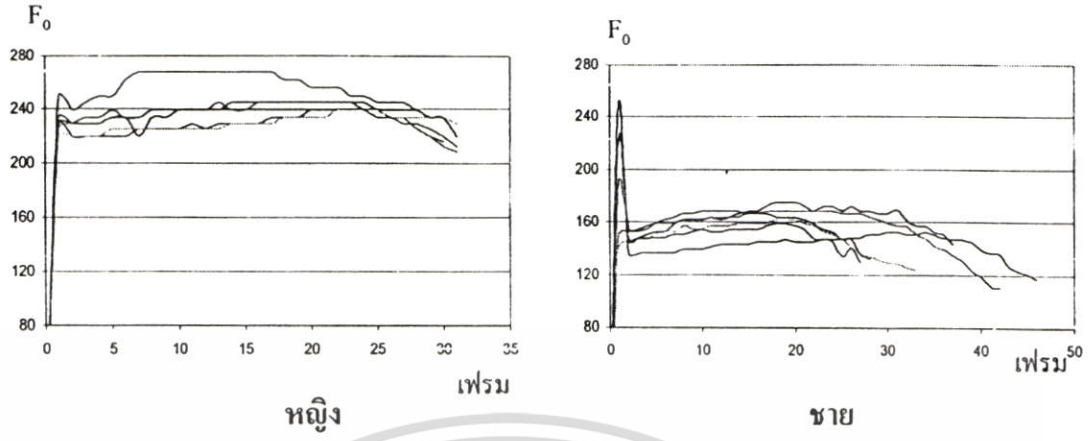


(a) ระดับเสียง กลาง (mid tone) หรือเสียงสามัญ



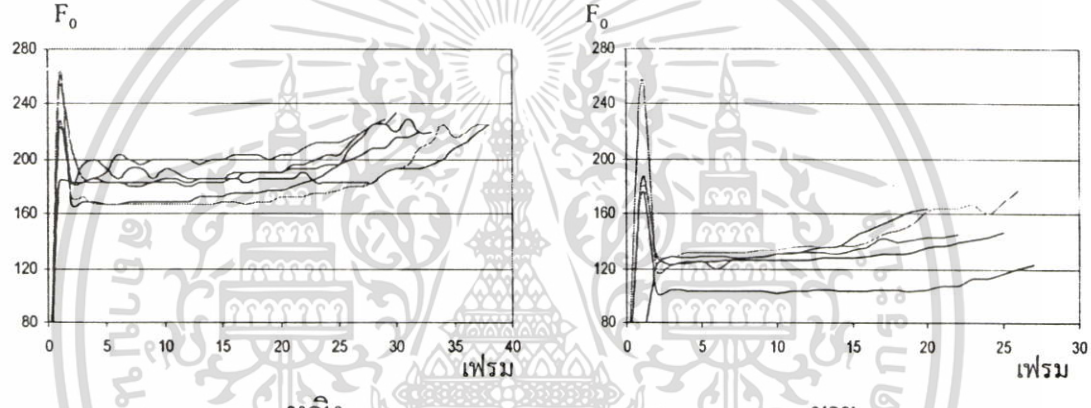
(b) ระดับเสียง สูง-ตก (high falling) หรือเสียงเอก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



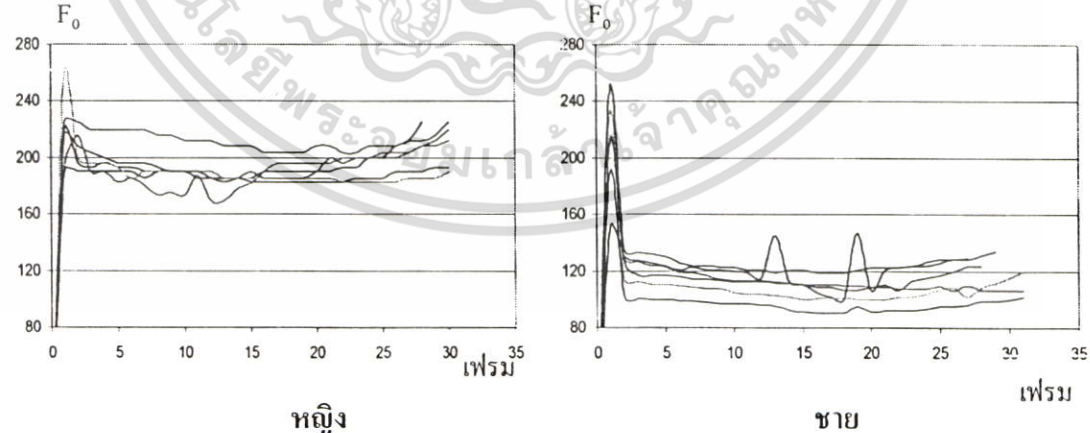
หญิง ชาย

(c) ระดับเสียง กลาง-ตก (mid-falling) หรือเสียงโทร



หญิง ชาย

(d) ระดับเสียง ต่ำ-ขึ้น (rising-tone) หรือเสียงตรี



หญิง ชาย

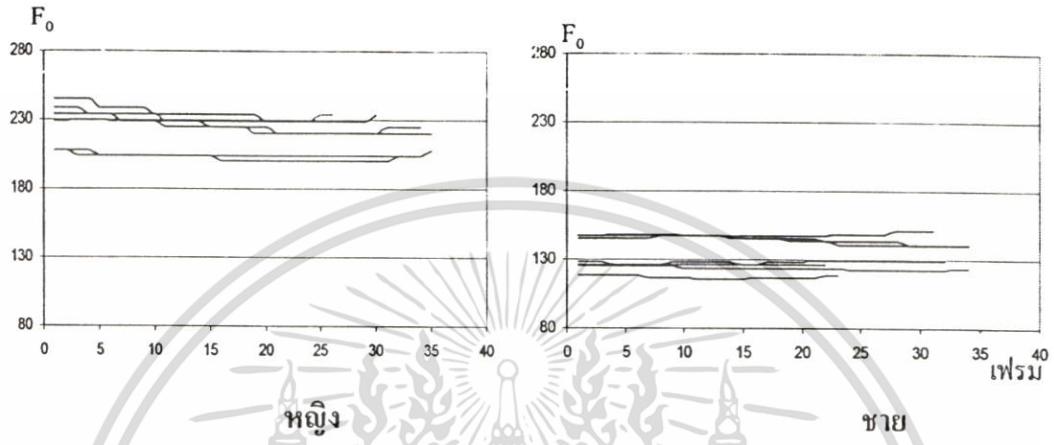
(e) ระดับเสียง กลาง-ขึ้น (mid-rising) หรือเสียงจัตวา

รูปที่ 5.2 (a-e) แสดงตัวอย่างการเปลี่ยนแปลงค่าความถี่มูลฐานของทางเดินเสียงวรรณยุกต์

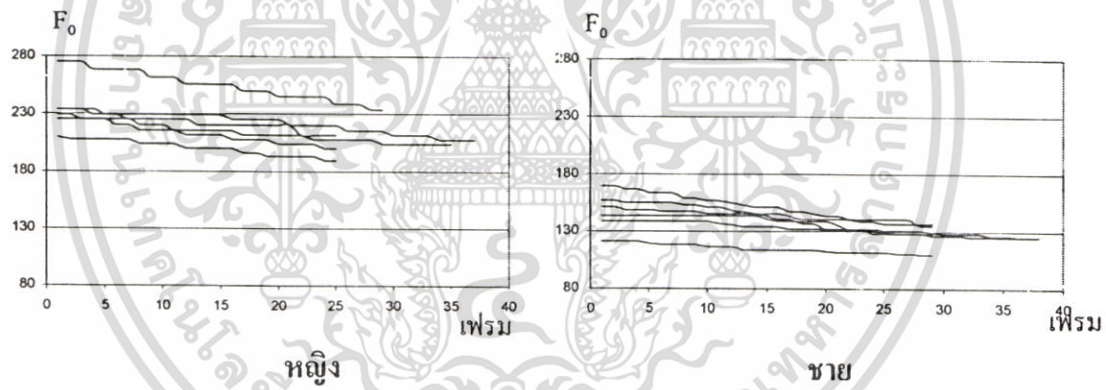
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.1.2 การกรองค่ากลาง

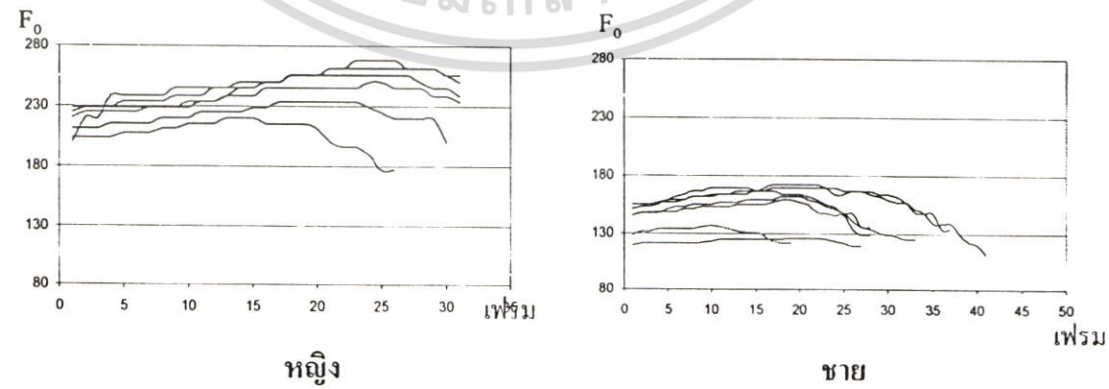
เป็นส่วนของการปรับปรุงข้อมูลให้มีความต่อเนื่องมากขึ้น ซึ่งจากรูปที่ 5.2 เมื่อนำมาผ่านขั้นตอนตามทฤษฎีการวิเคราะห์บทที่ 4 เมื่อนำมาผ่านขั้นตอนนี้จะได้กราฟที่มีลักษณะดังรูปที่ 5.3 ซึ่งเป็นผลจากการทดลองของผู้ออกเสียง หญิง 6 คน และชาย 6 คน



(a) ระดับเสียง กลาง (mid tone) หรือเสียงสามัญ

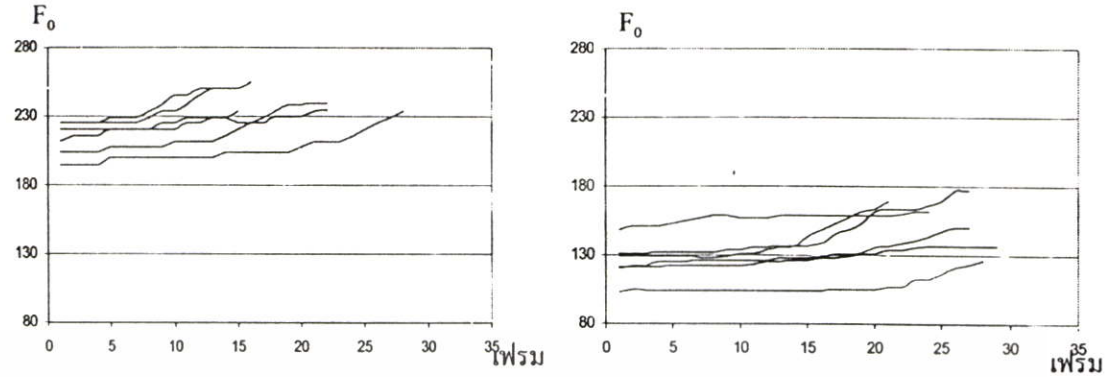


(b) ระดับเสียง สูง-ตก (high falling) หรือเสียงเอก



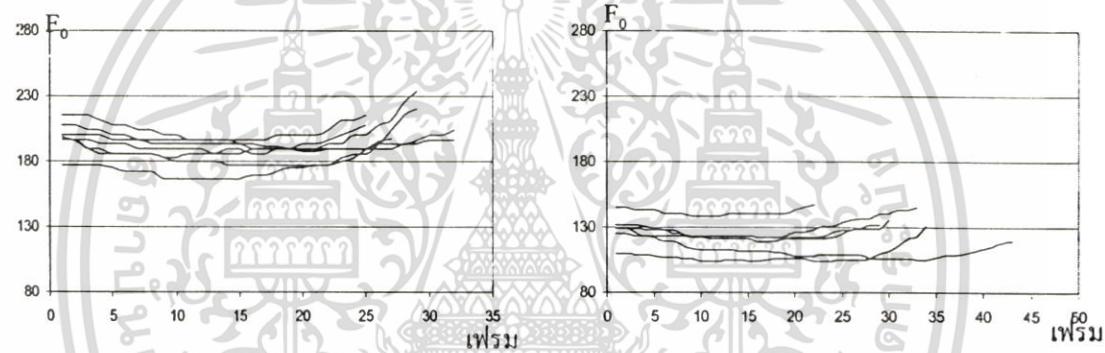
(c) ระดับเสียง กลาง-ตก (mid-falling) หรือเสียงโทร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



หญิง ชาย

(d) ระดับเสียง ต่ำ-ขึ้น (rising-tone) หรือเสียงตรี



หญิง ชาย

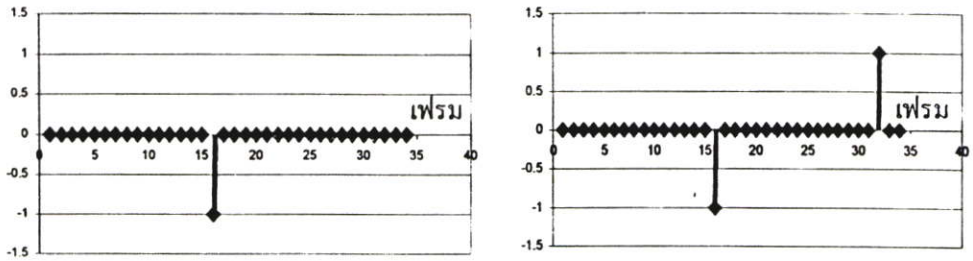
(e) ระดับเสียง กลาง-ขึ้น (mid-rising) หรือเสียงจัตวา

รูปที่ 5.3 (a-e) แสดงตัวอย่างการกองค่ากลางความถี่มูลฐานของทางเดินเสียงวรรณยุกต์

5.2.2 การควอนไตซ์พิทช์

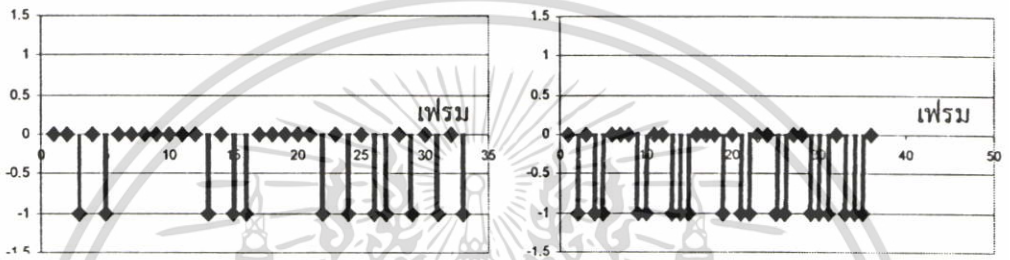
การควอนไตซ์ เป็นการเตรียมข้อมูลเพื่อใช้เป็นข้อมูลฝึกสอนในกระบวนการสร้างแบบจำลองด้วยฮิดเดนมาร์คอฟโมเดล ซึ่งข้อมูลที่ผ่านการกรองค่ากลางจะถูกนำมาทำการจัดระดับออกเป็น 3 ระดับตามทิศทางการเปลี่ยนแปลงของความถี่มูลฐานต่อเวลาโดยแทนค่าเป็น 1 เมื่อความ ถี่เพิ่มขึ้นเป็น 0 เมื่อความถี่ไม่เปลี่ยนแปลงและ –1 เมื่อความถี่ลดลงจากข้อมูลก่อน ดังตัวอย่างสัญญาณข้อมูลในรูปที่ 5.2 เมื่อนำมาผ่านการควอนไตซ์ออกเป็น 3 ระดับสามารถแสดงได้ดังรูปที่ 5.3 เมื่อผ่านขั้นตอนนี้ แล้วข้อมูลจะอยู่ในรูปของ ลำดับการเปลี่ยนแปลงของความถี่ ซึ่งมีสมาชิกของลำดับเป็น{-1,0, 1}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



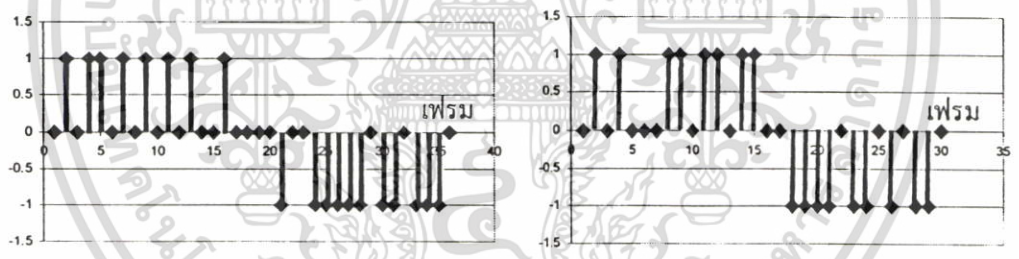
หญิง ชาย

(a) เสียง กลาง (mid tone)



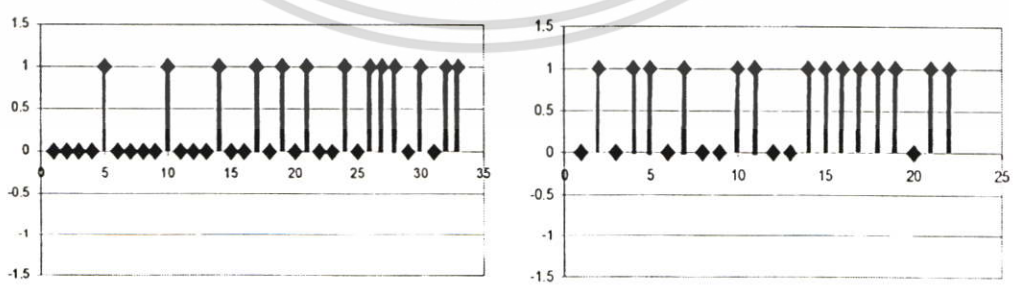
หญิง ชาย

(b) เสียง สูง-ตก (high -falling)



หญิง ชาย

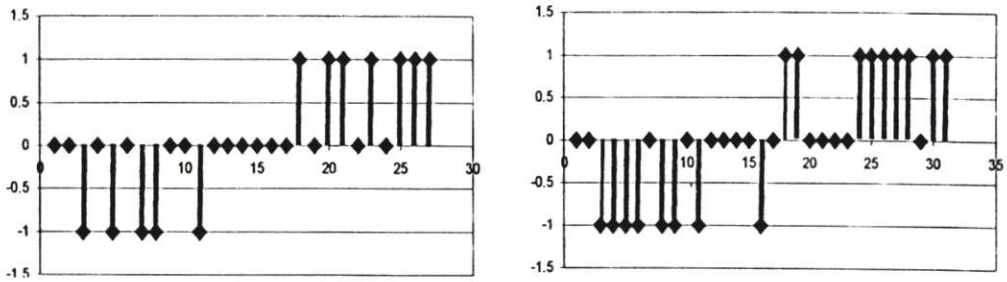
(c) เสียง กลาง-ตก (mid-falling)



หญิง ชาย

(d) เสียง ต่ำ-ขึ้น (rising-tone)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(e) เสียง กลาง-ขึ้น (mid-rising)

รูปที่ 5.4 (a-e) แสดงตัวอย่างการการเปลี่ยนแปลงค่าของพิทช์จากการทำควอนไตซ์

โดยค่าที่ผ่านการควอนไตซ์ตามการเปลี่ยนแปลงของความถี่ จะเป็นคุณลักษณะเฉพาะตามแต่ระดับเสียงของวรรณยุกต์ดังนี้

ระดับเสียง "สูง-ตก" จะมีค่าเป็น -1 จำนวนมาก สลับกับ 0 จากเฟรมแรกถึงเฟรมสุดท้ายของเสียง ระดับเสียง "กลาง-ตก" จะมีค่า 1 จำนวนมากสลับกับ 0 ถึงระยะกลางของจำนวนเฟรมเสียงและมีค่า 0 สลับกับ -1 จำนวนมากแต่ระยะกลางถึงระยะสุดท้ายของจำนวนเฟรม ระดับเสียง "กลาง-ขึ้น" จะมีค่า -1 จำนวนมากสลับกับ 0 ถึงระยะกลางของจำนวนเฟรมเสียงและมีค่า 0 สลับกับ 1 แต่ระยะกลางถึงระยะสุดท้ายของจำนวนเฟรมเสียง ระดับเสียง "ต่ำ-ขึ้น" จะมีค่าเป็น 0 สลับกับ 1 จำนวน มากแต่เฟรมแรกถึงเฟรมสุดท้ายของเสียง ระดับเสียง "กลาง" จะมีค่า 0 จำนวนมากแต่เฟรมแรก ถึงเฟรมสุดท้าย

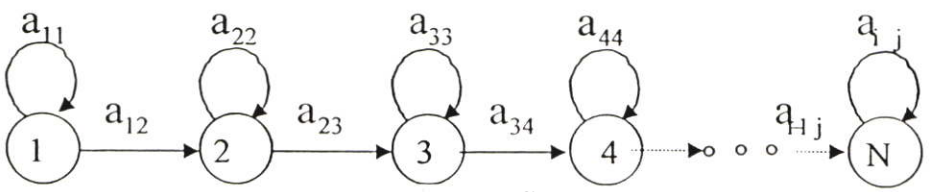
ผลการทดลองจากการทำควอนไตซ์ของคำพูดพยางค์เดียวจากคำพูดในตารางที่ 5.2 ผู้ออกเสียง 12 หญิง 6 คน และชาย 6 คน

5.2.3 การทดสอบหารูปแบบ HMM ที่เหมาะสม

ในวิทยานิพนธ์นี้ได้ทำการทดสอบหารูปแบบของ Left-Right Model ที่เหมาะสมสำหรับอัลกอริธึมที่ได้พัฒนาขึ้นดังกล่าวมาก่อนหน้านี้ โดยพารามิเตอร์ที่มีความสัมพันธ์เกี่ยวข้อง ที่ต้องคำนึงถึงมีอยู่ 2 ค่า คือ

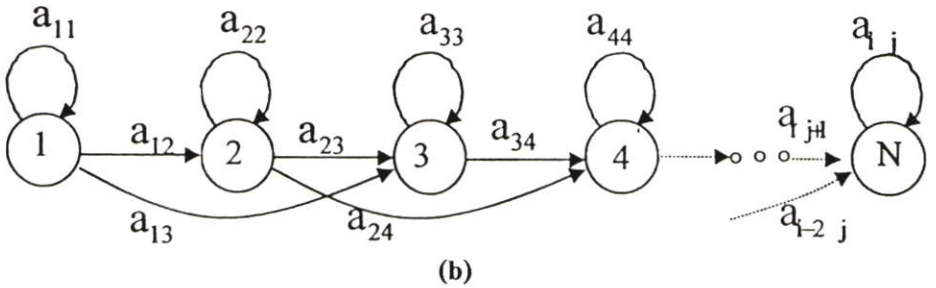
5.2.2.1 จำนวนสเตท

5.2.2.2 การย้ายข้ามสเตท

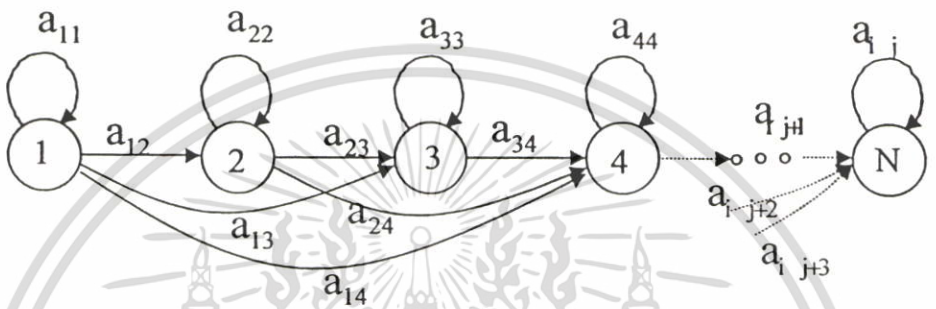


(a)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(b)



(c)

รูปที่ 5.5 Left-Right Model N สเตท

(a) การย้ายข้ามสเตทสูงสุดได้ไม่เกิน 1 สเตท (Single Transition)

(b) การย้ายข้ามสเตทสูงสุดได้ไม่เกิน 2 สเตท (Double Transition)

(c) การย้ายข้ามสเตทสูงสุดได้ไม่เกิน 3 สเตท (Triple Transition)

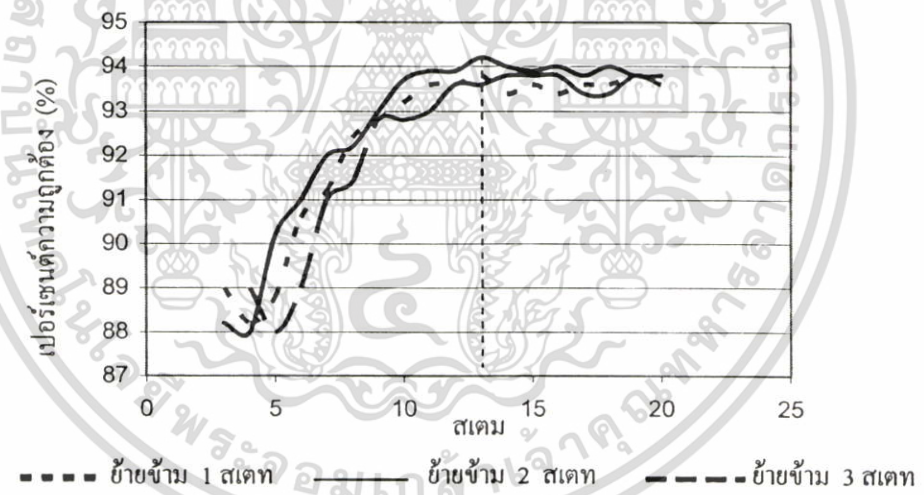
รูปที่ 5.5 เป็นตัวอย่างของแบบจำลอง HMM ที่มีจำนวนสเตท N สเตท และมีค่าการย้ายข้ามสเตทที่เป็นไปได้ 3 แบบ ซึ่งจากลักษณะเดียวกันนี้ วิทยานิพนธ์นี้ได้ทำการศึกษาทดลองหาจำนวนขนาดสเตท และการย้ายข้ามสเตทที่เหมาะสม เพื่อให้แบบจำลองการรู้จำระดับเสียงที่สร้างขึ้นมีความถูกต้องแม่นยำสูงสุด โดยทำการศึกษา HMM ตั้งแต่ 3 สเตท จนถึง 18 สเตท และเพิ่มขึ้นทีละ 1 สเตท ในขณะเดียวกันก็กำหนดให้ HMM แต่ละขนาดมีการย้ายข้ามสเตทสูงสุดได้ทั้ง 3 แบบ คือมีการย้ายข้ามสเตทสูงสุดได้ 1สเตท, 2 สเตท และ 3 สเตท ซึ่งในการทดลองนี้ได้ทำการสร้างแบบจำลองขึ้นมาทั้งหมด 47 แบบจำลอง (แบบจำลองที่มี 3 สเตทมี 2 แบบจำลอง) แล้วทำการสร้างแบบจำลองตามโฟลว์ชาร์ตรูปที่ 4.16

โดยในทุกๆแบบจำลองจะถูกสร้างขึ้นจากคำต้นแบบจำนวน 50 คำ ดังแสดงไว้ในตารางที่ 5.1 จากผู้ออกเสียง 10 คน เป็นชาย 5 คนและหญิง 5 คนออกเสียงคนละ 1 ครั้ง รวมเป็นเสียงต้นแบบจำนวนทั้งสิ้น 500 เสียง และในการทดสอบการรู้จำเพื่อหารูปแบบ HMM ที่เหมาะสมนี้จะวัดผลโดยนำเสียงต้นแบบเดิมทั้ง 500 เสียง เข้าไปทำการทดสอบการรู้จำอีกครั้ง ในทุกๆแบบจำลองที่สร้างขึ้นทั้ง 53 แบบจำลอง โดยผลการทดสอบแสดงได้ดังรูปที่ 5.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 กลุ่มคำที่ 1 ใช้ในการทดสอบหาแบบจำลอง HMM ที่เหมาะสม

กลาง[—]	สูง-ตก [\]	กลาง-ตก[∧]	ต่ำ-ขึ้น [/]	กลาง-ขึ้น [∨]
ยาว	บ้	เก้า	ปัก	สาม
นิ	โตะ	ทั้งๆ	หมิด	ส่าม
พอม	ทุ	ท้าว	อิด	เส่ง
แซว	ลิด	ได้	เท่า	โต
บา	เตะ	มั้	สิ	ตุ
มิม	เจ้ง	ท้า	หิก	ติ่
ลาม	ก็ม	ไล่	หมู่	เถ้า
กอม	ล้าม	ม้า	โภะ	ถิ
ปา	องก	แล้ว	โข	เป้า
นัด	ใส่	นั้น	ติอะ	เหิอ

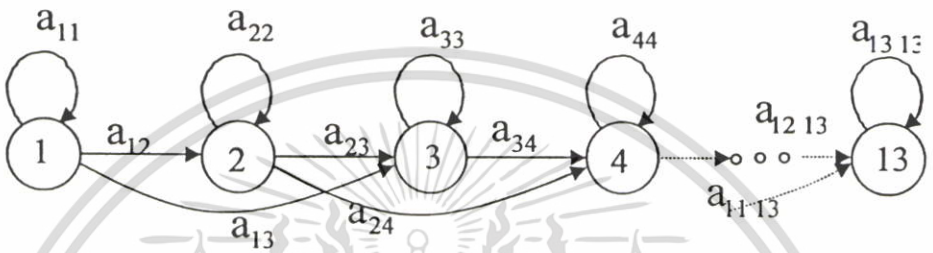


รูปที่ 5.6 เปอร์เซนต์ความถูกต้องเมื่อมีการเปลี่ยนแปลงสเตท และการย้ายข้ามสเตทของ HMM

จากรูปที่ 5.6 แสดงให้เห็นถึงแบบจำลองที่มีจำนวนสเตทเพิ่มมากขึ้นจะให้ผลการรู้จำที่แม่นยำขึ้นและมีค่าค่อนข้างคงที่ตั้งแต่สเตทที่ 17 ขึ้นไป โดยแบบจำลองที่สร้างจาก HMM 4 หรือ 5 สเตท ให้ผลการรู้จำแม่นยำน้อยที่สุด ทั้งนี้เนื่องจากจำนวนสเตทนั่นมีผลโดยตรงต่อค่าพารามิเตอร์ B (ค่าความน่าจะเป็นของค่าปรากฏที่สามารถเป็นไปได้ภายในสเตท {-1, 0, 1}) ซึ่งจะเป็นตัวกำหนดรายละเอียดของข้อมูล ดังนั้นถ้าสเตทมีจำนวนน้อยเมตริกซ์ B จะมีขนาดเล็ก ซึ่งนั่นหมายถึงรายละเอียดของข้อมูล ในแบบจำลองจะน้อยตามลงไปด้วยเป็นผลให้การรู้จำมีความแม่นยำลดลง ในขณะเดียวกันแบบจำ ลอง HMM ที่สร้างจากสเตทจำนวนมาก จากรูปจะสังเกตุได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตั้งแต่สเตทที่ 15 ถึง 20 ให้ค่าการรู้จำสูงและค่อนข้างคงที่ ทั้งนี้เป็นผลเนื่องมาจากข้อมูลอินพุท คือ ค่าการ ควอนไตซ์ความถี่มีระดับเพียงแค่ 3 ระดับซึ่งให้ค่ารายละเอียดของข้อมูลน้อยเกินไปเมื่อเทียบกับจำนวนสเตท ทำให้ไม่ว่าจะเพิ่มจำนวนสเตทขึ้นเท่าใดก็ไม่ทำให้ผลการรู้จำแม่นยำเพิ่มขึ้น ซึ่งจากการ ทดลอง แบบจำลองที่ให้ผลการรู้จำดีที่สุดอยู่ที่ 13 สเตท โดยมีรูปแบบของการย้ายข้ามสเตทได้สูง สุด 2 สเตท แสดงได้ดังรูปที่ 5.6 โดยให้ผลการแม่นยำสูงสุดถึง 94.20 % เมื่อทำการทดสอบกับเสียงต้นแบบเดิม



รูปที่ 5.7 แบบจำลอง HMM ที่เหมาะสมกับอัลกอริธึมที่พัฒนาขึ้น 13 สเตท

5.2.4 การสร้างแบบจำลองอ้างอิงเพื่อใช้ในการรู้จำระดับเสียงวรรณยุกต์

จากการทดลองในส่วนแรก เราจะได้ขั้นตอนในการหาพารามิเตอร์ที่ต้องการซึ่งก็คือรูปแบบการเปลี่ยนแปลงของค่าความถี่มูลฐาน (เป็นตัวกำหนดระดับเสียงวรรณยุกต์ในคำหรือพยางค์ในภาษาลาว) และรูปแบบของ HMM ที่เหมาะสมกับการสร้างแบบจำลองขึ้น

การทดลองในส่วนนี้จึงเป็นการสร้างแบบจำลองอ้างอิงที่ใช้ในการรู้จำระดับเสียง โดยใช้เสียงต้นแบบจำนวนทั้งสิ้น 1000 เสียง แบ่งออกเป็นแบบจำลองระดับเสียงสูง-ตก (high falling) กลาง-ตก (mid-falling) กลาง-ขึ้น (mid-rising) ต่ำ-ขึ้น (rising-tone) และกลาง (mid tone) ระดับเสียงละ 100 (20 ×5) เสียง โดยรูปแบบของคำที่ใช้เป็นคำพยางค์เดียว ดังแสดงไว้ในตารางที่ 5.2 ซึ่งเสียงต้นแบบที่ใช้ ได้จากผู้ออกเสียงจำนวนทั้งสิ้น 10 คน แบ่งเป็นชาย 5 คน และหญิง 5 คน ออกเสียงคำในตารางที่ 5.2 คนละ 1ครั้ง

ตารางที่ 5.2 กลุ่มคำที่ 2 ใช้ในการสร้างแบบจำลองอ้างอิง

กลาง[—]	สูง-ตก [\ ]	กลาง-ตก[∧]	ต่ำ-ขึ้น [/]	กลาง-ขึ้น [∨]
ยาว	ທີ	ເຈົ້າ	ป่า	สูบ
มี	โตะ	ย้าว	ย่า	ฝน
พอม	ทุ	มี้	ไห	ส่า
แซว	ลิด	ปั้น	แส่ว	ห้าว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.2 กลุ่มคำที่ 2 ใช้ในการสร้างแบบจำลองอ้างอิง (ต่อ)

ขา	เตะ	ง้า	ดั่ง	เถิ่ง
มิน	เจ้ง	ห้างๆ	ต่า	ทิน
ลาน	และ	ไอ้	ห่า	เธิอ
กอน	เกาะ	ป้า	สา	ซื้
ปา	ฝ่า	บ้าน	โถ	หยัง
นัด	ปั้น	อ้อย	ข่า	ผัน
หางๆ	ปัก	เก้า	สู่	สาม
ลิน	หมิด	ทั้งๆ	ด่า	ส่าน
ไฟ	อิด	ห้าอ	ม่างๆ	เส่ง
ปืน	เห่า	ได้	ด่วอ	โต
ขาย	สี	มื้	เป่า	ตู
ถิน	หิก	ห้า	บ่าย	ติ่
พู	หมู่	ไล่	บ่	เถ้า
ลา	โภะ	ม้า	ภาด	ถิ
ซาอ	โห	แล้อ	จี่	เป่า
ไล	ติอะ	มั้น	ตู่	เหื้อ

ผลที่ได้จากการทดลองในส่วนนี้ คือ แบบจำลองอ้างอิงจำนวน 5 แบบจำลอง ได้แก่

แบบจำลองอ้างอิง เสียง กลาง แทนด้วยพารามิเตอร์ $\lambda_1 = (A_1, B_1, \pi)$

แบบจำลองอ้างอิง สูง-ตก แทนด้วยพารามิเตอร์ $\lambda_2 = (A_2, B_2, \pi)$

แบบจำลองอ้างอิง กลาง-ตก แทนด้วยพารามิเตอร์ $\lambda_3 = (A_3, B_3, \pi)$

แบบจำลองอ้างอิง ต่ำ-ขึ้น แทนด้วยพารามิเตอร์ $\lambda_4 = (A_4, B_4, \pi)$

แบบจำลองอ้างอิง กลาง-ขึ้น แทนด้วยพารามิเตอร์ $\lambda_5 = (A_5, B_5, \pi)$

โดยรายละเอียดของค่าพารามิเตอร์ของแต่ละแบบจำลอง มีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองการรู้จำเสียง สูง-ตก (high falling) แทนด้วยพารามิเตอร์ $\lambda_1 = (A_1, B_1, \pi)$

ค่าความน่าจะเป็นในการย้ายข้ามสเตท : A_1

$\mathbf{A}_{13x13}$

$$\begin{bmatrix}
0.2862 & 0.3589 & 0.3549 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.2890 & 0.3616 & 0.3494 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.2996 & 0.3646 & 0.3358 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.3106 & 0.3670 & 0.3224 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3198 & 0.3697 & 0.3105 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3207 & 0.3711 & 0.3082 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3214 & 0.3723 & 0.3063 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3252 & 0.3756 & 0.2992 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3288 & 0.3763 & 0.2949 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3285 & 0.3771 & 0.2944 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3297 & 0.3789 & 0.2914 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4959 & 0.5041 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000
\end{bmatrix}$$

ค่า ความน่าจะเป็นในการเกิดค่าปรากฏในแต่ละสเตท : B_1

$\mathbf{B}^{T}_{3x13} =$

$$\begin{bmatrix}
0.0017 & 0.3851 & 0.3040 & 0.1877 & 0.1805 & 0.1025 & 0.0927 & 0.0513 & 0.0293 & 0.0206 & 0.0102 & 0.0096 & 0.0054 \\
0.9982 & 0.5444 & 0.6541 & 0.7848 & 0.8011 & 0.8176 & 0.8272 & 0.8286 & 0.8306 & 0.8593 & 0.8598 & 0.8602 & 0.8665 \\
0.0001 & 0.0705 & 0.0419 & 0.0275 & 0.0184 & 0.0799 & 0.0801 & 0.1201 & 0.1401 & 0.1201 & 0.1300 & 0.1302 & 0.1281
\end{bmatrix}$$

ค่าความน่าจะเป็นในการเป็นสเตทเริ่มต้น (π_{1xN})

$$\pi_{1x13} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองการรู้จำเสียง กลาง-ตก (mid-falling) แทนด้วยพารามิเตอร์ $\lambda_2 = (A_2, B_2, \pi)$

ค่าความน่าจะเป็นในการย้ายข้ามสเตท : A_2

A_{13x13} =

0.3581	0.3732	0.2687	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.3148	0.3738	0.3114	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.3540	0.3082	0.3378	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.3554	0.3154	0.3292	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.3002	0.3838	0.3160	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.3067	0.3800	0.2133	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4079	0.3794	0.2127	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3104	0.3765	0.3131	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3084	0.3838	0.3078	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3110	0.3557	0.3333	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2991	0.3637	0.3372
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4733	0.4267
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

ค่าความน่าจะเป็นในการเกิดค่าปรากฏในแต่ละสเตท : B_2

B^T_{3x13} =

0.0001	0.0003	0.0003	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003	0.0009	0.0073	0.5751
0.9082	0.3402	0.7801	0.7476	0.6116	0.6189	0.6533	0.6721	0.6900	0.7089	0.7346	0.8249	0.4198
0.0917	0.6595	0.2196	0.2522	0.3882	0.3810	0.3466	0.3278	0.3099	0.2908	0.2645	0.1678	0.0051

ค่าความน่าจะเป็นในการเป็นสเตทเริ่มต้น (π_{1xN})

π_{1x13} = [1 0 0 0 0 0 0 0 0 0 0 0 0]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองการรู้จำเสียง เสียง กลาง-ขึ้น (mid-rising) แทนด้วยพารามิเตอร์ $\lambda_3 = (A_3, B_3, \pi)$

ค่าความน่าจะเป็นในการย้ายข้ามสเตท : A_3

$\mathbf{A}_{13x13}$ =

$$\begin{bmatrix}
0.3048 & 0.3484 & 0.3468 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.3783 & 0.3846 & 0.2371 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.4188 & 0.3423 & 0.2389 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.4399 & 0.3362 & 0.2239 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4483 & 0.3294 & 0.2223 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4592 & 0.3231 & 0.2177 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4868 & 0.3125 & 0.2007 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4979 & 0.3028 & 0.1993 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5226 & 0.2890 & 0.1884 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5620 & 0.2522 & 0.1858 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.6246 & 0.2035 & 2.1719 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.8771 & 0.1229 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000
\end{bmatrix}$$

ค่าความน่าจะเป็นในการเกิดค่าปรากฏในแต่ละสเตท : B_3

$\mathbf{B}^{T}_{3x13}$ =

$$\begin{bmatrix}
0.0031 & 0.1228 & 0.2661 & 0.3719 & 0.3743 & 0.4001 & 0.3824 & 0.2902 & 0.2657 & 0.3574 & 0.4502 & 0.3785 & 0.0022 \\
0.9967 & 0.8769 & 0.7337 & 0.6280 & 0.6255 & 0.5996 & 0.6172 & 0.7089 & 0.7315 & 0.6350 & 0.5304 & 0.5608 & 0.5479 \\
0.0001 & 0.0003 & 0.0002 & 0.0001 & 0.0001 & 0.0003 & 0.0004 & 0.0009 & 0.0028 & 0.0076 & 0.0193 & 0.0607 & 0.4499
\end{bmatrix}$$

ค่าความน่าจะเป็นในการเป็นสเตทเริ่มต้น ($\pi_{1 x N}$)

$$\pi_{1x13} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองการรู้จำ เสียง ต่ำ-ขึ้น (rising-tone) แทนด้วยพารามิเตอร์ $\lambda_4 = (A_4, B_4, \pi)$

ค่า ความน่าจะเป็นในการย้ายข้ามสเตท : A_4

A_{13x13} =

$$\begin{bmatrix}
0.3188 & 0.3767 & 0.3045 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.3776 & 0.3115 & 0.3109 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.2656 & 0.4085 & 0.3259 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.3350 & 0.3751 & 0.2899 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3337 & 0.4210 & 0.2453 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3677 & 0.4093 & 0.2230 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3525 & 0.3997 & 0.2478 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3082 & 0.4816 & 0.2102 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3167 & 0.4268 & 0.2565 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.2512 & 0.3572 & 0.3916 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3207 & 0.3263 & 0.3529 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5201 & 0.4799 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000
\end{bmatrix}$$

ค่า ความน่าจะเป็นในการเกิดค่าปรากฏในแต่ละสเตท : B_4

B^T_{3x1} =

$$\begin{bmatrix}
0.0003 & 0.0011 & 0.0023 & 0.0016 & 0.0016 & 0.0028 & 0.0127 & 0.1381 & 0.2859 & 0.2336 & 0.0338 & 0.0032 & 0.1659 \\
0.9900 & 0.8692 & 0.7603 & 0.7764 & 0.7633 & 0.8113 & 0.8589 & 0.6957 & 0.4486 & 0.4899 & 0.6913 & 0.8749 & 0.5158 \\
0.0097 & 0.1297 & 0.2374 & 0.2220 & 0.2350 & 0.1859 & 0.1284 & 0.1662 & 0.2655 & 0.2765 & 0.2749 & 0.1219 & 0.3283
\end{bmatrix}$$

ค่า ความน่าจะเป็นในการเป็นสเตทเริ่มต้น (π_{1xN})

$\pi_{1x13} = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองการรู้จำเสียง กลาง (mid tone) แทนด้วยพารามิเตอร์ $\lambda_5 = (A_5, B_5, \pi)$

ค่า ความน่าจะเป็นในการย้ายข้ามสเตท : A_5

$\mathbf{A}_{10x10}$ =

$$\begin{bmatrix}
0.2878 & 0.3650 & 0.3172 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.2133 & 0.3954 & 0.3913 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.2717 & 0.3954 & 0.3329 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.3372 & 0.3680 & 0.2946 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3501 & 0.3540 & 0.2958 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3499 & 0.3571 & 0.2929 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3557 & 0.3525 & 0.2918 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3550 & 0.3523 & 0.2926 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3539 & 0.3622 & 0.2939 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3571 & 0.3294 & 0.3134 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.3879 & 0.4038 & 0.2083 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5560 & 0.4440 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000
\end{bmatrix}$$

ค่า ความน่าจะเป็นในการเกิดค่าปรากฏในแต่ละสเตท : B_5

$\mathbf{B}^{T}_{3x10}$ =

$$\begin{bmatrix}
0.0002 & 0.0075 & 0.0315 & 0.2676 & 0.3753 & 0.2634 & 0.1622 & 0.1976 & 0.2679 & 0.2004 & 0.0614 & 0.0098 & 0.0949 \\
0.9990 & 0.9910 & 0.9673 & 0.7321 & 0.6244 & 0.7363 & 0.8374 & 0.8020 & 0.7315 & 0.7991 & 0.9380 & 0.9898 & 0.8576 \\
0.008 & 0.0015 & 0.0012 & 0.0003 & 0.0003 & 0.0003 & 0.0004 & 0.0004 & 0.0006 & 0.0005 & 0.0006 & 0.0004 & 0.0475
\end{bmatrix}$$

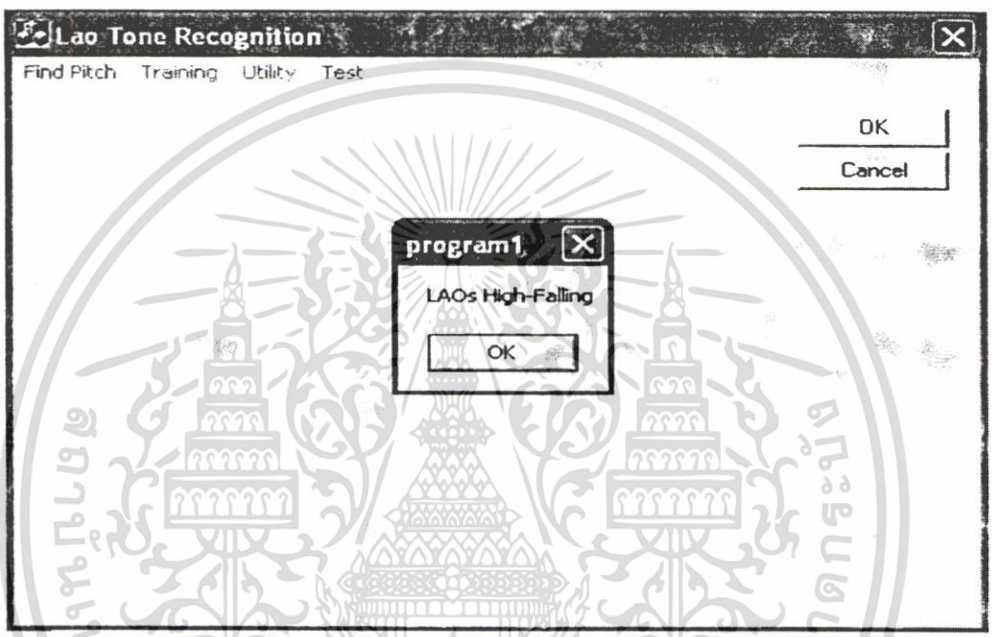
ค่า ความน่าจะเป็นในการเป็นสเตทเริ่มต้น (π_{1xN})

$\pi_{1x13} = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แสดงตัวอย่าง การแสดงผลของโปรแกรม ในการทดสอบการรู้จำ

ในรูปข้างล่าง เป็นตัวอย่างการทดสอบคำพยางค์เดียวที่มีเสียงวรรณยุกต์ โดยใช้คำตัวอย่างว่า " ป่า " ซึ่งเป็นระดับเสียง สูง-ตก (high falling) เมื่อนำเสียงนี้มาทำการทดสอบไปคำนวณโดยใช้ค่าพารามิเตอร์ของแบบจำลองอ้างอิงของเสียงวรรณยุกต์ทั้ง 5 แบบจำลอง ซึ่งจะได้ค่าความน่าจะเป็นจำนวน 5 ค่า โดยแบบจำลองอ้างอิงใดให้ค่าความน่าจะเป็นสูงสุดจะถือว่าเสียงที่นำมาทดสอบมีระดับเสียงวรรณยุกต์เดียวกันกับค่าปรากฏนั้น



รูปที่ 5.8 แสดงผลของโปรแกรมในการทดสอบการรู้จำ

5.3 การทดสอบแบบจำลองอ้างอิง

ในการทดสอบแบบจำลองอ้างอิงที่สร้างขึ้นนี้ ได้แบ่งการทดสอบออกเป็น 2 กลุ่ม คือ

5.3.1 ทดสอบกับกลุ่มคำต้นแบบ โดยใช้ เสียงต้นแบบเดิม

5.3.2 ทดสอบกับกลุ่มคำใหม่ โดยใช้ เสียงจากผู้ออกเสียงต้นแบบและใช้ เสียงจากผู้ออกเสียงกลุ่มใหม่

5.3.1 การทดสอบแบบจำลองอ้างอิงโดยใช้กลุ่มคำต้นแบบ

แบบจำลองอ้างอิงเสียงวรรณยุกต์ λ_1 - λ_5 ที่ถูกสร้างขึ้น จะถูกนำมาทดสอบการทำงานเพื่อพิสูจน์ว่าสามารถรู้จำระดับเสียงได้จริง โดยนำมาทดสอบกับกลุ่มคำต้นแบบและใช้เสียงต้นแบบเดิมซึ่งแต่ละระดับเสียงใช้จำนวนเสียงในการทดสอบเท่ากันคือ 100 เสียงจากผู้ออกเสียง 5 คน และจากการทดสอบ ให้ผลการรู้จำถูกต้องเฉลี่ย 94.24 เปอร์เซ็นต์ ดังแสดงไว้ในตารางที่ 5.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.3 ผลการทดสอบแบบจำลองอ้างอิง โดยใช้กลุ่มคำและเสียงต้นแบบจำนวน 500 เสียง

ผู้ออกเสียง	การรู้จำระดับเสียงถูกต้อง (%)					(%) ถูกต้อง
	กลาง	สูง-ตก	กลาง-ตก	ต่ำ-ขึ้น	กลาง-ขึ้น	เฉลี่ย / คน
M1	98	97	94	95	80	**92.80**
M2	95	96	93	94	82	**92.00**
W1	100	97	95	95	86	**94.60**
W2	100	99	95	96	88	**95.60**
W3	99	99	98	98	87	**96.20**
เฉลี่ย	**98.40**	**97.60**	**95.00**	**95.60**	**84.60**	**94.24**

หมายเหตุ M1- M2 แทนผู้ออกเสียงต้นแบบที่เป็นชายคนที่ 1-2

W1- W3 แทนผู้ออกเสียงต้นแบบที่เป็นหญิง คนที่ 1-3

5.3.2 การทดสอบแบบจำลองอ้างอิงโดยใช้กลุ่มคำใหม่

กลุ่มคำใหม่เป็นกลุ่มคำพยางค์เดียวจำนวนทั้งสิ้น 50 คำ ที่ไม่ซ้ำกับกลุ่มคำต้นแบบดังแสดงในตารางที่ 5.4 โดยถูกนำมาใช้ในการทดสอบ เพื่อพิสูจน์ว่าแบบจำลองอ้างอิงเสียงวรรณยุกต์ λ_1 - λ_5 ที่สร้างขึ้นนี้สามารถใช้ได้กับคำพยางค์เดียวทั่วไป และไม่จำกัดตัวบุคคล

ตารางที่ 5.4 กลุ่มคำที่ 3 ใช้ในการทดสอบแบบจำลองอ้างอิง

กลาง[—]	สูง-ตก [\ ]	กลาง-ตก[∧]	ต่ำ-ขึ้น [/]	กลาง-ขึ้น [∨]
ນາມ	ໂຕະ	ກ້າ	ດ່ຳ	ຂ່າຍ
ນິ	ຫົວ	ລ້ານ	ກອດ	ສູ່
ຄືນ	ທຸ	ພ້ຳ	ອ່ຳໆ	ຖ່ຳມ
ພາຍ	ລົດ	ບ້ຳ	ຈ່ຳ	ຈ່ຳ
ນຳ	ເຕະ	ຮູ້	ຢ່ຳ	ຖຳ
ໄປ	ເຈ້ຍ	ຮ້ານ	ຜ່ານ	ຖ່ຳຍ
ຊາຍ	ກົມ	ໄມ້	ຕ່ອນ	ໂຂໆ
ປີ	ຄ້ານ	ຟ້ຳ	ເພື່ອ	ຫ່ວໆ
ທຳ	ວຽກ	ຊ້ຳໆ	ກ່ຳ	ຖັນ
ມຳ	ໃສ່	ຍ້ຳຍ	ຊ໋	ຂອໆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยในการทดสอบ ได้แบ่งการทดสอบออกเป็น 2 กรณี คือ

5.3.2.1 ทดสอบโดยใช้เสียงจากผู้ออกเสียงต้นแบบ ออกเสียงตามตารางที่ 5.4 จำนวน 300 เสียง จากผู้ออกเสียง 6 คน เป็นชาย 3 คน และหญิง 3 คน ออกเสียงกลุ่มคำใหม่ จำนวน 50 คำ คนละ 1 ครั้ง โดยในการทดสอบให้ผลการรู้จำถูกต้องเฉลี่ย 92.47 เปอร์เซนต์ ดังแสดงไว้ในตารางที่ 5.5

5.3.2.2 ทดสอบโดยใช้เสียงจากผู้ออกเสียงกลุ่มใหม่ จำนวน 3 เสียง จากผู้ออกเสียง 6 คน เป็นชาย 3 คน และหญิง 3 คน ออกเสียงกลุ่มคำใหม่ในตารางที่ 5.4 จำนวน 50 คำ คนละ 1 ครั้ง โดยในการทดสอบ ให้ผลการรู้จำถูกต้องเฉลี่ย 90.67 เปอร์เซนต์ ดังแสดงไว้ในตารางที่ 5.6

ตารางที่ 5.5 ผลการรู้จำระดับเสียงวรรณยุกต์ จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงต้นแบบ

ผู้ออกเสียง	การรู้จำระดับเสียงถูกต้อง (%)					(%) ถูกต้อง
	กลาง	สูง-ตก	กลาง-ตก	ต่ำ-ขึ้น	กลาง-ขึ้น	เฉลี่ย / คน
M1	98	96	82	96	78	**90.00**
M2	98	98	88	98	80	**92.40**
M3	98	100	90	98	76	**92.40**
W1	100	98	84	98	80	**92.00**
W2	100	100	86	98	82	**93.20**
W3	100	100	90	100	84	**94.80**
เฉลี่ย	**99.00**	**98.67**	**86.67**	**98.00**	**80.00**	**92.47**

หมายเหตุ M1- M3 แทนเสียงจากกลุ่มผู้ออกเสียงต้นแบบที่เป็นชาย คนที่ 1-3
W1- W3 แทนเสียงจากกลุ่มผู้ออกเสียงต้นแบบที่เป็นหญิง คนที่ 1-3

ตารางที่ 5.6 ผลการรู้จำระดับเสียงวรรณยุกต์ จากการทดสอบโดยใช้เสียงจากผู้ออกเสียงกลุ่มใหม่

ผู้ออกเสียง	การรู้จำระดับเสียงถูกต้อง (%)					(%) ถูกต้อง
	กลาง	สูง-ตก	กลาง-ตก	ต่ำ-ขึ้น	กลาง-ขึ้น	เฉลี่ย / คน
NM1	98	98	80	96	68	**88.00**
NM2	100	98	88	98	70	**90.80**
NM3	100	100	86	98	70	**90.80**
NW1	100	100	88	98	72	**91.60**
NW2	100	98	86	100	74	**91.60**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.6 (ต่อ)

NW3	100	100	84	98	74	**91.20**
เฉลี่ย	**99.67**	**99.00**	**85.34**	**98.00**	**71.34**	**90.67**

หมายเหตุ NM1- NM3 แทนเสียง จากกลุ่มผู้ออกเสียงกลุ่มใหม่ที่เป็นชาย

NW1- NW3 แทนเสียง จากกลุ่มผู้ออกเสียงกลุ่มใหม่ที่เป็นหญิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
สรุปผล และข้อเสนอแนะ

วิทยานิพนธ์นี้เป็นการเสนอการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์ สำหรับภาษาลาวที่เป็นคำพยางค์เดียว เสียงของคนลาวภาคกลาง(เสียงคนเวียงจันทน์)แบบไม่จำกัดบุคคล โดยทำการวิเคราะห์ในโดเมนของเวลา และเนื่องจากหน่วยเสียงวรรณยุกต์เป็นหน่วยเสียงซ้อนวางตัวอยู่ในหน่วยเสียงก้อง ซึ่งคุณสมบัติเด่นที่สำคัญของเสียงก้องก็คือความเป็นคาบ ดังนั้นการวิเคราะห์ระดับเสียงในโดเมนของเวลาจึงเป็นการหาลักษณะการเปลี่ยนแปลงของคาบเวลา หรืออีกนัยหนึ่งคือหาแนวทางการเปลี่ยนแปลงของค่าความถี่มูลฐานในเสียงนั้นๆ การนำเสนอในงานวิจัยนี้ได้ทำตั้งแต่ขั้นตอนการหาค่าความถี่มูลฐานจากสัญญาณเสียงพูด จนถึงการสร้างแบบจำลองอ้างอิงในการรู้จำระดับเสียงด้วยฮิดเดนมาร์คอฟโมเดล และทำการทดสอบการรู้จำโดยใช้ทั้งกลุ่มคำต้นแบบที่ใช้ในการสร้างแบบจำลอง และกลุ่มคำใหม่มาทำการทดสอบ

6.1 การทดลอง

การทดลองแบ่งออกเป็น 3 ตอน คือ

6.1.1 การพัฒนาอัลกอริธึมที่ใช้ในการสร้างแบบจำลองการรู้จำระดับเสียงวรรณยุกต์

การหาค่าคาบพิทช์ของสัญญาณเสียง จะหาได้โดยใช้วิธีออโตโครีเลชั่นที่มีการคลิปสัญญาณ จากนั้นค่าพิทช์จะถูกเปลี่ยนให้อยู่ในรูปของค่าความถี่มูลฐาน รูปแบบการเปลี่ยนแปลงค่าความถี่มูลฐาน (หรือเรียกว่าเส้นทางเดินเสียงวรรณยุกต์) ที่แตกต่างกัน จะเป็นตัวบ่งบอกถึงระดับเสียงของวรรณยุกต์ที่แตกต่างกันในแต่ละคำหรือพยางค์ โดยแบบจำลองการรู้จำที่สร้างขึ้นมุ่งเน้นให้สามารถใช้ได้แบบไม่จำกัดเพศและตัวบุคคล ดังนั้นจึงใช้วิธีการคอวนไตซ์โดยทำการจัดกลุ่มค่าความถี่มูลฐานออกเป็น 3 ระดับ (-1 0 1) ตามแนวทางการเปลี่ยนแปลงของความถี่มูลฐานที่เพิ่มขึ้น คงที่ หรือลดลง ตามเวลาเปลี่ยนไป เพื่อที่จะกำจัดข้อจำกัดอันเนื่องมาจากความถี่มูลฐานที่แตกต่างกันระหว่างผู้ออกเสียงที่เป็นชายและหญิง อีกทั้งยังช่วยลดเนื้อที่ของหน่วยความจำในการจัดเก็บข้อมูลและทำให้การคำนวณทำได้เร็วขึ้นเมื่อเทียบกับการใช้ช่วงความถี่มูลฐานทั้งหมดมาสร้างแบบจำลอง

ทำการศึกษารูปแบบของ HMM ที่เหมาะสม ซึ่งจากการทดลองพบว่า HMM ขนาด 13 สเตทและมีการย้ายข้ามสเตทได้สูงสุดไม่เกิน 2 สเตท เป็นรูปแบบที่เหมาะสมกับอัลกอริธึมที่พัฒนาขึ้นมามากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.1.2 การสร้างแบบจำลองอ้างอิง

การเลือกคำต้นแบบจะพยายามใช้คำที่มี หน่วยเสียงพยัญชนะต้น ความสั้น-ยาวของหน่วยเสียงสระ และหน่วยเสียงพยัญชนะสะกด ที่แตกต่างกัน เพื่อให้ได้แบบจำลองอ้างอิงที่ครอบคลุมความหลากหลายมากที่สุด โดยใช้คำจำนวนทั้งหมด 100 คำ จากผู้ออกเสียง 10 คน เป็น ชาย 5 คน และหญิง 5 คน ออกเสียงเพื่อใช้เป็นข้อมูลฝึกสอนในการสร้างแบบจำลองอ้างอิงจำนวนทั้งสิ้น 1000 เสียง

6.1.3 ผลการทดสอบแบบจำลองอ้างอิงที่สร้างขึ้น

แบ่งการทดสอบออกเป็น 3 กรณี คือ ใช้กลุ่มคำต้นแบบ และเสียงต้นแบบเดิมที่นำมาสร้างแบบจำลอง ให้ผลการรู้จำถูกต้องเฉลี่ยคิดเป็นร้อยละ 94.24 และ ให้ผลการรู้จำในแต่ละระดับเสียง กลาง เสียง สูง-ตก กลาง-ตก ต่ำ-ขึ้น กลาง-ขึ้น และ คิดเป็นร้อยละ 98.40 , 97.60, 90.00, 95.60 และ84.00 ตามลำดับ

ใช้กลุ่มคำใหม่ ที่ออกเสียงโดยผู้ที่ออกเสียงต้นแบบ 6 คน เป็นชาย 3 คน หญิง 3 คน ให้ผลการรู้จำถูกต้องเฉลี่ยคิดเป็นร้อยละ 92.47 และให้ผลการรู้จำในแต่ละระดับเสียง เสียง กลาง สูง-ตก กลาง-ตก ต่ำ-ขึ้นและเสียงกลาง-ขึ้น คิดเป็นร้อยละ 99.00, 98.67, 86.67, 98.00 และ 80.00 ตามลำดับ

ใช้กลุ่มคำใหม่ ที่ออกเสียงโดยผู้ที่ออกเสียงกลุ่มใหม่ 6 คน เป็นชาย 3 คน หญิง 3 คน ให้ผลการรู้จำถูกต้องเฉลี่ยคิดเป็นร้อยละ 90.67 และผลการรู้จำในแต่ละระดับเสียง เสียง กลาง สูง-ตก กลาง-ตก ต่ำ-ขึ้นและเสียงกลาง-ขึ้น คิดเป็นร้อยละ 99.67, 99.00, 84.34, 98.00 และ71.34 ตามลำดับ

6.2 ข้อสังเกตปัญหาที่พบในการทดลองและข้อแสนอแนะ

การใช้วิธีการควอนไตซ์การเปลี่ยนแปลงค่าความถี่มูลฐานออกเป็น 3 ระดับ พบว่าให้ผลดีในการรู้จำระดับเสียงวรรณยุกต์แบบไม่จำกัดเพศและบุคคล โดยให้ผลการรู้จำเฉลี่ยมากกว่า 90 % ในทุกกรณี ซึ่งจากการทดลองพบว่า สำเนียง หรือท่วงทำนองในการออกเสียงมีส่วนสำคัญมากต่อการรู้จำหน่วยเสียงทั้ง 5 และจากการทดลองได้แบ่งระดับหน่วยเสียงวรรณยุกต์ออกเป็น 2 ประเภท คือ

หน่วยเสียงวรรณยุกต์เปลี่ยนระดับ ได้แก่ระดับเสียงกลาง-ตก และ กลาง-ขึ้น ซึ่งจากการทดสอบพบว่าหน่วยเสียงวรรณยุกต์เปลี่ยนระดับให้การรู้จำไม่ถูกต้องเมื่อเทียบกับ สามระดับเสียงที่เหลือ โดยผลการรู้จำที่ผิดพลาด สาเหตุหนึ่งมาจากการออกเสียงไม่ชัดเจน การผันเสียงไม่เป็นไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตามกับภาษาเขียน ทำให้การออกเสียงบางพยางค์ มีช่วงต้นเสียงหรือท้ายมีค่าคงที่ทำให้ เส้นทางเดินเสียงวรรณยุกต์มีลักษณะใกล้เคียงกับระดับเสียงที่เหลือทำให้การรู้จำไม่ถูกต้องเท่าที่ควร

หน่วยเสียงวรรณยุกต์คงระดับ ได้แก่ระดับเสียง สูง-ตก ต่ำ-ขึ้นและเสียงกลาง ซึ่งจากการทดลองพบว่า เมื่อออกเสียงเพี้ยนไปหรือออกเสียงเอื้อนในส่วนท้ายพยางค์มากไปจะทำให้การรู้จำผิดพลาด ซึ่งในการพัฒนาให้มีความแม่นยำการรู้จำสูงขึ้น อาจจะใช้ความแตกต่างของค่าความถี่มูลฐานในช่วงต้นพยางค์และช่วงท้ายของพยางค์เป็นพารามิเตอร์ร่วมในการรู้จำ ตัวอย่างเช่นในกรณีของเสียงกลางและเสียง สูง-ตก บางเสียงที่มีการเปลี่ยนแปลงค่าความถี่มูลฐานลดลงเหมือนกันแต่ในระดับเสียง สูง-ตกจะมีการลดลงในช่วงท้ายพยางค์ต่ำกว่าเสียงกลางมากและในเสียงต่ำ-ขึ้นช่วงท้ายพยางค์จะมีค่าความถี่มูลฐานสูงกว่าช่วงต้นพยางค์ ดังนั้นถ้านำเอาค่าความแตกต่างของความถี่มูลฐานในช่วงต้นและท้ายพยางค์มาเป็นพารามิเตอร์ร่วมในการรู้จำ คาดว่าอาจทำให้การรู้จำระดับเสียง สูง-ตก ต่ำ-ขึ้นและเสียงกลาง มีความแม่นยำสูงขึ้น

ดังที่ได้กล่าวมาแล้วคือสำเนียง หรือท่วงทำนองในการออกเสียงมีส่วนสำคัญมากต่อการรู้จำหน่วยเสียง ดังนั้นแบบจำลองนี้จึงใช้ได้แต่เฉพาะเสียงสำเนียงที่นำมาสร้างเป็นต้นแบบเท่านั้น (เสียงเวียงจันทน์ภาคกลาง) ส่วนสำเนียงภาคอื่นๆมาใช้กับแบบจำลองนี้ไม่ได้ แต่ด้วยวิธีเดียวกันนี้ก็สามารถนำไปใช้สร้างกับแบบจำลองเสียงภาษาลาวกับสำเนียงต่างๆได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] จิตรลดา จารุมิศร์ **การออกแบบ แบบจำลองในการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทย โดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling.** ปริญญานิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542.

[2] กรุณา แก้วสมศรี และคณะ. **การต่อหมายเลขโทรศัพท์โดยใช้เสียง.** ปริญญานิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมโทรคมนาคม, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2539.

[3] ธันวา ศรีประโมง. **การวิเคราะห์เสียงพูดภาษาไทยในแกนความถี่ฮาร์โมนิค.** วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2537.

[4] ณัฐกร ทับทอง. **การรู้จำคำพูดภาษาไทย โดยใช้ลักษณะบ่งความต่างของหน่วยเสียง.** วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2538.

[5] ทัศเวท วีระวัฒน์. **การรู้จำเสียงคำไทยเฉพาะบุคคล.** วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2541.

[6] มีทอง สุวันวิไช แบบเรียนภาษาลาวและวรรณกดี กระทรวงศึกษาธิการ สถาบันค้นคว้าวิทยาศาสตร์ การศึกษา แห่งชาติ. 1999

[7] R. Suzuki. **The Lao Language**. 1999. Tokyo: Express.(in Japanese), pp. 18-19.

[8] Rabiner L.R., Schafer R.W. **Digital Processing of Speech Signals.** New Jersey : Prentice-Hall, Inc. 1978.

[9] Rabiner L.R., Cheng M.J. et. Al. "**A Comparative Performance Study of Several Pitch Detection Algorithms.**", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-24, no.5, Oct. 1976. pp. 399-418.

[10] John R. Deller, Jr. HL, Hansen, G. Preakis. Discrete-Time Processing of Speech Signals. New York: Electrical and Electronics Engineer, Inc. 2000.

[11] A. M. Kondoz **Digital Speech**. John Wiley & Sons Ltd, Inc. 1982.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

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

[13] Thomas W. **Voice and Speech Processing**. Newyork : McGraw-Hill, Inc. 1987.

[14] Rabiner L.R., Juang B.H. **Fundamentals of Speech Recognition.** New Jersey : Prentice Hall, Inc. 1993.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

ผลงานวิจัยที่ได้รับการตีพิมพ์

ผลงานวิจัยเรื่อง " แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิค การควอนไตซ์พิทช์และ Hidden Markov Modeling" ได้รับตีพิมพ์ใน วิศวสาร ลาดกระบัง คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปีที่ 19 ฉบับที่ 4 ธันวาคม 2545

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลองเสียงวรรณยุกต์สำหรับภาษาลาว โดยใช้เทคนิค การควอนไตซ์พิทช์และ Hidden Markov Modeling

Tone Recognition Model for Lao Language Using Pitch Quantization and Hidden Markov Modeling Techniques

คำ ขันทะวีวอน ไกรสิน ส่งวัฒนา

ภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

บทความนี้นำเสนอวิธีการการสร้างแบบจำลองระดับเสียงวรรณยุกต์สำหรับภาษาลาว (เฉพาะเสียงพูดคนลาวที่เป็นเสียงเวียงจันทน์) ซึ่งเสียงพูดในภาษาลาวมี 5 ระดับด้วยกัน เสียงจากการบันทึกแต่ละคำได้ถูกแบ่งให้เป็นส่วนย่อยๆแล้วนำ แต่ละส่วนมาคำนวณหาคาบเวลาพิทช์โดยใช้วิธีออโตโครีเลชั่น ตามด้วยทำการแปลงคาบเวลาพิทช์มาเป็นค่าความถี่มูลฐาน ค่าความถี่มูลฐานที่ได้ถูกนำเข้ากระบวนการควอนไตซ์การเบี่ยงเบนเพื่อเป็นข้อมูลฝึกสอนสำหรับการสร้าง Hidden Markov Model (HMM) ซึ่งจะได้แบบจำลองเสียงวรรณยุกต์ออกมา เราได้ทำการทดลองตรวจสอบประสิทธิภาพการรู้จำระดับเสียง วรรณยุกต์ของแบบจำลองดังกล่าว โดยการป้อนข้อมูลเสียงที่ได้จากเพศชาย 5 คนและเพศหญิง 5 คนผลปรากฏมีความแม่นยำ เฉลี่ยมากกว่า 84 เปอร์เซนต์

Abstract

This paper presents a design of tone recognition modeling for spoken Laotian language (local Vientiane pronunciation). Laotian language has 5 tone levels. The spoken word is divided into frames. Then the auto-correlation method using center clipping is applied to each frame of speech to determine the pitch period and it fundamental frequency. The sequence of fundamental frequency is improved by making smoothing data using median filter. The observed sequence of pitch levels is preprocessed to find the pitch differences and the sequence of pitch differences are then grouped into three quantized levels. The resultant sequence is used as bases for training a Hidden Markov Model (HMM) and recognition of the 5 tones. A model for 5 tones was generated and tests were then conducted from the speech data of 5 males and 5 females to evaluate the capability of such model. The results showed an average accuracy of recognition above 84 percent.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.บทนำ

จากที่มีความต้องการให้เครื่องคอมพิวเตอร์สามารถเรียนรู้จำคำพูดได้ จึงทำให้เกิดศาสตร์แขนงหนึ่งขึ้น เรียกว่า Speech Recognition โดยใช้การสร้างแบบจำลองของคำพูดเพื่อสอนให้ระบบจำคำพูดเหล่านั้นเสียก่อน แต่การรู้จำเสียงพูดในภาษาลาวยังมีความหยุ่งยากซับซ้อน เนื่องจากเป็นภาษาที่มีระดับเสียงหลายระดับ โดยมีวรรณยุกต์เป็นตัวบังคับระดับเสียงของพยางค์ ซึ่งระดับเสียงที่แตกต่างกันนี้จะมีผลต่อความหมายของคำนั้นๆ ดังนั้นในการสร้างระบบการรู้จำคำพูดจึงจำเป็นต้องมีส่วนของการรู้จำวรรณยุกต์ด้วย เพื่อทำให้การรู้จำนั้นมีความเที่ยงตรงและแม่นยำสูงขึ้น เช่นเดียวกับการศึกษาการรู้จำระดับเสียงของภาษาจีน[1] และภาษาไทย [2]

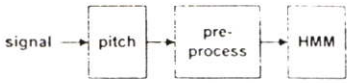
บทความวิจัยนี้ทำการศึกษาระบบการรู้จำเสียงของภาษาลาว 5 ระดับเสียง[3] ได้แก่ระดับเสียง

สูง-ตก (high falling) [\]
กลาง-ตก (mid-falling) [^]
กลาง-ขึ้น (mid-rising) [v]
กลาง (mid tone) [-] และ
ต่ำ-ขึ้น (rising-tone) [/]

จากสัญญาณข้อมูลที่ผ่านการแซมปลิ้งแล้วได้ถูกนำมาวิเคราะห์หาค่าความถี่มูลฐาน F_0 โดยใช้วิธีออโตโครีเลชั่น [4] เพื่อคำนวณหาคาบของสัญญาณเสียงอยู่ในรูปของค่าพิทช์ แล้วแปลงจากค่าพิทเป็นค่าความถี่มูลฐาน F_0 จากนั้นเราได้นำค่า F_0 ที่ได้มาผ่าน median filtering ด้วยวิธีการหาค่าการเปลี่ยนแปลงของความถี่ F_0 เทียบกับเวลา [5] แล้วนำค่าที่ได้ไปใช้เป็นข้อมูลสร้าง HMM ต่อไป

2 . ขั้นตอนการวิเคราะห์และสร้างแบบจำลองเสียงวรรณยุกต์

ในการวิเคราะห์และสร้างแบบจำลองเสียงวรรณยุกต์ได้แบ่งขั้นตอนออกเป็น 3 ขั้นตอนดังแสดงในรูปที่ 1 [2]



รูปที่ 1 ขั้นตอนในการวิเคราะห์

2.1 การหาค่าพิทช์

เริ่มจากนำสัญญาณเสียง ได้จากการแซมปลิ้งที่ความถี่ 11.025 kHz ผ่านขบวนการกรองความถี่ต่ำผ่านที่ความถี่ 900 Hz จากนั้นทำการแบ่งสัญญาณออกเป็นเฟรม (frame) เพื่อคำนวณหาพารามิเตอร์อยู่ในช่วงเวลานั้นๆ ออกมา ซึ่งแต่ละเฟรมกำหนดให้มีตัวอย่างสัญญาณ 300 แซมเปิล การวิเคราะห์ได้ทำทีละเฟรมโดยกำหนดให้มีช่วงของการเลื่อนเฟรมครั้งละ 100 แซมเปิล นั่นคือแต่ละเฟรมมีช่วงของการซ้อนทับกัน 2 ใน 3 เฟรม จากนั้นทำการหาค่าคาบเวลาพิทช์โดยใช้วิธี Modified Auto-correlation Method using Clipping (AUTOC) [6-7] ซึ่งสัญญาณ คลิบ (clip) อยู่ใน ช่วง 65 เปอร์เซนต์ ของ Absolute Amplitude Peak จากนั้นนำสัญญาณที่ผ่านขบวนการคลิบ คือค่า y(n) มาคำนวณออโตโครีเลชั่นตามสมการต่อไปนี้

$$R(k) = \sum_{n=0}^{N-1-k} y(n) \times y(n+k) \qquad (1)$$

เมื่อ k = การเลื่อนไปของเวลา

N = จำนวนข้อมูลในเฟรม 300

จากคุณสมบัติของออโตโครีเลชั่นฟังก์ชั่น ถ้าสัญญาณมีความเป็นคาบที่ระยะ P จะได้ว่า R(k) มีความเป็นคาบที่ระยะ P เช่นเดียวกันโดยค่าที่มากที่สุดของ R(k)จะเกิดที่ตำแหน่ง $k = 0, \pm P, \pm 2P, \ldots$ จากนั้นทำการหาตำแหน่งที่มี Auto-correlation Peak สูงสุดเมื่อเทียบกับ R(0) ซึ่งระยะที่ได้คือคาบเวลาพิทช์นั่นเอง โดยค่าที่ได้จะเอานำมาหาค่าความถี่มูลฐาน F_0 ได้จากความสัมพันธ์คือ

$$F_0 = \frac{F_s}{P} \qquad (2)$$

F_s = ความถี่ที่ใช้ในการสุ่มสัญญาณ ในบทความนี้ใช้ 11.025 kHz

P = คาบเวลาพิทช์

2.2 การ Pre-process ข้อมูล

ค่า F_0 ที่คำนวณได้จากการหาค่าพิทช์ในขั้นตอนแรกนั้นยังมีความไม่ต่อเนื่องของความถี่ ซึ่งอาจจะเกิดขึ้นจากการคำนวณหาค่าพิทช์ผิดพลาด ด้วยเหตุนี้จึงได้นำ median filter[8] มาใช้เพื่อแก้ปัญหาดังกล่าว เมื่อผ่านกระ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บวนการนี้จะได้ค่าความถี่มูลฐานชุดใหม่(F_0') ซึ่งจะนำมาหาค่าเปลี่ยนแปลงของความถี่ F_0' ตามเวลา โดยได้ทำการ ควอนไตซ์ออกเป็น 3 ระดับ คือ {-1, 0, 1}[2] ตามทิศทาง เปลี่ยนแปลงเพิ่มขึ้น คงที่หรือลดลงของ F_0' ค่าความเปลี่ยน แปลงที่ได้จะถูกนำไปใช้เป็นข้อมูล Training เพื่อสร้างแบบ จำลองของเสียงวรรณยุกต์ต่อไป

2.3 การสร้างแบบจำลองการรู้จำเสียงวรรณยุกต์โดยใช้เทคนิค Hidden Markov Model

แบบจำลอง HMM เป็นแบบจำลองทางสถิติ ซึ่งพัฒนามาใช้สำหรับจัดกลุ่มของสัญญาณที่ยังไม่รู้จัก ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่งแบบจำลอง HMM นี้ได้ถูกนำมาประยุกต์ใช้ในการรู้จำเสียงพูด[9]

ในบทความนี้ใช้ HMM แบบ Left-Right Model[9] เนื่องจากมีความเหมาะสม กับสัญญาณที่มีลักษณะการเปลี่ยนแปลงตามเวลาอย่างต่อเนื่องเช่นเสียงพูด โดยมีรูปแบบในการย้ายสถานะที่สามารถเป็นไปได้ดังรูปที่ 2

2.3.1 ส่วนประกอบของ HMM

1. N คือจำนวนสถานะในแบบจำลอง

ถ้าเราให้เซ็ตของสเตทเป็น $\{1, 2, \ldots, N\}$ ในบทความวิจัยนี้กำหนดให้ $N = 5$ [2] และแทนสถานะที่เปลี่ยนไปตามเวลา t ด้วยเซ็ตของ $Q = \{q_1, q_2, \ldots, q_N\}$

M คือจำนวนของเหตุการณ์ที่สามารถเป็นไปได้ใน 1 สถานะ แทนสัญลักษณ์ด้วย $V = \{v_1, v_2, \ldots v_M\}$ ซึ่งจากการจัดระดับการควอนไตซ์ออกเป็น 3 ระดับจะได้เซ็ตของเหตุการณ์ที่สามารถเป็นไปได้ในแต่ละสถานะมีค่าเป็น

$V = \{-1, 0, 1\}$ ค่าความน่าจะเป็นในการย้ายสถานะ $A = \{a_{ij}\}$

$a_{ij} = P[q_t = j \mid q_{t-1} = i]$ เมื่อ $1 \le i, j \le N$

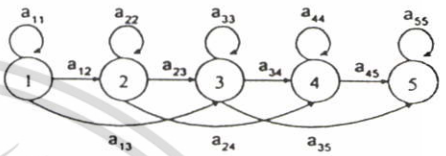
ค่าความน่าจะเป็นของเหตุการณ์ที่สามารถเป็นไปได้ภายในสถานะทนด้วยเมตริกซ์ B เมื่อ $B = \{b_j(k)\}$ โดยที่ $b_j(k) = P[v_k$ ที่เวลา $t \mid q_j$ ที่เวลา $t]$ เมื่อ $1 \le j \le N$ และ $1 \le k \le M$

ความน่าจะเป็นของการเป็นสถานะเริ่มต้น $\pi = \{\pi_i\}$

$\pi_i = P[q_i$ ที่เวลา $t = 1]$; $1 \le i \le N$

จะเห็นว่า Hidden Markov Model ต้องการพารามิเตอร์ของแบบจำลองคือ N , M และกลุ่มของความน่าจะเป็น A,B, π ดังนั้นในการแสดงเซ็ตของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองจะแทนด้วยสัญลักษณ์

$\lambda = (A, B, \pi)$



รูปที่ 2 Left-Right Model 5 state

ซึ่งแบบจำลองนี้มีคุณสมบัติในการย้ายสเตทดังนี้ ไม่มีการย้ายสเตทไปยังสเตทที่ต่ำกว่าและจะย้ายสถานะไปยังสถานะที่อยู่สูงกว่าได้ไม่เกิน Δi สถานะ

$a_{ij} = 0$; $j < i$

$a_{ij} = 0$; $i > i + \Delta i$

ดังรูป $\Delta i = 2$ คือจะไม่มีการย้ายข้ามสเตทเกิน 2 สถานะ มีค่าความน่าจะเป็นของสภานะเริ่มต้นคือ

$$\pi_i = \begin{cases} 1 & : i = 1 \\ 0 & : i \ne 1 \end{cases}$$

ที่ทุกๆความน่าจะเป็นจะมีค่าอยู่ในช่วง 0 ถึง 1 และผล รวมของความน่าจะเป็นในการย้ายสถานะจากสถานะใดๆจะต้องมีค่าเท่ากับ 1 เสมอคือ

$0 \le a_{ij} \le 1$; $\forall i, j$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad ; \quad \forall i$$

โดยในการสร้างและทดสอบแบบจำลองแบ่งการพิจารณาออกเป็น 3 ขั้นตอนคือ

1. การสร้างแบบจำลอง โดยใช้วิธี Forward-Backward Procedure และ Baum-Welch[9] โดยลำดับข้อมูลที่ได้จากควอนไตซ์จากการวิเคราะห์เสียง จะถูกนำมาเข้ากระบวนการสร้างแบบจำลอง เรียกลำดับความถี่นี้ว่า "ลำดับเทรนนิ่ง" โดยลำดับเทรนนิ่งนี้จะถูกนำมาทำการคำนวณหาพารามิเตอร์ของแบบจำลองที่เหมาะสมกับเสียงนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

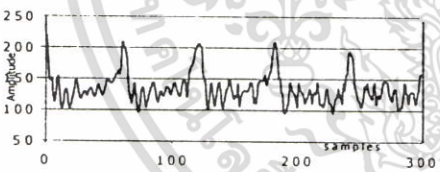
2. กำหนดลำดับสถานะ ที่เป็นส่วนพยายามจะระบุสถานะให้กับแต่ละลำดับเทรนนิ่งของคำโดยใช้ Viterbi Algorithm [8] เพื่อให้แบบจำลองมีความสามารถจำลองเสียงที่พูดเข้าได้

3. เป็นขั้นตอนในการนำเสียงที่ต้องการทดสอบ มาเทียบกับแบบจำลองของเสียงที่มีอยู่ทั้งหมด ซึ่งเสียงที่นำมาทดสอบต้องผ่านการวิเคราะห์ให้อยู่ในรูปของความถี่ ที่ถูกควอนไตซ์เสียก่อน แล้วจึงนำมาเทียบกับแบบจำลองเพื่อดูความน่าจะเป็นว่าเสียงที่นำมาทดสอบเป็นเสียงใด

3. ผลการทดลอง

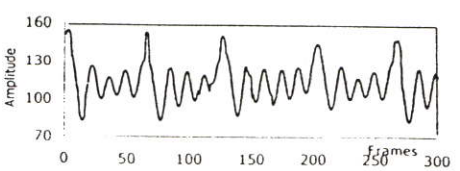
ทำการวิเคราะห์เสียงคำพยางค์เดียวจำนวน 50 คำ จากคำที่พูดออก 5 ระดับเสียงคือ สูง-ตก (high falling) กลาง-ตก(mid-falling) กลาง-ขึ้น(mid-rising) กลาง(mid tone) และ ต่ำ-ขึ้น (rising-tone) นำทั้ง 5 ระดับเสียงจากเพศหญิง 5 คนและเพศชาย 5 คน มาสร้างแบบจำลองตามลักษณะการเปลี่ยนแปลงความถี่ของระดับเสียงทั้ง 5

รูปที่ 3 แสดงลักษณะสัญญาณเสียงที่ได้จากการแซมปลิ้งเสียง "บา" ของผู้หญิง โดยในการทดลองใช้ความถี่แซมปลิ้ง 11.025 kHz

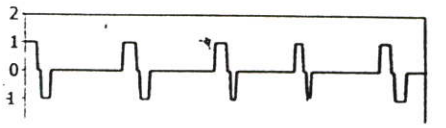


รูปที่ 3 สัญญาณที่ได้จากการ แซมปลิ้ง

เมื่อนำสัญญาณผ่าน Low-pass filter จะได้สัญญาณที่มีลักษณะเรียบขึ้นดังรูปที่ 4 จากนั้นนำสัญญาณผ่านขบวน Clipping Level ของสัญญาณ โดยใช้ 65% ของ Absolute Amplitude Peak ในแต่ละเฟรมดังรูปที่ 5

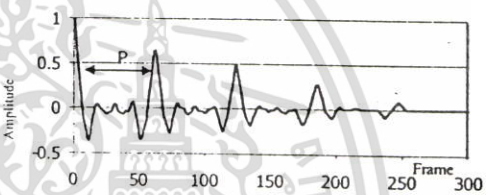


รูปที่ 4 สัญญาณที่ผ่าน lowpass filter



รูปที่ 5 แสดงจุด Clipping Level ของสัญญาณ

จากนั้นนำสัญญาณที่ผ่านการคลิบมากกว่า 65% ไปทำการคำนวน Auto-correlation Function ตามสมการที่(1) จะได้สัญญาณที่มีลักษณะดังรูปที่ 6 ซึ่งระยะห่างระหว่างจุดยอดที่สูงที่สุดกับจุดสูงสุดถัดไปคือค่าพิทช์ โดยสามารถนำมาหาค่าความถี่มูลฐานได้ตามสมการที่(2) และจากรูปที่ 6 ค่าความถี่มูลฐานช่วงที่ P มีระยะเท่ากับ 68 หน่วยเท่ากับ 11025/68=162.13 Hz



รูปที่ 6 ค่าที่ผ่านการคำนวณ Normalized Autocorrelaion Function

รูปที่ 7 แสดงเส้นกราฟการเปลี่ยนแปลงความถี่ F_0 ของ 5 ระดับเสียง (สูง-ตก กลาง-ตก กลาง-ขึ้น กลาง ต่ำ-ขึ้น) นำค่าความถี่ F_0 ผ่าน median filter ค่า F_0 จะถูกปรับให้เรียบขึ้น โดยค่าไม่ต่อเนื่องในช่วงต้นจะถูกกำจัดออกดัง ที่แสดงในรูปที่ 8 และรูปที่ 9 แสดงตัวอย่างของ 5 ระดับเสียงที่ผ่านการควอนไตซ์แล้วจะเห็นได้ว่า ค่าความถี่มูลฐานถูกแบ่งออกเป็น {-1, 0, 1} ตามการเปลี่ยนแปลงของความถี่ที่มีการเปลี่ยนแปลงตามแต่ละระดับเสียงดังนี้

ระดับเสียง "สูง-ตก" จะมีค่าเป็น -1 จำนวนมากสลับกับ 0 จากเฟรมแรกถึงเฟรมสุดท้ายของเสียง

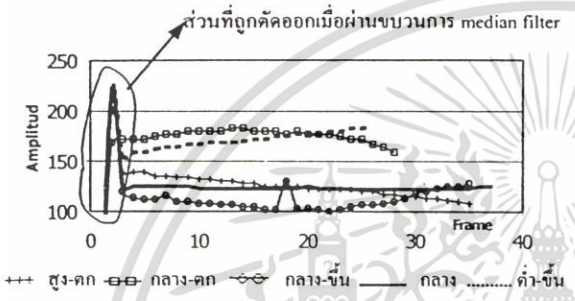
ระดับเสียง "กลาง-ตก" จะมีค่า 1 จำนวนมากสลับกับ 0 ถึงระยะกลางของจำนวนเฟรมเสียงและมีค่า 0 สลับกับ -1 จำนวนมากแต่ระยะกลางถึงระยะสุดท้ายของจำนวนเฟรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

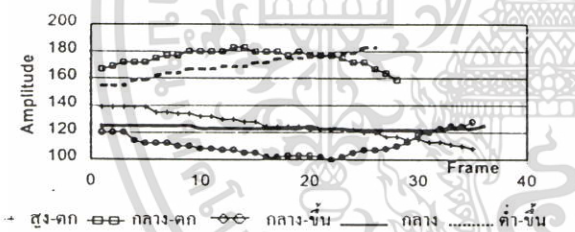
ระดับเสียง "กลาง-ขึ้น" จะมีค่า -1 จำนวนมากสลับกับ 0 ถึงระยะกลางของจำนวนเฟรมเสียงและมีค่า 0 สลับกับ 1 แต่ระยะกลางถึงระยะสุดท้ายของจำนวนเฟรมเสียง

ระดับเสียง "กลาง" จะมีค่า 0 จำนวนมากแต่เฟรมแรกถึงเฟรมสุดท้าย

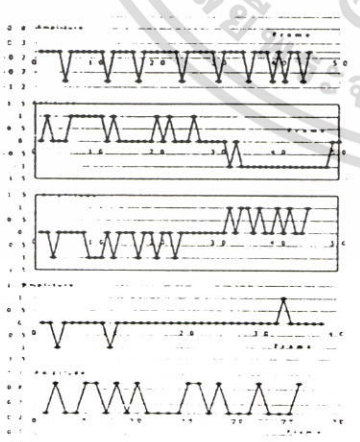
ระดับเสียง "ต่ำ-ขึ้น" จะมีค่าเป็น 0 สลับกับ 1 จำนวนมากแต่เฟรมแรกถึงเฟรมสุดท้ายของเสียง



รูปที่ 7 ลักษณะของการเปลี่ยนแปลงความถี่มูลฐานในวรรณยุกต์ทั้ง 5 เสียง



รูปที่ 8 ความถี่มูลฐานที่ผ่าน median filtering



รูปที่ 9 เสียงวรรณยุกต์ทั้ง 5 เสียงที่ผ่านการควอนไตซ์

4. การทดลองสร้างแบบจำลองและผลการทดสอบ

จากข้อมูลที่ผ่านขบวนการควอนไตซ์พิทช์ เราได้นำเอามาทำการสร้าง เป็นแบบจำลองเสียงวรรณยุกต์ภาษาลาว โดยนำใช้เสียงคนเวียงจันทน์ มาทำการสร้าง

การทดลองสร้างแบบจำลอง และผลของการทดสอบแบ่งออกเป็น 3 กรณี

1. นำเสียงต้นแบบของชาย 5 คน ที่ออกเสียงพูดครบ 5 ระดับ ระดับละ 5 คำ คนละ 25 คำ รวม 125 คำ มาทำการสร้างเป็นแบบจำลองที่ 1 และทำการทดสอบด้วยเสียงเดิมแทนด้วย 5M ในตารางที่ 1 ผลที่ได้คือความถูกต้องเฉลี่ยต่อจำนวนคำทดสอบ 91.20 เปอร์เซนต์ โดยมีระดับเสียง สูง-ตก ถูกต้องมากที่สุด 100 เปอร์เซนต์ และ เสียง กลาง-ขึ้น มีความถูกต้องน้อยที่สุดเท่า 84 เปอร์เซนต์

2. นำเสียงของหญิงเท่ากับจำนวนเสียงของชาย 5 คนดังที่ได้กล่าวมา ทำการสร้างเป็นแบบจำลองที่ 2 และทำการทดสอบโดยใช้เสียงต้นแบบเดิมของหญิง 5 คน แทนด้วย 5F ในตารางที่ 1 ผลที่ได้คือความถูกต้องเฉลี่ยต่อจำนวนคำที่ทดสอบ 92.80 เปอร์เซนต์ โดยมีระดับเสียง สูง-ตก ถูกต้องมากที่สุด 100 เปอร์เซนต์ และเสียง กลาง-ขึ้น มีความถูกต้องน้อยที่สุดเท่ากับ 84 เปอร์เซนต์

3. นำเสียงของหญิง 3 คน และชาย 2 คน ที่ออกเสียงครบ 5 ระดับ รวมหญิง 75 และชาย 50 คำ มาทำการสร้างเป็นแบบจำลองที่ 3 และทำการทดสอบ โดยใช้เสียงต้นแบบของหญิง 3 คน ชาย 2 คน แทนด้วย 3F + 2M ในตารางที่ 1 ผลที่ได้คือ ความถูกต้องเฉลี่ยต่อจำนวนคำทดสอบ 94.40 เปอร์เซนต์ โดยมีระดับเสียง สูง-ตก ถูกต้องที่สุดและเสียง กลาง-ขึ้นมีความถูกต้องน้อยที่สุดเท่ากับ 88 เปอร์เซนต์

ตารางที่ 1

ผู้ออกเสียง	(%) การรู้จำเสียงถูกต้องต่อจำนวนเสียงทดสอบกับเสียงแต่ละระดับ					(%)ถูกต้องเฉลี่ย
	[\]	[^]	[v]	[-]	[/]	
5M	100	92.00	84.00	88.00	92.00	91.20
5F	100	92.00	84.00	92.00	96.00	92.80
3F+2M	100	100	88.00	92.00	92.00	94.40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลได้รับจากการทดสอบแบบจำลองเสียงเห็นว่าค่าความถูกต้องของเสียงแต่ละระดับไม่เท่ากัน โดยแบบจำลองมีความถูกต่อระดับเสียงที่ทำการทดสอบ 3 กรณีที่กล่าวมาข้างต้น ระดับเสียงถูกมากที่สุดคือ สูง-ตก มีความถูกต้อง 100 เปอร์เซนต์ ต่อกับเสียงที่นำมาทดสอบ และที่มีความถูกต้องน้อยที่สุดคือระดับเสียง กลาง-ขึ้น เพราะว่าระดับเสียง สูง-ตก และ กลาง-ขึ้น มีส่วนต้นของข้อมูลคล้ายกัน ดังนั้นจึงทำให้เสียงพูดที่ออกระดับเสียง กลาง-ขึ้น ไม่ชัดเจนจึงออกมาผิดพลาด ซึ่งเป็นปัญหาที่จะต้องทำการแก้ไขต่อไป

5. สรุป

จากการวิเคราะห์ลักษณะของเสียงวรรณยุกต์พบว่าการเปลี่ยนแปลงความถี่มูลฐานมีลักษณะเฉพาะตัว สามารถนำมาสร้างแบบจำลองในการรู้จำเสียงวรรณยุกต์ แต่ละเสียงได้ การนำ median filter มาใช้นอกจากช่วยลดความไม่ต่อเนื่องของความถี่แล้ว ยังช่วยปรับความเรียบความถี่ในช่วงต้นเสียงด้วย การสร้างแบบจำลองจากการจัดระดับของการเปลี่ยนแปลงความถี่ทำให้แบบจำลองนี้สามารถใช้ได้ทั้ง ผู้พูดที่เป็นเพศชายและเพศหญิง จากการทดสอบการรู้จำจะพบว่าแบบจำลองของกลุ่มทดสอบในทุกกรณี ให้เปอร์เซนต์ความแม่นยำในการรู้จำมากกว่า 84 เปอร์เซนต์ โดยแบบจำลองสร้างจากผู้พูดหลายคนจะให้ผลการรู้จำ แม่นยำมากกว่าแบบจำลองที่สร้างจากผู้พูดเพียงคนเดียว เนื่องจากมีความหลากหลายของข้อมูลต้นแบบมากกว่า

นอกจากนี้การจัดแบ่งข้อมูลออกเป็น 3 ระดับยัง ทำให้การประมวลผล ของแบบจำลองลดเวลาที่ต้องใช้ในการคำนวณแบบจำลองอีกด้วย จากผลการทดลองสร้างแบบจำลองที่ใช้ในการรู้จำระดับเสียงวรรณยุกต์ จากเสียงต้นแบบ 25 เสียง ใช้เวลาประมาณ 15 วินาทีโดยการประมวลผลบนเครื่องคอมพิวเตอร์ PC Pentium III (800 MHz)

การสร้างแบบจำลองและผลที่ได้ในบทความนี้ เป็นเฉพาะแต่เสียงคนลาวที่เป็นเสียงเวียงจันทน์ ภาคกลางเท่านั้น ส่วนเสียงคนลาวในภาคต่างๆอาจจะใช้กับแบบจำลองนี้ไม่ได้ แต่ด้วยวิธีการเดียวกันนี้ สามารภพัฒนาสร้างแบบจำลองของเสียงเฉพาะของสำเนียงใช้กับแต่ละภาคนั้นๆได้ต่อไป

6. เอกสารอ้างอิง

[1] W.J. Yang, J.C. Lee, Y.C. Chang and H.C. Wang, "Hidden Markov Model for Mandarin Lexical Tone Recognition," IEEE Trans. Acoust., Speech, and Signal Processing, vol. 36, pp.988-992, July 1988.

[2] จิตรลดา จารุมิศร "แบบจำลองเสียงวรรณยุกต์สำหรับภาษาไทย โดยใช้เทคนิคการควอนไตซ์พิทช์และ Hidden Markov Modeling" งานประชุมวิชาการทางวิทยาการคอมพิวเตอร์และวิศวกรรมคอมพิวเตอร์แห่งชาติ 2541 คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

[3] R. Suzuki "The Lao Language". 1999. Tokyo: Express.(in Japanese), pp. 18-19.

[4] ธันวา ศรีประโมง "การวิเคราะห์เสียงพูดภาษา ไทยในแกนความถี่ฮาร์โมนิค" วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย สถาบันเทคโนโลยี พระจอมเกล้า เจ้าคุณทหารลาดกระบัง 2537

[5] อภิชาติ ตั้งทางธรรม "การเปลี่ยนความเร็วของ เสียงพูด" การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 17 พ.ศ. 2537

[6] L.R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," IEEE Trans. Acoust., Speech, and Signal Processing , vol. ASSP-25, pp. 24- 33, Feb. 1977.

[7] L.R. Rabiner and R.W. Schafer, "Digital Process of Speech Signal," New jersey: Prentice Hall, 1978.

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

[9] L.R. Rabiner and B.H. Juang, "Fundamental of Speech Recognition," New jersey: Prentice Hall, 1993.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

นาย คำ ขันทะวีวอน เกิดเมื่อวันที่ 18 ตุลาคม 2519 ที่เวียงจันทน์ ประเทศลาว จบการศึกษาระดับอนุปริญญาวิศวกรรมศาสตร์บัณฑิต สาขาวิชาอิเล็กทรอนิกส์ ภาควิชาอิเล็กทรอนิกส์ มหาลัยแห่งชาติลาว ปี พ.ศ. 2540 และ จบการศึกษาระดับปริญญาวิศวกรรมศาสตร์บัณฑิต สาขาวิชาอิเล็กทรอนิกส์ ภาควิชาอิเล็กทรอนิกส์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (KMITL: King Mongkut's Institute of Technology at Ladkrabang) ปี พ.ศ. 2544

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้