

การรู้จำเสียงตัวเลขภาษาไทย โดยการวิเคราะห์เสียงสระและเสียงวรรณยุกต์  
ด้วยวิธีความเข้มแถบวิกฤติและควอนไทซ์พิทช์

NUMERIC SPEECH RECOGNITION FOR THAI SPOKEN LANGUAGE BY  
ANALIZING VOWELS AND TONES USING CRITICAL BAND INTENSITY  
AND QUANTIZED PITCH

ภักดี จันทรเกษ  
PAKDEE CHANTRAKET

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมโทรคมนาคม

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2550

**สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง**

**การรู้จำเสียงตัวเลขภาษาไทย โดยการวิเคราะห์เสียงสระและเสียงวรรณยุกต์  
ด้วยวิธีความเข้มแถบวิกฤติและควอนไทซ์พิทช์**

**NUMERIC SPEECH RECOGNITION FOR THAI SPOKEN LANGUAGE BY  
ANALIZING VOWELS AND TONES USING CRITICAL BAND INTENSITY  
AND QUANTIZED PITCH**



**ภักดี จันทร์เกษ**

**PAKDEE CHANTRAKET**

**วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต**

**สาขาวิชาวิศวกรรมโทรคมนาคม**

**บัณฑิตวิทยาลัย**

**สถาบันเทคโนโลยีเจ้าคุณทหารลาดกระบัง**

**พ.ศ. 2550**

**NUMERIC SPEECH RECOGNITION FOR THAI SPOKEN LANGUAGE BY  
ANALIZING VOWELS AND TONES USING CRITICAL BAND INTENSITY  
AND QUANTIZED PITCH**

**PAKDEE CHANTRAKET**

**A THESIS SUBMITTED IN PARTIAL FULFILMENT  
OF THE REQUIREMENT FOR THE DEGREE OF  
MASTER OF ENGINEERING IN TELECOMMUNICATION ENGINEERING  
SCHOOL OF GRADUATE STUDIES  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**2007**

**COPYRIGHT 2007**

**SCHOOL OF GRADUATE STUDIES**

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

|                             |   |
|-----------------------------|---|
| หัวข้อวิทยานิพนธ์           | การรู้จำเสียงตัวเลขภาษาไทย โดยการวิเคราะห์เสียงสระและเสียงวรรณยุกต์ ด้วยวิธีความเข้มแถบวิกฤติ และควอน ไตซ์พิทซ์ |
| นักศึกษา                    | นายภักดี จันทรเกษ   |
| รหัสประจำตัว                | 45061090  |
| ปริญญา                      | วิศวกรรมศาสตรมหาบัณฑิต  |
| สาขาวิชา                    | วิศวกรรมโทรคมนาคม   |
| พ.ศ.                        | 2550  |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์ | รศ. ดร. ไกรสิน ส่งวัฒนา   |

### บทคัดย่อ

งานวิจัยนี้เป็นการนำเสนอวิธีการรู้จำตัวเลขภาษาไทย โดยการจำแนกเสียงสระและเสียงวรรณยุกต์ ด้วยวิธี K-Nearest Neighborhood (KNN) และ Hidden Markov Modeling (HMM) โดยจำแนกเสียงสระ จำนวน 9 เสียง และเสียงวรรณยุกต์ จำนวน 2 เสียง คลื่นเสียงของตัวเลขภาษาไทย จะผ่านกระบวนการพิจารณาใน 2 ขั้นตอนก่อนเข้าสู่กระบวนการตัดสินใจ ในขั้นตอนแรกจะพิจารณาเสียงสระ โดยคลื่นเสียงจะถูกแทนด้วยเวกเตอร์คุณสมบัติขนาด 18 มิติของความเข้มแถบวิกฤติ (Critical Band Intensity : CBI) เพื่อจำแนกเสียงสระด้วยวิธี KNN และการพิจารณาเสียงวรรณยุกต์ของสัญญาณเสียงสระ “อา” โดยการวิเคราะห์คาบพิทซ์เพื่อนำมาหาค่าความถี่มูลฐาน เพื่อให้ได้ค่าควอน ไตซ์พิทซ์ผ่านกระบวนการรู้จำ โดยใช้วิธี HMM ผลลัพธ์ของการจำแนกเสียงสระและเสียงวรรณยุกต์จะผ่านกระบวนการตัดสินใจเพื่อจำแนกเสียงตัวเลขภาษาไทยต่อไป



## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ดีจากการให้คำแนะนำ คำปรึกษา และความช่วยเหลือ จาก รศ.ดร. ไกรสิน สงวัฒนา ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้เขียนจึงขอขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

- ขอขอบคุณ คุณพ่อ คุณแม่ พี่สาว และพี่ชาย ที่ให้การสนับสนุน ส่งเสริม คอยให้กำลังใจเป็นห่วงเป็นใย ช่วยเหลือดูแลด้วยดีตลอดมา

- ขอขอบคุณอาจารย์ทุกท่านที่ให้คำแนะนำและให้คำปรึกษา ตลอดระยะเวลาที่ศึกษาอยู่

- ขอขอบคุณพี่ๆ เพื่อนๆ น้องๆ ทุกท่านทั้งที่ห้องวิจัย ที่ทำงาน และที่เรียนด้วยกัน ที่คอยเป็นกำลังใจ ดูแลเอาใจใส่อย่างสม่ำเสมอ

- สุดท้ายขอขอบคุณเจ้าของเสียงที่เสียสละเวลามานับทีกเสียงให้

คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบแด่ผู้มีพระคุณทุกท่าน

ภักดี จันทร์เกษ

# สารบัญ

|   | หน้า |
|---|------|
| บทคัดย่อภาษาไทย.....                              | I    |
| บทคัดย่อภาษาอังกฤษ.....                           | II   |
| กิตติกรรมประกาศ.....                              | III  |
| สารบัญ.....                                       | IV   |
| สารบัญตาราง.....                                  | VIII |
| สารบัญรูป.....                                    | IX   |
| บทที่ 1 บทนำ.....                                 | 1    |
| 1.1 ความเป็นและความสำคัญของปัญหา.....             | 1    |
| 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....   | 2    |
| 1.3 สมมุติฐานของการศึกษา.....                     | 2    |
| 1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย.....          | 3    |
| 1.5 ขอบเขตการวิจัย.....                           | 4    |
| 1.6 ขั้นตอนการศึกษา.....                          | 4    |
| บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....        | 5    |
| 2.1 กล่าวนำ.....                                  | 5    |
| 2.2 ทฤษฎีภาษาศาสตร์.....                          | 5    |
| 2.2.1 อวัยวะออกเสียง (The Organs of Speech) ..... | 6    |
| 2.2.2 กระบวนการผลิตเสียงพูด.....                  | 9    |
| 2.2.3 เสียงพูดของมนุษย์.....                      | 10   |
| 2.2.4 ความยาวของเสียง (Length) .....              | 11   |
| 2.2.5 ระดับเสียง สูง – ต่ำ (Pitch) .....          | 11   |
| 2.2.6 ความดัง (Loudness) .....                    | 11   |
| 2.2.7 การลงน้ำหนัก (Stress) .....                 | 11   |
| 2.2.8 ช่วงต่อของเสียง (Juncture) .....            | 11   |

# สารบัญ

หน้า

|   |    |
|---|----|
| 2.3 หน่วยเสียงสำคัญในภาษาไทย.....                                       | 12 |
| 2.3.1 หน่วยเสียงสระ.....  | 12 |
| 2.3.2 หน่วยเสียงพยัญชนะ.....  | 14 |
| 2.3.3 หน่วยเสียงวรรณยุกต์.....  | 16 |
| 2.4 ลักษณะ พยางค์ และคำในภาษาไทย.....                                   | 17 |
| 2.5 กระบวนการรู้จำเสียงพูด.....   | 18 |
| 2.6 แบบจำลองระบบกำเนิดเสียงพูด.....                                     | 19 |
| 2.6.1 แบบจำลองระบบกำเนิดเสียงพูดเชิงเลข.....                            | 20 |
| 2.6.2 พารามิเตอร์ (Parameter) ของระบบกำเนิดเสียงพูด.....                | 20 |
| 2.7 การวิเคราะห์สัญญาณเสียงพูดเบื้องต้น.....                            | 21 |
| 2.7.1 กรรมวิธีการปรับบรรทัดฐานแอมพลิจูด (Amplitude Normalization) ..... | 22 |
| 2.7.2 กรรมวิธีการเน้นล่วงหน้า (Pre-emphasis) .....                      | 22 |
| 2.7.3 การแบ่งช่วงสัญญาณ (Frame Blocking) .....                          | 23 |
| 2.7.4 การวินโดว์ (Windowing) .....                                      | 24 |
| 2.8 หลักการของการประมาณเชิงเส้น.....                                    | 25 |
| 2.9 แถบความถี่วิกฤต และความเข้มสัญญาณในแถบความถี่วิกฤต.....             | 31 |
| 2.9.1 วิธีการในการหาความกว้างแถบความถี่วิกฤต.....                       | 32 |
| 2.9.2 Critical Band Rate Scale.....                                     | 37 |
| 2.9.3 Critical Band Intensity.....                                      | 40 |
| 2.10 การหา K-Nearest Neighbor.....                                      | 41 |
| 2.10.1 ความหมายของ Nearest Neighbor.....                                | 41 |
| 2.10.2 การหา Nearest Neighbor โดยตรง.....                               | 42 |
| 2.10.3 K-Nearest Neighbor.....  | 43 |
| 2.11 ทฤษฎีการประมาณค่าพิทซ์โดยใช้ฮอโตคอร์รัเลชันฟังก์ชัน.....           | 44 |
| 2.12 การสร้างแบบจำลองการรู้จำด้วยวิธี Hidden Markov Model .....         | 46 |
| 2.12.1 ส่วนประกอบของแบบจำลองมาร์คอฟ.....                                | 46 |
| 2.12.2 คุณสมบัติการย้ายสเตทของ HMM.....                                 | 47 |
| 2.12.3 ปัญหาพื้นฐานของแบบจำลอง มาร์คอฟ.....                             | 48 |

## สารบัญ

|  | หน้า |
|--|------|
| บทที่ 5 สรุปผลงานวิจัยและข้อเสนอแนะ..... | 95   |
| 5.1 สรุปผลการทดลอง.....                  | 95   |
| 5.2 อภิปรายผลความผิดพลาด.....            | 96   |
| 5.3 ข้อเสนอแนะ.....                      | 96   |
| 5.4 การประยุกต์ใช้งาน.....               | 96   |
| <br>                                     |      |
| เอกสารอ้างอิง.....                       | 97   |
| <br>                                     |      |
| ภาคผนวก ก. ....                          | 99   |
| <br>                                     |      |
| ประวัติผู้เขียน.....                     | 106  |

# สารบัญตาราง

| ตารางที่  | หน้า |
|---|------|
| 1.1 ตารางสรุป เสียงสระ และเสียงวรรณยุกต์ของตัวเลข .....   | 3    |
| 2.1 อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูด.....   | 7    |
| 2.2 ลักษณะเสียงพูดของมนุษย์.....  | 12   |
| 2.3 เสียงสระเดี่ยวในภาษาไทย.....  | 13   |
| 2.4 เสียงสระเดี่ยวในภาษาไทยแบ่งตามความสั้น-ยาว.....   | 14   |
| 2.5 เสียงพยัญชนะจำแนกตามการออกเสียง.....  | 15   |
| 2.6 เสียงพยัญชนะในภาษาไทย.....  | 15   |
| 2.7 แสดงหน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย.....   | 16   |
| 2.8 แสดงลักษณะของเสียงวรรณยุกต์ในภาษาไทย.....   | 17   |
| 2.9 ขอบเขตและความถี่ศูนย์กลาง Critical Band Rate scale.....   | 39   |
| 3.1 ตารางสรุป เสียงสระ และเสียงวรรณยุกต์ของตัวเลขภาษาไทย.....   | 63   |
| 4.1 ผลการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด โดยทดสอบแบบเฟรม.....  | 88   |
| 4.2 ผลการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด โดยทดสอบแบบคำตัวเลขภาษาไทย.....   | 89   |
| 4.3 ผลการทดสอบแบบอ้างอิงเสียงวรรณยุกต์คำตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูด.....   | 90   |
| 4.4 ผลการทดลองแบบอ้างอิงเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลองอ้างอิงเสียงสระ<br>และเสียงวรรณยุกต์ แบบไม่ขึ้นกับผู้พูด.....   | 92   |
| 4.5 การเปรียบเทียบผลการทดลองแบบอ้างอิงเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลอง<br>อ้างอิงเสียงสระ และเสียงวรรณยุกต์ แบบไม่ขึ้นกับผู้พูด กับวิธีค่าสัมประสิทธิ์เซปตรัม<br>(MFCC) ..... | 94   |

# สารบัญรูป

| รูปที่   | หน้า |
|--|------|
| 1.1 ส่วนประกอบของคำในภาษาไทยโดยแยกตามลักษณะของหน่วยเสียง.....  | 2    |
| 2.1 วงจรการสื่อสารภาษาพูด.....   | 5    |
| 2.2 แสดงอวัยวะในระบบการพูดของมนุษย์.....   | 9    |
| 2.3 แผนภาพระบบเสียงพูดของมนุษย์.....   | 10   |
| 2.4 แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง.....  | 13   |
| 2.5 แสดงการเกิดเสียงสระผสมในภาษาไทย.....   | 14   |
| 2.6 แสดงองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์.....   | 18   |
| 2.7 บล็อกไดอะแกรมกระบวนการรู้จำเสียงพูด.....   | 19   |
| 2.8 บล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น.....  | 19   |
| 2.9 แบบจำลองระบบสร้างสัญญาณเสียงพูดเชิงเลข.....  | 20   |
| 2.10 ขั้นตอนการวิเคราะห์สัญญาณเสียงพูดเบื้องต้น.....   | 21   |
| 2.11 แสดงขนาดสเปกตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส.....  | 23   |
| 2.12 แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์.....   | 23   |
| 2.13 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์ในกรอบของเฟรม.....  | 24   |
| 2.14 แสดงวินโดว์แบบสี่เหลี่ยม (ก) ในโดเมนเวลา (ข) ในโดเมนความถี่.....  | 25   |
| 2.15 แสดงวินโดว์แบบแฮมมิง (ก) ในโดเมนเวลา (ข) ในโดเมนความถี่.....  | 25   |
| 2.16 ขั้นตอนการหาค่าพารามิเตอร์เสียงพูดเชิงเลข.....  | 30   |
| 2.17 ระดับของสัญญาณ โทนทดสอบที่ถูกคบบังโดยสัญญาณ White Noise<br>ที่ระดับสัญญาณต่างๆ และเส้นประเป็นระดับ Threshold in quiet.....  | 31   |
| 2.18 ความสัมพันธ์ระหว่างระดับ Threshold in Quiet และจำนวนสัญญาณ โทนทดสอบ<br>โดยจุดที่มีลูกศรอยู่เป็นจุดที่ใช้ในการคาดคะเนความกว้างแถบความถี่วิกฤติ.....                              | 33   |
| 2.19 ทั้งสองรูปเป็นผลต่อเนื่องจากรูปที่ 2.18 โดยการเพิ่มผลที่ได้จากการใช้.....   | 34   |
| 2.20 ระดับของสัญญาณ โทนทดสอบที่ถูกคบบังโดย Uniform Masking Noise ที่ระดับ<br>สัญญาณต่างๆ โดยกราฟด้านบนเป็นค่าการลดทอนที่ให้กับ White Noise เพื่อ<br>สร้าง Uniform Masking Noise..... | 35   |

# สารบัญรูป

| รูปที่  | หน้า |
|---|------|
| 2.21 ระดับ Threshold ของสัญญาณรบกวนแถบความถี่แคบที่มีความถี่ศูนย์กลางอยู่ระหว่างสองโทนที่ใช้ในการบังคับสัญญาณ โดยแสดงเปรียบเทียบกับความแตกต่างของความถี่ของสองโทน.....                                  | 36   |
| 2.22 ระดับ Threshold ที่สัญญาณโทนทดสอบถูกบังคับโดยสัญญาณรบกวนสองแถบ (Bandpass Noise) โดยเปรียบเทียบกับความแตกต่างความถี่ของ Cutoff ของทั้งสองแถบ.....   | 37   |
| 2.23 ความกว้างแถบความถี่วิกฤติเทียบกับความถี่ และเส้นประแสดงการประมาณ.....  | 38   |
| 2.24 ความสัมพันธ์ระหว่างลำดับของ Critical Band กับความถี่.....  | 40   |
| 2.25 โพลวซาร์ทการคำนวณหา Nearest Neighbor โดยตรง.....   | 42   |
| 2.26 การจัดกลุ่มโดยอาศัยการตัดสินใจของ K-Nearest Neighbor.....  | 43   |
| 2.27 แสดงการหาค่าพิทช์จากการคำนวณออก โอโต โครี เลขชั้นฟังก์ชันของสัญญาณเสียงพูด.....  | 45   |
| 2.28 แบบจำลอง Left-Right Model ของ HMM.....   | 48   |
| 2.29 (a) และ (b) กระบวนการไปข้างหน้า.....   | 50   |
| 2.30 กระบวนการย้อนกลับ.....   | 52   |
| 2.31 ลำดับการคำนวณการเกิดค่าปรากฏร่วมซึ่งจะอยู่ที่สเตท $i$ ที่เวลา $t$ และอยู่ที่ สเตท $j$ ที่เวลา $t+1$ .....  | 55   |
| 3.1 แผนผังกระบวนการรู้จำตัวเลขภาษาไทย.....  | 64   |
| 3.2 ขั้นตอนในการวิเคราะห์หน่วยเสียงสระของตัวเลขภาษาไทย.....   | 65   |
| 3.3 ตัวอย่างของสเปกตรัมแอลพีซี และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระทั้งหมด จำนวน 25 เฟรม ที่ประกอบอยู่ในคำตัวเลข “1” เสียงสระทั้งหมดที่ประกอบอยู่ในคำตัวเลขภาษาไทย (เลข 0 ถึงเลข 9)..... | 71   |
| 3.4 สเปกตรัมแอลพีซี และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระทั้งหมด (ในแต่ละเฟรม) ที่ประกอบอยู่ในคำตัวเลขภาษาไทย (เลข 0 ถึงเลข 9).....   | 76   |
| 3.5 ขั้นตอนในการวิเคราะห์หน่วยเสียงสระของตัวเลขภาษาไทย.....   | 77   |
| 3.6 ขั้นตอนต่างๆในการคำนวณหาค่าพิทช์ และตัวอย่างการหาค่าพิทช์ของเลข “หนึ่ง” (ก)–(ค) คือเอาต์พุทของแต่ละขั้นตอน.....   | 78   |
| 3.7 ตัวอย่างพิทช์ของตัวเลขภาษาไทย (ตั้งแต่ 0 ถึง 9) .....   | 81   |
| 3.8 ตัวอย่างการควอนไทซ์พิทช์ของเสียงวรรณยุกต์ตัวเลขภาษาไทย (ตั้งแต่ 0 ถึง 9) .....  | 85   |
| 3.9 โพลวซาร์ท การคำนวณหาค่าพารามิเตอร์ของแบบจำลองอ้างอิง.....   | 87   |

# สารบัญรูป

รูปที่

หน้า

|   |    |
|---|----|
| 4.1 กระบวนการตัดสินใจการรู้จำตัวเลขภาษาไทย..... | 91 |
|---|----|

# บทที่ 1

## บทนำ

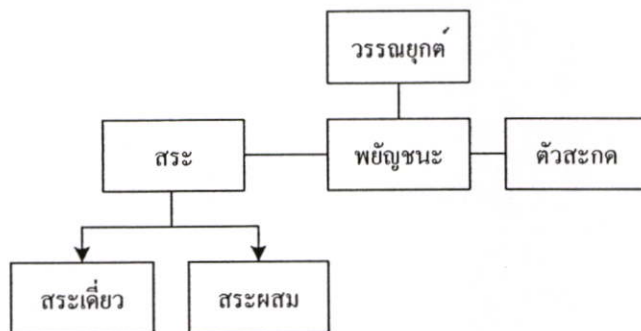
### 1.1 ความเป็นและความสำคัญของปัญหา

ในปัจจุบันการรู้จำเสียงพูด (Speech Recognition) ถูกนำมาประยุกต์ใช้งานกันอย่างกว้างขวาง โดยเฉพาะในระบบสื่อสาร โทรคมนาคม จุดมุ่งหมายหลักของการรู้จำเสียงพูดก็คือการเพิ่มพูนความสามารถของอุปกรณ์ต่างๆ ในการรับรู้และสื่อสารโต้ตอบกับมนุษย์ได้ เพื่อเพิ่มทางเลือกในการควบคุมสั่งการอุปกรณ์เครื่องมือต่างๆ โดยเฉพาะเครื่องคอมพิวเตอร์ ซึ่งการใช้เสียงพูดควบคุมสั่งการนี้ถือได้ว่าเป็นวิธีการที่เป็นธรรมชาติมากที่สุดของมนุษย์

การรู้จำเสียงพูดนั้นได้มีการศึกษาและวิจัยกันอย่างแพร่หลาย มีทั้งการรู้จำเสียงแบบขึ้นกับผู้พูด (Speaker-Dependent Speech Recognition) และการรู้จำเสียงแบบไม่ขึ้นกับผู้พูด (Speaker-Independent Speech Recognition) ซึ่งการรู้จำเสียงพูดนั้นมีความยุ่งยากและซับซ้อนมาก เนื่องจากความแตกต่างในการออกเสียงของผู้พูดแต่ละบุคคล จึงทำให้การพัฒนาทางด้านการรู้จำเสียงพูดเป็นไปอย่างล่าช้า การศึกษา และวิจัยการรู้จำเสียงพูดนั้นยังแบ่งตามลักษณะการรับรู้เสียงพูดออกเป็น 2 วิธีคือ

1. พิจารณาทั้งหน่วยภาษาที่เปล่งเสียงออกมาทั้งหมด มีทั้งระบบการรู้จำคำพูดเดี่ยว [1]-[2] (Isolated word Recognition) และระบบรู้จำคำพูดต่อเนื่อง (Continuous word Recognition) ซึ่งข้อดีของระบบเหล่านี้คือง่าย เนื่องจากมีการหลีกเลี่ยงผลกระทบอันเนื่องจากรากของเสียงภายในคำหรือกลุ่มคำนั้น แต่ข้อเสียคือสามารถรู้จำคำได้จำนวนคำที่จำกัด เนื่องจากต้องใช้เนื้อที่จำนวนมากในการจัดเก็บแบบจำลองอ้างอิง และต้องใช้เวลาในการคำนวณเพื่อเปรียบเทียบมากตามจำนวนของแบบจำลองอ้างอิงที่มีอยู่

2. พิจารณาโดยการแยกแยะรายละเอียดของหน่วยเสียง (Phonetic Recognition) [3]-[5] วิธีนี้จะพิจารณาลักษณะของหน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ ดังแสดงในรูปที่ 1.1 โดยจะใช้หน่วยเสียงย่อยเหล่านี้เป็นหลักในการรู้จำเสียงพูด ซึ่งวิธีนี้เหมาะสำหรับการพัฒนาไปสู่ระบบการรู้จำคำจำนวนมาก



รูปที่ 1.1 ส่วนประกอบของคำในภาษาไทย โดยแยกตามลักษณะของหน่วยเสียง

ตัวเลขในภาษาไทยมีความสำคัญเพื่อใช้ในการติดต่อสื่อสารกัน การรู้จำเสียงตัวเลขในภาษาไทยเป็นส่วนหนึ่งของการรู้จำเสียงพูดในภาษาไทย ทั้งนี้เนื่องจากการสื่อความหมายของหน่วยเสียงในภาษาไทยจำเป็นต้องใช้ทั้งเสียงพยัญชนะ, สระ และวรรณยุกต์จึงจะสื่อความได้สมบูรณ์ งานวิจัยนี้มุ่งเน้นที่จะจำแนกเสียงตัวเลขในภาษาไทย จำนวน 10 เสียง โดยการจำแนกเสียงสระ และเสียงวรรณยุกต์ เพื่อใช้ในการตัดสินใจ และจำแนกเสียงตัวเลขภาษาไทย

## 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อพัฒนาระบบการรู้จำเสียงพูดในภาษาไทย โดยใช้หน่วยเสียงสระ และเสียงวรรณยุกต์
2. เพื่อลดเนื้อที่หน่วยความจำในการเก็บแบบอ้างอิงและลดเวลาในการเปรียบเทียบกับแบบอ้างอิง
3. เพื่อศึกษาลักษณะของเสียงสระ และเสียงวรรณยุกต์ที่ประกอบขึ้นเป็นเสียงตัวเลขภาษาไทย
4. เป็นพื้นฐานในการรู้จำเสียงเพื่อพัฒนาไปสู่การรู้จำเสียงในภาษาไทยทั้งระบบคำพูดเดี่ยว และคำพูดต่อเนื่องต่อไป

## 1.3 สมมุติฐานของการศึกษา

ตัวเลขภาษาไทย ประกอบด้วยตัวเลขที่มีค่าตั้ง 0 ถึง 9 รวมทั้งสิ้นจำนวน 10 ตัว ตัวเลขแต่ละตัวจะประกอบด้วย 1 พยางค์ ที่รวมขึ้นจากหน่วยเสียงในภาษาไทย ประกอบด้วย เสียงพยัญชนะ, เสียงสระ และเสียงวรรณยุกต์ การพิจารณาตัวเลขไทยนั้น เราสามารถพิจารณาได้จากหน่วยเสียงสระ และเสียงวรรณยุกต์ ดังตารางต่อไปนี้

ตารางที่ 1.1 ตารางสรุป เสียงสระ และเสียงวรรณยุกต์ของตัวเลข

| ตัวเลข | เสียง | สระ<br>(Vowel) | วรรณยุกต์<br>(Tone) |
|--------|-------|----------------|---------------------|
| 1      | หนึ่ง | อี             | เอก                 |
| 2      | สอง   | ออ             | จัตวา               |
| 3      | สาม   | อา             | จัตวา               |
| 4      | สี่   | อี             | เอก                 |
| 5      | ห้า   | อา             | โท                  |
| 6      | หก    | โอะ            | เอก                 |
| 7      | เจ็ด  | เอะ            | เอก                 |
| 8      | แปด   | แอะ            | เอก                 |
| 9      | เก้า  | เอา            | โท                  |
| 0      | ศูนย์ | อู             | จัตวา               |

จากตารางตัวเลขภาษาไทยจะพบว่า จำนวนเสียงสระที่ใช้ในการออกเสียงตัวเลขภาษาไทย มีทั้งหมด 9 เสียง ได้แก่ สระอี, ออ, อา, อี, โอะ, เอะ, แอะ, เอา และอู และจำนวนเสียงของวรรณยุกต์ จำนวน 3 เสียง ได้แก่ วรรณยุกต์เอก, โท และจัตวา ซึ่งไม่มีตัวเลขภาษาไทยตัวใดที่มีหน่วยเสียงสระ และเสียงวรรณยุกต์ซ้ำกัน ดังนั้นการรู้จำเสียงภาษาไทยจะต้องเริ่มต้นจากการรู้จำเสียงสระ และเสียงวรรณยุกต์ที่ประกอบขึ้นเป็นเสียงตัวเลขภาษาไทย ซึ่งงานวิจัยนี้ได้้นำการรู้จำเสียงสระ และการรู้จำเสียงวรรณยุกต์มาสร้างให้เกิดการรู้จำเสียงตัวเลขภาษาไทย

#### 1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

งานวิจัยการรู้จำตัวเลขภาษาไทยนี้ เป็นการประยุกต์โดยใช้วิธีการจำแนกเสียงสระโดยคลื่นเสียงจะถูกแทนด้วยเวกเตอร์คุณสมบัติขนาด 18 มิติของความเข้มแถบวิกฤต (CBI) เพื่อจำแนกเสียงสระด้วยวิธี KNN และ การจำแนกเสียงวรรณยุกต์ โดยการวิเคราะห์คาบพิทซ์เพื่อนำมาหาค่าความถี่มูลฐานเพื่อให้ได้ค่าควอนไทซ์พิทซ์ผ่านกระบวนการรู้จำโดยใช้วิธี HMM หลังจากนั้นนำผลลัพธ์ดังกล่าวผ่านกระบวนการการตัดสินใจเพื่อจำแนกเสียงตัวเลขภาษาไทยได้

## 1.5 ขอบเขตการวิจัย

งานวิจัยนี้ได้เลือกศึกษาการจำแนกเสียงตัวเลขในภาษาไทย จำนวน 10 เสียง โดยใช้การจำแนกเสียงสระ และเสียงวรรณยุกต์ เพื่อใช้ในการตัดสินใจจำแนกเสียงตัวเลขภาษาไทย

## 1.6 ขั้นตอนการศึกษา

1. ศึกษากระบวนการรู้จำภาษาในลักษณะต่างๆ ที่ได้มีการศึกษามาแล้ว
2. ศึกษาทฤษฎีทางด้านภาษาศาสตร์ และอัลกอริทึมต่างๆที่ใช้ในการวิเคราะห์เสียงพูด
3. กำหนดขอบเขตการวิจัย
4. บันทึกข้อมูลเสียง โดยใช้ผู้พูดทั้งหมด 6 คน เพศชาย 3 คน เพศหญิง 3 คน
5. เก็บข้อมูล วิเคราะห์เสียง และทดสอบความถูกต้อง
6. สรุปผลการวิจัย และเสนอแนะแนวทางในการวิจัยต่อไป

## บทที่ 2

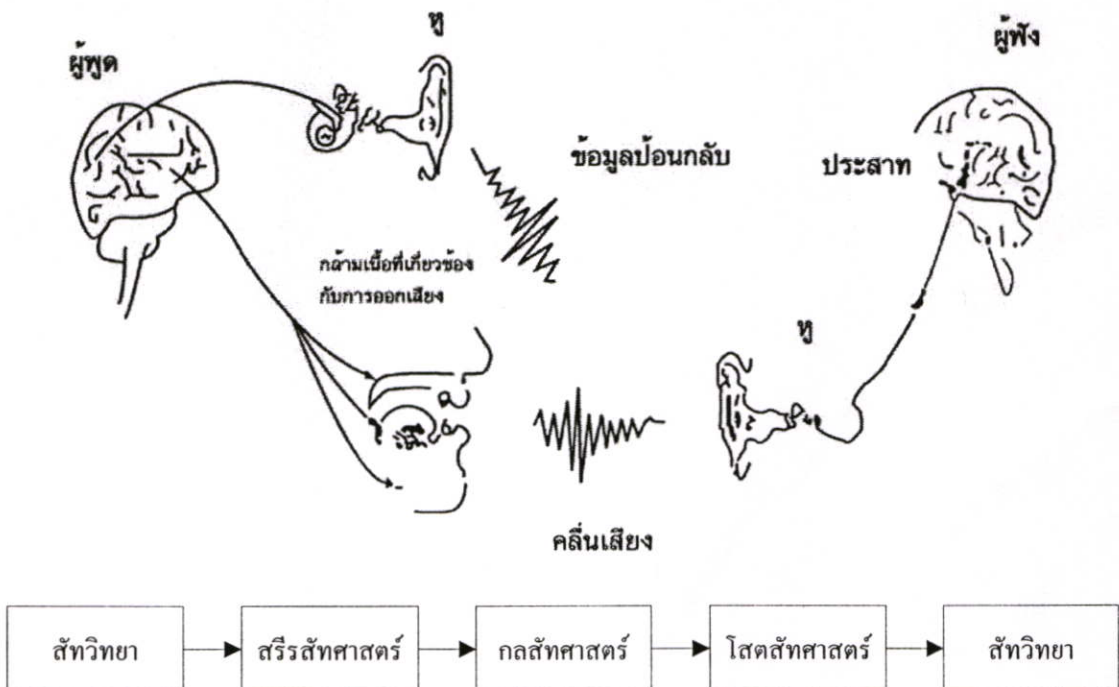
# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 กล่าวนำ

ในบทนี้จะกล่าวถึงทฤษฎีต่างๆที่เกี่ยวข้องกับงานวิจัยนี้ โดยจะเริ่มจากทฤษฎีทางภาษาศาสตร์ ซึ่งจะกล่าวถึงเรื่องทฤษฎีการสร้างเสียงพูดประกอบด้วย หน่วยเสียงสระ หน่วยเสียงพยัญชนะ หน่วยเสียงวรรณยุกต์ และลักษณะคำ และพยางค์ ในภาษาไทย จากนั้นจะกล่าวถึงทฤษฎีการวิเคราะห์เสียงประกอบด้วย การคำนวณสเปกตรัมแอลพีซี แถบความถี่วิกฤติ และความเข้มสัญญาณในแถบความถี่วิกฤติเพื่อจำแนกเสียงสระด้วยวิธี KNN และการวิเคราะห์คำพิตซ์เพื่อนำมาหาค่าความถี่มูลฐานเพื่อให้ได้ค่าควอนไทซ์พิตซ์ผ่านกระบวนการรู้จำโดยใช้วิธี HMM

### 2.2 ทฤษฎีภาษาศาสตร์

ในการสื่อสาร โดยภาษาพูด ผู้พูดเริ่มวงจรจากความคิดที่ต้องการจะสื่อข้อความใด แล้วสมองจึงสั่งการส่งสัญญาณให้อวัยวะต่างๆที่เกี่ยวข้องทำงานและออกเสียงทีละเสียงเป็นคำและถ้อยความตามที่ต้องการ คลื่นเสียงเดินทางผ่านตัวกลางคือ อากาศมากระทบหูผู้ฟัง ผู้ฟังรับคลื่นเสียงและถอดเสียงในถ้อยความและตีความหมายรับรู้อาร



รูปที่ 2.1 วงจรการสื่อสารภาษาพูด

วงจรมันเริ่มต้นที่สามัคยิยะทางภาษา (Linguistic Competence) คือความรู้ความสามารถทางภาษาที่เป็นธรรมชาติของมนุษย์ โดยเฉพาะในส่วนที่เกี่ยวข้องกับเสียงและระบบเสียงซึ่งเป็นเรื่องทาง สัทวิทยา (Phonology) การออกเสียงและเปล่งเสียงเป็นเรื่องทางสรีรศาสตร์ (Articulatory Phonetics) ภายภาพของเสียงหรือคลื่นเสียงเป็นเรื่องทางกลศาสตร์ (Acoustic phonetics) และการได้ยินและการรับรู้เสียงเป็นเรื่องทางโสตศาสตร์ (Auditory Phonetics) การถอดเสียงในถ้อยความของผู้ฟังเป็นเรื่องของสามัคยิยะทางภาษาที่เกี่ยวข้องกับเสียงและระบบเสียงซึ่งสัมพันธ์กับหน่วยคำหรือคำซึ่งเป็นเรื่องของสัทวิทยา โดยสัทวิทยาเป็นการศึกษาเรื่องเสียงและระบบเสียงของภาษาซึ่งเป็นนามธรรมและปรากฏเป็นพฤติกรรมทางภาษาที่เป็นรูปธรรมกล่าวคือสัทวิทยาศึกษาโครงสร้างและหน้าที่ของเสียงซึ่งสัมพันธ์กับความหมายในภาษา [6]

มนุษย์สร้างเสียงพูดขึ้นมาได้โดยมนุษย์จะมีเส้นเอ็นในลำคอที่เรียกว่าเส้นเสียง (Vocal Cords) ซึ่งอยู่ในช่องหลอดลม และจากการสั่นของเส้นเสียงจะทำให้เกิดคลื่นเสียงความถี่ต่างๆ ผ่านเข้าสู่ลำคอ ผ่านจากลำคอเข้าสู่ปากหรือช่องจมูกออกไปภายนอกซึ่งขนาดและรูปร่างของอวัยวะภายในช่องปากนี่จะเป็นสิ่งกำหนดว่าคลื่นเสียงความถี่ไหนจะยอมให้ปรากฏออกมาให้ได้ยิน (Passed Through) หรือคลื่นเสียงความถี่ไหนที่จะถูกดูดซับไว้ไม่ยอมให้ปรากฏออกมา (Absorbed) ปรากฏการณ์ที่ทำให้เกิดเสียงพูดออกมาได้นี้ เนื่องจากรูปร่างของช่องคอและช่องปาก ของมนุษย์ที่สามารถเปลี่ยนแปลงไปมาได้อย่างอิสระ ทำให้ผู้พูดมีความสามารถในการกำหนดคลื่นเสียงระดับต่างๆ เหล่านี้เองที่ทำให้เกิดการแปลงเป็นเสียงพูดออกมา [4]

### 2.2.1 อวัยวะออกเสียง (The Organs of Speech)

ธรรมชาติได้กำหนดตำแหน่งของอวัยวะที่ใช้ในการทำให้เกิดเสียงของมนุษย์เอาไว้ นั่นคือริมฝีปาก ฟัน เพดาน และ ลิ้น ที่อยู่ตามลำดับตลอดช่องของกล้ามเนื้อช่องคอ เพื่อให้สามารถเปลี่ยนขนาดและรูปร่างได้ตามต้องการ ดังรูปที่ 2.1 เราจะเปล่งคำพูดออกมาโดยการเปล่งออกมาเป็นลำดับขั้นของความถี่คลื่นเสียงที่ได้คัดเลือกไว้ออกมา นั่นคือ จะมีการเปลี่ยนแปลงตำแหน่งของช่องทางเดินเสียง (Vocal Tract) ของมนุษย์เป็นลำดับขั้นอย่างต่อเนื่อง

อวัยวะที่ใช้ในการออกเสียงนั้น อาจจะแบ่งคร่าวๆ ได้เป็น 2 พวก คือ (กาญจนา นาคสกุล, 2524)

2.2.1.1 อวัยวะที่เป็นส่วนกระทำอาการ (Articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลิตลมไปยังส่วนต่างๆ อวัยวะที่เป็นส่วนกระทำการที่สำคัญ คือลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

2.2.1.2 อวัยวะที่เป็นฐานที่เกิดเสียงต่างๆ (Point of Articulator) หมายถึง ตำแหน่งหรือฐานที่เกิดของเสียงต่างๆ เช่น ริมฝีปาก ฟัน เพดานส่วนต่างๆ เป็นต้น

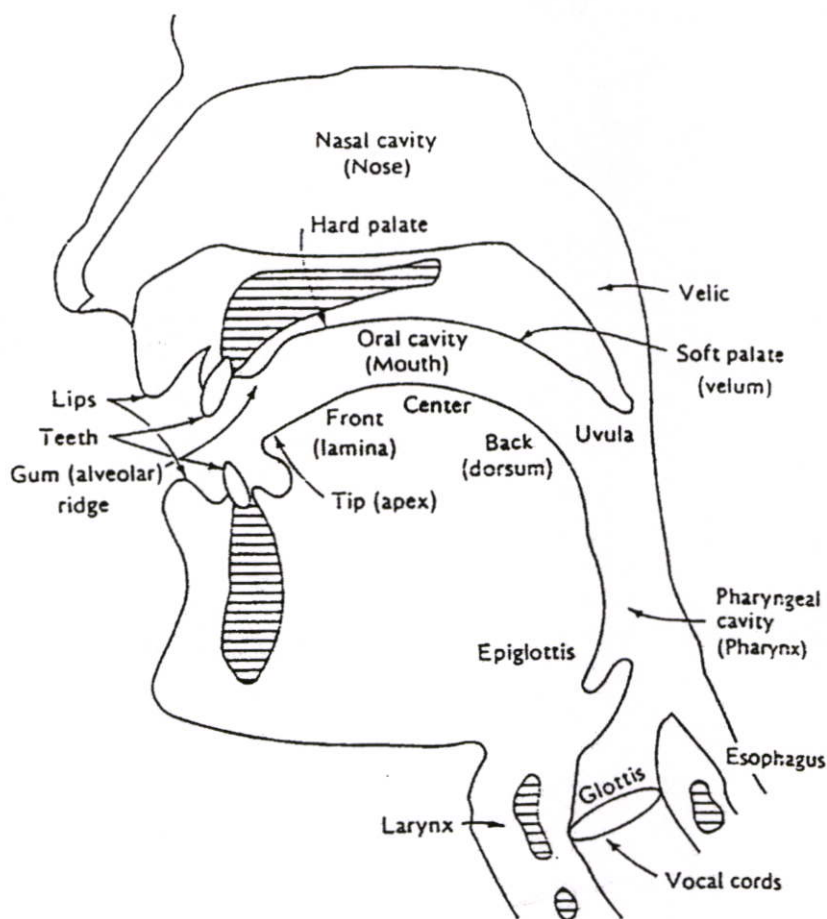
อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูดมีดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูด

| อวัยวะ                               | ลักษณะ / หน้าที่   |
|--------------------------------------|--|
| ริมฝีปาก (Lips)                      | เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกัน ได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ยื่นออกมา หรือ ห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่างๆ นี้ ล้วนมีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น  |
| ฟัน (Teeth)                          | เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่างหรือคดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอดผ่านได้ออกมาทำให้เกิดเป็นเสียงชนิดที่เรียกว่าเสียงเสียดแทรกที่เกิดที่ฟันเป็นต้น   |
| ปุ่มเหงือก (Gum Ridge)               | เป็นส่วนที่อยู่ต่อจากฟันบน   |
| เพดานแข็ง หรือเพดานปาก (Hard Palate) | เฉพาะเพดานที่โค้งเป็นกระดูกแข็งเท่านั้น  |
| เพดานอ่อน (Soft Palate)              | เป็นส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดกับหลังคอก นอกจากเวลาออกเสียงนาสิก เท่านั้นที่เพดานอ่อนจะลดระดับลงมา |
| ลิ้น (Tongue)                        | เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งเป็น 3 ส่วนด้วยกันตามหน้าที่ในการออกเสียง  |
| ปลายลิ้น (Blade of the Tongue)       | เป็นส่วนปลายซึ่งสามารถยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้ง่าย   |
| หน้าลิ้น (Front of the Tongue)       | หมายถึงลิ้นที่อยู่ตรงข้ามกับเพดานแข็ง ขณะที่วางลิ้นราบกับปากตอนที่ไม่ได้พูด  |
| หลังลิ้น (Back of the Tongue)        | หมายถึงส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ขณะที่วางลิ้นราบกับปาก  |

ตารางที่ 2.1 (ต่อ)

| อวัยวะ                     | ลักษณะ / หน้าที่  |
|----------------------------|---|
| ช่องปาก (Oral Cavity)      | ช่องปากนี้เป็นอวัยวะอีกส่วนหนึ่งที่มีความสำคัญในการพูด กล่าวคือ ช่องปากจะทำหน้าที่เป็นช่องกำทอน (Resonant Chamber) ซึ่งมีบทบาทในการกำทอนเสียงที่เดินทางผ่านมาถึงบริเวณนี้ ทั้งนี้เพราะช่องปากสามารถจะเปลี่ยนแปลงเป็นช่องกำทอนรูปร่างต่างๆ กัน เนื่องจากรูปร่างของมันแปรผันไปตามการจัดท่าทางของลิ้น ริมฝีปาก และขากรรไกร |
| โพรงจมูก (Nasal Cavity)    | คุณสมบัติลักษณะของเสียงที่พูดที่เกิดขึ้นจะแปรผันไปตามการปิด – เปิดของช่องทางออกสู่โพรงจมูก ซึ่งเป็นผลจากการยกขึ้นหรือเลื่อนลงของเพดานอ่อน   |
| ช่องคอ (Pharynx)           | หมายถึงโพรงคอซึ่งอยู่ถัดจากปากลงไปจนถึงเส้นเสียง  |
| ลิ้นปิดหลอดลม (Epiglottis) | เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่ อยู่ต่อจากโคนลิ้นลงไปในลำคอ มีหน้าที่ปิดช่องลมขณะรับประทานอาหาร และปิดช่องลมและจมูก   |
| เส้นเสียง (Vocal Cords)    | เป็นอวัยวะสำคัญที่เกิดของเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ 2 แผ่น ปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างระหว่างเส้นเสียง (glottis) เส้นเสียงทั้งสองสามารถจะดึงออกให้ห่างจากกัน หรือดึงเข้ามาประชิดกันก็ได้ เส้นเสียงเป็นส่วนทำให้เกิดเสียงพูดขึ้น                               |



รูปที่ 2.2 แสดงอวัยวะในระบบการพูดของมนุษย์

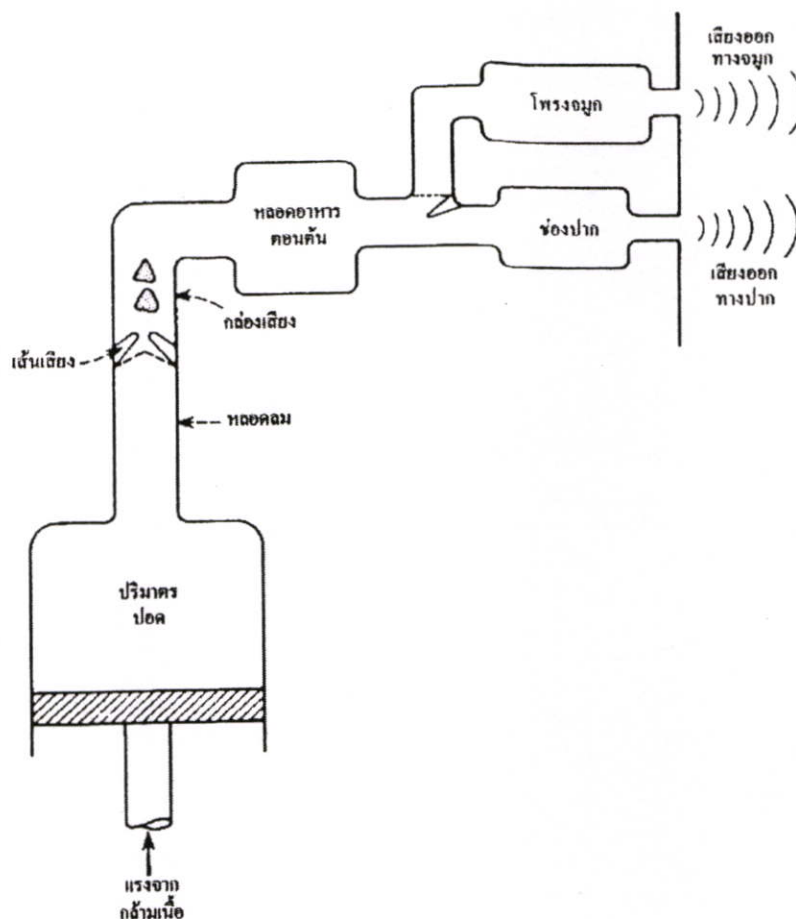
### 2.2.2 กระบวนการผลิตเสียงพูด

จากระบบเสียงพูดของมนุษย์ สามารถแสดงเป็นแผนภาพของระบบกำเนิดเสียงดังรูปที่ 2.2 ซึ่งเราสามารถจำแนกกลไกสร้างเสียงพูด ของมนุษย์ได้ 3 แบบ ดังนี้

2.2.2.1 อากาศไหลจากปอดจะถูกมอดูเลต (Modulate) โดยการสั่นของเส้นเสียงทำให้เกิดคลื่นเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (Quasi-periodic pulse – like Excitation)

2.2.2.2 อากาศไหลจากปอดถูกทำให้ปั่นป่วน ด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปากทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (Noise – like Excitation)

2.2.2.3 อากาศที่ไหลถูกกัก และเกิดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้อากาศที่มีแรงดันพุ่งออกไปอย่างรวดเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น (Transient Excitation)



รูปที่ 2.3 แผนภาพระบบเสียงพูดของมนุษย์

### 2.2.3 เสียงพูดของมนุษย์

เสียงพูดเป็นคลื่นความยาว (Longitudinal Wave) เกิดจากการสั่นของอนุภาคตัวกลาง นั่นคืออากาศ และทิศทางการสั่นของอนุภาคจะอยู่ในทิศเดียวกันกับทิศทางการเคลื่อนที่ คลื่นเสียงเป็นคลื่นที่เปลี่ยนแปลงไปตามเวลา

เสียงพูดแบ่งได้เป็น 2 ชนิด ตามการกำเนิดเสียงหรือโหมด (Mode) การกระตุ้น คือ

2.2.3.1 เสียงวอยซ์ หรือเสียงโฆษะ (Voiced Sound) เกิดจากการบังคับอากาศให้ผ่านสายเสียงทำให้มีการเปลี่ยนแปลงความตึงหย่อนของเส้นเสียง โดยเส้นเสียงจะสั่นและเกิดเป็นพัลส์ (Pulse) ของอากาศไปกระตุ้นอวัยวะกำเนิดเสียงวอยซ์ ตัวอย่างเสียงวอยซ์ ได้แก่ เสียงสระเสียงพยัญชนะ ที่ออกเสียงจากลำคอ (Voiced Consonants)

2.2.3.2 เสียงอันวอยซ์หรือเสียงอโฆษะ (Voiceless หรือ Unvoiced Sounds) เป็นเสียงที่เกิดในช่องปากหรือโพรงจมูกโดยอวัยวะภายในช่องปาก ริมฝีปาก ขวางการไหลของอากาศได้เป็นช่องเล็กๆ อากาศจึงไหลผ่านอย่างรวดเร็วและปั่นป่วนจนกระทั่งสร้างเป็นเสียงรบกวนช่วงความถี่

กว้าง (Broad – spectrum Noise) ตัวอย่างเสียงอันวอยซ์ ได้แก่ เสียงพยัญชนะที่ไม่ได้เกิดจากลำคอ (Voiceless Consonants)

#### 2.2.4 ความยาวของเสียง (Length)

การเกิดเสียงใดเสียงหนึ่งเปล่งออกมาได้นานเท่าใด เสียงพูดบางเสียงอาจจะเปล่งออกมาได้ติดต่อกันได้นาน เช่น เสียงสระ เสียงพยัญชนะ หรือเสียงพยัญชนะเสียดแทรก

ในภาษาไทย เสียงพูดที่มีความยาว – สั้น ก็มีเพียงเสียงสระเท่านั้น เช่น อะ อี อุ เป็นเสียงสั้น อา อี อุ เป็นเสียงยาวเป็นต้น

#### 2.2.5 ระดับเสียง สูง – ต่ำ (Pitch)

เสียงพูดจะมีระดับ สูงหรือต่ำ อยู่ที่ความถี่พื้นฐานของเสียง (Fundamental Frequency) ถ้าความถี่ต่ำเสียงก็จะต่ำ อวัยวะส่วนที่ทำให้เสียงมีระดับ สูง – ต่ำ คือ เส้นเสียง ดังนั้นระดับเสียงสูง – ต่ำ ก็คืออัตราการสั่นสะบัดของเส้นเสียงนั่นเอง

ในการพูดเสียงที่มีระดับสูง – ต่ำ ได้คือเสียงก้องเท่านั้น เพราะมีการสั่นสะเทือนของเส้นเสียง ที่ทำให้เกิดมีความถี่ระดับต่างๆ ได้ ในภาษาไทยระดับเสียงสูง – ต่ำ ของคำเราเรียกว่า “วรรณยุกต์”

#### 2.2.6 ความดัง (Loudness)

ความดังขึ้นอยู่กับปริมาณของลม ที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่งๆ

#### 2.2.7 การลงน้ำหนัก (Stress)

การออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นที่อยู่ข้างเคียง (เพื่อต้องการเรียกร้องความสนใจเป็นพิเศษ หรือแสดงอารมณ์ อย่างใดอย่างหนึ่ง)

#### 2.2.8 ช่วงต่อของเสียง (Juncture)

ช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งเสียงอีกเสียงหนึ่ง ซึ่งเรียงกันมาเป็นลำดับเสียงที่ประกบกันเข้าเป็นพยางค์ จะมีช่วงต่อของเสียงแนบสนิทจนไม่เห็นร่องรอย (Close Juncture) แต่ถ้าเสียงปรากฏอยู่คนละพยางค์ หรือคนละคำจะมีช่วงต่อ “ห่าง” จนสังเกตเห็นได้ชัด (Open Juncture) ดังนั้นช่วงต่อของเสียง โดยเฉพาะช่วงต่อห่างจะมีความสำคัญมากในการแบ่งคำในภาษา

## 2.3 หน่วยเสียงสำคัญในภาษาไทย

“หน่วยเสียง” (Phoneme) เป็นหน่วยเล็กที่สุดของภาษา หน่วยดังกล่าวได้แก่เสียงสำคัญๆ ในภาษาใดภาษาหนึ่งซึ่งทำหน้าที่ให้ความหมายของคำที่ใช้ในภาษานั้น และทำให้ความหมายของคำนั้นๆ มีความหมายแตกต่างจากคำอื่นๆ หน่วยเสียงสำคัญในภาษาไทยมี 3 ประเภทใหญ่ๆ คือ เสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์ หน่วยเสียงทั้ง 3 นี้เองที่ประกอบกันเข้าเป็นคำที่ใช้ในภาษาไทย

เสียงพูดของมนุษย์ซึ่งมีความแตกต่างกันมากมายนั้น ถ้าเราพิจารณาอย่างถ้วนๆ จะพบว่าสามารถแบ่งออกเป็น 2 ประเภทใหญ่ ดังตารางที่ 2.2

ตารางที่ 2.2 ลักษณะเสียงพูดของมนุษย์

| เสียง                                 | ลักษณะ  |
|---------------------------------------|---|
| เสียงเรียง (Segmental Sound)          | เป็นหน่วยเสียงที่แยกออกจากเสียงอื่นได้โดยเด็ดขาด เพราะมีลักษณะเด่นเฉพาะตัว ในภาษาไทยได้แก่เสียงสระ และเสียงพยัญชนะ                        |
| เสียงซ้อน (Supra – segmental Feature) | เป็นเสียงที่ทำหน้าที่เป็นส่วนประกอบของเสียงอื่น เพราะไม่สามารถแยกเปล่งเสียงได้ตามลำพัง ในภาษาไทยได้แก่เสียงวรรณยุกต์ และทำนองเสียงเป็นต้น |

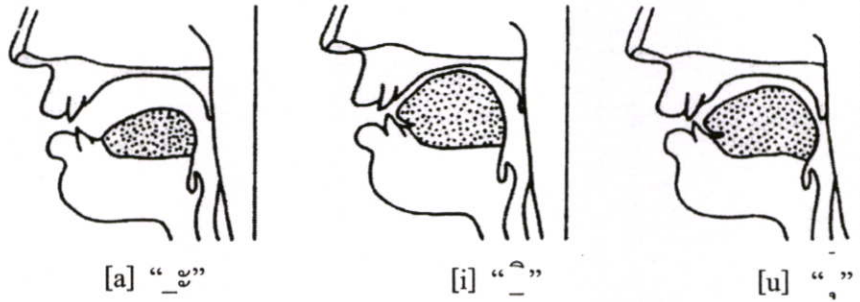
### 2.3.1 หน่วยเสียงสระ

#### 2.3.1.1 ลักษณะของเสียงสระ

ลักษณะสำคัญของเสียงสระก็คือ “เป็นเสียงก้องที่เปล่งเสียงออกมาโดยให้ลมออกทางช่องปากโดยไม่ถูกลิ้นกักหรือขัดขวาง” ดังนั้นเวลาเราออกเสียงสระจะออกเสียงได้สะดวกและออกเสียงได้นาน ทั้งนี้เพราะคุณสมบัติของเสียงสระมีความดังเด่นกว่าเสียงอื่นๆ ที่เรียงอยู่ข้างเสมอ อวัยวะที่เกี่ยวข้องกับการออกเสียงสระได้แก่ลิ้น ริมฝีปาก ถ้าลิ้นส่วนใดทำหน้าที่เพียงส่วนเดียว เสียงที่เกิดขึ้นก็จะมีเพียงเส้นเสียงเดียว เสียงเช่นนี้เรียกว่า “สระเดี่ยว” แต่ถ้าลิ้นส่วนอื่นทำหน้าที่ร่วมด้วยเสียงสระนั้นเรียกว่า “สระผสม”

ฉะนั้นการอธิบายเรื่องการออกเสียงสระจึงเป็นเรื่องที่ยากกว่าการออกเสียงพยัญชนะเพราะไม่มีจุดสัมผัสในปากที่เราสามารถใช้เป็นหลักในการอธิบายได้ การที่สระมีเสียงต่างๆ กันก็เพราะช่องคอ ช่องปากมีรูปร่างต่างกันในขณะที่อากาศกำลังผ่านออกจากปากไป สิ่งที่มีส่วนทำให้ช่องปากช่องคามีรูปร่างต่างๆ กันออกไปก็คือ ลิ้น และริมฝีปาก เพื่อความสะดวกในการอธิบายเสียงเราจะ

แบ่งลิ้นออกเป็น 3 ส่วน คือ ส่วนหน้า (Front) ส่วนกลาง (Center) และส่วนหลัง (Back) ในขณะที่เราหุบปากอยู่หนึ่งๆ ส่วนหน้าของลิ้นอยู่ใต้เพดานแข็ง (Hard Palate) ส่วนหลังอยู่ใต้เพดานอ่อน (Velum) และส่วนกลางอยู่ระหว่างกลางของสองส่วนนี้ รูปต่อไปนี้จะแสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง ดังรูปที่ 2.4



รูปที่ 2.4 แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง

สำหรับภาษาไทยมีหน่วยเสียงสระทั้งหมด 24 หน่วยเสียง แยกออกเป็นสระเดี่ยว 18 หน่วยเสียง และสระผสม 6 หน่วยเสียง [8]

#### 2.3.1.1.1 สระเดี่ยว

เสียงสระเดี่ยว 18 หน่วยเสียง พิจารณาการเกิดเสียงได้เป็น 2 กรณีใหญ่ๆ คือ

2.3.1.1.1.1 การเกิดจากส่วนต่างๆ ของลิ้นหมายถึง ลมผ่านส่วนหน้า ส่วนกลาง หรือส่วนหลังของลิ้น

2.3.1.1.1.2 การเกิดจากลมผ่านลิ้นในขณะที่ลิ้นอยู่ในระดับ สูง กลาง หรือ ต่ำ

#### ตารางที่ 2.3 เสียงสระเดี่ยวในภาษาไทย

| สระ  | ระดับลิ้น | หน่วยเสียง | สัญลักษณ์     |
|------|-----------|------------|---------------|
| หน้า | สูง       | อิ, อี     | “ i ”, “ i: ” |
|      | กลาง      | เอะ, เอ    | “ e ”, “ e: ” |
|      | ต่ำ       | แอะ, แอ    | “ ɛ ”, “ ɛ: ” |
| กลาง | สูง       | อี อีอ     | “ ω ”, “ ω: ” |
|      | กลาง      | เออะ เออ   | “ γ ”, “ γ: ” |
|      | ต่ำ       | อะ อา      | “ a ”, “ a: ” |
| หลัง | สูง       | อุ อุ      | “ u ”, “ u: ” |
|      | กลาง      | โอะ โอ     | “ o ”, “ o: ” |
|      | ต่ำ       | เอะ ออ     | “ ɔ ”, “ ɔ: ” |

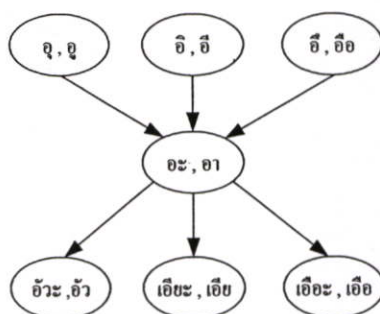
นอกจากนี้ หน่วยเสียงสระเดี่ยว 18 หน่วย สามารถแบ่งตามความสั้น - ยาว ของการออกเสียงได้เป็น

ตารางที่ 2.4 เสียงสระเดี่ยวในภาษาไทยแบ่งตามความสั้น-ยาว

| สระ                | จำนวน (หน่วย) | หน่วยเสียง |    |     |    |     |     |     |      |      |
|--------------------|---------------|------------|----|-----|----|-----|-----|-----|------|------|
| สระเดี่ยวเสียงสั้น | 9             | อะ         | อิ | อี  | อุ | เอะ | แอะ | โอะ | เอาะ | เออะ |
| สระเดี่ยวเสียงยาว  | 9             | อา         | อี | อือ | อู | เอ  | แเอ | โอ  | ออ   | เออ  |

#### 2.3.1.1.2 สระผสม

เสียงผสม 6 หน่วยเสียง เกิดจากลมผ่านกระแทบลิ้น 2 ส่วนคือ ส่วนบน และส่วนล่าง ซึ่งในขณะที่ออกเสียงสั้น จะอยู่ในระดับสูง แล้วลดลงต่ำโดยเสียงหลังเป็นสระ (อะ อา) เสมอ ดังรูปที่ 2.5 นี้



รูปที่ 2.5 แสดงการเกิดเสียงสระผสมในภาษาไทย

นอกจากนี้ในภาษาไทยยังมีสระที่เรียกว่าสระอักษร ได้แก่ อำ, ไอ, โอ, เอา ซึ่งแม้จะมีรูปเฉพาะ แต่ก็มีลักษณะเป็นสระเสียงสั้น และสระเสียงเบา [7] เพราะสระทั้ง 4 ตัวนี้ ออกเสียงมีตัวสะกดกำกับอยู่ด้วย

อำ = อัม (อะ+ม)

ไอ = อัย (อะ+ย)

โอ = อัย (อะ+ย)

เอา = อัว (อะ+ว)

เมื่อพิจารณาเสียงแล้วจะเหมือนกับสระปกติที่รวมกับพยัญชนะท้ายนั่นเอง

#### 2.3.1.2 หน้าที่ของหน่วยเสียงสระในภาษาไทย

หน่วยเสียงสระในภาษาไทยทั้ง 24 หน่วยเสียงนี้ ทำหน้าที่เป็นแกนกลางของพยางค์ หรือคำ กล่าวคือ คำทุกคำในภาษาไทยจะต้องมีเสียงสระอยู่ด้วย และเสียงสระในภาษาไทยจะสามารถเกิด

กับเสียงพยัญชนะต้นได้ทุกเสียง และสามารถเกิดกับหน่วยเสียงวรรณยุกต์ได้ทุกหน่วย แต่ไม่สามารถเกิดกับหน่วยเสียงพยัญชนะสะกดได้ทุกหน่วย หน่วยเสียงสระที่ทำให้เกิดคำหรือพยางค์ใช้ได้มากที่สุดในภาษามักเป็นหน่วยเสียงยาว

### 2.3.2 หน่วยเสียงพยัญชนะ

เสียงพยัญชนะในภาษาไทยมีทั้งหมด 21 หน่วยเสียง (44 รูป) หน่วยเสียงพยัญชนะออกเสียงได้ไม่สะดวกเท่าหน่วยเสียงชนิดอื่น เพราะเวลาออกเสียงลมหายใจที่พุ่งออกมาจากหลอดลมจะถูกขัดขวางตามส่วนต่างๆ ของปาก เสียงพยัญชนะจึงออกเสียงให้ยาวนานอย่างเสียงสระไม่ได้ และเสียงพยัญชนะก็ไม่ใช่เสียงก้องเสมอไป

#### 2.3.2.1 ลักษณะของเสียงพยัญชนะ

หน่วยเสียงพยัญชนะ 21 หน่วยเสียงนี้จำแนกตามลักษณะเสียงเป็น เสียงก้อง เสียงไม่ก้อง เสียงหนัก เสียงเบา ตามตารางที่ 2.5 และลักษณะการออกเสียง ตามตารางที่ 2.6

ตารางที่ 2.5 เสียงพยัญชนะจำแนกตามการออกเสียง

| ลักษณะเสียง          | พยัญชนะ                              |
|----------------------|--------------------------------------|
| เสียงก้อง (โฆณะ)     | /จ//ย//บ//ค//ม//น//ร//ล//ว/          |
| เสียงไม่ก้อง (อโฆณะ) | /ก//ค//จ//ช//ซ//ท//ต//ป//พ//ฟ//อ//ฮ/ |
| เสียงหนัก (ธนิต)     | /ค//ช//ท//พ/                         |
| เสียงเบา (ถวิล)      | /ก//จ//ต//ป/                         |

ตารางที่ 2.6 เสียงพยัญชนะในภาษาไทย

| ที่เกิดเสียง  |           | ริมฝีปาก | ปุ่มเหงือก | เพดานแข็ง | เพดานอ่อน | เส้นเสียง |
|---------------|-----------|----------|------------|-----------|-----------|-----------|
| พยัญชนะกัก    | ไม่พ่นลม  | ป*       | ฎ, ต*      | จ         | ก*        | อ*        |
|               | พ่นลม     | พ พภ     | ฐ ฑ ฒ ท ธ  | ฉ ช ฌ     | ข ค ฌ     |           |
|               | ก้อง      | บ        | ฎ ค        |           |           |           |
| พยัญชนะไม่กัก | นาสิก     | ม*       | น ฌ*       |           | ง*        |           |
|               | เสียดแทรก | ฝ ฟ      | ช ศ ษ ส    |           |           | ฮ ห       |
|               | กระทบ     |          | ร          |           |           |           |
|               | ข้างลิ้น  |          | ล ฬ        |           |           |           |
|               | กึ่งสระ   | ว*       |            |           | ฌ, ย*     |           |

หมายเหตุ หน่วยเสียง พยัญชนะที่มีเครื่องหมายดอกจัน (\*) คือหน่วยเสียงที่สามารถปรากฏที่ตำแหน่งท้ายพยางค์ได้

### 2.3.2.2 หน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย

เสียงพยัญชนะในภาษาไทย 21 หน่วยเสียงนี้สามารถทำหน้าที่ได้ตามตารางที่ 2.7

ตารางที่ 2.7 แสดงหน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย

| หน้าที่              | พยัญชนะ   |
|----------------------|---|
| พยัญชนะต้นเดี่ยว     | /ก//ค//จ//ย//บ//ค//ม//น//ร//ล//ว//จ//ช<br>//ช//ท//ต//ป//พ//ฟ//อ//ฮ/   |
| พยัญชนะต้นควบ        | หน่วยเสียงแรก /ก//ค//ต//ป/และ/พ/<br>หน่วยเสียงที่สอง /ร//ล/หรือ/ว/  |
| พยัญชนะสะกดของพยางค์ | /ป / (แม่กบ) / ต / (แม่ กค) / ก / (แม่กค) / ม / (แม่กม) / ง /<br>(แม่กง) / น / (แม่กน) / ว / (แม่เกว) / ย / (แม่เกย) และ ไม่มี<br>เสียงพยัญชนะ สะกด (แม่กา) |

### 2.3.3 หน่วยเสียงวรรณยุกต์

หน่วยเสียงวรรณยุกต์ คือ ระดับเสียงสูง – ต่ำ ของคำในภาษาไทย เช่นเดียวกับภาษาจีนและภาษาอื่นๆ ที่เป็นภาษาคำโดด ซึ่งมีการกำหนดเสียงสูง – ต่ำ ไว้ตายตัวในแต่ละคำ ถ้าออกเสียงสูง – ต่ำ ผิดไปความหมายย่อมผิดตามไปด้วย

ในภาษาไทยหน่วยเสียงวรรณยุกต์เป็นหน่วยเสียงสำคัญ มีความหมายต่างกัน ดังนั้นอาจกล่าวได้ว่าหน้าที่ ของหน่วยเสียงวรรณยุกต์ ก็คือ การทำให้เกิดคำขึ้นใช้ในภาษามากขึ้น และเป็นวิธีการสร้างคำขึ้นใช้เพิ่มในภาษาเป็นวิธีแรก ทั้งนี้เพราะถ้าเราเปลี่ยนเสียงวรรณยุกต์ ก็จะทำให้คำเกิดความหมายเพิ่มขึ้นใหม่ นั่นเอง

เสียงสูง – ต่ำ ในภาษาพูด เกิดจากการสั่นสะเทือนของเส้นเสียงในอัตราต่างๆ กัน โดยเสียงที่เปล่งออกมาในขณะที่เส้นเสียงสั่นนั้น จะต้องเป็นเสียงก้อง ดังนั้นหน่วยเสียงวรรณยุกต์ ในภาษาไทยจึงจัดเป็นหน่วยเสียงซ้อน กับอักษรที่ใช้จึงเป็นรูปเครื่องหมายเขียนซ้อนข้างบนหน่วยเสียงสระ (ซึ่งเป็นเสียงก้อง) ซึ่งรูปวรรณยุกต์อยู่ 4 รูป แทนเสียงวรรณยุกต์ทั้งหมด 5 หน่วยเสียง โดยเสียงสามัญไม่มีรูปวรรณยุกต์

ซึ่งลักษณะของเสียงวรรณยุกต์ สามารถแบ่งออกตามลักษณะระดับเสียง ได้ตามตารางที่ 2.8

ตารางที่ 2.8 แสดงลักษณะของเสียงวรรณยุกต์ในภาษาไทย

| หน่วยเสียง  |   | สัญลักษณ์ | เสียง | ลักษณะ   |
|---|---|-----------|-------|--|
| กลุ่ม<br>วรรณยุกต์<br>ระดับ (Level<br>tone)             | หน่วยเสียงวรรณยุกต์<br>ระดับต่ำ (Low tone)          | / \ /     | เอก   | หน่วยเสียงนี้จะปรากฏในพยางค์ของ<br>ภาษาไทย ได้ทุกแบบ   |
|   | หน่วยเสียงวรรณยุกต์<br>ระดับกลาง (Mid<br>tone)      | / - /     | สามัญ | หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ ที่มี<br>ตัวสะกดเป็น พยัญชนะกัก (พยางค์คำ<br>ตาย)            |
|   | หน่วยเสียงวรรณยุกต์<br>ระดับสูง (High tone)         | / / /     | ตรี   | หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่<br>ผสมด้วยสระเสียงยาวซึ่งมีตัวสะกดเป็น<br>เสียงกัก        |
| กลุ่ม<br>วรรณยุกต์<br>เปลี่ยนระดับ<br>(Contour<br>tone) | หน่วยเสียงวรรณยุกต์<br>เปลี่ยนตก (Falling<br>tone)  | / ^ /     | โท    | หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มี<br>สระเสียงสั้น และมีเสียงพยัญชนะสะกด<br>เป็นพยัญชนะกัก |
|   | หน่วยเสียงวรรณยุกต์<br>เปลี่ยนขึ้น (Rising<br>tone) | / v /     | จัตวา | หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มี<br>เสียงพยัญชนะสะกด เป็นเสียงกัก                        |

#### 2.4 ลักษณะ พยางค์ และคำในภาษาไทย

“พยางค์ หมายถึง จำนวนเสียงที่ชัดเจนซึ่งปรากฏในกลุ่มเสียงที่เรียงกันเป็นคำพูด ส่วนเสียงอื่นๆ ที่อยู่ข้างเคียงก็จะประกอบกันเข้าเป็นส่วนหนึ่งของพยางค์” เสียงที่ชัดเจนในกลุ่มเสียงก็คือ

เสียงสระ ซึ่งมีลักษณะประจำตัวก็คือเป็นเสียงก้อง ซึ่งดังเด่นกว่าเสียงอื่นๆ ดังนั้นเสียงสระจึงมักเป็นเสียงที่ทำให้เกิดพยางค์ ถ้ามีเสียงสระเด่นอยู่ที่เสียง พยางค์ก็จะมีจำนวนเท่านั้นด้วย

พยางค์ที่เปล่งออกมาครั้งหนึ่งๆ อาจมีความหมายหรือไม่ก็ได้ แต่เมื่อใดพยางค์ที่ประกอบขึ้นจากเสียงสระ พยัญชนะ และวรรณยุกต์ เป็นอย่างน้อยที่สุด และกลุ่มเสียงเหล่านี้มีความหมายและสามารถปรากฏได้โดยลำพัง พยางค์นั้นๆ ก็กลายเป็นคำในภาษาไทย

คำในภาษาไทยส่วนใหญ่จะเป็นคำพยางค์เดียว ซึ่งเป็นคำพื้นฐาน (Base Words) ของภาษา ภาษาไทยจึงจัดอยู่ในตระกูลภาษา คำโดด หรือ คำพยางค์เดียว (Monosyllabic Language) หน่วยเสียงที่ประกอบกันเข้าเป็นพยางค์ จะต้องมียังน้อย 3 หน่วยคือ หน่วยเสียงพยัญชนะต้น 1 หน่วย หน่วยเสียงสระ 1 หน่วย และหน่วยเสียงวรรณยุกต์ 1 หน่วย และมีหน่วยเสียงอย่างมากไม่เกิน 5 หน่วย คือ เพิ่มหน่วยเสียงพยัญชนะต้นที่เป็นเสียงควบกล้ำอีก 1 หน่วย และหน่วยเสียงพยัญชนะสะกด อีก 1 หน่วย โดยมีองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์ แสดงได้ ดังรูปที่

2.6

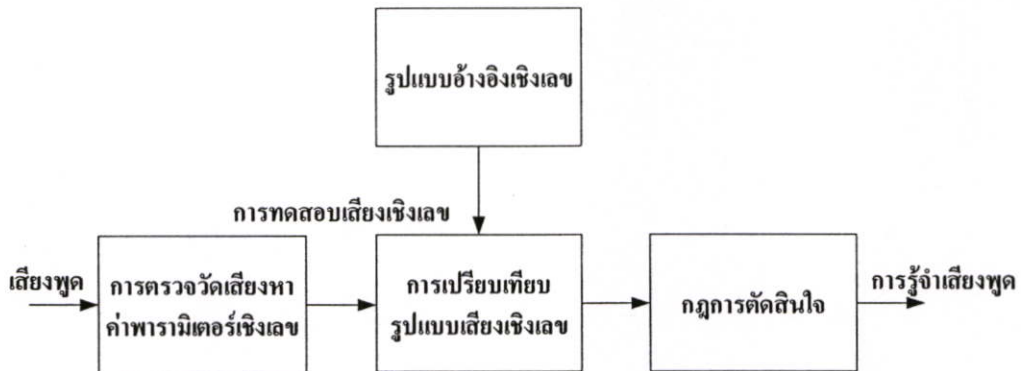
|            |       |           |               |
|------------|-------|-----------|---------------|
|            |       | วรรณยุกต์ |               |
| พยัญชนะต้น | (ควบ) | สระ       | (พยัญชนะสะกด) |

รูปที่ 2.6 แสดงองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์

## 2.5 กระบวนการรู้จำเสียงพูด

ในงานวิจัยนี้เป็นการศึกษาการรู้จำเสียงตัวเลขภาษาไทย โดยใช้กระบวนการรู้จำเสียงสระเสียงเดียวและสระเสียงประสมในภาษาไทยโดยการใช้สเปกตรัมแอลพีซี บนสเกลบาร์ก [8] และกระบวนการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทยโดยใช้เทคนิค การควอนไทซ์พิทช์ และ Hidden Markov Modeling [9] ซึ่งจากหัวข้อที่ผ่านมาเราได้ทราบถึงการทำงานของระบบเสียงพูดของมนุษย์ และหน่วยเสียงในภาษาไทยมาแล้ว สำหรับการรู้จำเสียงตัวเลขภาษาไทยเป็นส่วนหนึ่งของกระบวนการรู้จำเสียงพูด ดังนั้นในหัวข้อนี้เราจะกล่าวถึง กระบวนการรู้จำเสียงพูดของตัวเลขภาษาไทย ทำการศึกษาแบบจำลองระบบกำเนิดเสียงพูด การวิเคราะห์สัญญาณเสียงพูดเชิงเลขเบื้องต้น การคำนวณสเปกตรัมแอลพีซี แถบความถี่วิกฤติและความเข้มแถบความถี่วิกฤติ การหาค่า K-Nearest Neighbor การประมาณค่าพิทช์โดยใช้โอโตคอร์รีเลชันฟังก์ชัน และการสร้างแบบจำลองการรู้จำเสียงหน่วยวรรณยุกต์ด้วยวิธี Hidden Markov Model ซึ่งเราสามารถแสดงบล็อกไดอะแกรมกระบวนการรู้จำเสียงพูด ดังรูปที่ 2.7 จากรูปมีการแบ่งขั้นตอนการรู้จำเสียงออกเป็น 3 ขั้นตอน คือ 1) ขั้นตอนการตรวจวัดเสียงหาค่าพารามิเตอร์เสียงเชิงเลข ซึ่งจะเป็นการสร้างรูปแบบเสียงเชิงเส้น

สำหรับทดสอบการรู้จำเสียงด้วย 2) ขั้นตอนการเปรียบเทียบรูปแบบเสียงเชิงเลข และ 3) ขั้นตอนการตัดสินใจผลของการรู้จำเสียง [10] นอกจากนี้

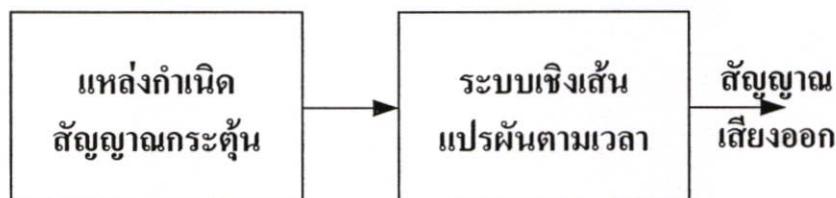


รูปที่ 2.7 บล็อกไดอะแกรมกระบวนการรู้จำเสียงพูด

จากรูปที่ 2.7 ขั้นตอนการตรวจวัดเสียงหาค่าพารามิเตอร์เชิงเลขเป็นการทำหน้าที่ค้นหาแสดงส่วนต่างๆ ทางกายภาพของเสียงพูด พารามิเตอร์เชิงเลขนี้มีความจำเป็นที่จะต้องทำให้มีความกะทัดรัด และมีประสิทธิภาพ ซึ่งการทำให้พารามิเตอร์เชิงเลขได้ลักษณะนั้นมีหลายวิธี เช่น การเพิ่มประสิทธิภาพของการคำนวณ ชนิดของอุปกรณ์ที่ใช้ ชนิดของวิธีการที่ใช้ หน่วยความจำที่เป็นประโยชน์ทั้งขนาด และคุณภาพ ทั้งหมดที่กล่าวมาอยู่บนทางเลือกของระบบประมวลสัญญาณ (Signal Processing) [10]

## 2.6 แบบจำลองระบบกำเนิดเสียงพูด

เราสามารถแสดงบล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้นดังรูปที่ 2.8 จากรูปมีการแยกภาคแหล่งกำเนิดสัญญาณกระตุ้นออกจากส่วนกำเนิดเสียงซึ่งแทนด้วยระบบเชิงเส้นแปรผันตามเวลา [5]



รูปที่ 2.8 บล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น

จากรูปที่ 2.8 แหล่งกำเนิดสัญญาณกระตุ้นทำหน้าที่แทนการทำงานของปอดและกล่องเสียงส่วนนี้จะผลิตขบวนพัลส์ที่มีความยาวพิชชณะเปล่งเสียง โฆษะ (Voiced) และให้กำเนิดเสียงซึ่งคล้ายเสียงรบกวนขณะเปล่งเสียงอ โฆษะ (Voiceless)

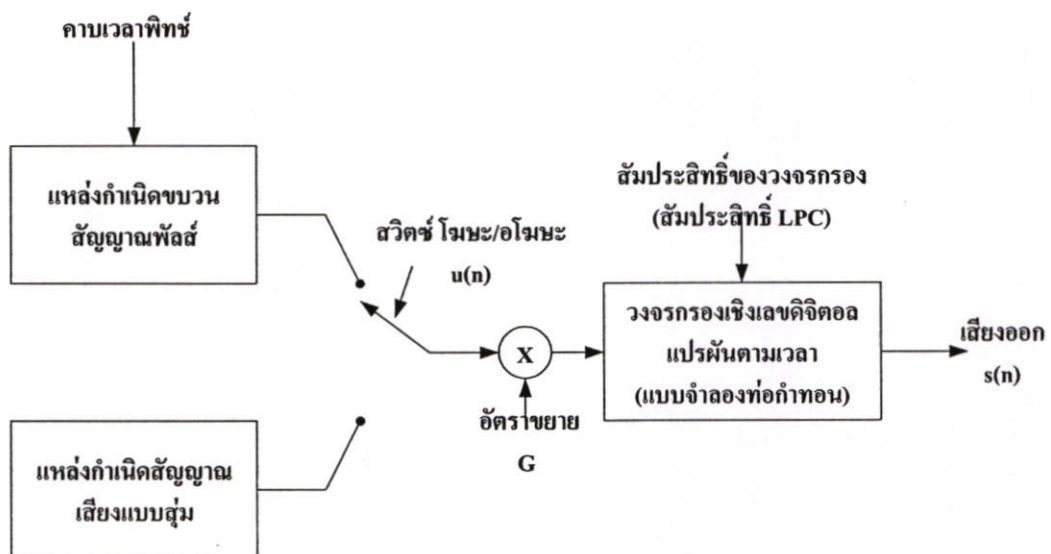
ส่วนที่สองเป็นท้อกำทอนเสียง จะแทนการทำงานของช่องปากและโพรงจมูก ทำหน้าที่เสมือนตัวกรองสัญญาณ (Filter) ที่ยอมให้ความถี่ฟอร์แมนท์ผ่านได้ ซึ่งสามารถแทนด้วยระบบเชิงเส้นแปรผันตามเวลา (Time-varying Linear System)

### 2.6.1 แบบจำลองระบบกำเนิดเสียงพูดเชิงเลข

จากแบบจำลองรูปที่ 3.2 ประกอบกับการใช้เทคนิคการประมวลสัญญาณเชิงเลข (Digital Signal Processing) สามารถแสดงเป็นแบบจำลองระบบกำเนิดสัญญาณเสียงพูดเชิงเลขดังรูปที่ 2.9

2.6.1.1 แหล่งกำเนิดสัญญาณกระตุ้นแบ่งเป็น 2 ชนิด ตามประเภทของเสียง โดยชนิดแรกเป็นแหล่งกำเนิดเสียง โฆษะ (Voiced) ทำหน้าที่ผลิตขบวนสัญญาณพัลส์ด้วยคาบเวลาคงที่ และชนิดที่สองเป็นแหล่งกำเนิดสัญญาณเสียงรบกวนแบบสุ่ม ในกรณีที่เกิดเสียงอ โฆษะ (Voiceless)

2.6.1.2 แบบจำลองอวัยวะกำทอนเสียง เป็นวงจรกรองเชิงดิจิทัลซึ่งเป็นชนิดที่มีแคโพลทั้งหมด (All-pole Digital Filter) หรือ วงจรกรองเชิงเลขแบบป้อนกลับ (Recursive Digital Filter) [10]



รูปที่ 2.9 แบบจำลองระบบสร้างสัญญาณเสียงพูดเชิงเลข

### 2.6.2 พารามิเตอร์ (Parameter) ของระบบกำเนิดเสียงพูด

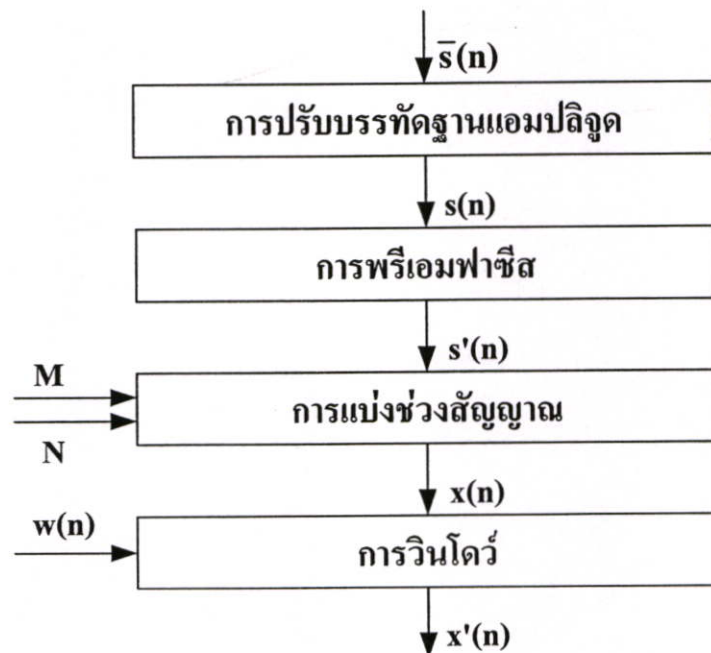
การสร้างระบบเสียงพูดจำเป็นต้องวิเคราะห์หาค่าตัวแปรหรือพารามิเตอร์ที่ต้องใช้ตามแบบจำลองและระบบประมวลสัญญาณเสียง ซึ่งประกอบไปด้วย

- ดัชนีหรือพารามิเตอร์บอกชนิดของเสียงว่าเป็นเสียง โฆษะ (Voiced) หรือเสียงอ โฆษะ (Voiceless)

- คาบเวลาพิทช์ของขบวนพัลส์  $u(n)$  ในกรณีเสียงโฆษะ (Voiced)
- อัตราการขยาย  $G$  (gain) ของวงจรกรอง หรือค่ารากกำลังสองเฉลี่ยของเสียง
- พารามิเตอร์ของแบบจำลองท่อกำเนิดหรือสัมประสิทธิ์ของวงจรกรอง
- อัตราการสุ่มตัวอย่างสัญญาณเสียงหรืออัตราเร็วข้อมูล

## 2.7 การวิเคราะห์สัญญาณเสียงพูดเบื้องต้น

โดยธรรมชาติของสัญญาณเสียงพูดจะไม่เสถียรและเปลี่ยนแปลงตามเวลา (Non-Stationary) ดังนั้น เมื่อต้องการนำสัญญาณเสียงพูดมาประมวลผลสัญญาณดิจิทัล (Digital Signal Processing) จึงจำเป็นต้องแบ่งสัญญาณพูดออกเป็นช่วงเวลาสั้น ๆ (Short Time) เพื่อให้สัญญาณเสียงมีความเสถียรและไม่เปลี่ยนแปลงเวลา (Stationary) จากนั้นจึงจะสามารถนำสัญญาณเสียงไปประมวลผลต่อไปได้ กรอบเสียงพูด (Speech Frame) ความยาวประมาณ 20-30 มิลลิวินาที ทำให้สัญญาณเสียงพูดในแต่ละกรอบเสียงพูดเป็นสัญญาณที่มีความเสถียรและไม่เปลี่ยนแปลงตามเวลา การเหลื่อมกรอบเสียงพูด (Frame Overlap) จะทำให้รอยต่อของลักษณะสำคัญของเสียงพูด จากกรอบเสียงพูดหนึ่งไปยังอีกกรอบเสียงพูดหนึ่งเรียบ (Smooth) ขึ้น กรรมวิธีการวิเคราะห์สัญญาณเสียงพูดเบื้องต้นดังรูปที่ 2.10 ประกอบด้วย [11]



รูปที่ 2.10 ขั้นตอนการวิเคราะห์สัญญาณเสียงพูดเบื้องต้น

### 2.7.1 กรรมวิธีการปรับบรรทัดฐานแอมพลิจูด (Amplitude Normalization)

กรรมวิธีการปรับบรรทัดฐานแอมพลิจูดของสัญญาณเสียงพูดเป็นการเพิ่มหรือลดขนาดของสัญญาณเสียงพูด เพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน จึงจำเป็นต้องปรับให้มีขนาดของสัญญาณเสียงพูดอยู่ในบรรทัดฐานเดียวกัน เพื่ออำนวยความสะดวกในการวัดคุณลักษณะและเปรียบเทียบสัญญาณเสียง การปรับบรรทัดฐานแอมพลิจูดแสดงดังสมการที่ 2.1

$$S[i] = \frac{\bar{S}[i]}{2^{N-1}} \quad (2.1)$$

เมื่อ  $S[i]$  คือสัญญาณเสียงที่ปรับบรรทัดฐานแอมพลิจูดแล้ว  
 $\bar{S}[i]$  คือสัญญาณเสียงพูดดิจิทัล  
 $N$  คือ จำนวนบิต (Bit) ที่ใช้แทนค่าของสัญญาณเสียงพูด

### 2.7.2 กรรมวิธีการเน้นล่วงหน้า (Pre-emphasis)

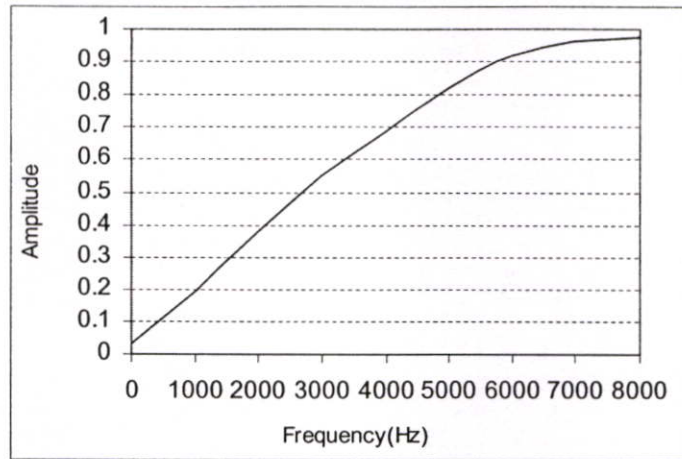
เนื่องจากสัญญาณเสียงพูดของมนุษย์ จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ เมื่อเทียบกับแถบความถี่ปฏิบัติงาน (Bandwidth) ไม่เกิน 5 กิโลเฮิร์ตซ์ ดังนั้น เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal to Noise Ratio : SNR) มีค่าค่อนข้างคงที่ตลอดช่วงความถี่ปฏิบัติงานนี้ เราจึงต้องมีการพรีเอมฟาซิส (Pre-emphasis) โดยเน้นความถี่สูงให้มีขนาดสูงขึ้น ดังนั้นการพรีเอมฟาซิสก็คือการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (High Pass Filter) ซึ่งมักนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชันถ่ายโอนเป็น

$$H(z) = 1 - a \cdot z^{-1} \quad \text{โดยที่ } 0.9 < a < 1.0 \quad (2.2)$$

เมื่อเทียบกับรูปที่ 2.10 เราจะได้ว่า

$$s'[i] = s[i] - a \cdot s[i-1] \quad (2.3)$$

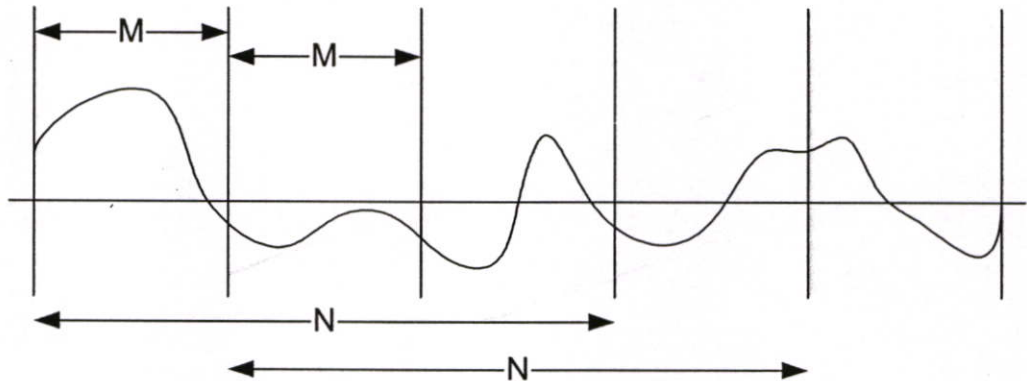
ยิ่งค่า  $a$  ใกล้ 1 เท่าใด ความถี่สูงก็จะยิ่งถูกขยายมากขึ้นเท่านั้น ค่า  $a$  ที่นิยมสำหรับใช้ในการหาพารามิเตอร์ของ LPC คือ  $15/16 = 0.9375$  [10] เมื่อนำฟังก์ชันถ่ายโอนมาพล็อตกราฟของขนาดเทียบกับความถี่จะได้ดังรูปที่ 2.11



รูปที่ 2.11 แสดงขนาดสเปกตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส

### 2.7.3 การแบ่งช่วงสัญญาณ (Frame Blocking)

สัญญาณที่ผ่านการพรีเอมฟาซิสแล้ว  $S'[i]$  จะถูกตัดแบ่งออกเป็นช่วงๆ หรือ เฟรม ช่วงละ  $N$  ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ  $N$  ตัวอย่างสัญญาณ ดังรูปที่ 2.12



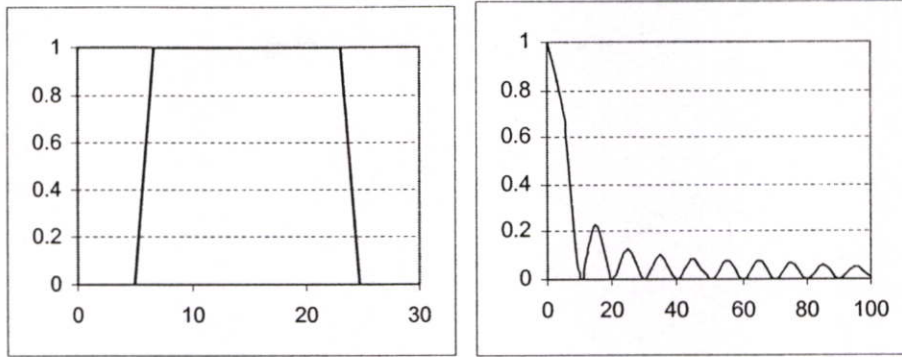
รูปที่ 2.12 แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์

โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ  $M$  ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า  $M$  โดกว่าค่า  $N$  ในการเลื่อนของช่วงในการวิเคราะห์จะทำให้บางสัญญาณไม่ถูกให้ในการวิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า  $M$  เล็กกว่า  $N$  ก็จะ ทำให้ตัวอย่างสัญญาณ ทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า  $M$  เล็กเท่าใดความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่ก็จะทำให้การคำนวณช้าลง

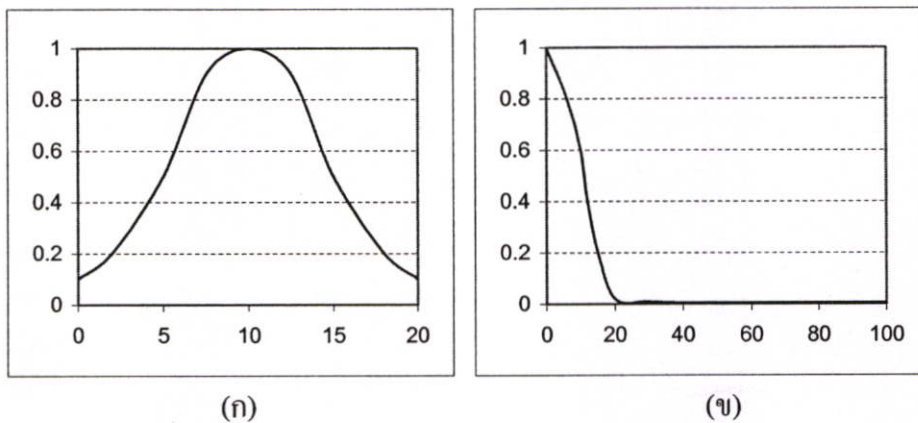
หลักการกำหนดขนาดกรอบของเฟรมในการวิเคราะห์ คือ

2.7.3.1 กรอบของเฟรมจะต้องสั้นพอที่คุณสมบัติของเสียงที่เราสนใจ จะวิเคราะห์นั้นยังไม่เปลี่ยนแปลงในกรอบของเฟรมนั้น





รูปที่ 2.14 แสดงวินโดว์แบบสี่เหลี่ยม (ก) ใน โดเมนเวลา (ข) ใน โดเมนความถี่



รูปที่ 2.15 แสดงวินโดว์แบบแฮมมิง (ก) ใน โดเมนเวลา (ข) ใน โดเมนความถี่

ดังนั้นจากรูปที่ 3.4 เราจะได้ว่า

$$x'(n) = x(n) \cdot w(n) \quad (2.5)$$

## 2.8 หลักการของการประมาณเชิงเส้น

เนื่องจากสัญญาณเสียง มีพารามิเตอร์ที่สำคัญต่างๆดังที่กล่าวมาแล้ว การวิเคราะห์การเข้ารหัสโดยการประมาณเชิงเส้น (Linear Predictive Coding: LPC) จึงถูกใช้ในการหาค่าพารามิเตอร์พื้นฐานต่างๆ ที่กล่าวมา เพื่อใช้สำหรับเข้ารหัสสัญญาณเสียงเพื่อการรู้จำเสียง หรือการส่งสัญญาณเสียงด้วยความเร็วบีตค่า ด้วยเหตุที่ว่าวิธีนี้เป็นการวิเคราะห์พารามิเตอร์ต่างๆ ที่ให้ความแม่นยำสูง และสามารถย่อข้อมูลได้อย่างมีประสิทธิภาพ

หลักการพื้นฐานของการประมาณเชิงเส้น[10] [12] คือ การประมาณค่าสัญญาณจากผลรวมเชิงเส้นของสัญญาณก่อนหน้า

สมมติว่าสัญญาณเดิมเป็น  $s(n)$  การประมาณค่าสัญญาณเป็น  $s'(n)$  ดังนั้นเราสามารถอธิบายการประมาณเชิงเส้นได้ด้วยสมการต่อไปนี้

$$s'(n) = \sum_{k=1}^p \alpha_k s(n-k) \quad (2.6)$$

เมื่อ  $\alpha_k$  เป็นค่าคงที่ วิธีการนี้เรียกว่าการประมาณเชิงเส้นอันดับ  $p$  โดยมีเงื่อนไขว่า ค่า  $\alpha_k$  ที่ใช้ในการประมาณจะต้องทำให้ผลรวมของกำลังสองของความคลาดเคลื่อน  $\{s(n) - s'(n)\}^2$  มีค่าน้อยที่สุด นั่นคือ  $\sum e^2(n) = \sum \{s(n) - s'(n)\}^2$  มีค่าต่ำที่สุด ซึ่งการประมาณเชิงเส้นนี้มีหลายวิธี [10] ได้แก่

- วิธีโควาเรียนซ์ (Covariance Method)
- วิธีออโตคอร์รีเลชัน (Autocorrelation Method)
- วิธีแลตทิซ (Lattice Method)
- วิธีวงจรรองผกผัน (Inverse Filter Method)
- วิธีการประมาณสเปกตรัม (Spectrum Estimation Method)

และอื่นๆอีกหลายวิธี แต่ในที่นี้จะเลือกใช้วิธีออโตคอร์รีเลชัน หรือ วิธีออคสัมพันธ์

จากหลักการพื้นฐานของการประมาณเชิงเส้น และรูปที่ 2.9 ซึ่งเป็นแบบจำลองระบบสร้างสัญญาณเสียง เราสามารถเขียนสมการได้เป็น

$$s(n) = G \cdot u(n) + \sum_{k=1}^p \alpha_k s(n-k) \quad (2.7)$$

โดยที่  $\alpha_k$  คือสัมประสิทธิ์ของวงจรรองเชิงเลข

ส่วนการประมาณเชิงเส้นโดยใช้สัมประสิทธิ์  $\{\alpha_k\}$  คือ

$$s'(n) = \sum_{k=1}^p \alpha_k s(n-k) \quad (2.8)$$

ดังนั้นความคลาดเคลื่อนคือ

$$\begin{aligned} e(n) &= s(n) - s'(n) \\ &= s(n) - \sum_{k=1}^p \alpha_k s(n-k) \end{aligned} \quad (2.9)$$

ฟังก์ชันถ่ายโอนระหว่าง  $e(n)$  และ  $s(n)$  คือ

$$\begin{aligned} A(z) &= \frac{E(z)}{S(z)} \\ &= 1 - \sum_{k=1}^p \alpha_k z^{-k} \end{aligned} \quad (2.10)$$

จากสมการที่ 2.2 ถึง 2.5 จะเห็นได้ว่าถ้า  $\{\alpha_k\} = \{a_k\}$  แล้ว

$$e(n) = G \cdot u(n) \quad (2.11)$$

ในการประมวลสัญญาณจะมีการแบ่งสัญญาณออกเป็นช่วงสั้นๆ (Windowing) [5] [10] ช่วงละประมาณ 20-40 มิลลิวินาที ซึ่งก็ได้มีการอธิบายผ่านมา ดังนั้นค่าผลรวมของกำลังสองของความคลาดเคลื่อน

$$\begin{aligned} E &= \sum_m e_n^2(m) \\ &= \sum_m [s_n(m) - s'_n(m)]^2 \\ &= \sum_m [s_n(m) - \sum_{k=1}^p \alpha_k s(n-k)]^2 \end{aligned} \quad (2.12)$$

โดยที่  $n$  คือช่วงที่  $n$  ของสัญญาณที่ใช้คำนวณ เพราะฉะนั้นเพื่อให้ได้ค่า  $E_n$  ค่าที่ต่ำที่สุดจะต้องมีเงื่อนไขว่า

$$\frac{\partial E_n}{\partial \alpha_i} = 0 \quad \text{เมื่อ } i=1,2,\dots,p$$

จากสมการที่ 2.12

$$\begin{aligned} \frac{\partial E_n}{\partial \alpha_i} &= -2s_n(m-i) \sum_m \left[ s_n(m) - \sum_{k=1}^p \alpha_k s_n(m-k) \right] \\ &= -2 \left[ \sum_m s_n(m)s_n(m-i) - \sum_{k=1}^p \alpha_k \sum_m s_n(m-k)s_n(m-i) \right] \quad \text{เมื่อ } i=1,2,3,\dots,p \end{aligned}$$

$$\frac{\partial E_n}{\partial \alpha_i} = 0 \quad \text{ก็ต่อเมื่อ}$$

$$\sum_{k=1}^p \alpha_k \sum_m s_n(m-k) s_n(m-i) = \sum_m s_n(m) s_n(m-i) \text{ เมื่อ } i=1,2,3,\dots,p \quad (2.13)$$

ถ้าเรากำหนดให้  $\phi_n(i,k) = \sum_m s_n(m-k) s_n(m-i)$  เพราะฉะนั้น

$$\sum_{k=1}^p \alpha_k \phi_n(i,k) = \phi_n(i,0) \quad (2.14)$$

จากสมการที่ 2.12-2.14 จะได้ว่า

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^p \alpha_k \sum_m s_n(m) s_n(m-i)$$

และจาก  $\phi_n(i,k) = \sum_m s_n(m-k) s_n(m-i)$

$$E_n = \phi_n(0,0) - \sum_{k=1}^p \alpha_k \phi_n(0,k) \quad (2.15)$$

สมมติว่าใน 1 เฟรมของสัญญาณที่ตัดมาจำนวนมี  $N$  ตัวอย่างคือ  $s_n(0), s_n(1), s_n(2), \dots, s_n(N-1)$  ในที่นี้เราให้  $s_n(m) = 0$  เมื่อ  $m < 0$  หรือ  $m > N-1$  เพราะฉะนั้น

$$\begin{aligned} \phi_n(i,k) &= \sum_m s_n(m-k) s_n(m-i) \\ &= \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m+i-k) \quad 0 \leq k \leq p, k \leq i \leq p \end{aligned}$$

ให้ 
$$R_n(k) = \sum_{m=0}^{N-i-k} s_n(m) s_n(m+k) \quad (2.16)$$

$$\therefore \phi_n(i,k) = R_n(|i-k|) \text{ เมื่อ } i=1,2,\dots,p; k=0,1,2,\dots,p$$

จากสมการที่ 2.14 จะได้ว่า

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i) \text{ เมื่อ } i=1,2,\dots,p \quad (2.17)$$

และจากสมการที่ 2.15

$$E_n = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) \quad (2.18)$$

จากสมการที่ 2.17 เขียนให้อยู่ในรูปเมตริกซ์ได้เป็น

$$\begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p-1) \\ R_n(1) & R_n(2) & \cdots & R_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{bmatrix} \bullet \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix} \quad (2.19)$$

หรือ  $R_n \bullet \alpha = r_n \quad (2.20)$

เมื่อ  $R_n = \begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p-1) \\ R_n(1) & R_n(2) & \cdots & R_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{bmatrix}$ ,  $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$  และ  $r_n = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix}$

จากสมการที่ 2.7 และ 2.9

$$G \bullet u(n) = s(n) - \sum_{k=1}^p a_k s(n-k)$$

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k)$$

เมื่อ  $a_k = \alpha_k$  จะได้ว่า

$$e(n) = G \bullet u(n)$$

$$\therefore E_n = \sum_{m=0}^{N-1} e^2(m) = G \sum_{m=0}^{N-1} u^2(m) \quad (2.21)$$

$$= R_n(0) - \sum_{k=1}^p \alpha_k R_n(k)$$

จากสมการที่ 2.20 เราสามารถหา  $\alpha$  โดยตรงดังนี้

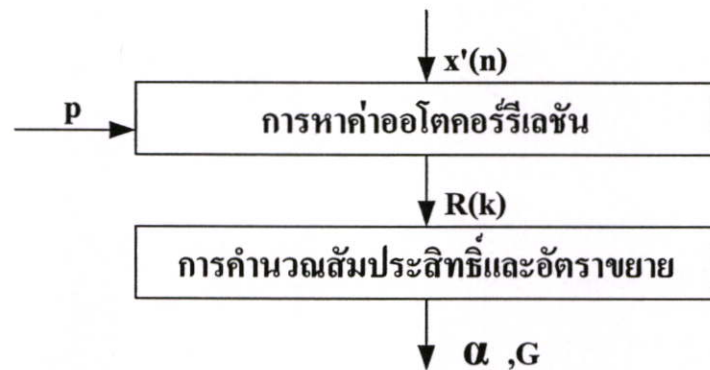
$$\alpha = R_n^{-1} \bullet r_n \quad (2.22)$$

และจากสมการที่ 2.21 เราสามารถหาค่า  $G$  โดยตรงจาก

$$G = \frac{R_n(0) - \sum_{k=1}^p \alpha_k R_n(k)}{\sum_{m=0}^{N-1} u^2(m)} \quad (2.23)$$

### การคำนวณสเปกตรัมแอลพีซี

การวิเคราะห์สัญญาณเสียงเบื้องต้น ดังรูปที่ 2.10 จะเป็นการกำหนดขอบเขตเฟรมเสียงเท่านั้น การจะได้พารามิเตอร์เสียงพูดเชิงเลขจำเป็นต้องมาผ่านการประมวลผลเชิงเส้น และหาฟังก์ชันถ่ายโอนของอวัยวะกำทอนเสียง หรือเรียกอีกอย่างว่าสเปกตรัมแอลพีซี ต่อไปแสดงขั้นตอนการหาค่าพารามิเตอร์เสียงพูดเชิงเลข ดังรูปที่ 2.16



รูปที่ 2.16 ขั้นตอนการหาค่าพารามิเตอร์เสียงพูดเชิงเลข

จากรูปที่ 2.16 การคำนวณอโตคอร์รีเลชัน โดยแทนค่าสมการที่ 2.11 ที่ได้จากการวิเคราะห์เสียงเบื้องต้น ในสมการที่ 2.16 จะได้ว่า [10]

$$R_n(k) = \sum_{m=0}^{N-1-k} x'(m)x'(m+k) \quad (2.24)$$

เมื่อได้ค่า  $R_n(0), R_n(1), R_n(2), \dots, R_n(p)$  แล้วก็สามารถหาค่าสัมประสิทธิ์  $\alpha$  และค่าอัตราขยาย  $G$  ได้จากสมการที่ 2.22 และ 2.23 ตามลำดับ

จากแบบจำลองระบบสร้างสัญญาณเสียงพูดเชิงเลข ในรูปที่ 2.9 แสดงให้เห็นว่าสเปกตรัมแอลพีซีก็คือ ฟังก์ชันถ่ายโอนของอวัยวะกำทอนเสียง ดังนั้นจากสมการที่ 2.7 ทำการแปลงให้อยู่ใน  $Z$  โดเมนได้เป็น

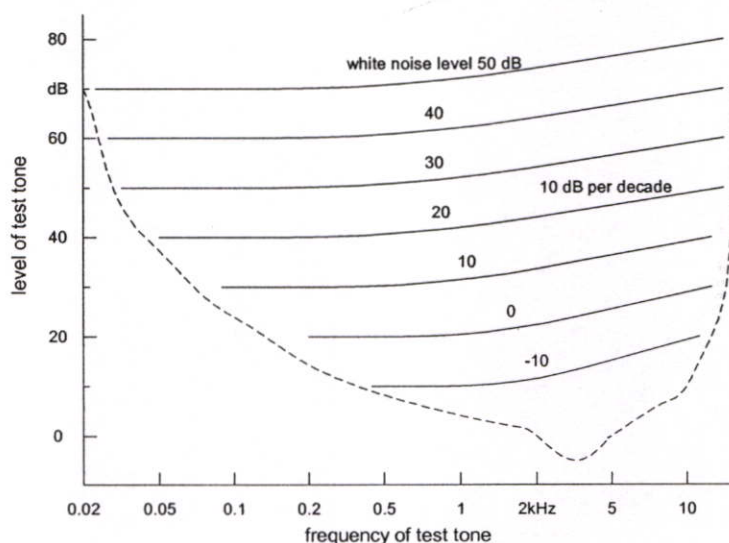
$$S(z) = \sum_{k=1}^p a_k z^{-k} S(z) + GU(z)$$

ซึ่งจัดรูปให้เป็นฟังก์ชันถ่ายโอนของอวัยวะกำทอนเสียงทำให้สามารถคำนวณ สเปกตรัมแอลพีซีได้ ดังนี้

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2.25)$$

## 2.9 แถบความถี่วิกฤต และความเข้มสัญญาณในแถบความถี่วิกฤต

คำจำกัดความของแถบความถี่วิกฤตนั้นได้ถูกกล่าวขึ้นโดย Fletcher [13] โดยที่เขาได้ตั้งสมมติฐานไว้ว่าในกรณีที่มีสัญญาณรบกวน(Noise) ทับซ้อนจนสามารถบดบัง(Mask) สัญญาณ โทนทดสอบ(Test Tone) ได้ นั่น ส่วนของสัญญาณรบกวนที่มีผลในการบดบังสัญญาณ โทนทดสอบจริงๆ จะเป็นเพียงส่วนที่มีความถี่ใกล้เคียงกับสัญญาณ โทนทดสอบเท่านั้น ในการที่จะทำการหาค่าสัมประสิทธิ์ต่างๆของแถบความถี่วิกฤตได้นั้นยังต้องอาศัยสมมติฐานที่ว่า การบดบังสัญญาณจะเป็นผลก็ต่อเมื่อกำลังงานของสัญญาณรบกวนในแถบสัญญาณเดียวกับสัญญาณ โทนทดสอบนั้น เท่ากับกำลังงานของสัญญาณ โทนทดสอบ และถือว่ากำลังงานของสัญญาณรบกวนที่อยู่ นอกเหนือจากแถบสัญญาณที่ใกล้กับสัญญาณ โทนทดสอบไม่มีผลต่อการบดบังสัญญาณ



รูปที่ 2.17 ระดับของสัญญาณ โทนทดสอบที่ถูกบดบังโดยสัญญาณ White Noise ที่ระดับสัญญาณ ต่างๆ และเส้นประเป็นระดับ Threshold In Quiet

รูปที่ 2.17 แสดงการบดบังสัญญาณโทนทดสอบโดยใช้สัญญาณรบกวนแถบความถี่กว้าง (Broadband Noise) ซึ่งในที่นี้ใช้ White Noise ที่ระดับความเข้มสัญญาณต่างๆ โดยเส้นประเป็นระดับ Threshold In Quiet ซึ่งหมายถึงระดับของสัญญาณโทนที่หูของมนุษย์เริ่มที่จะได้ยินเมื่อไม่มีสัญญาณรบกวนใดๆ เช่นหากมีสัญญาณโทนที่ความถี่ 2 kHz โดยไม่มีสัญญาณรบกวนใดๆ เราสามารถได้ยินสัญญาณโทนนั้นได้ก็ต่อเมื่อ สัญญาณมีระดับความเข้มประมาณ  $0 \text{ dB}(10^{-12} \text{ w/m}^2)$  ขึ้นไป ส่วนเส้นทึบเป็นเส้นที่แสดงระดับ Masking Threshold สำหรับความเข้มสัญญาณ White Noise ที่ระดับต่างๆ ซึ่ง Threshold นี้บ่งบอกว่าหากสัญญาณโทนมีระดับความเข้มน้อยกว่า Threshold นี้ เราจะไม่สามารถได้ยินสัญญาณโทนนี้ได้ เช่น สัญญาณโทนที่ 2 kHz ขณะที่มีความเข้มสัญญาณมากกว่าประมาณ 40 dB ซึ่งจะเห็นได้ว่าในการบดบังสัญญาณโดย White Noise นั้นจะมีระดับ Threshold ที่ไม่เป็นอิสระกับความถี่ต่างๆที่ตัว White Noise เองนั้นความเข้มสัญญาณของมันไม่ขึ้นกับความถี่ แต่จะเห็นได้ว่าระดับการบดบังสัญญาณจะมีลักษณะที่ไม่ขึ้นกับความถี่ ในช่วงที่สัญญาณมีความถี่ต่ำกว่า 500 Hz ส่วนที่ความถี่ประมาณ 1 kHz ขึ้นไปนั้นจะมีอัตราการเปลี่ยนแปลงอยู่ที่ประมาณ 10 dB/decade

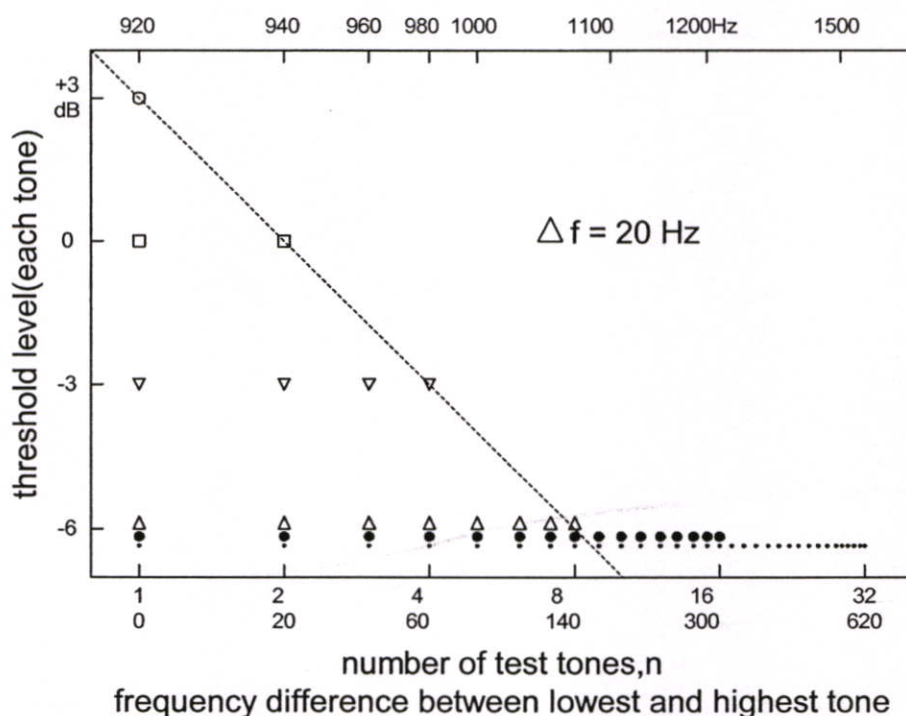
ระบบการรับฟังเสียงของคนเรามีระดับของการถูกบดบังสัญญาณที่ไม่ขึ้นกับความถี่ ในช่วงที่ต่ำกว่า 500 Hz ในช่วงความถี่นี้แถบความถี่วิกฤติจะมีความกว้างเท่าๆ กันด้วย สำหรับความถี่ที่สูงกว่านั้นจะมีระดับการบดบังสัญญาณเพิ่มขึ้นประมาณ 10 dB/decade นั่นก็หมายความว่าระดับความเข้มของสัญญาณนั้นเพิ่มขึ้นตามความถี่ ดังนั้นความกว้างแถบความถี่ของช่วงสัญญาณนี้เพิ่มขึ้น 10 dB/decade ด้วยเช่นกัน

หากเราเชื่อว่าสมมติฐานของ Fletcher นั้นถูกต้องแสดงว่าเราต้องสามารถได้ยินเสียงสัญญาณโทนในขณะที่สัญญาณรบกวนที่มีแถบสัญญาณใกล้เคียงกับแถบความถี่วิกฤติที่มีความถี่ของสัญญาณโทนเป็นศูนย์กลางนั้นมีกำลังเท่ากับสัญญาณโทน ซึ่งจะทำให้เราสามารถที่จะคาดคะเนความกว้างของแถบความถี่วิกฤติได้โดย ที่ความถี่ต่ำกว่า 500 Hz นั้นระดับความเข้มของโทนที่ถูกบดบังสัญญาณจะสูงกว่าระดับความหนาแน่นของ White Noise ที่บดบังสัญญาณโทนนั้นอยู่ประมาณ 17 dB ซึ่งจากสมมติฐานนี้ เราสามารถคำนวณความกว้างแถบความถี่วิกฤติจาก  $17 = 10 \cdot \text{Log}(BW)$  ได้เป็น  $10^{1.7}$  หรือประมาณความกว้าง 50 Hz [13]

แต่ในความเป็นจริงแล้ว เราสามารถได้ยินเสียงโทนในขณะที่กำลังของเสียงโทนเป็นเพียงครึ่งหนึ่งหรือเพียงหนึ่งในสี่ของกำลังของสัญญาณรบกวนในแถบความถี่ที่ใกล้กับความถี่ของเสียงโทนนั้นเท่านั้น [13] [14] และการศึกษาโดยใช้ข้อมูลอื่นๆ เพิ่มเติม เราสามารถคาดคะเนความกว้างของแถบความถี่วิกฤติได้ใกล้เคียงยิ่งขึ้น ซึ่งในย่านความถี่ต่ำกว่า 500 Hz นั้นความกว้างของแถบความถี่วิกฤติจะมีค่าประมาณ 100 Hz ส่วนที่ความถี่สูงกว่า 500 Hz นั้นจะมีความกว้างของแถบความถี่วิกฤติประมาณ 20 เปรอร์เซ็นต์ ของความถี่ศูนย์กลางนั้นๆ

### 2.9.1 วิธีการในการหาความกว้างแถบความถี่วิกฤติ

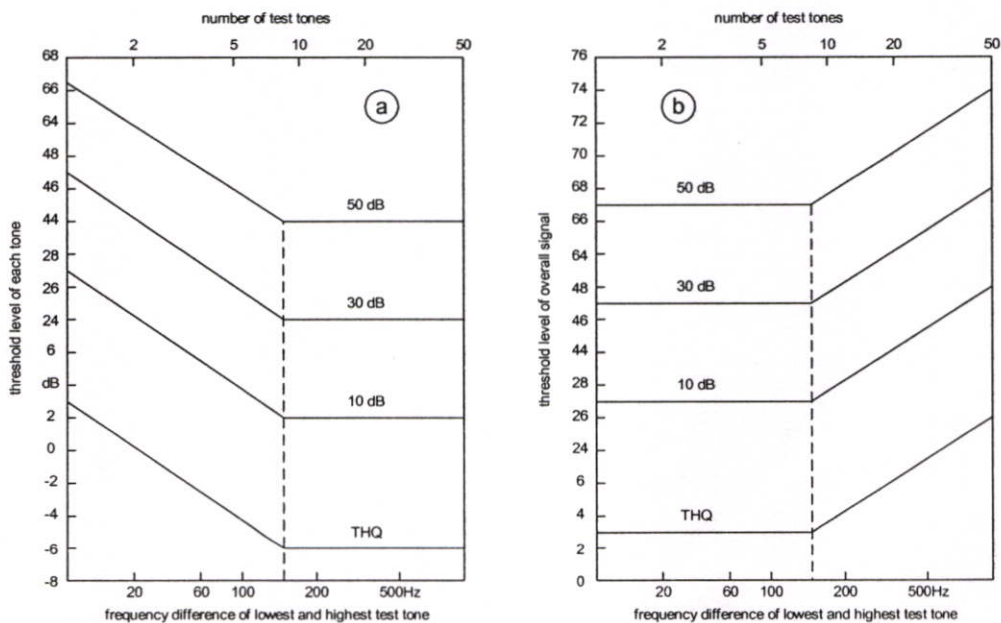
วิธีการแรกที่ใช้ในการหาความกว้างของแถบความถี่วิกฤติคือ การวัดจากระดับ Threshold โดยตรงซึ่งมีลักษณะเช่นเดียวกับวิธีการอื่นๆที่ใช้การวัดระดับสัญญาณโดยตรงคือ ความกว้างแถบความถี่ หรือระดับสัญญาณที่มีความเกี่ยวข้องกับแถบกว้างความถี่ จะต้องเป็นตัวแปรซึ่งในกรณีนี้ Threshold In Quiet ของสัญญาณโทนความถี่ต่างๆซึ่งมีแอมพลิจูดเท่ากันถูกใช้ในการประมาณความกว้างแถบความถี่วิกฤติที่อยู่ในช่วงประมาณ 1 kHz



รูปที่ 2.18 ความสัมพันธ์ระหว่างระดับ Threshold In Quiet และจำนวนสัญญาณ โทณทดสอบ โดยจุดที่มีลูกศรอยู่เป็นจุดที่ใช้ในการคาดคะเนความกว้างแถบความถี่วิกฤติ

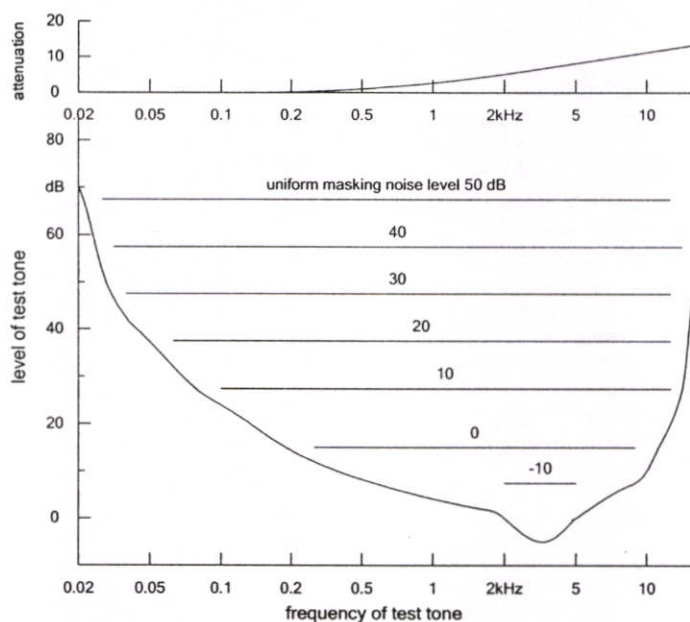
รูปที่ 2.18 เป็นระดับที่หูมนุษย์เริ่มได้ยินเมื่อไม่มีสัญญาณรบกวน (Threshold In Quiet) ของแต่ละ โทณเมื่อเปลี่ยนจำนวนของสัญญาณ โทณความถี่ต่างๆและความแตกต่างระหว่างความถี่สูงสุดและความถี่ต่ำสุดของ โทณ จากรูปเป็นการวัดระดับของ Threshold In Quiet โดยอาศัยวิธีการ Tracking สัญญาณ โดยเริ่มต้นที่สัญญาณ โทณที่ความถี่ 920 Hz เพียงสัญญาณเดียวจะได้ Threshold In Quiet ที่ +3dB เมื่อเพิ่มสัญญาณ โทณอีกค่าก็คือ 940 Hz แล้วทำการ Tracking เพื่อวัดระดับ Threshold In Quiet ของสัญญาณที่มี โทณทั้งสองรวมอยู่ก็จะ ได้ระดับ Threshold In Quiet ของแต่ละ สัญญาณที่ 0 dB และเมื่อทำตามลักษณะนี้ไปเรื่อยๆโดยการเพิ่มจำนวนสัญญาณ โทณทีละสองเท่า แล้วทำการ Tracking เพื่อวัดระดับ Threshold In Quiet จะ ได้ -3dB และ -6dB สำหรับจำนวน โทณ 4 และ 8 โทณตามลำดับซึ่งแสดงให้เห็นว่าระดับ Threshold In Quiet ของแต่ละ โทณจะลดลงเมื่อมี

การรวมกันของสัญญาณ โทนมามากขึ้น แต่เมื่อเพิ่มจำนวน โทนอนไปถึงจุดหนึ่ง ระดับ Threshold In Quiet ของแต่ละ โทนก็จะไม่ลดลงอีกดังที่สังเกตได้ชัดในรูปที่ 2.18 ที่จำนวน โทน 16 และ 32 โทน ก็จะไม่มี ความแตกต่างกันของระดับ Threshold In Quiet ของแต่ละ โทน กับที่ระดับ Threshold In Quiet ของการใช้จำนวน โทนเท่ากับ 8 มากนัก ซึ่งค่านี้เองเป็นค่าที่เราใช้ในการวัดความกว้างแถบ ความถี่วิกฤติ เป็นที่น่าสังเกตว่าช่วงของจำนวน โทนตั้งแต่ 1 ถึง 8 นั้นจะมีระดับ Threshold In Quiet ลดลง 3 dB ทุกๆการเพิ่มขึ้นสองเท่าของจำนวน โทน หมายความว่าที่ระดับ Threshold In Quiet ค่าใดค่าหนึ่งนั้น ระดับแรงดันเสียงรวมทุก โทนก่อนข้างคงที่ไม่ขึ้นกับจำนวน โทน แต่กฎนี้ใช้ได้กับ เฉพาะจำนวน โทนที่ไม่เกิน 8 เท่านั้นถ้ามากกว่านี้ระดับ Threshold In Quiet ของแต่ละ โทนไม่ได้ ลดลงตามการเพิ่มขึ้นของจำนวน โทนจึงทำให้ ระดับแรงดันเสียงโดยรวมเพิ่มขึ้น ซึ่งหมายความว่า ระดับ Threshold In Quiet ของระดับการได้ยินเสียงของเราสามารถประมาณ ได้โดยความเข้มของ เสียง โทนทั้งหมด トラบใดที่ โทนต่างๆอยู่ในแถบความถี่วิกฤติ ส่วนที่อยู่ นอกเหนือแถบกว้างความถี่ วิกฤตินี้จะไม่นำมา มีผลกับค่า Threshold In Quiet หรืออีกนัยหนึ่งคือ สัญญาณ โทนความถี่ 920 Hz ได้ก็ต่อเมื่อมันมีความเข้ม +3 dB ขึ้นไป หรือสัญญาณ ใดๆที่มีความถี่อยู่ในแถบความถี่วิกฤติที่มีความถี่ศูนย์กลางที่ 920 Hz แล้วมีความเข้มสัญญาณรวมกัน 3 dB ขึ้นไป การคำนวณความกว้าง แถบความถี่วิกฤติโดยวิธีนี้สามารถคำนวณได้จากจำนวนของ โทน และระยะห่างระหว่าง โทนซึ่งอยู่ใน รูปจะได้ค่า 140 Hz ซึ่งมาจาก  $(8-1)*20=140$  Hz เป็นความกว้างแถบความถี่วิกฤติสำหรับความถี่ ศูนย์กลางที่ 920 Hz



รูปที่ 2.19 ทั้งสองรูปเป็นผลต่อเนื่องจากรูปที่ 2.18 โดยการเพิ่มผลที่ได้จากการใช้ Uniform

Masking Noise ด้านซ้ายเป็นระดับ Threshold ของแต่ละ โทนส่วนด้านขวาเป็นของ สัญญาณรวมทุก โทน

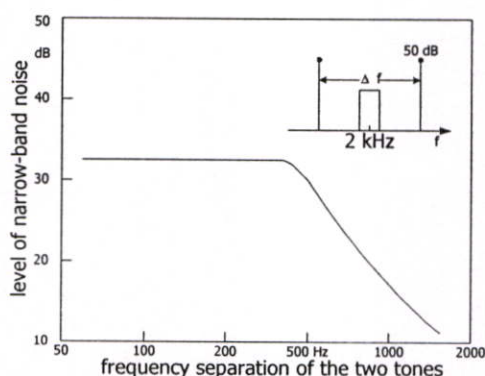


**รูปที่ 2.20** ระดับของสัญญาณ โทนทดสอบที่ถูกบดบังโดย Uniform Masking Noise ที่ระดับสัญญาณต่างๆ โดยกราฟด้านบนเป็นค่าการลดทอนที่ให้กับ White Noise เพื่อสร้าง Uniform Masking Noise

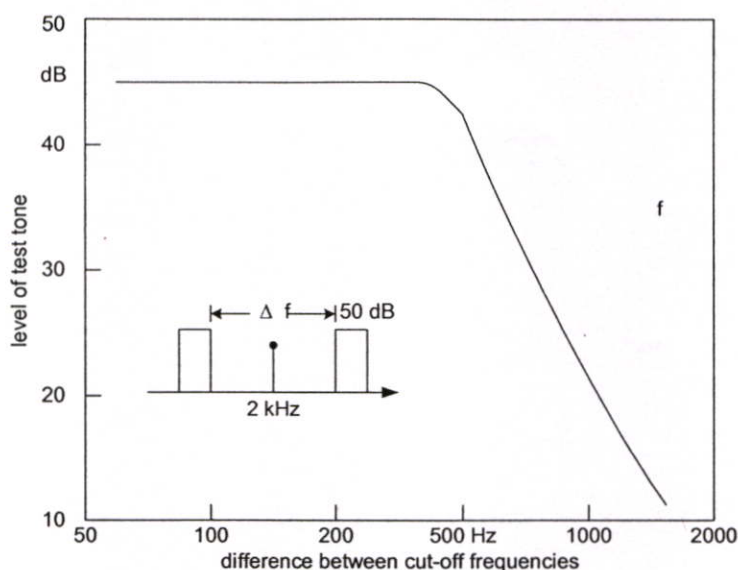
การหาค่าความกว้างแถบความถี่วิกฤตโดยใช้ Threshold In Quiet นี้จะให้คำตอบที่ถูกต้องก็ต่อเมื่ออยู่ในย่านความถี่ซึ่งระดับ Threshold In Quiet ไม่ขึ้นกับความถี่ ซึ่งช่วงความถี่นั้นอยู่ในช่วงที่ไม่เกิน 500 Hz ถึง 2KHz เท่านั้น อย่างไรก็ตามหากดูในรูปที่ 2.20 ในรูปบนจะแสดงการลดทอนซึ่งเป็นการรวมของ White Noise ซึ่งการลดทอนมีค่าเป็นศูนย์ที่ความถี่เสียงต่ำกว่า 500 Hz จะเห็นได้ว่า Uniform Masking Noise จะสามารถบดบังสัญญาณซึ่งจะไม่ขึ้นกับความถี่ของโทน ถ้าหากว่าปรากฏการณ์ที่เกิดขึ้นซึ่งอธิบายในรูปที่ 2.18 นั้นเกิดขึ้นไม่ใช่เพียงแค่ที่ Threshold In Quiet เท่านั้น แต่ยังเกิดขึ้นกับระดับของการบดบังสัญญาณโดย Uniform Masking Noise แล้วหมายความว่ามันเป็นไปได้ที่จะเกิดปรากฏการณ์นี้ตลอดช่วงความถี่ของการได้ยินของมนุษย์ได้ด้วย ผลที่ได้จากการวัดจะใกล้เคียงกับรูปที่ 2.18 สำหรับความถี่ของสัญญาณ โทนที่ใกล้เคียงกับ 920 Hz ส่วนการใช้ Uniform Masking Noise ที่ระดับต่างๆ แสดงในรูปที่ 2.19 จำนวนสัญญาณ โทนซึ่งมีระดับสัญญาณเท่ากันจะเป็นค่าในแกนด้านบน ส่วนความกว้างแถบสัญญาณที่ได้จากการเปลี่ยนแปลงจำนวน โทนแสดงอยู่ในแกนด้านล่าง ผลแสดงให้เห็นชัดว่าวิธีที่กล่าวผ่านมาทั้งสองวิธีนั้น ระดับ Threshold ของแต่ละ โทนจะลดลงในช่วงที่ความแตกต่างของความถี่มีค่าน้อยกว่าค่าๆ หนึ่ง ส่วนที่ความแตกต่างของความถี่มีค่าสูงกว่าค่านั้นระดับสัญญาณของแต่ละ โทนจะคงที่ ในรูปที่ 2.19a นั้นจะแสดงระดับ Threshold ของแต่ละ โทนในขณะที่รูปที่ 2.19b จะแสดงเปรียบเทียบกับระดับแรงดันเสียงรวมทุก โทนแทนที่จะแยกออกสำหรับแต่ละ โทน และจะเห็นได้ว่าระดับแรงดันเสียงรวมนั้นจะคงที่ที่ระดับ Threshold In Quiet หรือ Masking Threshold ในบริเวณที่ความแตกต่างของความถี่

น้อยกว่าความกว้างแถบความถี่วิกฤต ระดับแรงดันของเสียงรวมจะเพิ่มขึ้นเมื่อความแตกต่างของความถี่มากกว่าความกว้างแถบความถี่วิกฤต ซึ่งแสดงให้เห็นว่าองค์ประกอบของเสียงส่วนที่เกินจากแถบความถี่วิกฤตนั้นจะไม่ถูกนำมามีผลในการวัด Threshold In Quiet หรือระดับการบดบังสัญญาณ โดย Uniform Masking Noise จึงสามารถกล่าวได้ว่าสิ่งที่มีผลต่อ Threshold In Quiet และระดับของการบดบังสัญญาณคือ ความเข้มของสัญญาณในแถบความถี่วิกฤตนั้นๆ เท่านั้น

อีกวิธีที่ใช้ในการประมาณความกว้างแถบความถี่วิกฤตคือการใช้สัญญาณ โทนสองความถี่ที่มีระดับสัญญาณเท่ากันทำหน้าที่เป็น Masker และให้สัญญาณรบกวนแถบสัญญาณแคบทำหน้าที่เป็นสัญญาณทดสอบ โดยจะทำการวัด Threshold ของการที่สัญญาณทดสอบถูกบดบังเทียบกับความแตกต่างของความถี่ของสัญญาณ โทนทั้งสองซึ่งมีสัญญาณทดสอบอยู่ที่ความถี่กึ่งกลาง ดังนั้นช่วงความถี่ของสัญญาณทดสอบจะต้องแคบกว่าความกว้างแถบสัญญาณความถี่วิกฤต ในรูปที่ 2.21 ภาพเล็กแสดงสัญญาณ โทนทั้งสองที่ใช้ในการบดบังสัญญาณ และสัญญาณรบกวนแถบสัญญาณแคบที่ใช้เป็นสัญญาณทดสอบในแกนความถี่ และในรูปใหญ่เป็นข้อมูลระดับ Threshold เมื่อถูกบดบังด้วยสัญญาณ โทนที่มีระดับ 50 dB สองค่าและสัญญาณทดสอบมีความถี่ศูนย์กลางที่ 2 kHz โดยระดับ Threshold ของสัญญาณรบกวนแถบความถี่แคบที่ถูกบดบังสัญญาณ โดยสัญญาณ โทนทั้งสองนั้นจะแสดงเปรียบเทียบกับความแตกต่างความถี่ของสัญญาณ โทนทั้งสอง สำหรับที่ความแตกต่างของความถี่น้อยๆ นั้นระดับ Threshold จะไม่ขึ้นกับความแตกต่างของความถี่ แต่เมื่อความแตกต่างของความถี่เกินกว่าระดับหนึ่งไปแล้วระดับ Threshold จะลดลงซึ่งจุดที่ความแตกต่างความถี่ค่านั้นก็คือ ความกว้างแถบความถี่วิกฤต เมื่อทำการวัดโดยเปลี่ยนค่าความถี่ศูนย์กลางของสัญญาณทดสอบ ไปเรื่อยๆ ก็จะทำให้เราได้ความสัมพันธ์ระหว่างความกว้างแถบความถี่วิกฤตกับความถี่



รูปที่ 2.21 ระดับ Threshold ของสัญญาณรบกวนแถบความถี่แคบที่มีความถี่ศูนย์กลางอยู่ระหว่างสอง โทนที่ใช้ในการบดบังสัญญาณ โดยแสดงเปรียบเทียบกับความแตกต่างของความถี่ของสอง โทน



รูปที่ 2.22 ระดับ Threshold ที่สัญญาณ โทนทดสอบถูกบังคับ โดยสัญญาณรบกวนสองแถบ (Bandpass Noise) โดยเปรียบเทียบกับความแตกต่างความถี่ของ Cutoff ของทั้งสองแถบ

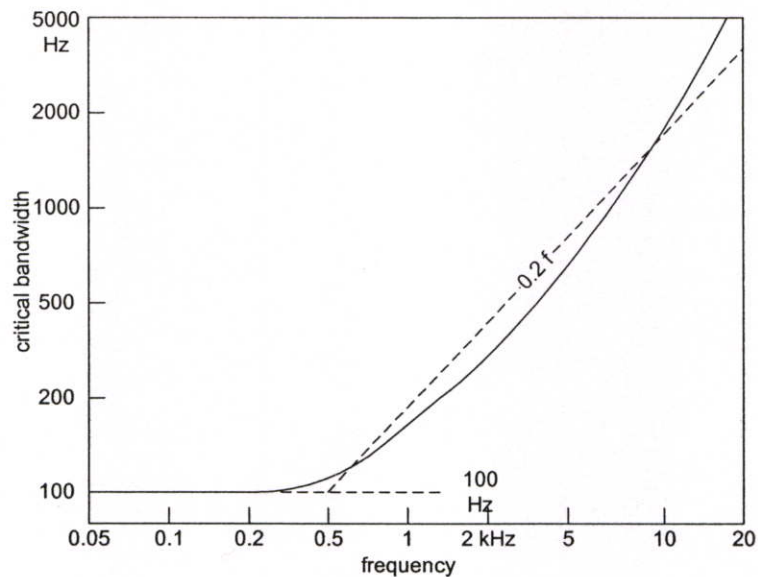
อีกวิธีการหนึ่งที่คล้ายกันจะแสดงในรูปที่ 2.22 เราจะใช้สัญญาณรบกวนแถบสัญญาณแคบสองช่วงความถี่มาเป็น Masker แล้วใช้สัญญาณ โทนเป็นสัญญาณทดสอบ โดยให้ความแตกต่างระหว่าง Higher และ Lower cutoff ของสัญญาณรบกวนแถบความถี่แคบทั้งสองเป็นความแตกต่างความถี่และให้มันเปลี่ยนค่าไปเรื่อยๆ ขณะที่ทำการวัดระดับ Threshold โดยสัญญาณทดสอบเป็นสัญญาณ โทนความถี่ 2 kHz และระดับของสัญญาณรบกวนทั้งสองแถบที่ใช้เป็น Masker มีระดับ 50 dB และสัญญาณรบกวนแถบความถี่แคบที่มีแถบความถี่กว้าง 200 Hz ซึ่งจะเห็นได้ว่าได้ผลเช่นเดียวกันกับที่ได้จากวิธีการก่อนหน้านี้

นอกจากนี้ยังมีวิธีการอื่นๆอีกที่ใช้ในการหาความกว้างแถบความถี่วิกฤติ เช่นการใช้การตรวจสอบการเปลี่ยนเฟส การวัดระดับความดังของเสียง และการวัดโดยใช้ Short Impulse เป็นต้น ซึ่งวิธีการทั้งหมดนี้จะให้ผลที่ใกล้เคียงกัน

### 2.9.2 Critical Band Rate Scale

มีวิธีการหลายอย่างที่ใช้ในการประมาณค่าความกว้างแถบความถี่วิกฤติซึ่งได้กล่าวผ่านมาในหัวข้อที่แล้ว ซึ่งแถบความถี่วิกฤติแรกที่อยู่ในช่วงความถี่ต่ำที่สุดของระบบการได้ยินของมนุษย์จะอยู่ที่ประมาณ 20-100 Hz แต่จะเป็นการง่ายกว่าหากเรารวมเอาช่วงความถี่ 0-20 Hz ที่เราไม่สามารถได้ยินได้รวมเข้าไว้ด้วย และกำหนดให้แถบความถี่วิกฤติแรกของการได้ยินของมนุษย์เป็น 0-100 Hz ในรูปที่ 2.23 เป็นข้อมูลที่ได้จากการหาค่าเฉลี่ยจากการใช้วิธีการต่างๆกับกลุ่มตัวอย่างจำนวนมาก แสดงให้เห็นว่าที่ความถี่ต่ำกว่า 500 Hz นั้นแถบความถี่วิกฤติจะมีความกว้างประมาณ 100 Hz ส่วนที่ความถี่สูงกว่านั้นแถบความถี่วิกฤติจะกว้างประมาณ 20 เปอร์เซ็นต์ของ

ความถี่ศูนย์กลาง ส่วนค่าที่ได้จากผลการทดลองจริงก็นั้นอยู่ในตารางที่ 2.9 ที่แสดงทั้งของเขตนบน และล่างของแถบความถี่วิกฤติ

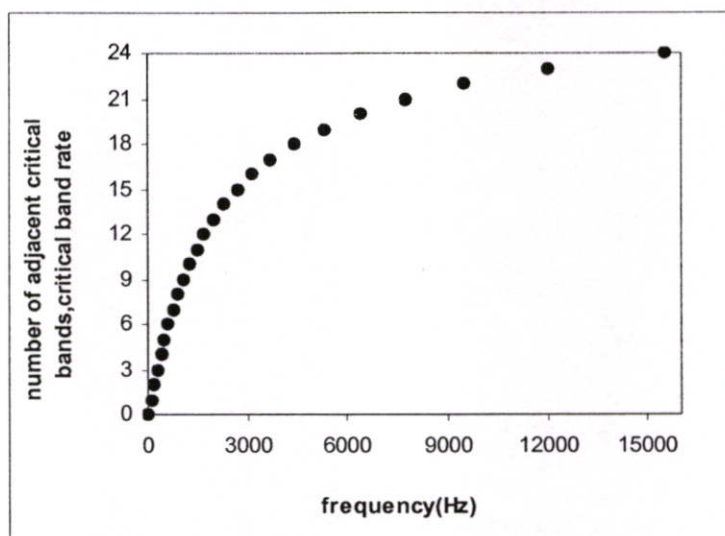


รูปที่ 2.23 ความกว้างแถบความถี่วิกฤติเทียบกับความถี่ และเส้นประแสดงการประมาณ

ความหมายของแถบความถี่วิกฤตินั้น มีความสำคัญต่อการอธิบายลักษณะในการได้ยินของมนุษย์อย่างมาก โดยจะถูกใช้ในแบบจำลองต่างๆ ในรูปแบบของสเกลที่เรียกว่า Critical Band Rate Scale ซึ่งสเกลนี้อ้างอิงจากหลักความจริงที่ว่า ระบบการรับฟังเสียงของคนเรานั้นจะทำการแบ่งการวิเคราะห์เสียงออกเป็นส่วนต่างๆตามแถบความถี่วิกฤติ และถ้าเราขีดเอาแถบความถี่วิกฤติแรกเป็นจุดอ้างอิงแล้วทำการคำนวณแถบความถี่วิกฤติต่อมาเรื่อยๆ โดยให้ขอบเขตนบนของแถบความถี่วิกฤติปัจจุบันเป็นขอบเขตนล่างของแถบความถี่วิกฤติถัดไปก็จะทำให้ได้สเกลที่เรียกว่า Critical Band Rate Scale ซึ่งจากตารางที่ 2.9 จะเห็นได้ว่าช่วงความถี่ถึง 16 kHz นั้นจะถูกแบ่งออกเป็นแถบความถี่วิกฤติ 24 แถบ ส่วนในรูปที่ 2.24 เป็นแผนภูมิที่นำข้อมูลในตารางที่ 2.9 มาแสดงให้เห็น ซึ่งใน Critical Band Rate Scale นั้นอันดับของ Critical Band ถูกเรียกเป็น “Bark”

ตารางที่ 2.9 ขอบเขตและความถี่ศูนย์กลาง Critical Band Rate scale [13]

| $Z$<br>Bark | $f_{Lower} / f_{Upper}$<br>Hz | $f_c$<br>Hz | $f_c$<br>Bark | Bandwidth<br>Hz | $Z$<br>Bark | $f_{Lower} / f_{Upper}$<br>Hz | $f_c$<br>Hz | $f_c$<br>Bark | Bandwidth<br>Hz |
|-------------|-------------------------------|-------------|---------------|-----------------|-------------|-------------------------------|-------------|---------------|-----------------|
|             | 0                             |             |               |                 |             | 1720                          |             |               |                 |
| 0           |                               | 50          | 0.5           | 100             | 12          |                               | 1850        | 12.5          | 280             |
|             | 100                           |             |               |                 |             | 2000                          |             |               |                 |
| 1           |                               | 150         | 1.5           | 100             | 13          |                               | 2150        | 13.5          | 320             |
|             | 200                           |             |               |                 |             | 2320                          |             |               |                 |
| 2           |                               | 250         | 2.5           | 100             | 14          |                               | 2500        | 14.5          | 380             |
|             | 300                           |             |               |                 |             | 2700                          |             |               |                 |
| 3           |                               | 350         | 3.5           | 100             | 15          |                               | 2900        | 15.5          | 450             |
|             | 400                           |             |               |                 |             | 3150                          |             |               |                 |
| 4           |                               | 450         | 4.5           | 110             | 16          |                               | 3400        | 16.5          | 550             |
|             | 510                           |             |               |                 |             | 3700                          |             |               |                 |
| 5           |                               | 570         | 5.5           | 120             | 17          |                               | 4000        | 17.5          | 700             |
|             | 630                           |             |               |                 |             | 4400                          |             |               |                 |
| 6           |                               | 700         | 6.5           | 140             | 18          |                               | 4800        | 18.5          | 900             |
|             | 770                           |             |               |                 |             | 5300                          |             |               |                 |
| 7           |                               | 840         | 7.5           | 150             | 19          |                               | 5800        | 19.5          | 1100            |
|             | 920                           |             |               |                 |             | 6400                          |             |               |                 |
| 8           |                               | 1000        | 8.5           | 160             | 20          |                               | 7000        | 20.5          | 1300            |
|             | 1080                          |             |               |                 |             | 7700                          |             |               |                 |
| 9           |                               | 1170        | 9.5           | 190             | 21          |                               | 8500        | 21.5          | 1800            |
|             | 1270                          |             |               |                 |             | 9500                          |             |               |                 |
| 10          |                               | 1370        | 10.5          | 210             | 22          |                               | 10500       | 22.5          | 2500            |
|             | 1480                          |             |               |                 |             | 12000                         |             |               |                 |
| 11          |                               | 1600        | 11.5          | 240             | 23          |                               | 13500       | 23.5          | 3500            |
|             | 1720                          |             |               |                 |             | 15500                         |             |               |                 |



รูปที่ 2.24 ความสัมพันธ์ระหว่างลำดับของ Critical Band กับความถี่

ซึ่งความสัมพันธ์ระหว่าง Critical Band Rate scale และ สเกลความถี่มีการประมาณการแปลงระหว่างความถี่ และ Critical Band Rate ตลอดช่วงความถี่เสียงที่เราสามารถได้ยินดังต่อไปนี้ [13]

$$z / \text{Bark} = 13 \arctan(0.76 f / \text{kHz}) + 3.5 \arctan(f / 7.5 \text{kHz})^2 \quad (2.26)$$

และ

$$\Delta f_G / \text{Hz} = 25 + 75 [1 + 1.4 (f / \text{kHz})^2]^{0.69} \quad (2.27)$$

### 2.9.3 Critical Band Intensity

ความสามารถในการรับรู้เสียงในย่านความถี่ต่างๆของระบบการได้ยินเสียงของมนุษย์นั้นสามารถอธิบายอย่างคร่าวๆได้โดยความเข้มของสัญญาณเสียงที่ถูกคำนวณโดยแบ่งเป็นย่านความถี่ตามแถบความถี่วิกฤติซึ่งเรียกค่านี้ว่าเป็น Critical Band Intensity โดยค่า Critical Band Intensity สามารถคำนวณได้โดยใช้สมการดังนี้

$$I_G(f) = \int_{f-0.5\Delta f_G(f)}^{f+0.5\Delta f_G(f)} \frac{dI}{df} df \quad (2.28)$$

และสามารถเขียนได้เป็น

$$I_G(z) = \int_{z-0.5\text{Bark}}^{z+0.5\text{Bark}} \frac{dI}{dz} dz \quad (2.29)$$

สำหรับการแสดงค่าในสเกล Logarithm นั้นเราใช้  $I_0$  เป็นค่าอ้างอิงดังนั้น Critical Band Level ( $L_G$ ) สามารถหาได้จาก

$$L_G = 10 \cdot \log \frac{I_G}{I_0} \text{dB} \quad (2.30)$$

จากสมการเห็นได้ว่า Critical Band Intensity นั้นเป็นค่าความเข้มส่วนหนึ่งของเสียง ในช่วงความถี่ของแถบความถี่วิกฤติและการแปลงจากสเกลความถี่ให้เป็น Critical Band Rate Scale นั้นคือการเปลี่ยนจากขนาด Window ที่ขึ้นกับความถี่เป็นขนาด Window ที่มีขนาดเดียวคือ 1 Bark ไม่ขึ้นกับช่วงความถี่ใดๆ

## 2.10 การหา K-Nearest Neighbor

### 2.10.1 ความหมายของ Nearest Neighbor

กำหนดให้มีสเปซแบบหลายมิติ (Multi-dimensional Space) ให้เวกเตอร์ข้อมูลสมาชิก (Exemplar) เป็นสมาชิกในเซตของเวกเตอร์ข้อมูลสมาชิก  $E$  และให้เซตของเวกเตอร์ข้อมูลสมาชิก ประกอบไปด้วยเวกเตอร์ข้อมูลสมาชิกจำนวนจำกัดจำนวนหนึ่ง

$E$  เป็นเซตของเวกเตอร์ข้อมูลสมาชิกและ  $d$  เป็นเวกเตอร์เป้าหมายที่สนใจและต้องการหา Nearest Neighbor ถ้าให้ Nearest Neighbor ของ  $d$  คือ  $d'$  ซึ่ง  $d' \in E$  หมายความว่าไม่มีจุดใดๆใน  $E$  ที่ใกล้กับ  $d$  มากกว่า  $d'$  หรือ  $\text{None-nearer}(E, d, d')$  ก็ต่อเมื่อ

$$\text{None-nearer}(E, d, d') \Leftrightarrow \forall d'' \in E \quad |d - d'| \leq |d - d''| \quad (2.31)$$

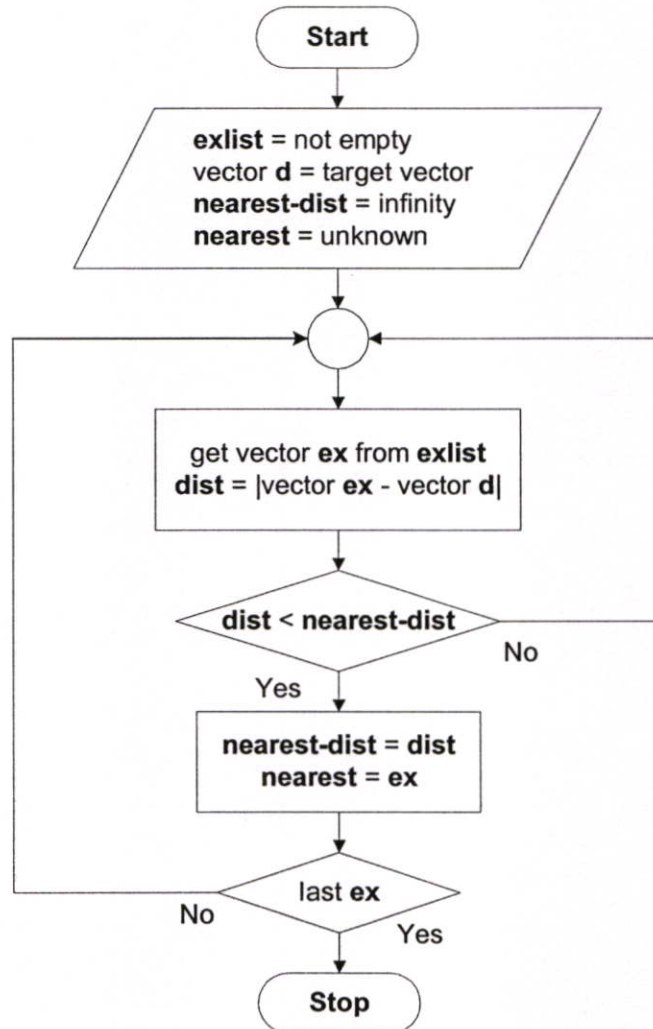
ในสมการที่ 3.31 ระยะห่างที่ใช้คือระยะทางแบบ Euclidean ซึ่งสามารถหาได้โดยสมการที่ 3.32 [15] [16]

$$|d - d'| = \sqrt{\sum_{i=1}^{i=k_d} (d_i - d'_i)^2} \quad (2.32)$$

โดยที่  $d_i$  เป็นองค์ประกอบในมิติหนึ่งของเวกเตอร์  $d$

### 2.10.2 การหา Nearest Neighbor โดยตรง

วิธีการนี้ทำโดยการแจกแจงเซตของเวกเตอร์ข้อมูลสมาชิกออกเป็นลำดับรายการของเวกเตอร์ข้อมูลสมาชิกทั้งหมด ในรูปที่ 2.25 เป็นโฟลวชาร์ท (Flow Chart) แสดงขั้นตอนในการหา Nearest Neighbor โดยตรง โดยการตรวจดูเวกเตอร์ข้อมูลสมาชิกในรายการทั้งหมดซึ่งจะทำให้วิธีการนี้มีการคำนวณเป็น  $O(N)$  โดยที่  $N$  เป็นขนาดของ  $E$  (เซตของเวกเตอร์ข้อมูลสมาชิก) ใช้ในการคำนวณหาระยะทางจากทุกเวกเตอร์ข้อมูลสมาชิกต่อไป [8] [15]

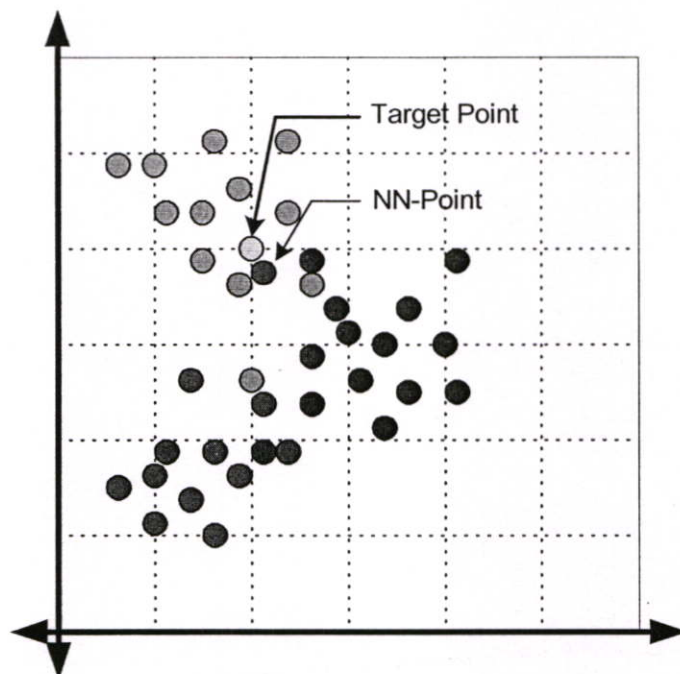


- exlist** = เซตของเวกเตอร์ข้อมูลสมาชิก  
**vector d** = เวกเตอร์ที่ต้องการหา Nearest Neighbor  
**vector ex** = เวกเตอร์ข้อมูลสมาชิก  
**dist** = ระยะห่างระหว่างเวกเตอร์ ex และ เวกเตอร์ d  
**nearest** = เวกเตอร์ที่เป็น Nearest Neighbor  
**nearest-dist** = ระยะห่างระหว่างเวกเตอร์ d กับเวกเตอร์ที่เป็น Nearest Neighbor

รูปที่ 2.25 โฟลวชาร์ทการคำนวณหา Nearest Neighbor โดยตรง

### 2.10.3 K-Nearest Neighbor

การประยุกต์ที่นิยมมากอย่างหนึ่งของการทำ Nearest Neighbor คือการทำ K-Nearest Neighbor เราสามารถใช้วิธีนี้ในการจัดกลุ่มให้กับเวกเตอร์ใดๆเข้ากับกลุ่มที่ปรากฏมากที่สุดของเวกเตอร์ที่ใกล้กับเวกเตอร์เป้าหมายนั้นมากที่สุดจำนวน  $k$  ตัว หรืออีกนัยหนึ่งก็คือการตัดสินใจในการจัดกลุ่มให้กับเวกเตอร์ใดๆ จะทำโดยการโหวตของเวกเตอร์ที่ใกล้ที่สุดจำนวน  $k$  ตัว



รูปที่ 2.26 การจัดกลุ่มโดยอาศัยการตัดสินใจของ K-Nearest Neighbor

จากรูปที่ 2.26 จะเห็นได้ว่าหากอาศัยการตัดสินใจของ Nearest Neighbor เพียงค่าเดียวแล้วอาจได้ผลที่ผิดพลาดได้ เนื่องจาก Nearest Neighbor นั้นอาจเป็นค่าที่เกิดจากการเบี่ยงเบนอย่างมากก็เป็นได้ แต่ถ้าหากใช้การตัดสินใจของ Nearest Neighbor ในลำดับถัดมาด้วยก็จะช่วยให้มีความถูกต้องในการตัดสินใจมากขึ้น

## 2.11 ทฤษฎีการประมาณค่าพิทช์โดยใช้ออโตคอร์รีเลชันฟังก์ชัน

ออโตคอร์รีเลชันฟังก์ชัน [10] เป็นวิธีหนึ่งที่ได้ถูกนำมาใช้ตรวจหาคาบพิทช์ของสัญญาณเสียงพูด โดยออโตคอร์รีเลชันจะทำหน้าที่ในการแสดงยอดกราฟหลัก (Prominent Peak) ที่เป็นคาบในแต่ละส่วนของเสียง (Section) ซึ่งจะทำให้เราสามารถหาระยะความเป็นคาบของสัญญาณตามแต่ละช่วงของเส้นเสียงได้

### 2.11.1 การจัดแบ่งการวิเคราะห์สัญญาณออกเป็นช่วงสั้นๆ

กำหนดให้สัญญาณเสียงแทนด้วย  $s(m)$  ซึ่งออโตคอร์รีเลชันฟังก์ชันของ Discrete-time Deterministic Signal โดยทั่วไปเขียนได้เป็น

$$\phi(\tau) = \sum_{m=-\infty}^{\infty} s(m) s(m+\tau) \quad (2.33)$$

ถ้าสัญญาณ  $s(m)$  มีความเป็นคาบที่แน่นอนด้วยระยะ  $P$  นั่นคือ

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

ดังนั้นสามารถเขียนได้ว่า

$$\phi(\tau) = \phi(\tau+p) \quad (2.34)$$

นั่นคือ ออโตคอร์รีเลชันฟังก์ชันก็มีความเป็นคาบด้วยระยะคาบเดียวกัน หรือในทางกลับกันก็คือ ความเป็นคาบในออโตคอร์รีเลชันฟังก์ชัน เป็นตัวบ่งชี้ให้เห็นถึงความเป็นคาบในสัญญาณ โดยคุณสมบัติของออโตคอร์รีเลชันฟังก์ชันที่สำคัญคือ

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

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

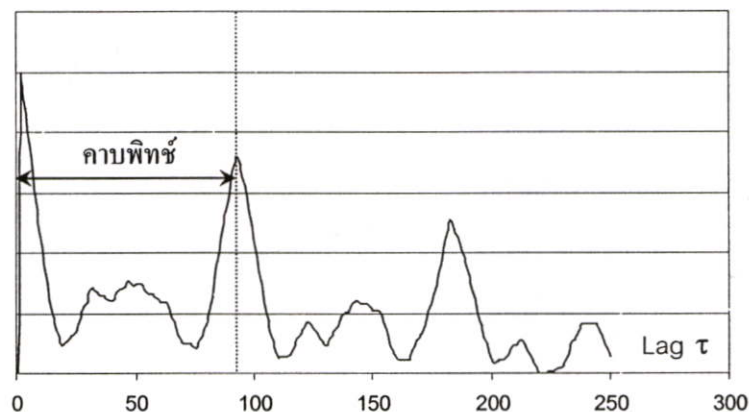
ถ้าพิจารณาสมการที่ 4.2 ควบคู่ไปกับคุณสมบัติในข้อ 1 และ 2 จะพบความเป็นคาบของสัญญาณ โดยแฉกเปิดของออโตคอร์รีเลชันจะมีค่ามากที่สุดที่  $0, \pm p, \pm 2p, \dots$  โดยไม่ต้องคำนึงถึงเวลาเริ่มต้น (Time Origin) ของสัญญาณ การคำนวณหาคาบของสัญญาณสามารถประมาณได้จากตำแหน่งแรกที่มีค่ามากที่สุด ในออโตคอร์รีเลชันฟังก์ชัน และตำแหน่งถัดไปจากยอดกราฟ ซึ่งจากคุณสมบัติเหล่านี้ทำให้ออโตคอร์รีเลชันฟังก์ชันเป็นหลักการพื้นฐาน ที่ใช้ประมาณค่าความเป็นคาบในสัญญาณได้

สำหรับสัญญาณที่มีลักษณะเปลี่ยนแปลงอยู่ตลอดเวลาเช่น สัญญาณเสียงพูดจะต้องทำการแบ่งสัญญาณออกเป็นช่วงสั้นๆเพื่อหาข้อมูล (Information) ที่ต้องการ โดย Short-time Auto-Correlation Function ซึ่งสามารถนิยามได้เป็น

$$R(\tau) = \sum_{m=0}^{N-1-\tau} s(m)s(m+\tau) \quad (2.35)$$

เมื่อ  $N$  คือ จำนวนตัวอย่างสัญญาณ (Sample) ต่อเฟรม  
 $\tau$  คือ จุดเลื่อนไปในการคำนวณออโตโครีเรลชัน

โดยในการเลือกค่าของจำนวนตัวอย่าง ( $N$ ) ที่ใช้ในแต่ละเฟรม รูปคลื่นสัญญาณจะต้องมีความเป็นคาบที่สมบูรณ์ (Complete Period) อย่างน้อย 2-3 คาบ ซึ่งในความเป็นจริงแล้วความยาวของสัญญาณเสียงพูดมีผลต่อการคำนวณของ  $R_n(\tau)$  เนื่องจากค่าของ  $R_n(\tau)$  จะลดลงเรื่อยๆเมื่อ  $\tau$  มีค่าเพิ่มมากขึ้น



รูปที่ 2.27 แสดงการหาค่าพิทช์จากการคำนวณออโตโครีเรลชันฟังก์ชันของสัญญาณเสียงพูด

พิจารณารูป 2.27 แสดงตัวอย่างการคำนวณออโตโครีเรลชันฟังก์ชันของสัญญาณเสียงพูดที่มีอัตราการแซมปลิงด้วยความถี่ 11.025 kHz โดยใช้สมการ 4.3 ด้วย  $N = 300$  ในหนึ่งเฟรมและการเลื่อนของเวลา (Lag) เป็น  $0 \leq \tau \leq 250$  จากรูปจะเห็นว่าจุดสูงสุด (Peak) ยกเว้นที่ตำแหน่ง  $\tau = 0$  เกิดที่ตำแหน่ง  $\tau = 83$  นั่นคือสัญญาณมีคาบที่ระยะ 83 นาที่หรือมีค่าความถี่มูลฐานประมาณ 133 Hz ( $11.025 \text{ kHz} / 83$ )

## 2.12 การสร้างแบบจำลองการรู้จำด้วยวิธี Hidden Markov Model

แบบจำลองมาร์คอฟเป็นแบบจำลองทางสถิติ ซึ่งพัฒนามาเพื่อแบ่งกลุ่มของอนุกรมทางเวลาหรือสัญญาณที่ไม่คงที่ นั่นคือใช้สำหรับจัดกลุ่มของสัญญาณที่ไม่รู้จัก (Unknown Signal) ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่งแบบจำลองมาร์คอฟได้ถูกนำมาประยุกต์ใช้ในการรู้จำเสียงพูด [9] [16] และเป็นวิธีการที่วิทยานิพนธ์นี้เลือกใช้

### 2.12.1 ส่วนประกอบของแบบจำลองมาร์คอฟ

พารามิเตอร์สำคัญที่เกี่ยวข้องในการสร้างแบบจำลองอ้างอิง ที่ต้องรู้จักได้แก่

2.12.1.1  $T$  คือ ความยาวของลำดับข้อมูลที่ได้จากการควอนไทซ์ค่าความถี่มูลฐาน ซึ่งจะใช้เป็นข้อมูลอินพุตในส่วนของ HMM โดยต่อไปจะเรียกแทนว่า “ลำดับของค่าปรากฏ”(Observation Sequence) ซึ่งมีขนาดความยาวของลำดับ เท่ากับจำนวนเฟรมทั้งหมดในเสียงแต่ละเสียง

2.12.1.2  $N$  คือจำนวนสเตตในแบบจำลอง ถ้ากำหนดให้เซตของสเตตเป็น  $\{1, 2, \dots, N\}$  จะสามารถแทนสเตตที่เปลี่ยนไปตามเวลา  $t$  ด้วยเซตของ  $Q = \{q_1, q_2, q_3, \dots, q_N\}$

2.12.1.3  $M$  คือจำนวนของค่าปรากฏที่สามารถเป็นไปได้ต่อหนึ่งสเตต แทนสัญลักษณ์ด้วย  $V = \{v_1, v_2, v_3, \dots, v_M\}$  ซึ่งจากการจัดระดับของการเปลี่ยนแปลงของความถี่ ( $\Delta F$ ) ออกเป็น 3 ระดับ จะได้เซตของค่าปรากฏที่สามารถเป็นไปได้ในแต่ละสเตตมีค่าเป็น  $V = \{-1, 0, 1\}$

2.12.1.4 ค่าความน่าจะเป็นในการย้ายสเตต  $A = \{a_{ij}\}$  โดย  $a_{ij}$  แทนการย้ายสเตตจาก  $i$  ไป  $j$

$$\text{เมื่อ } a_{ij} = P[q_i = j | q_{i-1} = i] \quad ; 1 \leq i, j \leq N \quad (2.36)$$

2.12.1.5 การกระจายความน่าจะเป็น ของค่าปรากฏที่สามารถเป็นไปได้ภายในสเตต  $B$  ซึ่ง  $B = \{b_j(\tau)\}$

$$\text{โดยที่ } b_j(\tau) = P[v_\tau \text{ ที่เวลา } t | q_j \text{ ที่เวลา } t]; 1 \leq \tau \leq M \quad (2.37)$$

ตามการกระจายสัญลักษณ์ในสเตต  $j$  เมื่อ  $j = 1, 2, \dots, N$

2.12.1.6 ค่าความน่าจะเป็นของการเป็นสเตตเริ่มต้น

$$\pi_i = P[q_i \text{ ที่เวลา } t = 1] \quad ; 1 \leq i \leq N \quad (2.38)$$

จะเห็นว่า Hidden Markov Model ต้องการพารามิเตอร์ของแบบจำลองคือ  $N, M$  และ กลุ่มของความน่าจะเป็น  $A, B, \pi$  ดังนั้นในการแสดงเซตของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองอ้างอิงจะแทนด้วยสัญลักษณ์

$$\lambda = (A, B, \pi) \quad (2.39)$$

### 2.12.2 คุณสมบัติการย้ายสแตทของ HMM

การสร้างแบบจำลองนี้ใช้ Left-Right Model ซึ่งเป็นแบบจำลองแบบที่เหมาะสมกับสัญญาณที่มีลักษณะเปลี่ยนแปลงตามเวลาอย่างต่อเนื่องเช่น เสียงพูด โดยมีคุณสมบัติของสัมประสิทธิ์ในการย้ายสแตทดังนี้

การย้ายสแตทจะย้ายจากซ้ายไปขวาหมายความว่า  $a_{ij} = 0, j < i$  แสดงว่าจะไม่มีการย้ายสแตทไปยังสแตทที่ต่ำกว่าสแตทปัจจุบันและนอกจากนี้ก็ยังมีความน่าจะเป็นของสแตทเริ่มต้น

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

ลำดับของสแตทจะต้องเริ่มที่สแตทที่ 1 เสมอ และ Left-Right Model นี้มักมีกฎบังคับการย้ายสแตทเพื่อไม่ให้มีการเปลี่ยนแปลงดัชนีของสแตทมากนักกล่าวคือ

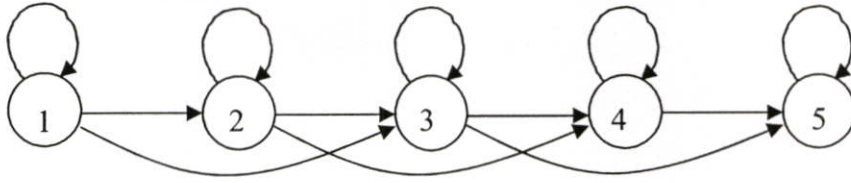
$$a_{ij} = 0, j > i + \Delta i$$

ดังรูปที่ 4.12 ค่าของ  $\Delta i = 2$  คือจะไม่มีการย้ายข้ามสแตทไปเกิน 2 สแตท และมีเมตริกซ์ในการย้ายสแตทเป็น

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

จะเห็นว่าสแตทสุดท้าย สัมประสิทธิ์การย้ายสแตทจะเป็น

$$\begin{aligned} a_{NN} &= 1 \\ a_{Ni} &= 0, i < N \end{aligned}$$



รูปที่ 2.28 แบบจำลอง Left-Right Model ของ HMM

### 2.12.3 ปัญหาพื้นฐานของแบบจำลอง มาร์คอฟ

ปัญหาของ HMM มี 3 ข้อ ซึ่งต้องใช้อัลกอริทึมวิธีต่างๆ ในการคำนวณเพื่อแก้ปัญหานี้

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

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

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

### 2.12.4 การคำนวณเพื่อแก้ปัญหของ HMM

#### 2.12.4.1 การแก้ปัญหที่ 1

เป็นการคำนวณหาว่าแบบจำลอง  $\lambda$  ใดๆ มีโอกาสจะให้ค่าลำดับเป็นไปตามลำดับของค่าปรากฏนั้นด้วยค่าของความน่าจะเป็นมากหรือน้อยเท่าใด การแก้ปัญหสามารถทำได้โดยระบุสแตตให้กับลำดับของค่าปรากฏซึ่งยาว  $T$  (โดยที่ค่าปรากฏหนึ่งตัวมีความเป็นไปได้ที่จะอยู่ในสแตตได้  $N$  สแตต) ซึ่งสามารถเป็นไปได้ถึง  $N^T$  แบบ โดยให้สแตตต่างๆ แทนด้วย

$$q = q_1, q_2, q_3, \dots, q_T \quad (2.40)$$

เมื่อ  $q_t$  เป็นสแตตเริ่มต้นที่เวลา  $t = 1$  ความน่าจะเป็นของลำดับของค่าปรากฏ  $O$  ที่กำหนดคือ

$$P(O|q, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda) \quad (2.41)$$

ความน่าจะเป็นในการเกิดค่าปรากฏคือ

$$P(O/q, \lambda) = b_{q_1}(O_1)b_{q_2}(O_2)...b_{q_t}(O_t) \quad (2.42)$$

และความน่าจะเป็นในการย้ายข้ามสแตต q จะเป็น

$$P(q/\lambda) = \pi_{q_1}a_{q_1q_2}a_{q_2q_3}a_{q_3q_4}\dots a_{q_{T-1}q_T} \quad (2.43)$$

ดังนั้นเมื่อนำความน่าจะเป็นของการเกิดค่าปรากฏ O และความน่าจะเป็นในการย้ายสแตต q มารวมกัน ซึ่งนั่นก็คือความน่าจะเป็นที่ O และ q จะเกิดขึ้นพร้อมกันจะได้

$$\begin{aligned} P(O, q/\lambda) &= P(O/q, \lambda)P(q/\lambda) \\ &= (b_{q_1}(O_1)b_{q_2}(O_2)...b_{q_t}(O_t))(\pi_{q_1}a_{q_1q_2}a_{q_2q_3}\dots a_{q_{T-1}q_T}) \end{aligned} \quad (2.44)$$

โดยความน่าจะเป็นของ O ได้มาจากผลรวมของความน่าจะเป็น O และ q เกิดขึ้นพร้อมกัน โดยคิดจากทุกสแตต q ที่จะเป็นไปได้ดังนี้

$$\begin{aligned} P(O/\lambda) &= \sum_{allq} P(O/q, \lambda)P(q/\lambda) \\ &= \sum_{q_1q_2\dots q_T} \pi_{q_1}b_{q_1}(O_1)a_{q_1q_2}b_{q_2}(O_2)\dots a_{q_{T-1}q_T}b_{q_T}O_T \end{aligned} \quad (2.45)$$

ที่เวลาเริ่มต้น ( $t = 1$ ) เราจะอยู่ที่สแตต  $q_1$  ด้วยค่าความน่าจะเป็น  $\pi_{q_1}$  และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏ  $O_1$  ที่สแตตนี้ด้วย  $b_{q_1}(O_1)$  ที่เวลาเพิ่มขึ้นจาก  $t \rightarrow t+1$  ( $t=2$ ) เราแทน การย้ายสแตตจากสแตต  $q_1$  ไปยัง  $q_2$  ด้วยค่าความน่าจะเป็น  $a_{q_1q_2}$  และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏเป็น  $O_2$  ด้วยค่าความน่าจะเป็น  $b_{q_2}(O_2)$  จนกระทั่ง ที่เวลา T เราแทนการย้ายสแตตจากสแตต  $q_{T-1}$  ไปยัง  $q_T$  ด้วยค่าความน่าจะเป็น  $a_{q_{T-1}q_T}$  และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏเป็น  $O_T$  ด้วยค่าความน่าจะเป็น  $b_{q_T}(O_T)$

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

#### 2.12.4.1.1 กระบวนการไปข้างหน้า (Forward Procedure)

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

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = i | \lambda) \quad (2.46)$$

คือความน่าจะเป็นของการเกิดลำดับของค่าปรากฏ  $O_1 O_2 \dots O_T$  และอยู่ที่สแตต  $q_i$  ณ เวลา  $t$  โดยมีแบบจำลองเป็น  $\lambda$  เราสามารถหา  $\alpha_t(i)$  ได้ดังนี้

2.12.4.1.1.1 การเริ่มต้น (Initialization)

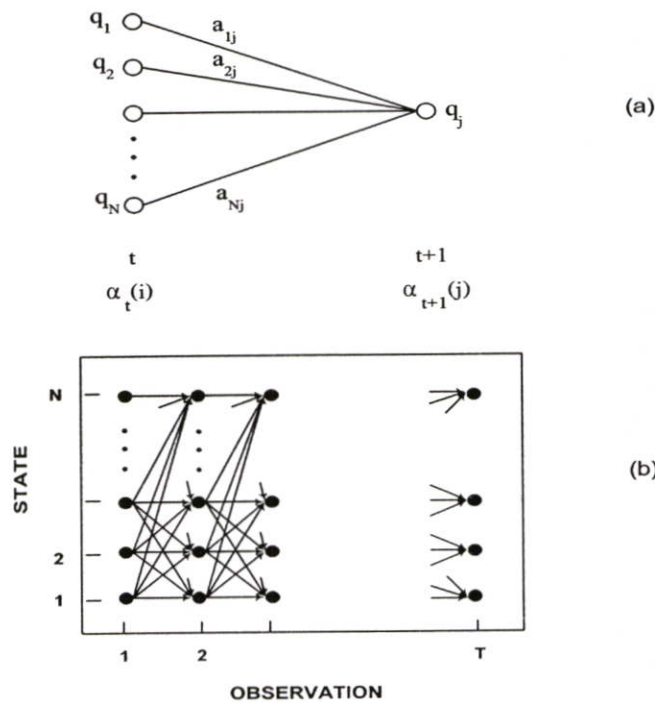
$$\alpha_1(i) = \pi_i b_i(O_1) \quad ; \quad 1 \leq i \leq N \tag{2.47}$$

เริ่มด้วยการกำหนดความน่าจะเป็นไปข้างหน้าซึ่งเป็นความน่าจะเป็นร่วมของสแตต  $i$  และมีเหตุการณ์เริ่มต้นเป็น  $O_1$

2.12.4.1.1.2 การเหนี่ยวนำ (Induction)

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad ; \quad \begin{matrix} 1 \leq i \leq T-1 \\ 1 \leq j \leq N \end{matrix} \tag{2.48}$$

หมายความว่าสแตต  $j$  ที่เวลา  $t+1$  สามารถมาได้จากสแตตก่อนหน้านั้นซึ่งเป็นไปได้ถึง  $N$  สแตต (สแตต  $i$  ณ เวลา  $t$  โดยที่) ดังรูป 2.29 (a)



รูปที่ 2.29 (a) และ (b) กระบวนการไปข้างหน้า

จากรูป 2.29 (b) แสดงให้เห็นว่าการคำนวณค่าความน่าจะเป็นแบบไปข้างหน้า (Forward Probability) มีโครงสร้างการคำนวณคล้ายๆกับลักษณะของโครงผลึก และเนื่องจากมีจำนวนสแตทเพียง  $N$  สแตท (แทนด้วยจำนวนโหนดในแต่ละช่วงเวลา  $t$  ใดๆในโครงผลึก) จำนวนลำดับสแตทจะถูกจัดเรียงลงในโหนดเหล่านี้ โดยในเวลา  $t = 1$  จะทำการคำนวณค่าของ  $\alpha_t(i)$  ในทุกๆสแตท  $1 \leq i \leq N$  และที่เวลา  $t = 2, 3, \dots, T$  จะทำการคำนวณค่าของ  $\alpha_t(j)$  ในทุกๆสแตท,  $1 \leq j \leq N$  โดยในแต่ละค่าจะทำการคำนวณมาจาก  $\alpha_{t-1}(i)$  จำนวน  $N$  ค่าก่อนหน้านี้

2.12.4.1.1.3 การสิ้นสุด (Termination) คือการหา  $P(O|\lambda)$  ได้จากผลรวมของ  $\alpha_t(i)$  จากทุกๆสแตท

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad ; \quad 1 \leq i \leq N \quad (2.49)$$

2.12.1.1.2 กระบวนการย้อนกลับ (Backward Procedure)

นิยาม  $\beta_t(i)$  = Backward variable

$$\beta_t(i) = P(O_{t+1}O_{t+2}O_{t+3}\dots O_T | i_t = q_i, \lambda) \quad (4.50)$$

คือ ความน่าจะเป็นของลำดับค่าปรากฏส่วนหลังจากเวลา  $t+1$  ไปจนจบโดยกำหนดว่าต้องอยู่ที่สแตท  $i$  ที่เวลา  $t$  และมีแบบจำลองเป็น เราจะคำนวณหาค่าได้ดังนี้

2.12.1.1.2.1 การเริ่มต้น (Initialization)

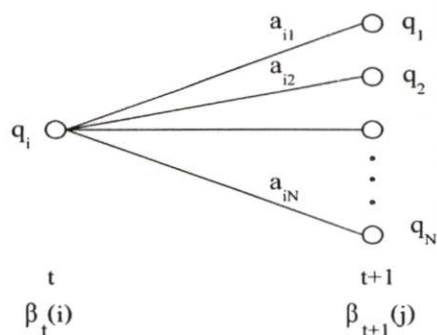
$$\beta_T(i) = 1 \quad ; \quad 1 \leq i \leq N \quad (2.51)$$

2.12.1.1.2.2 การเหนี่ยวนำ (Induction)

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (2.52)$$

เมื่อ  $t = T-1, T-2, \dots, 1, 1 \leq i \leq N$

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



รูปที่ 2.30 กระบวนการย้อนกลับ

#### 2.12.4.2 การแก้ปัญหาที่ 2

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

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1, q_2, \dots, q_{t-1}, q_t = i, O_1, O_2, \dots, O_t | \lambda] \quad (2.53)$$

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

$$\delta_{t+1}(j) = \left[ \max_i \delta_t(i) a_{ij} \right] b_j(O_{t+1}) \quad (2.54)$$

โดยกำหนดให้เป็นอาร์เรย์ที่เก็บตำแหน่งของสแตทที่ให้ค่าความน่าจะเป็นสูงสุด ที่คำนวณได้ในแต่ละเวลา  $t$  และแต่ละลำดับ  $j$  ซึ่งจะสามารถหาลำดับสแตทที่ดีที่สุดได้โดยใช้กระบวนการต่อไปนี้

##### 2.12.4.2.1 การเริ่มต้น (Initialization)

$$\delta_1(i) = \pi_i b_i(O_1) \quad ; 1 \leq i \leq N \quad (2.55)$$

$$\psi_1(i) = 0 \quad (2.56)$$

## 2.12.4.2.2 การย้อนกลับ (Recursion)

$$\delta_t(j) = \left[ \max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij} b_j(o_t) \right] \quad ; \quad \begin{array}{l} 2 \leq t \leq T \\ 1 \leq j \leq N \end{array} \quad (2.57)$$

$$\psi_t(j) = \left[ \arg \max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij} b_j \right] \quad ; \quad \begin{array}{l} 2 \leq t \leq T \\ 1 \leq j \leq N \end{array} \quad (2.58)$$

## 2.12.4.2.3 การสิ้นสุด (Termination)

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (2.59)$$

$$q_T = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (2.60)$$

## 4.5.4.2.4 เส้นทางเดินย้อนกลับ (Backtracking)

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad ; \quad t = T-1, T-2, \dots, 1 \quad (2.61)$$

## 2.12.4.3 การแก้ปัญหาที่ 3

จากที่กล่าวมาแล้วข้างต้นว่าแบบจำลองเสียงจะแทนด้วยค่าพารามิเตอร์  $\lambda = (A, B, \pi)$  ดังนั้นเมื่อมีลำดับของค่าปรากฏจำนวนหนึ่ง เพื่อที่จะนำมาสร้างแบบจำลองอ้างอิง จะต้องทำการคำนวณหาค่าพารามิเตอร์ ของแบบจำลองซึ่งจะอยู่ในรูปของค่าความน่าจะเป็น โดยวิธีที่เลือกใช้ก็คือ วิธีของ บาม-เวลช์ (Baum-Welch Method) หรือเรียกอีกชื่อหนึ่งว่า EM (Expectation-Maximization Method) โดยมี

นิยาม 1. คือ

$$\gamma_t(i) = P(q_t = i | O, \lambda) \quad (2.62)$$

เมื่อ  $\gamma_t(i)$  คือ ค่าความน่าจะเป็นที่จะอยู่ที่สเตต  $i$  ที่ขณะเวลา  $t$  โดยให้ลำดับของค่าปรากฏด้วยโมเดล  $\lambda$  โดยที่กำหนดลำดับของค่าปรากฏให้ สามารถแสดงค่า  $\gamma_t(i)$  ได้ดังนี้

$$\begin{aligned}
 \gamma_t(i) &= P(q_t = i | O, \lambda) \\
 &= \frac{P(O, q_t = i | \lambda)}{P(O | \lambda)} \\
 &= \frac{P(O, q_t = i | \lambda)}{\sum_{i=1}^N P(O, q_t = i | \lambda)} \quad (2.63)
 \end{aligned}$$

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

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)} \quad (2.64)$$

โดย  $\alpha_t(i)$  เริ่มจาก  $O_1, O_2, \dots, O_t$  จนถึงสแตต  $i$  ที่เวลา  $t$

และ  $\beta_t(i)$  เริ่มจาก  $O_{t+1}, O_{t+2}, \dots, O_T$  จนถึงสแตต  $q_t = i$  ที่เวลา  $t$

**นิยาม 2.** คือ

$$\varepsilon_t(i,j) = P(q_t = i, q_{t+1} = j | O, \lambda) \quad (2.65)$$

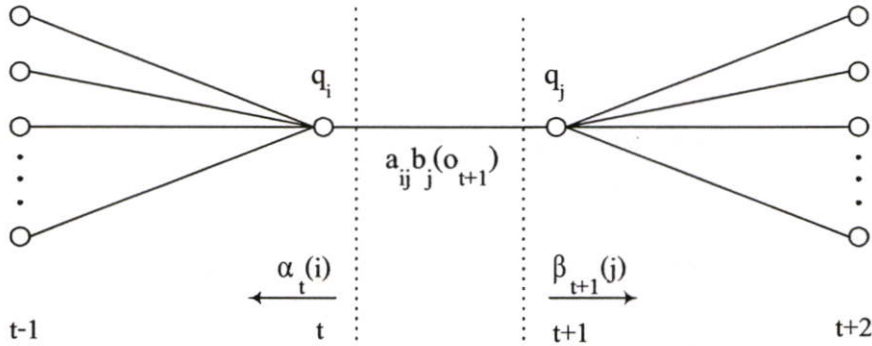
เมื่อ  $\varepsilon_t(i,j)$  คือความน่าจะเป็นที่จะอยู่ที่สแตต  $i$  ที่เวลา  $t$  และสแตต  $j$  ที่เวลา  $t+1$  เมื่อกำหนดแบบจำลองและลำดับค่าปรากฏให้

จากรูปที่ 2.31 แสดง ลำดับการคำนวณการเกิดค่าปรากฏร่วม ซึ่งระบบจะอยู่ในสแตต  $i$  ที่เวลา  $t$  และอยู่ที่ สแตต  $j$  ที่เวลา  $t+1$  โดย  $\alpha_t(i)$  เริ่มจากเวลา  $t = 1$  ที่ค่าปรากฏแรก จนถึงสแตต  $q_t$  ที่เวลา  $t$  และ  $a_{ij}b_{jO_{t+1}}$  เป็นการเปลี่ยนสแตตที่เวลา  $t$  ไปเป็น  $q_j$  ที่เวลา  $t+1$  และให้ค่าปรากฏเป็น  $O_{t+1}$

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

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

$$\begin{aligned}
 &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\
 &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (2.66)
 \end{aligned}$$



รูปที่ 2.31 ลำดับการคำนวณการเกิดค่าปรากฏร่วมซึ่งจะอยู่ที่สแตต  $i$  ที่เวลา  $t$  และอยู่ที่ สแตต  $j$  ที่เวลา  $t+1$

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

$$\gamma_t(i) = \sum_{j=1}^N \varepsilon_t(i,j) \quad (2.67)$$

เมื่อ  $\sum_{t=1}^{T-1} \gamma_t(i) =$  จำนวนของการย้ายสแตตจากสแตต  $i$  ในลำดับค่าปรากฏ  $O$  (2.68)

$\sum_{t=1}^{T-1} \varepsilon_t(i,j) =$  จำนวนของการย้ายสแตตจากสแตต  $i$  ไป  $j$  ในลำดับค่าปรากฏ  $O$  (2.69)

ดังนั้นสามารถคำนวณหาค่าของพารามิเตอร์ได้ดังนี้

$\pi'_i =$  จำนวนครั้งในการอยู่ที่สแตต  $i$  ที่เวลา  $t=1$

$$\pi'_i = \gamma_1(i) \quad ; \quad 1 \leq i \leq N \quad (2.70)$$

$$a'_{ij} = \frac{\text{จำนวนครั้งที่คาดไว้ของการย้ายสแตทจาก } i \text{ ไป } j}{\text{จำนวนครั้งที่คาดว่าจะย้ายจากสแตท } i}$$

$$a'_{ij} = \frac{\sum_{t=1}^{T-1} \epsilon_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (2.71)$$

$$b'_{j(k)} = \frac{\text{จำนวนครั้งที่คาดว่าจะอยู่ในสแตท } j \text{ และเกิดค่าปรากฏเป็น } V_K}{\text{จำนวนครั้งที่คาดว่าจะอยู่ที่สแตท } j}$$

$$b'_{j(k)} = \frac{\sum_{t=1, O_t=V_K}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (2.72)$$

จากกระบวนการข้างต้นถ้าให้  $\lambda = (A, B, \pi)$  เป็นแบบจำลองปัจจุบัน และใช้  $\lambda$  นี้คำนวณในด้านขวาของสมการที่ (2.70-2.72) และให้แบบจำลองได้จากการคำนวณซ้ำเป็น  $\lambda' = (A', B', \pi')$  เป็นแบบจำลองที่ได้จากด้านซ้ายของสมการที่ (2.70-2.72) ซึ่งจะได้จุดวิกฤติของฟังก์ชันความน่าจะเป็นในกรณีที่  $\lambda' = \lambda$  หรือถ้า  $\lambda'$  มีความน่าจะเป็นมากกว่าแบบจำลอง  $\lambda$  [ $P(O|\lambda') > P(O|\lambda)$ ] นั่นคือจะได้แบบจำลอง  $\lambda'$  ใหม่ที่น่าจะทำให้เกิดลำดับของค่าปรากฏ  $O$  ที่ดีกว่า

## 2.12.5 การปรับปรุงค่าพารามิเตอร์ของ HMM

### 2.12.5.1 การสเกลลิง (Scaling)

พิจารณาค่าจำกัดความของ  $\alpha_{t(i)}$  ในสมการที่ 2.46 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 (Dynamic Range) ของการ

ค่านวม  $\alpha_{t(i)}$  มีค่าสูงเกินขอบเขตการทำงานของเครื่องคำนวณทำให้ค่าที่ได้ไม่ถูกต้อง ซึ่งเราสามารถแก้ไขปัญหานี้ได้โดยใช้กระบวนการสเกลลิง (Scaling Procedure)

การสเกลลิงทำได้โดยการคูณ  $\alpha_{t(i)}$  ด้วยสัมประสิทธิ์การสเกลลิง ซึ่งไม่ขึ้นกับ  $i$  (นั่นคือขึ้นอยู่กับค่าของเวลา  $t$  เท่านั้น) เพื่อให้  $\alpha_{t(i)}$  ที่ผ่านการสเกลลิงแล้วมีค่าอยู่ในช่วง Dynamic Range ของเครื่องคำนวณในทุกๆค่าเวลาภายใต้  $1 \leq t \leq T$  และในทำนองเดียวกันจะต้องทำการคำนวณค่าสัมประสิทธิ์การสเกลลิงของค่า  $\beta_{t(i)}$  ด้วย ซึ่งในขั้นตอนสุดท้ายของการคำนวณ ค่าสัมประสิทธิ์ของการสเกลลิงจะตัดกันหมดไป

เพื่อให้เข้าใจการทำงานของกระบวนการสเกลลิงดีขึ้น เราจะพิจารณาสมการของการย้ายสเทท ( $a_{ij}$ ) ที่อยู่ในเทอมของตัวแปรไปข้างหน้าและตัวแปรย้อนกลับ

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_{t(i)} a_{ij} b_j(O_{t+1}) \beta_{t+1(j)}}{\sum_{t=1}^T \sum_{j=1}^N \alpha_{t(i)} a_{ij} b_j(O_{t+1}) \beta_{t+1(j)}} \quad (2.73)$$

พิจารณาสัญลักษณ์ในการคำนวณ  $\alpha_{t(i)}$  เมื่อกำหนดให้

|                             |  |
|-----------------------------|--|
| $\alpha_{t(i)}$ แทน         | $\alpha$ ที่ยังไม่ผ่านการสเกล                |
| $\hat{\alpha}_{t(i)}$ แทน   | $\alpha$ ที่สเกลแล้ว                         |
| $\tilde{\alpha}_{t(i)}$ แทน | $\alpha$ แทนเวอร์ชันของ $\alpha$ ก่อนการสเกล |

เมื่อเวลาเริ่มต้น  $t=1$

ค่านวม  $\alpha_{t(i)}$  ตามสมการที่ 2.47 และกำหนดให้  $\hat{\alpha}_{1(i)} = \alpha_{1(i)}$

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

และ 
$$\hat{\alpha}_{1(i)} = c_1 \alpha_{1(i)} \text{ เมื่อเวลา } 2 \leq t \leq T$$

เริ่มแรกทำการคำนวณหา  $\hat{\alpha}_{t(i)}$  ตามสมการการเหนี่ยวนำ สมการที่ 2.48 โดยใช้เทอมของค่าที่ผ่านการสเกลแล้ว  $\hat{\alpha}_{t-1(i)}$  จะได้ดังนี้

$$\hat{\alpha}_{t(i)} = \sum_{j=1}^N \alpha_{t-1(j)} a_{ji} b_i(O_t) \quad (2.73a)$$

เมื่อกำหนดค่าสัมประสิทธิ์การสเกลลิง ;  $c_t$  เป็น

$$c_t = \frac{1}{\sum_{i=1}^N \hat{\alpha}_t(i)} \quad (2.73b)$$

เมื่อให้

$$\tilde{\alpha}_t(i) = c_t \hat{\alpha}_t(i) \quad (2.73c)$$

จากสมการที่ 2.74a-c สามารถเขียนสมการได้เป็น

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^N \hat{\alpha}_{t-1}(j) a_{ji} b_i(O_t)}{\sum_{i=1}^N \sum_{j=1}^N \hat{\alpha}_{t-1}(j) a_{ji} b_i(O_t)} \quad (2.75)$$

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

$$\hat{\alpha}_{t-1}(j) = \left( \prod_{T=1}^{t-1} c_T \right) \alpha_{t-1}(j) \quad (2.76)$$

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

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^N \alpha_{t-1}(j) \left( \prod_{T=1}^{t-1} c_T \right) a_{ji} b_i(O_t)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_{t-1}(j) \left( \prod_{T=1}^{t-1} c_T \right) a_{ji} b_i(O_t)} = \frac{\alpha_t(i)}{\sum_{i=1}^N \alpha_t(i)} \quad (2.77)$$

นั่นคือการสเกลลิงจะทำได้โดยนำ  $\alpha_t^{(i)}$  แต่ละค่ามาหารด้วยผลรวมของ  $\alpha_t^{(i)}$  ทุกสแตท จากนั้นทำการคำนวณลักษณะเดียวกันนี้กับเทอมของตัวแปรย้อนกลับ  $\beta_t^{(i)}$  โดยใช้สเกลแฟกเตอร์เดียวกันในรูปของ

$$\widehat{\beta}_t^{(i)} = c_t \beta_t^{(i)} \quad (2.78)$$

พิจารณาสมการ 2.75 ในเทอมของตัวแปรที่ผ่านการสเกล จะได้เป็น

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \widehat{\alpha}_t^{(i)} a_{ij} b_j(O_{t+1}) \widehat{\beta}_{t+1}^{(j)}}{\sum_{t=1}^{T-1} \sum_{j=1}^N \widehat{\alpha}_t^{(i)} a_{ij} b_j(O_{t+1}) \widehat{\beta}_{t+1}^{(j)}} \quad (2.79)$$

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

$$\widehat{\alpha}_t^{(i)} = \left[ \prod_{s=1}^T c_s \right] \alpha_t^{(i)} = C_t \alpha_t^{(i)} \quad (2.80)$$

$$\widehat{\beta}_{t+1}^{(j)} = \left[ \prod_{s=t+1}^T c_s \right] \beta_{t+1}^{(j)} = D_{t+1} \beta_{t+1}^{(j)} \quad (2.81)$$

ดังนั้นสมการ 2.79 สามารถเขียนได้เป็น

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} C_t \alpha_t^{(i)} a_{ij} b_j(O_{t+1}) D_{t+1} \beta_{t+1}^{(j)}}{\sum_{t=1}^{T-1} \sum_{j=1}^N C_t \alpha_t^{(i)} a_{ij} b_j(O_{t+1}) D_{t+1} \beta_{t+1}^{(j)}} \quad (2.82)$$

โดยเทอม  $C_t D_{t+1}$  สามารถเขียนให้อยู่ในรูปของ

$$C_t D_{t+1} = \prod_{s=1}^t c_s \prod_{s=t+1}^T c_s = \prod_{s=1}^T c_s = C_T \quad (2.83)$$

จะเห็นว่าเทอม  $C_{D_{t+1}}$  เป็นค่าที่ไม่ขึ้นกับเวลา ดังนั้นสามารถตัดออกจากทั้งเศษและส่วนของสมการ 2.82 ได้ ซึ่งจะทำให้ได้สูตรของการคำนวณซ้ำ กระบวนการสเกลลิงดังกล่าวนี้จะถูกนำไปใช้กับสัมประสิทธิ์  $\pi$  และ  $\beta$  การสเกลลิงนี้จะทำให้การคำนวณค่า  $P(O|\lambda)$  เปลี่ยนไปเราจะไม่สามารถหาได้จากการรวมเทอมของ  $\hat{\alpha}_{T(i)}$  เนื่องจากเป็นค่าที่ถูกสเกลแล้ว แต่เราสามารถคำนวณได้จากคุณสมบัติ

$$\prod_{t=1}^T c_t \prod_{i=1}^N \alpha_{T(i)} = C_T \sum_{i=1}^N \alpha_{T(i)} = 1 \quad (2.84)$$

ดังนั้นจะได้

$$\prod_{t=1}^T c_t \cdot P(O|\lambda) = 1 \quad (2.85)$$

หรือ

$$P(O|\lambda) = \frac{1}{\prod_{t=1}^T c_t} \quad (2.86)$$

หรือ

$$\log [ P(O|\lambda) ] = -\sum_{t=1}^T \log C_t \quad (2.87)$$

นั่นคือ การคำนวณค่า  $P$  จะอยู่ในรูป  $\log$  ของ  $P$  เพื่อไม่ให้เกินช่วงไดนามิก (Dynamic Range) ของเครื่องคำนวณ

### 2.12.5.2 ลำดับของค่าปรากฏหลายลำดับ (Multiple Observation Sequences)

ในการสร้างแบบจำลองด้วย Left-Right Model จำเป็นจะต้องใช้จำนวนลำดับของเหตุการณ์หลายๆลำดับเข้ามาแทนเพื่อให้การประมาณค่าพารามิเตอร์ ของแบบจำลองที่ได้มีความน่าเชื่อถือที่สุด ถ้ากำหนดให้  $k$  แทนเซตของลำดับค่าปรากฏ ดังนี้

$$O = [ O^{(1)}, O^{(2)}, \dots, O^{(k)} ] \quad (2.88)$$

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

$$\begin{aligned}
 P(O|\lambda) &= \prod_{k=1}^K P(O^{(k)}|\lambda) \\
 &= \prod_{k=1}^K P_k
 \end{aligned}
 \tag{2.89}$$

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

$$\bar{a}_{ij} = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^{k(i)} a_{ij} b_j(O_{t+1}^{(k)}) \beta_{t+1}^k(j)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^{k(i)} \beta_t^k(i)}
 \tag{2.90}$$

และ

$$\bar{b}_j(l) = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1, O_t=v_l}^{T_k-1} \alpha_t^{k(i)} \beta_t^k(i)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^{k(i)} \beta_t^k(i)}
 \tag{2.91}$$

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

จากสมการของการสเกลถึงสมการที่ 2.90-2.91 เราสามารถเขียนสมการที่อยู่ในเทอมของตัวแปรที่สเกลแล้วได้เป็น

$$\bar{a}_{ij} = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^{k(i)} a_{ij} b_j(O_{t+1}^{(k)}) \hat{\beta}_{t+1}^k(j)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^{k(i)} \hat{\beta}_t^k(i)}
 \tag{2.92}$$

$$\bar{b}_j(k) = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1, O_t=v_l}^{T_k-1} \hat{\alpha}_t^{k(i)} \hat{\beta}_t^k(i)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^{k(i)} \hat{\beta}_t^k(i)}
 \tag{2.93}$$

## บทที่ 3

# วิธีการดำเนินการวิจัย และการรู้จำเสียงตัวเลขภาษาไทย

### 3.1 วิธีการดำเนินการวิจัย

#### 3.1.1 การสร้างรายการคำทดสอบ

การสร้างรายการคำทดสอบในวิทยานิพนธ์นี้ จะใช้ผู้บอกภาษาทั้งหมด จำนวน 6 คน ประกอบด้วยผู้ชาย 3 คน และผู้หญิง 3 คน โดยจะให้ผู้ทดสอบแต่ละคนอ่านออกเสียงตัวเลขภาษาไทย ตั้งแต่ 0 ถึง 9 จำนวน 10 ตัว คนละ 500 คำ ซึ่งเป็นคำตัวเลข 0 ถึง 9 จำนวน 50 ชุด ดังนั้นวิทยานิพนธ์นี้จะมีคำที่ใช้ทั้งหมด 3,000 คำ ซึ่งจะนำไปใช้สร้างแบบอ้างอิงจำนวน 1,500 คำ และใช้สำหรับสร้างแบบทดสอบ จำนวน 1,500 คำ

#### 3.1.2 การคัดเลือกผู้บอกภาษา

การคัดเลือกผู้บอกภาษาในวิทยานิพนธ์นี้ จะใช้ผู้บอกภาษาที่เป็นคนไทยมีอายุระหว่าง 20-40 ปี เป็นเพศชาย จำนวน 3 คน และเป็นเพศหญิง จำนวน 3 คน โดยภาษาที่ใช้จะเป็นภาษาไทยของคนภาคกลาง

#### 3.1.3 การเก็บข้อมูล

การเก็บข้อมูลในวิทยานิพนธ์นี้ จะใช้คำพูดเป็นภาษาไทย โดยจะให้ผู้ทดสอบแต่ละคนอ่านออกเสียงตัวเลขภาษาไทย ตั้งแต่เลข 0 ถึง 9 (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) จำนวน 50 ชุด 500 รอบ โดยการอ่านจะให้อ่านออกเสียงแบบไม่ต่อเนื่องคือจะให้อ่านแล้วหยุดสลับกัน เพื่อให้คำที่พูดมีลักษณะไม่ต่อเนื่องกัน ซึ่งในเสียงตัวเลขภาษาไทยที่ทำการบันทึกจะประกอบด้วยเสียงของสระและวรรณยุกต์รวมอยู่ด้วยกัน ในการบันทึกเสียงซึ่งข้อมูลที่ถูกรับจะถูกบันทึกอยู่ในรูปแบบของไฟล์ข้อมูล “.wav” ซึ่งข้อมูลหนึ่งตัวอย่างของเสียงจะถูกแทนด้วยข้อมูลขนาด 8 บิต โดยใช้ความถี่ในการแซมปลิงเสียงเท่ากับ 11.025 kHz ไฟล์ข้อมูล “.wav” จะถูกนำไปใช้เป็นอินพุตสำหรับโปรแกรมการคำนวณวิเคราะห์เสียงตัวเลขภาษาไทยที่เขียนขึ้น

#### 3.1.4 เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียง

เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียง ในวิทยานิพนธ์นี้ จะใช้คอมพิวเตอร์ จำนวน 1 เครื่อง ไมโครโฟน จำนวน 1 ตัว และลำโพงจำนวน 1 ชุด โปรแกรมบันทึกเสียงใช้ Wave Lab version 4.0 และ โปรแกรมที่ใช้ในการทำงานวิจัยใช้โปรแกรม MATLAB เวอร์ชัน 6.1

### 3.1.5 ข้อจำกัดในการดำเนินการวิจัย

ข้อจำกัดในการดำเนินการวิจัยในวิทยานิพนธ์นี้ ผู้บอกภาษาจะต้องออกเสียงภาษาไทยได้ชัดเจน และต้องพูดภาษาไทยของคนภาคกลางได้ชัดเจน

## 3.2 การรู้จำเสียงตัวเลขภาษาไทย

### 3.2.1 หลักการรู้จำเสียงภาษาไทย

ตัวเลขภาษาไทย ประกอบด้วยตัวเลขที่มีค่าตั้ง 0 ถึง 9 รวมทั้งสิ้นจำนวน 10 ตัว ตัวเลขแต่ละตัวจะประกอบด้วย 1 พยางค์ ที่รวมขึ้นจากหน่วยเสียงในภาษาไทย ประกอบด้วย เสียงพยัญชนะ, เสียงสระ และเสียงวรรณยุกต์ การพิจารณาส่วนประกอบของเสียงสระ และเสียงวรรณยุกต์จากตัวเลขภาษาไทยสามารถแสดงได้ดังตารางที่ 4.1

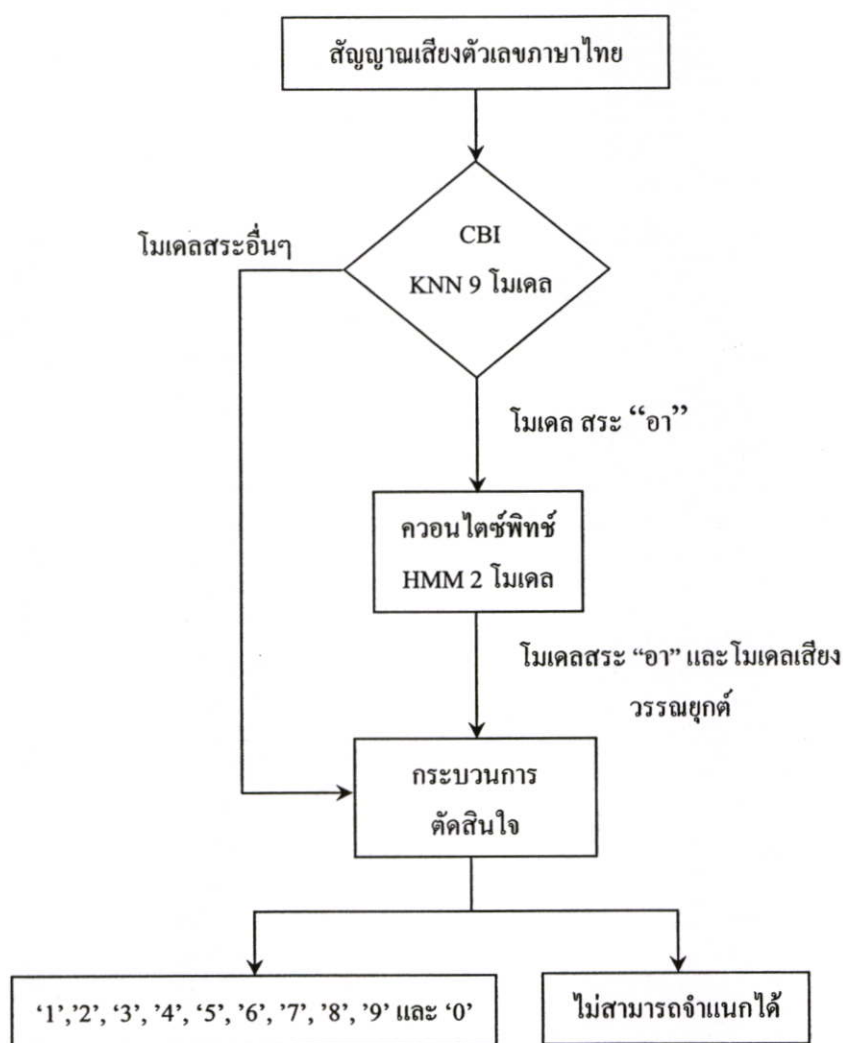
ตารางที่ 3.1 ตารางสรุป เสียงสระ และเสียงวรรณยุกต์ของตัวเลขภาษาไทย

| ตัวเลข | เสียง | สระ<br>(Vowel) | วรรณยุกต์<br>(Tone) |
|--------|-------|----------------|---------------------|
| 1      | หนึ่ง | อี             | เอก                 |
| 2      | สอง   | ออ             | จัตวา               |
| 3      | สาม   | อา             | จัตวา               |
| 4      | สี่   | อี             | เอก                 |
| 5      | ห้า   | อา             | โท                  |
| 6      | หก    | โอะ            | เอก                 |
| 7      | เจ็ด  | เอะ            | เอก                 |
| 8      | แปด   | แอะ            | เอก                 |
| 9      | เก้า  | เอา            | โท                  |
| 0      | ศูนย์ | อุ             | จัตวา               |

จากตารางที่ 3.1 จะพบว่า จำนวนเสียงสระที่ใช้ในการออกเสียงตัวเลขภาษาไทย มีทั้งหมด 9 เสียง ได้แก่ สระอี, ออ, อา, อี, โอะ, เอะ, แอะ, เอา และอุ และจำนวนเสียงของวรรณยุกต์จำนวน 3 เสียง ได้แก่ วรรณยุกต์เอก, โท และจัตวา

จากรูปแบบของเสียงสระ (9 เสียง) และเสียงวรรณยุกต์ (3 เสียง) ที่ประกอบขึ้นเป็นคำตัวเลขภาษาไทยนำมาผ่านกระบวนการการจำแนกเสียงสระ โดยการแทนคลื่นเสียงของตัวเลข

ภาษาไทยด้วยเวกเตอร์คุณสมบัติขนาด 18 มิติของความเข้มแถบวิกฤติ (CBI) เพื่อจำแนกเสียงสระด้วยวิธี KNN และการจำแนกเสียงวรรณยุกต์โดยการวิเคราะห์คาบพิทซ์เพื่อนำมาหาค่าความถี่มูลฐานเพื่อให้ได้ค่าควอนไทซ์พิทซ์ ผ่านกระบวนการรู้จำโดยใช้วิธี HMM และนำผลลัพธ์ของการจำแนกเสียงสระและเสียงวรรณยุกต์จะผ่านกระบวนการตัดสินใจ เพื่อให้ได้การจำแนกเสียงตัวเลขภาษาไทย ดังแสดงตามแผนผังการรู้จำตัวเลขภาษาไทย รูปที่ 3.1



รูปที่ 3.1 แผนผังกระบวนการรู้จำตัวเลขภาษาไทย

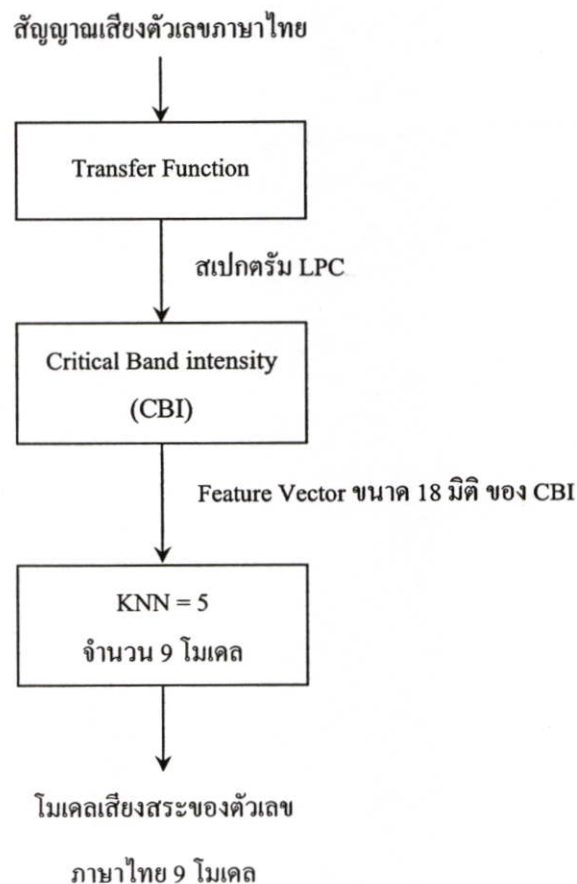
### 3.2.2 การหาค่าพารามิเตอร์ ที่ใช้ในการแยกหน่วยเสียงภาษาไทย

จากรูปที่ 3.1 การรู้จำตัวเลขภาษาไทยเราจะต้องพิจารณาใน 2 ขั้นตอน ได้แก่ การหาพารามิเตอร์การแยกหน่วยเสียงสระ และการหาพารามิเตอร์การแยกหน่วยเสียงวรรณยุกต์ ดังนี้

3.2.2.1 การหาพารามิเตอร์ที่ใช้ในการแยกหน่วยเสียงสระของตัวเลขภาษาไทย  
ในหัวข้อนี้จะกล่าวถึงวิธีการแยกหน่วยเสียงสระของตัวเลขภาษาไทย และขั้นตอนใน  
การวิเคราะห์เสียงสระของตัวเลขภาษาไทย

หน่วยเสียงสระมีลักษณะเป็นเสียงก้อง (Voice) และมีลักษณะสัญญาณเป็นคาบ ซึ่งเป็น  
ลักษณะที่เห็นได้ชัดในโดเมนของเวลา ดังนั้นการกำหนดขอบเขตของคำเพื่อที่จะวิเคราะห์หน่วย  
เสียงสระที่ประกอบขึ้นเป็นตัวเลขภาษาไทยนั้น จึงต้องกำหนดขอบเขตภายในช่วงที่มีความเป็น  
คาบทั้งหมดของคำตัวเลขภาษาไทยนั้นๆ รวมถึงการใช้ระดับพลังงานร่วมด้วย [8] จากนั้นนำเสียงที่  
ตัดได้มาทำการวิเคราะห์เสียงสระของตัวเลขภาษาไทย ซึ่งแบ่งออกเป็น 3 ขั้นตอน ตามรูปที่ 3.2  
ดังนี้

1. การหาสเปกตรัม LPC
2. การแปลงจากสเกลความถี่ ไปเป็นสเกลบาร์ก หรือ Critical Band Rate Scale และการ  
คำนวณหา Critical Band Intensity
3. การใช้ Critical Band Intensity เป็น Feature Vector ในการสร้างแบบอ้างอิง



รูปที่ 3.2 ขั้นตอนในการวิเคราะห์หน่วยเสียงสระของตัวเลขภาษาไทย

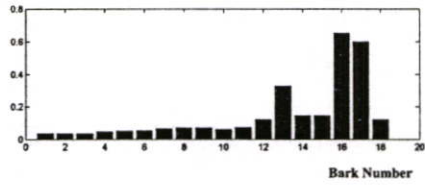
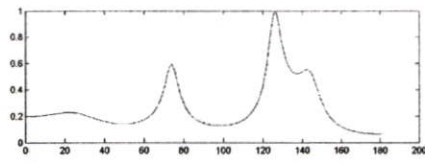
### 3.2.2.1.1 การหาสเปกตรัมแอลพีซี

วิธีการหาสเปกตรัมแอลพีซีนั้นได้กล่าวมาแล้วในบทที่ 2 ซึ่งจะเป็นการนำสัญญาณตัวอย่างที่ผ่านกระบวนการตัดเสียงแล้วมาคำนวณหาค่าตามขั้นตอนต่างๆ เพื่อให้ได้ค่าสเปกตรัมแอลพีซี ขั้นตอนในการคำนวณหาค่าแอลพีซีสเปกตรัมของสัญญาณเสียงโดยเริ่มจากสัญญาณเสียงที่ผ่านการตัดเสียงแล้วนำมาทำการพีเอ็มพีซี เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวนมีค่าค่อนข้างคงที่ตลอดช่วงความถี่ สัญญาณที่ผ่านการพีเอ็มพีซีจะถูกแบ่งออกเป็นเฟรม เฟรมละ 300 แซมเปิ้ล การวิเคราะห์สัญญาณเสียงจะทำการวิเคราะห์ทีละเฟรม โดยแต่ละเฟรมจะถูกเลื่อนไปทีละ 100 แซมเปิ้ล จากนั้นนำสัญญาณแต่ละเฟรมมาคูณด้วยฟังก์ชันวินโดว์ เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปกตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไป แล้วนำสัญญาณที่ผ่านการวินโดว์ไปคำนวณหาค่าสัมประสิทธิ์แอลพีซี เมื่อได้ค่าสัมประสิทธิ์แอลพีซีแล้วก็สามารถคำนวณหาสเปกตรัมของแอลพีซี [8] [15]

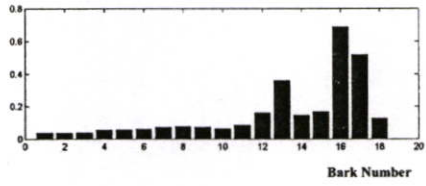
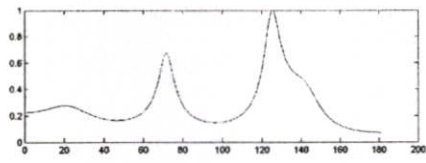
การคำนวณหาค่าสัมประสิทธิ์แอลพีซีนั้นจำเป็นต้องเลือกอันดับของแอลพีซีให้เหมาะสม ถ้าอันดับของแอลพีซีมีค่าต่ำจะทำให้คำนวณเร็วขึ้น แต่ค่าสเปกตรัมแอลพีซีที่คำนวณได้มีรายละเอียดน้อยส่งผลให้การคำนวณหาความเข้มแถบวิกฤติ (Critical Band Intensity) เกิดความผิดพลาด แต่ถ้าค่าระดับแอลพีซีมีค่าสูงจะใช้เวลานาน ทำให้การคำนวณยุ่งยากและซับซ้อนเกินความจำเป็น ค่าลำดับแอลพีซีที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์แอลพีซีในงานวิจัยนี้ ใช้ค่าแอลพีซีลำดับที่ 12

### 3.2.2.1.2 การแปลงจากสเกลความถี่ไปเป็นสเกลบาร์ก หรือ Critical Band Rate Scale และการคำนวณหาความเข้มแถบวิกฤติ

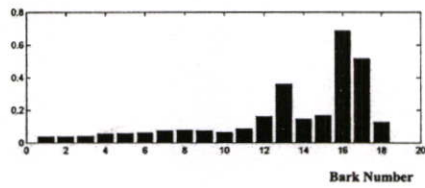
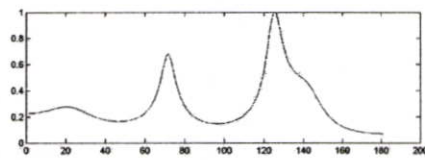
สเกลบาร์ก หรือ Critical Band Rate Scale นั้นเป็นตัวช่วยให้ความสามารถในการแสดงคุณสมบัติของการตอบสนองต่อความถี่ต่างๆ ของการได้ยินของมนุษย์เป็นเชิงเส้นตลอดช่วงความถี่ที่มนุษย์สามารถได้ยิน ดังนั้นจึงทำการแปลงสเปกตรัมแอลพีซีเป็นความถี่ไปบนสเกลบาร์ก ดังรูป 3.3 แล้วนำค่าสเปกตรัมแอลพีซีที่อยู่บนสเกลบาร์กไปคำนวณหาค่าความเข้มของสัญญาณในแต่ละความถี่วิกฤติ ซึ่งค่าที่ได้นี้จะนำไปใช้เป็นองค์ประกอบของเวกเตอร์ขนาด 18 มิติ [8] [15] ของความเข้มแถบวิกฤติสำหรับแทนหน่วยเสียงสระของตัวเลขภาษาไทยในแต่ละเฟรม ดังรูป 3.4 ค่าองค์ประกอบของเวกเตอร์แทนหน่วยเสียงสระของตัวเลขภาษาไทยทั้งหมดที่ได้ จะนำไปใช้ในการสร้างแบบอ้างอิงและแบบทดสอบของการรู้จำเสียงสระของตัวเลขภาษาไทยต่อไป



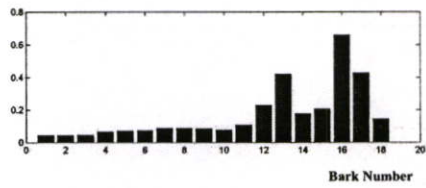
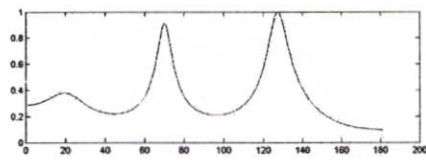
เฟรมที่ 1



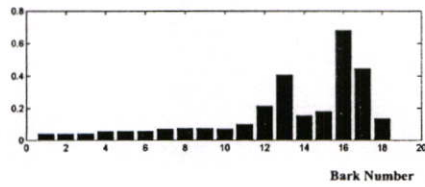
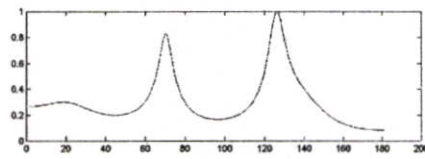
เฟรมที่ 2



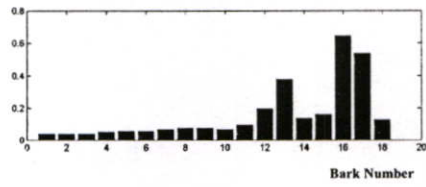
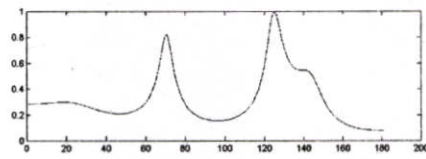
เฟรมที่ 3



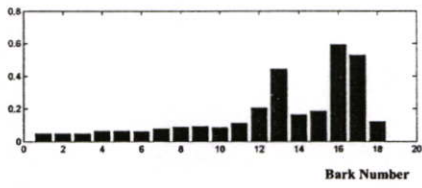
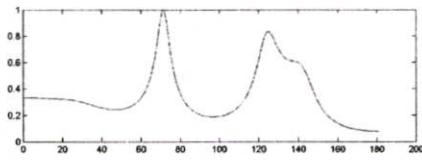
เฟรมที่ 4



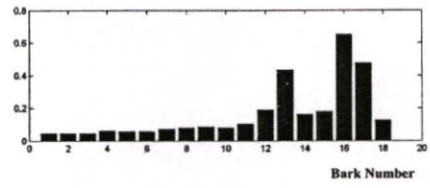
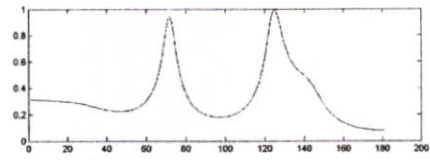
เฟรมที่ 5



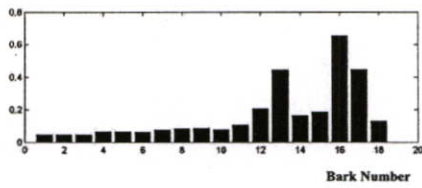
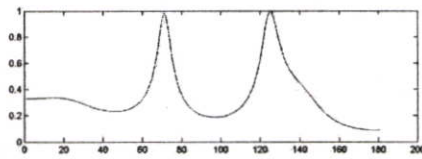
เฟรมที่ 6



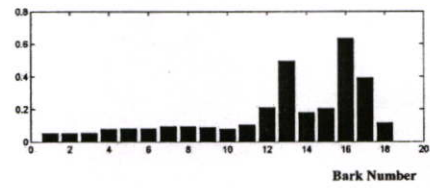
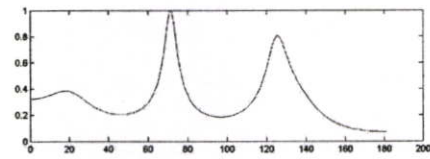
เฟรมที่ 7



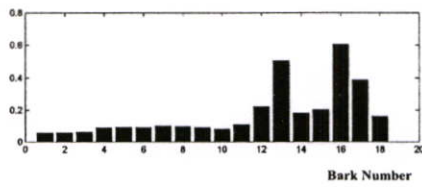
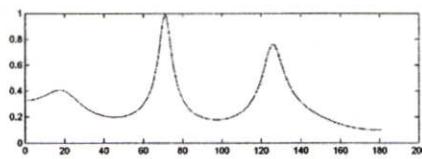
เฟรมที่ 8



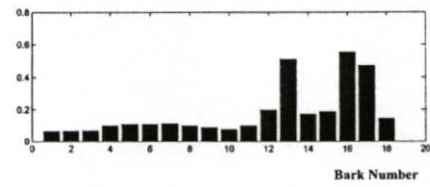
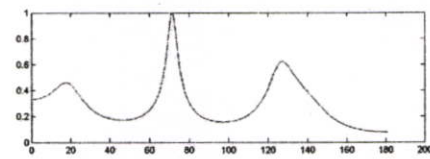
เฟรมที่ 9



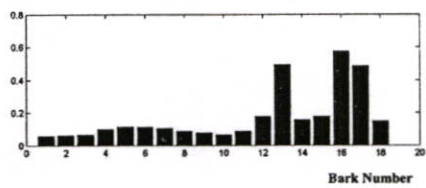
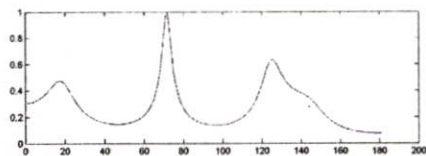
เฟรมที่ 10



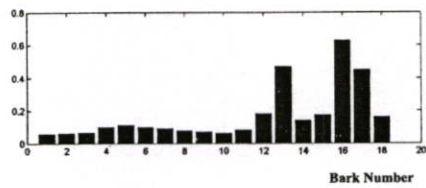
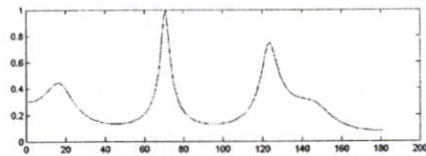
เฟรมที่ 11



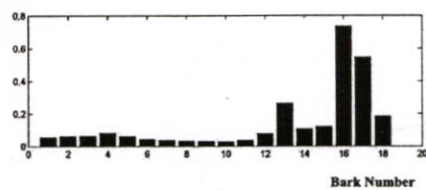
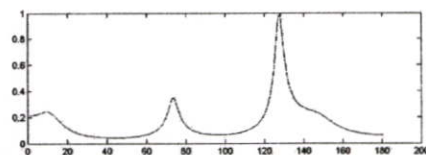
เฟรมที่ 12



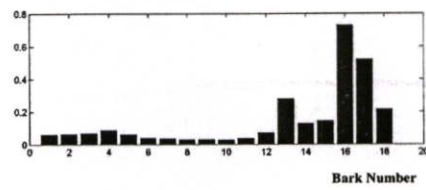
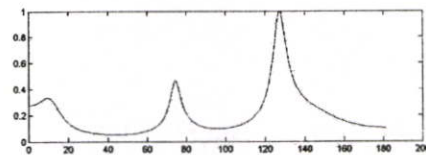
เฟรมที่ 13



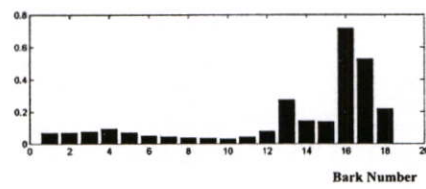
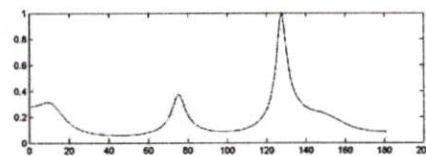
เฟรมที่ 14



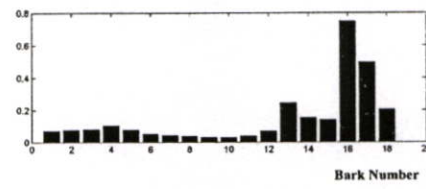
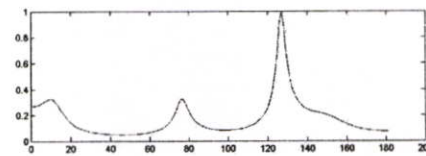
เฟรมที่ 15



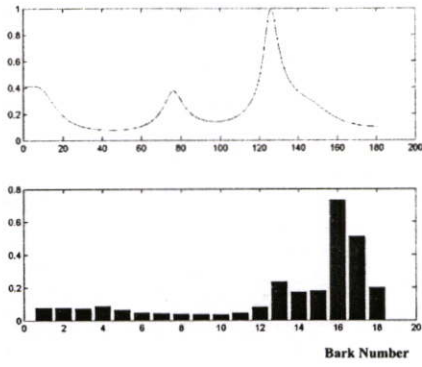
เฟรมที่ 16



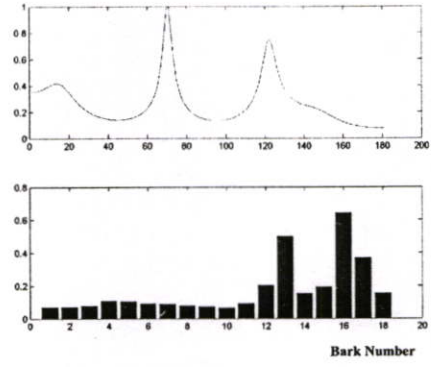
เฟรมที่ 17



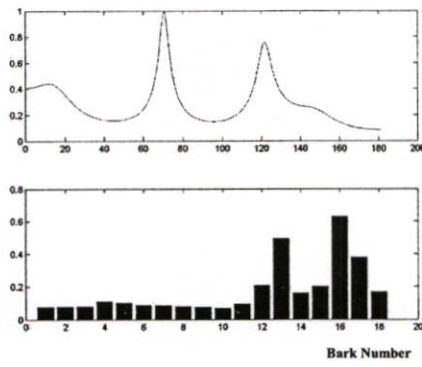
เฟรมที่ 18



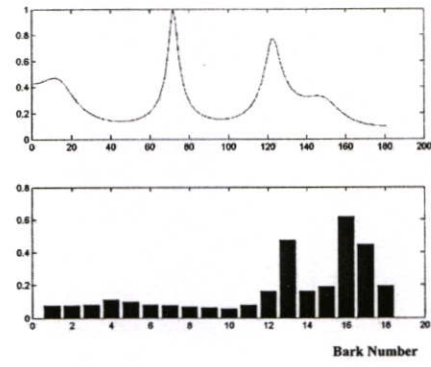
เฟรมที่ 19



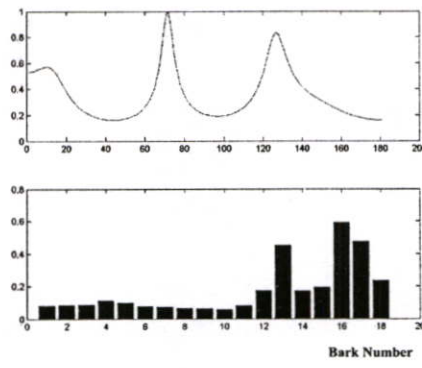
เฟรมที่ 20



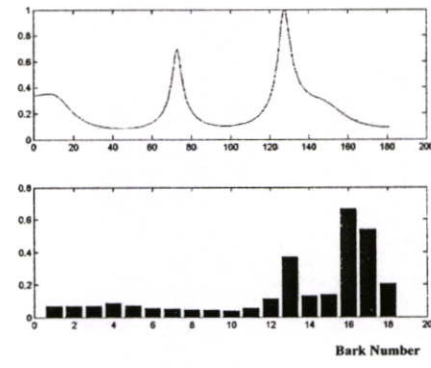
เฟรมที่ 21



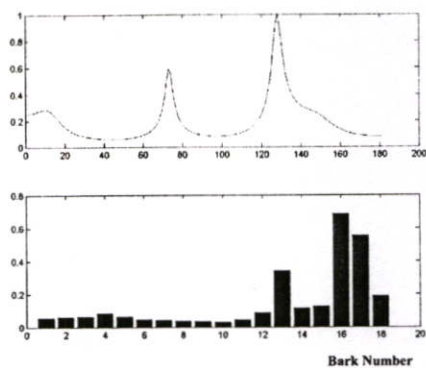
เฟรมที่ 22



เฟรมที่ 23

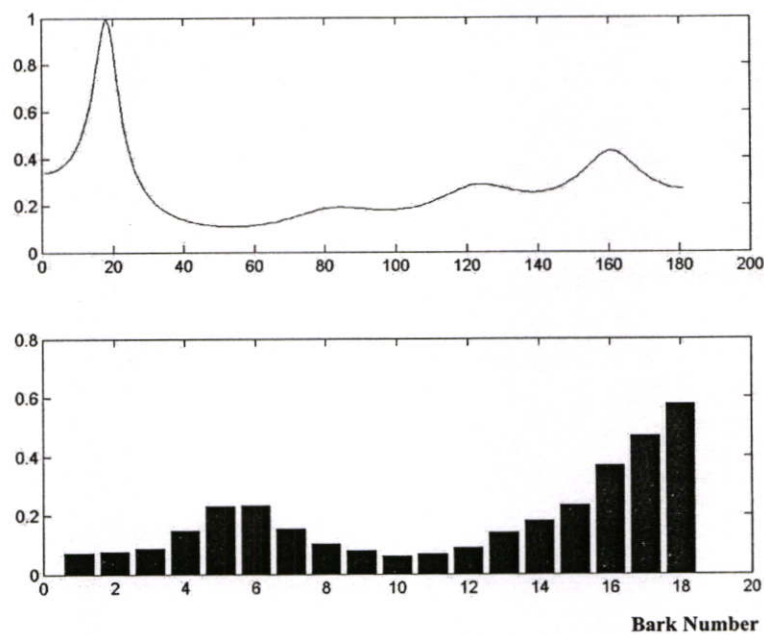


เฟรมที่ 24



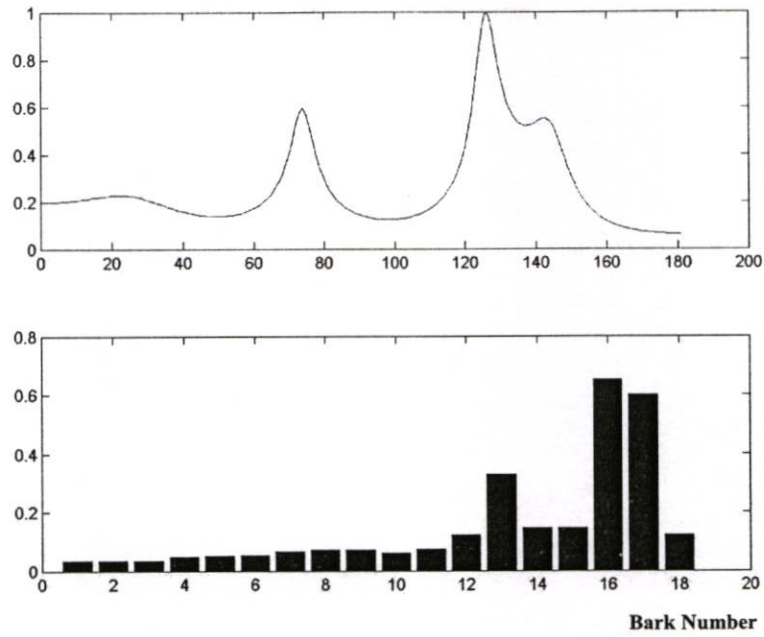
เฟรมที่ 25

รูปที่ 3.3 ตัวอย่างของสเปกตรัมแอลพีซี และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระทั้งหมด จำนวน 25 เฟรม ที่ประกอบอยู่ในคำตัวเลข “1”

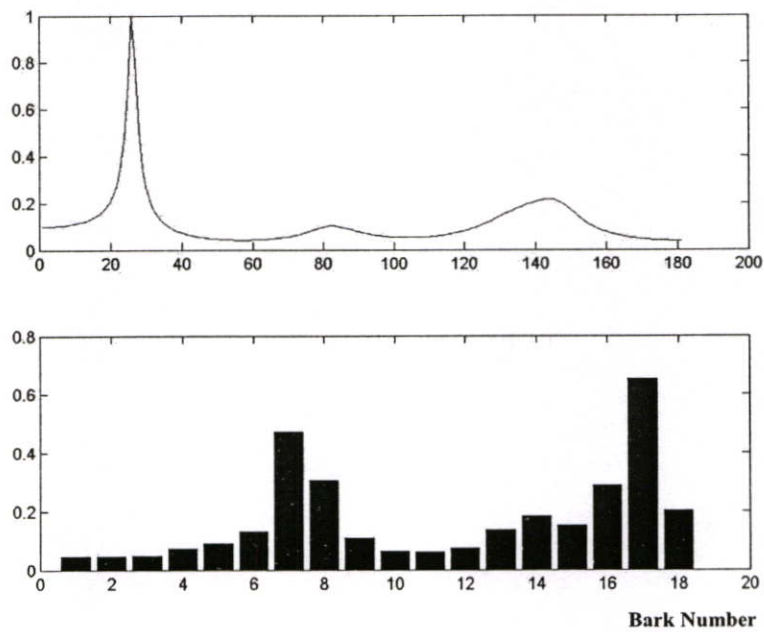


(ก) สเปกตรัมแอลพีซี (บน) และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง)

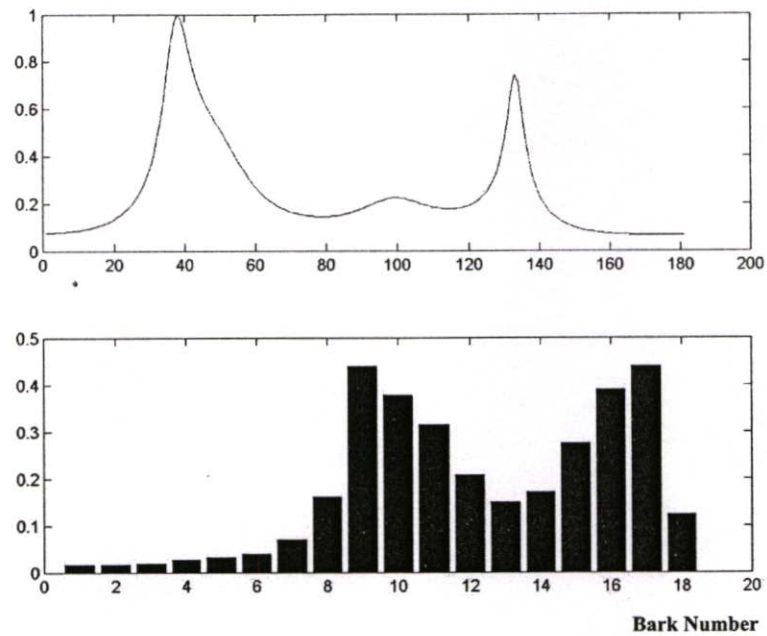
ที่ประกอบอยู่ในตัวเลข “0”



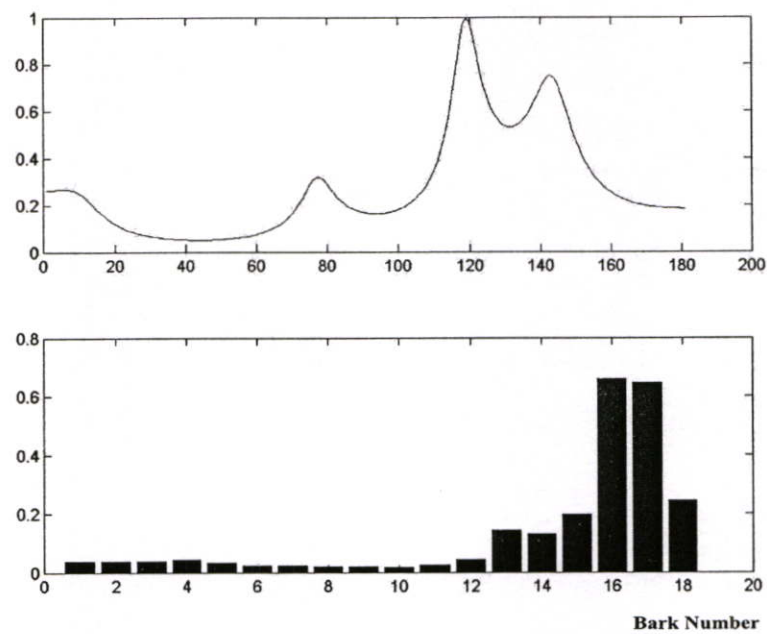
(ข) สเปกตรัมแอลพีซี (บน) และความเข้มสัมบูรณ์ในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “1”



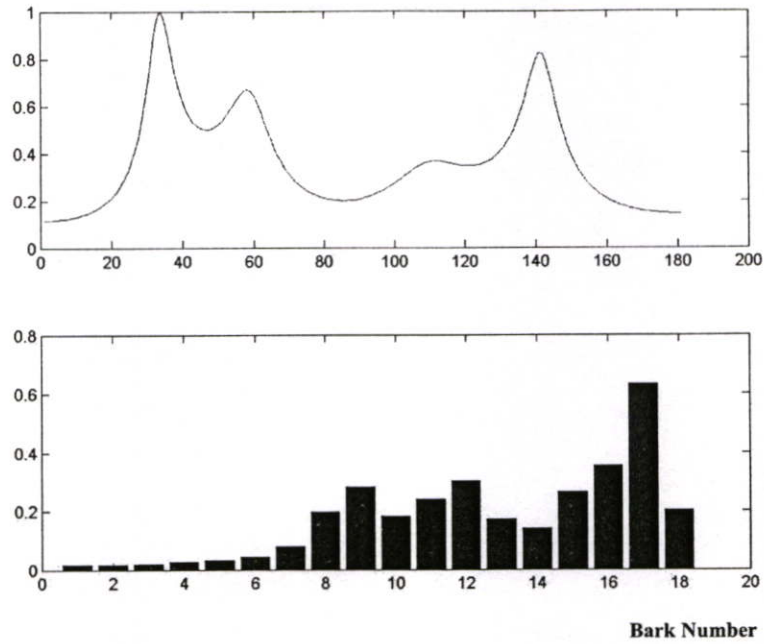
(ค) สเปกตรัมแอลพีซี (บน) และความเข้มสัมบูรณ์ในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “2”



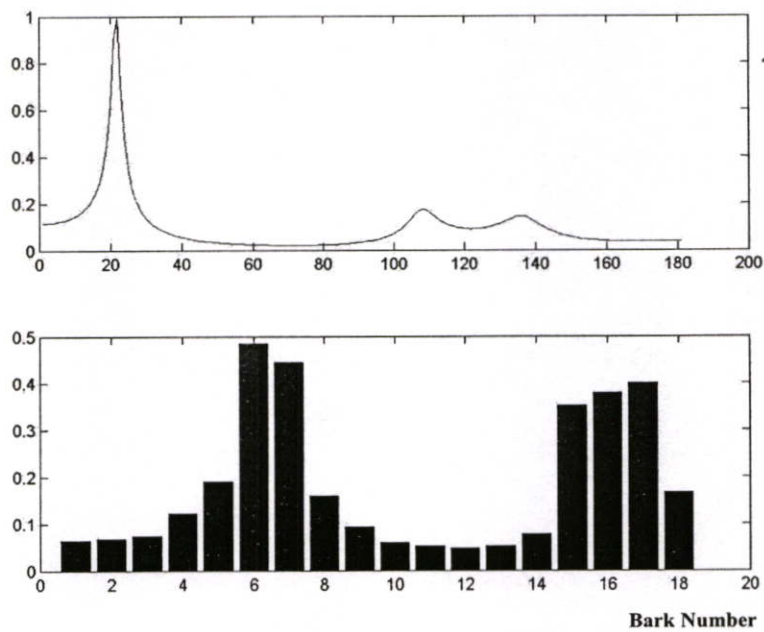
(ง) สเปกตรัมแอลพีซี (บน) และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “3”



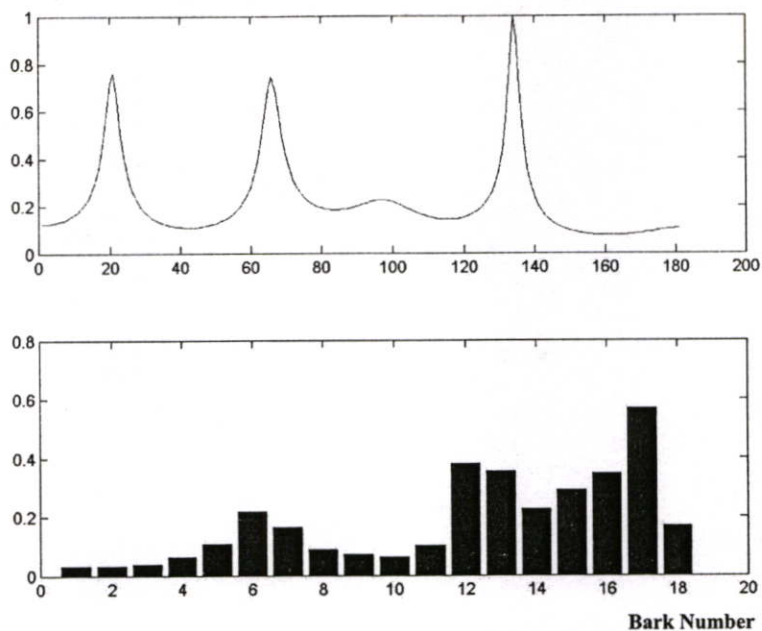
(จ) สเปกตรัมแอลพีซี (บน) และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “4”



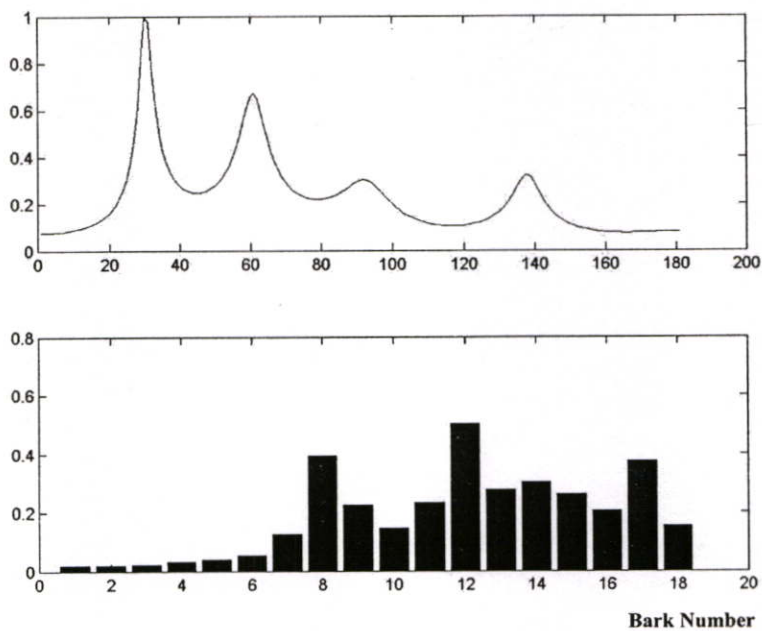
(จ) สเปกตรัมแอมพลิจูด (บน) และความเข้มสัมบูรณ์ในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “5”



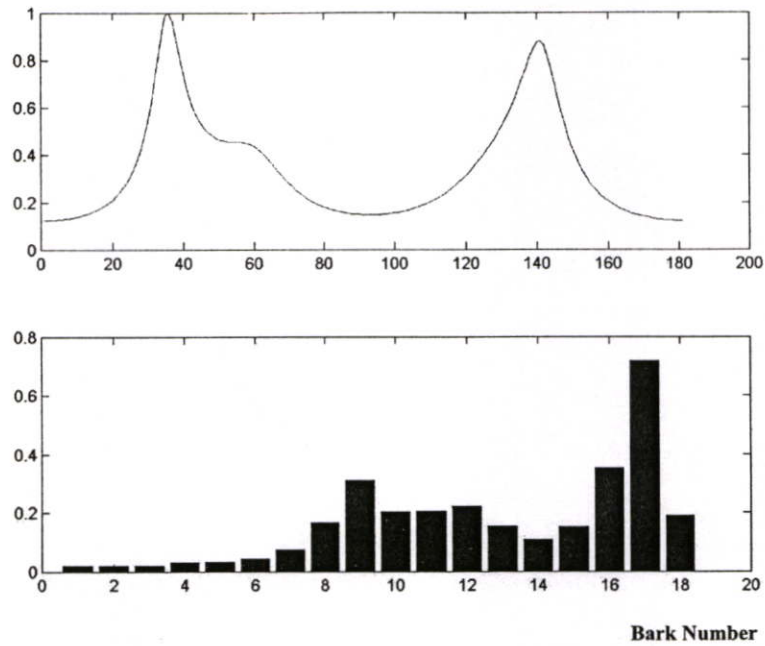
(ข) สเปกตรัมแอมพลิจูด (บน) และความเข้มสัมบูรณ์ในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “6”



(ข) สเปกตรัมแอมพลิจูด (บน) และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “7”



(ค) สเปกตรัมแอมพลิจูด (บน) และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “8”



(ญ) สเปกตรัมแอลพีซี (บน) และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระ (ล่าง) ที่ประกอบอยู่ในตัวเลข “9”

**รูปที่ 3.4** สเปกตรัมแอลพีซี และความเข้มสัญญาณในแต่ละแถบความถี่วิกฤติของเสียงสระทั้งหมด (ในแต่ละเฟรม) ที่ประกอบอยู่ในคำตัวเลขภาษาไทย (เลข 0 ถึงเลข 9)

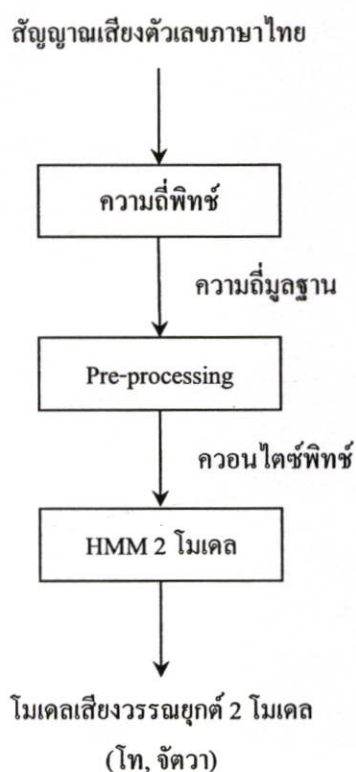
ค่าความเข้มสัญญาณในแต่ละความถี่วิกฤติ เป็นองค์ประกอบของเวกเตอร์ที่เราใช้เป็นตัวแทนของแต่ละเฟรมของสัญญาณเสียงสระในการสร้างแบบอ้างอิง และการรู้จำ จากรูปที่ 3.3 จะเห็นได้ว่าค่าเวกเตอร์ที่ได้มีความแตกต่างกันระหว่างเสียงสระ ซึ่งจะทำได้สามารถแยกแยะเสียงสระทั้งหมดออกจากกันได้ นำเวกเตอร์แทนหน่วยเสียงสระของแต่ละเฟรมในแต่ละคำของตัวเลขภาษาไทยที่คำนวณได้ของสระทั้งหมด 9 เสียงจากเสียงสระที่ประกอบอยู่กับตัวเลขภาษาไทย ตั้งแต่เลข 0 ถึง 9 จำนวน 10 ตัวมาสร้างโมเดลเสียงได้ทั้งหมด 9 โมเดลเสียง โดยใช้เสียงพูดของผู้ชาย 3 คน และผู้หญิง 3 คน จำนวน 500 คำต่อคน เสียงพูดของผู้ชาย และผู้หญิงชุดดังกล่าวจำนวน 250 คำต่อคน จะถูกนำมาใช้ในการสร้างแบบอ้างอิง และเสียงพูดผู้พูดชุดเดิมที่เหลือ จำนวน 250 คำต่อคน จะถูกใช้สำหรับการทดสอบ .

### 3.2.2.2 การหาพารามิเตอร์ที่ใช้ในการแยกหน่วยเสียงวรรณยุกต์ของตัวเลขภาษาไทย

ขั้นตอนการวิเคราะห์หน่วยเสียงวรรณยุกต์ จะเริ่มจากการพิจารณาคูณลักษณะของเสียงพูด ได้แก่ พิตช์ (Pitch) ซึ่งเป็นความถี่พื้นฐานที่เกิดจากการสั่นในหลอดเสียง และเป็น

ตัวกำหนดระดับความสูงต่ำของเสียงเนื่องจากภาษาไทยเป็นภาษาที่มีเสียงวรรณยุกต์ (Tonal Language) การกำหนดขอบเขตการวิเคราะห์เสียงวรรณยุกต์จะกำหนดขอบเขตภายในช่วงที่มีความเป็นคาบทั้งหมดของคำตัวเลขภาษาไทยนั้นๆ เช่นเดียวกับกับการวิเคราะห์หน่วยเสียงสระ จากนั้นนำเสียงที่ตัดได้มาทำการวิเคราะห์เสียงวรรณยุกต์ของตัวเลขภาษาไทย ซึ่งแบ่งออกเป็น 3 ขั้นตอนตามรูปที่ 3.5 ดังนี้ [9] [16]

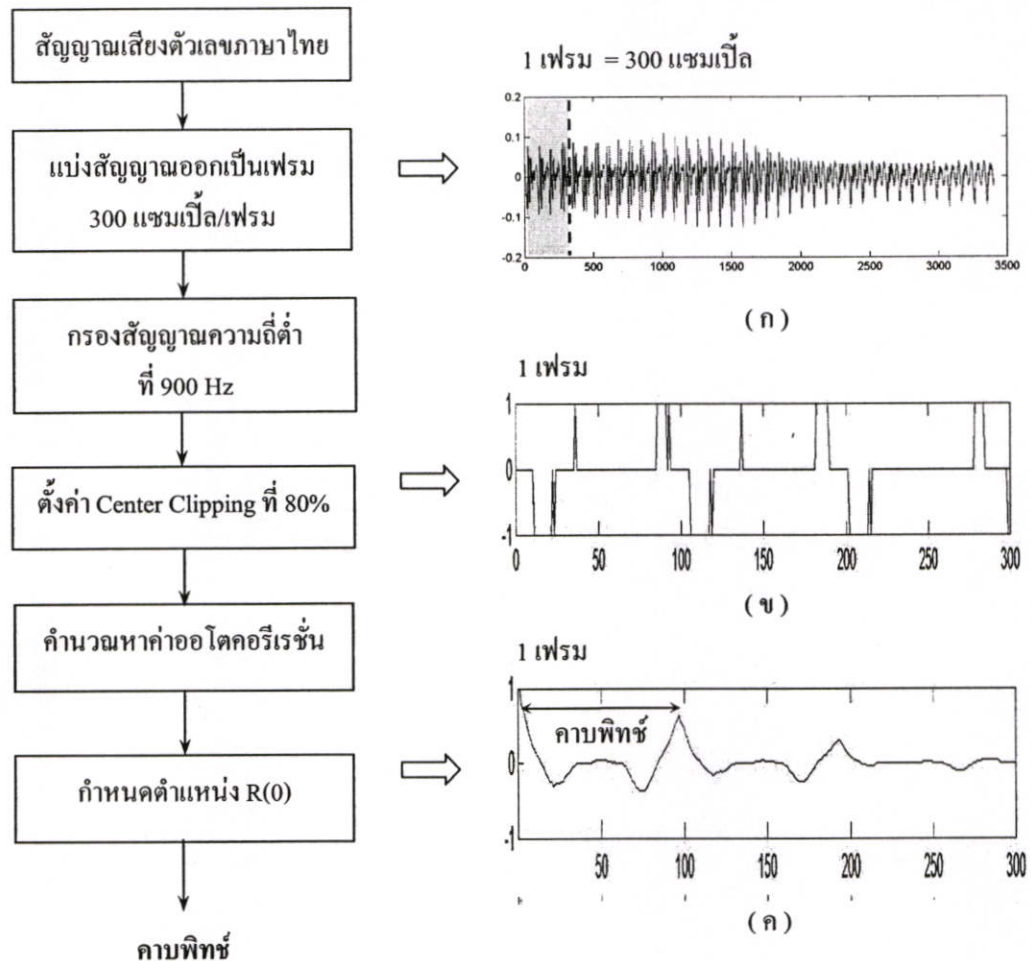
1. การหาความถี่พิทช์
2. การเตรียมข้อมูลความถี่พิทช์ ก่อนเข้าสู่กระบวนการในการสร้างแบบอ้างอิง (Pre-processing)
3. การใช้ Quantized Pitch เป็น Feature Vector ในการสร้างแบบอ้างอิง



รูปที่ 3.5 ขั้นตอนในการวิเคราะห์หน่วยเสียงสระของตัวเลขภาษาไทย

### 3.2.2.2.1 การคำนวณหาพิทช์

ขั้นตอนและวิธีการการคำนวณหาพิทช์นั้น แสดงดังรูปที่ 3.6 ซึ่งจะเป็นการนำสัญญาณตัวอย่างที่ผ่านกระบวนการตัดเสียงแล้วมาคำนวณหาค่าตามขั้นตอนต่างๆ เพื่อให้ได้ค่าพิทช์ [16]

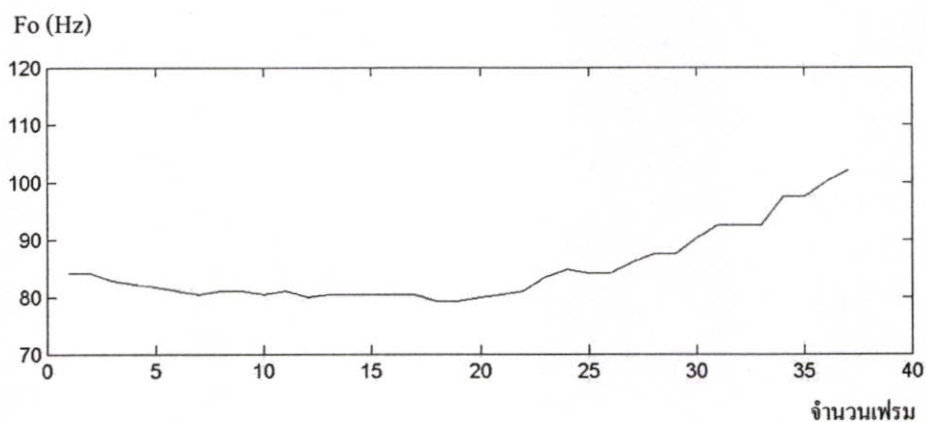


รูปที่ 3.6 ขั้นตอนต่างๆ ในการคำนวณหาพิทช์ และตัวอย่างการหาพิทช์ของเลข “หนึ่ง”

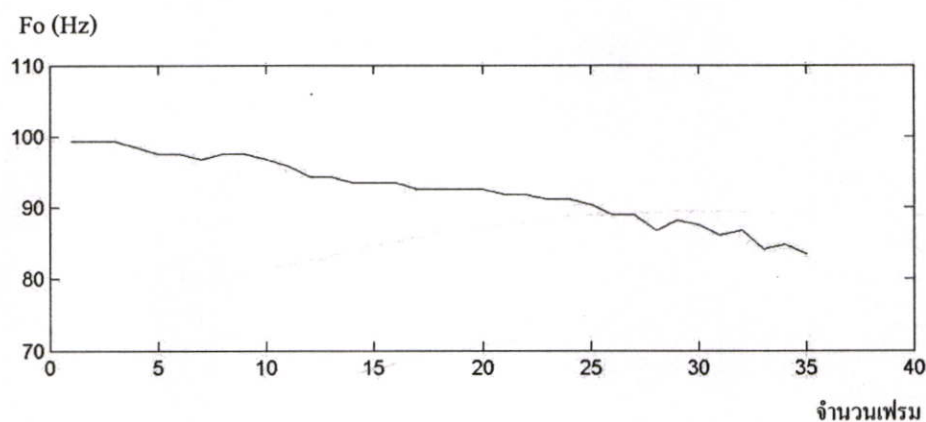
(ก) – (ค) คือเอาต์พุตของแต่ละขั้นตอน

จากรูปที่ 3.6 แสดงให้เห็นถึงขั้นตอนในการคำนวณหาพิทช์ ของสัญญาณเสียง โดยเริ่มจากสัญญาณเสียงที่ผ่านการตัดเสียง สัญญาณเสียงจะถูกแบ่งออกเป็นเฟรม เฟรมละ 300 แซมเปิ้ล การวิเคราะห์สัญญาณเสียงจะทำการวิเคราะห์ทีละเฟรม โดยแต่ละเฟรมจะถูกเลื่อน ไปทีละ 100 แซมเปิ้ล จากนั้น สัญญาณแต่ละเฟรมจำผ่านตัวกรองสัญญาณความถี่ ที่ 900 เฮิรท์ซ เพื่อลดผลที่เกิดขึ้นจากการสั่นของเส้นเสียงในลำคอ และทำการ Center Clipping เพื่อกำหนดตำแหน่งยอดของสัญญาณ โดยจะทำการตั้งค่าอยู่ที่ 80 % ของค่าสัญญาณสูงสุด จากนั้นจะทำการคำนวณหาค่าออดิโอรีเรชั่น เพื่อคำนวณหาพิทช์ในแต่ละเฟรม ความถี่พิทช์ที่จะได้จะผ่านเข้าสู่กระบวนการ Pre-processing

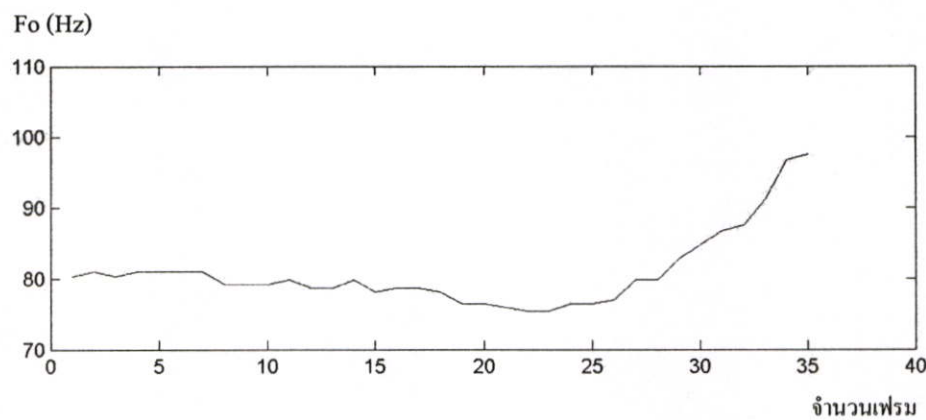
ซึ่งประกอบด้วย 2 ขั้นตอน ได้แก่ การกรองสัญญาณค่ากลาง (Median Filtering) เพื่อลดผลที่เกิดจากความไม่ต่อเนื่องของสัญญาณพิทช์อันเนื่องมาจากสัญญาณ Unvoiced ให้ค่ามีความต่อเนื่องและเรียบขึ้น ดังรูป 3.7



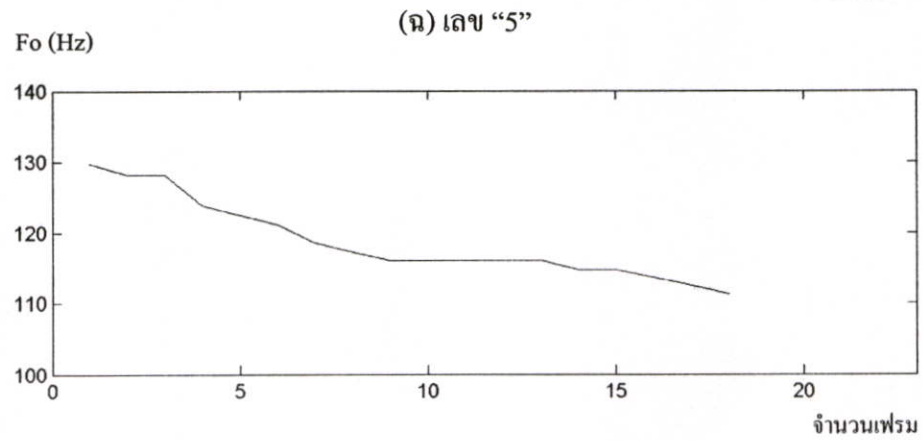
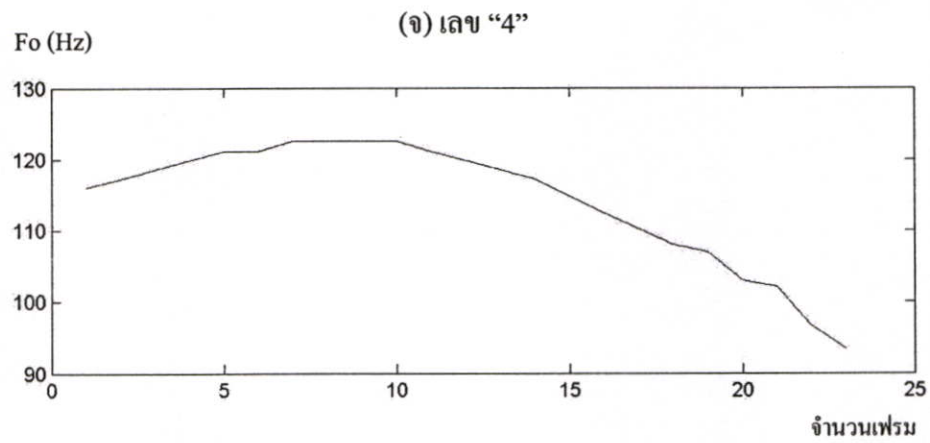
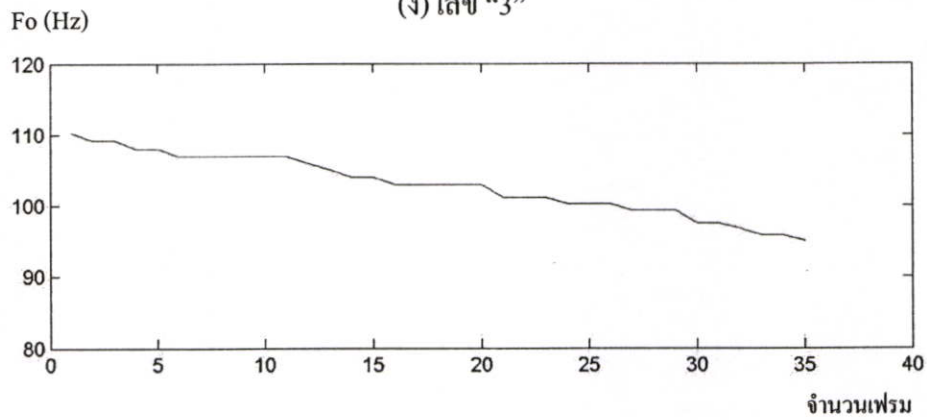
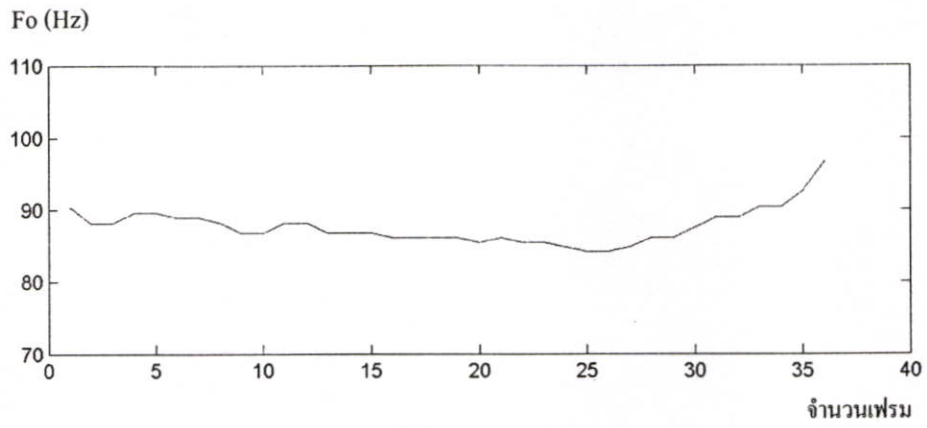
(ก) เลข "0"

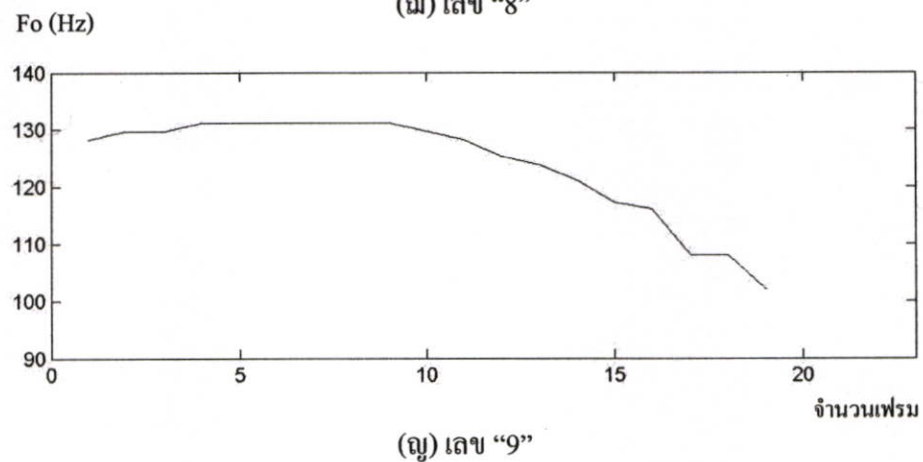
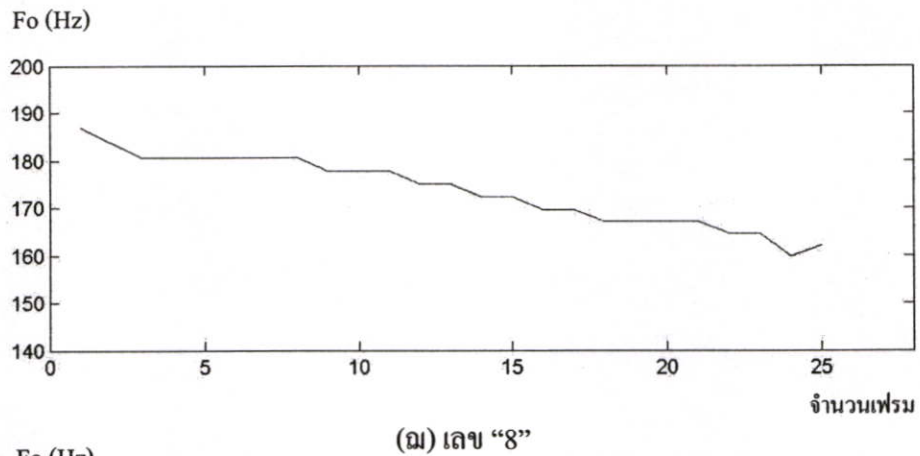
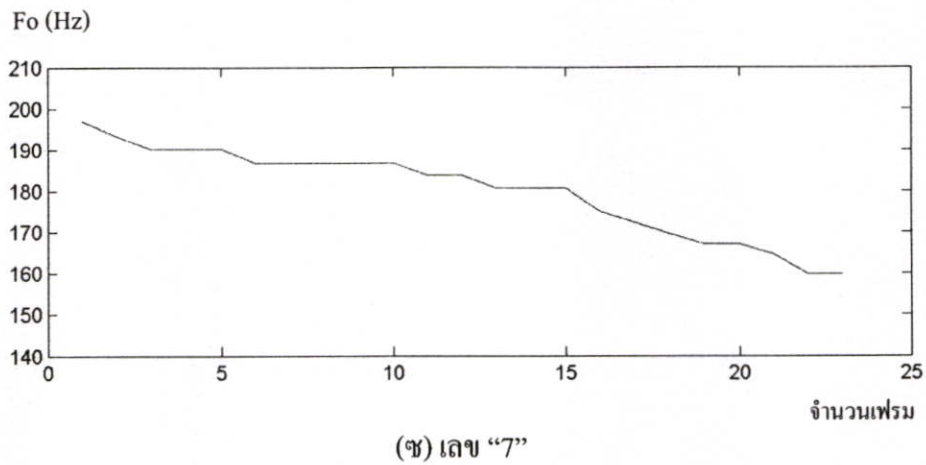


(ข) เลข "1"



(ค) เลข "2"

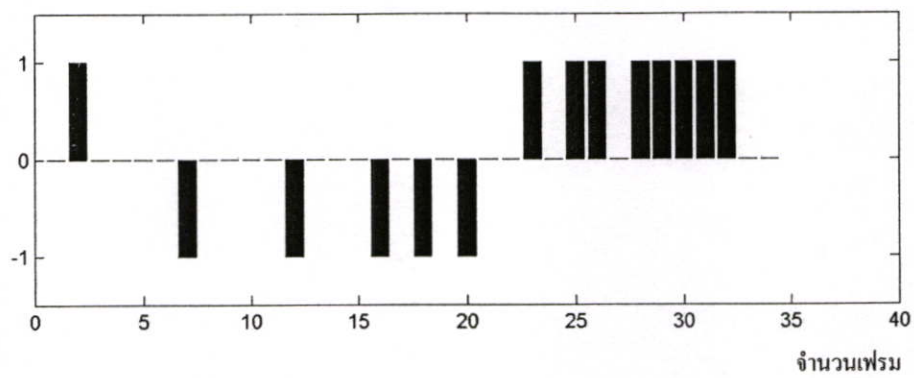




รูปที่ 3.7 ตัวอย่างพิทช์ของตัวเลขภาษาไทย (ตั้งแต่ 0 ถึง 9)

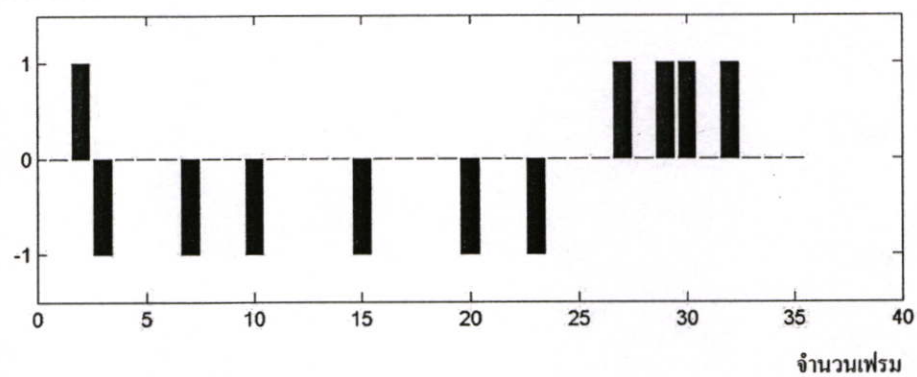


ควอนตัมพัทธ์



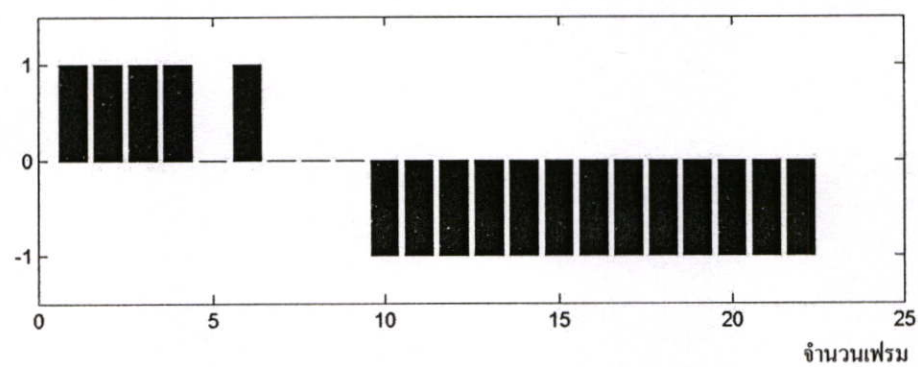
(ก) เลข "2"

ควอนตัมพัทธ์



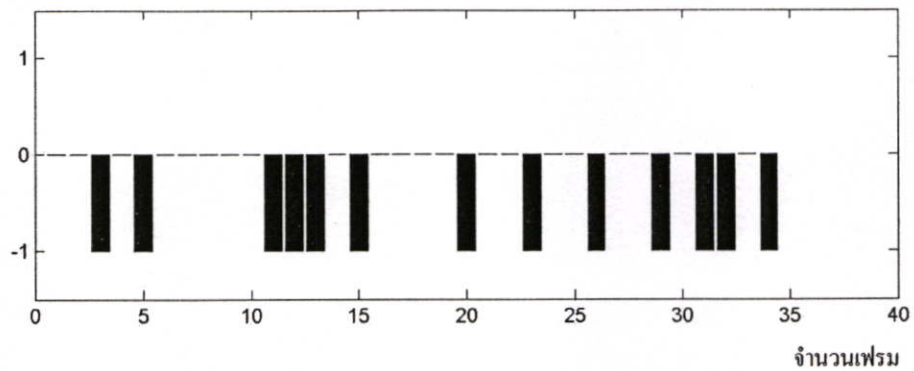
(ง) เลข "3"

ควอนตัมพัทธ์



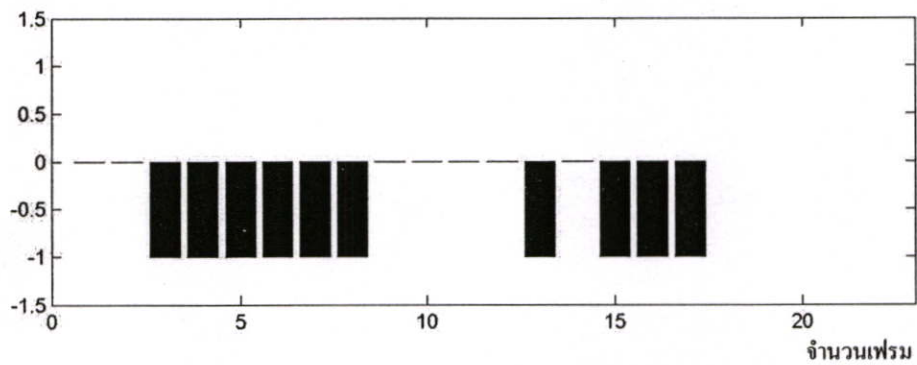
(จ) เลข "4"

ควอนตัมดิซพัทซ์



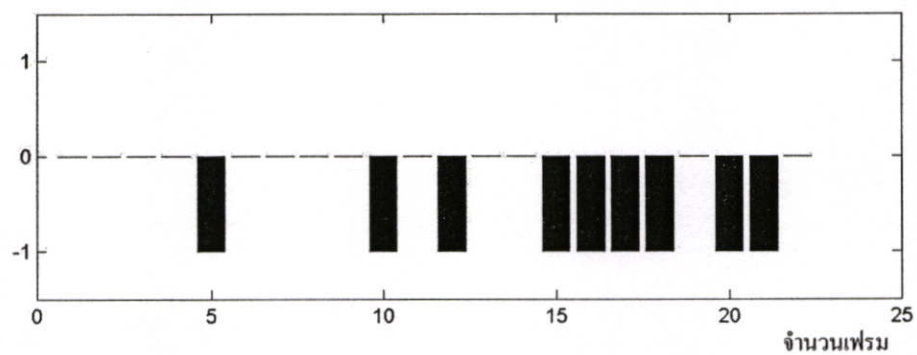
(ฉ) เลข "5"

ควอนตัมดิซพัทซ์



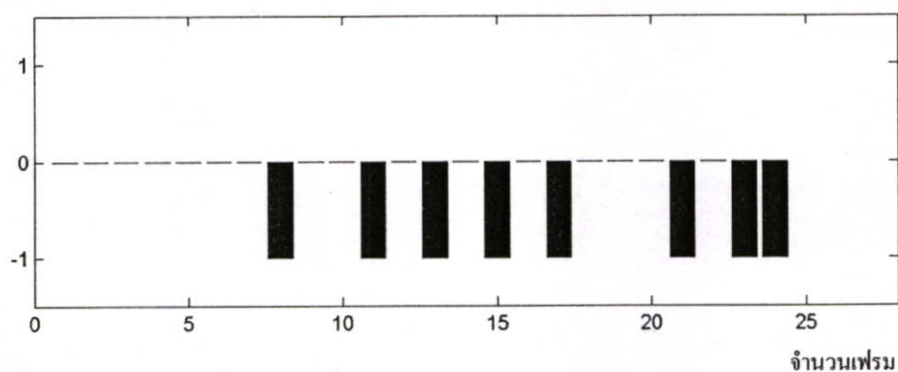
(ช) เลข "6"

ควอนตัมดิซพัทซ์



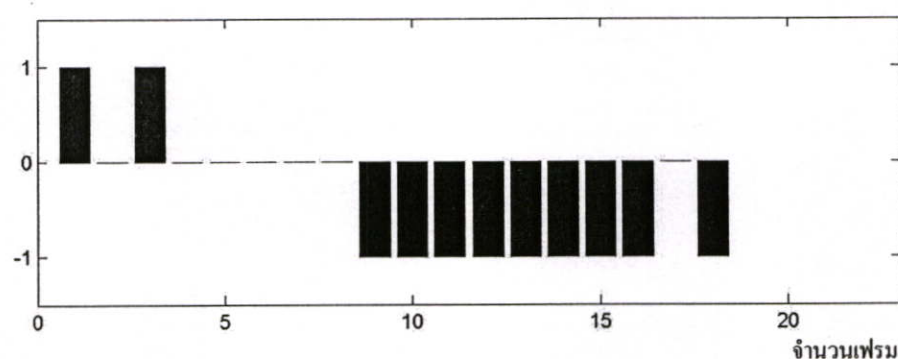
(ซ) เลข "7"

ควอนไทซ์พิทช์



(ฉ) เลข “8”

ควอนไทซ์พิทช์



(ญ) เลข “9”

รูปที่ 3.8 ตัวอย่างการควอนไทซ์พิทช์ของเสียงวรรณยุกต์ตัวเลขภาษาไทย (ตั้งแต่ 0 ถึง 9)

ค่าควอนไทซ์พิทช์ (Quantized Pitch) เป็นองค์ประกอบของเวกเตอร์ที่เราใช้เป็นตัวแทนของแต่ละเฟรมของสัญญาณเสียงวรรณยุกต์ในการสร้างแบบอ้างอิง และการรู้จำ จากรูปที่ 3.8 จะเห็นได้ว่าค่าควอนไทซ์พิทช์ ที่ได้มีความแตกต่างกันระหว่างเสียงวรรณยุกต์ที่เป็นส่วนประกอบของคำในตัวเลขภาษาไทย ซึ่งจะทำให้สามารถแยกแยะเสียงสระทั้งหมดออกจากกันได้ แต่เนื่องจากในตัวเลขภาษาไทย ประกอบด้วยวรรณยุกต์ ได้แก่วรรณยุกต์เอก, โท และวรรณยุกต์จัตวา ขณะที่ตัวเลขภาษาไทยมีคำที่มีสระซ้ำกัน ได้แก่ “3” และ “5” ซึ่งทำให้เราต้องใช้วรรณยุกต์ในการแยกแยะตัวเลขดังกล่าว เราจึงพิจารณาหน่วยเสียงวรรณยุกต์ของ “3” และ “5” ซึ่งได้แก่ วรรณยุกต์จัตวา และวรรณยุกต์โท มาสร้างโมเดลเสียงได้ทั้งหมด 2 โมเดลเสียง โดยเป็นข้อมูลเดียวกันกับที่ใช้สร้างแบบอ้างอิงเสียงสระ ซึ่งใช้เสียงพูดของผู้ชาย 3 คน และผู้หญิง 3 คน จำนวน 100 คำต่อคน เสียงพูดของผู้ชาย และผู้หญิงชุดดังกล่าวจำนวน 25 คำต่อคน จะถูกนำมาใช้ในการสร้างแบบอ้างอิง และเสียงพูดผู้พูดชุดเดิมที่เหลือ จำนวน 25 คำต่อคนจะถูกใช้สำหรับการทดสอบ

### 3.3 การสร้างแบบอ้างอิง และแบบทดสอบ

#### 3.3.1 การสร้างแบบอ้างอิง และแบบทดสอบในการแยกหน่วยเสียงสระของตัวเลข

ภาษาไทย

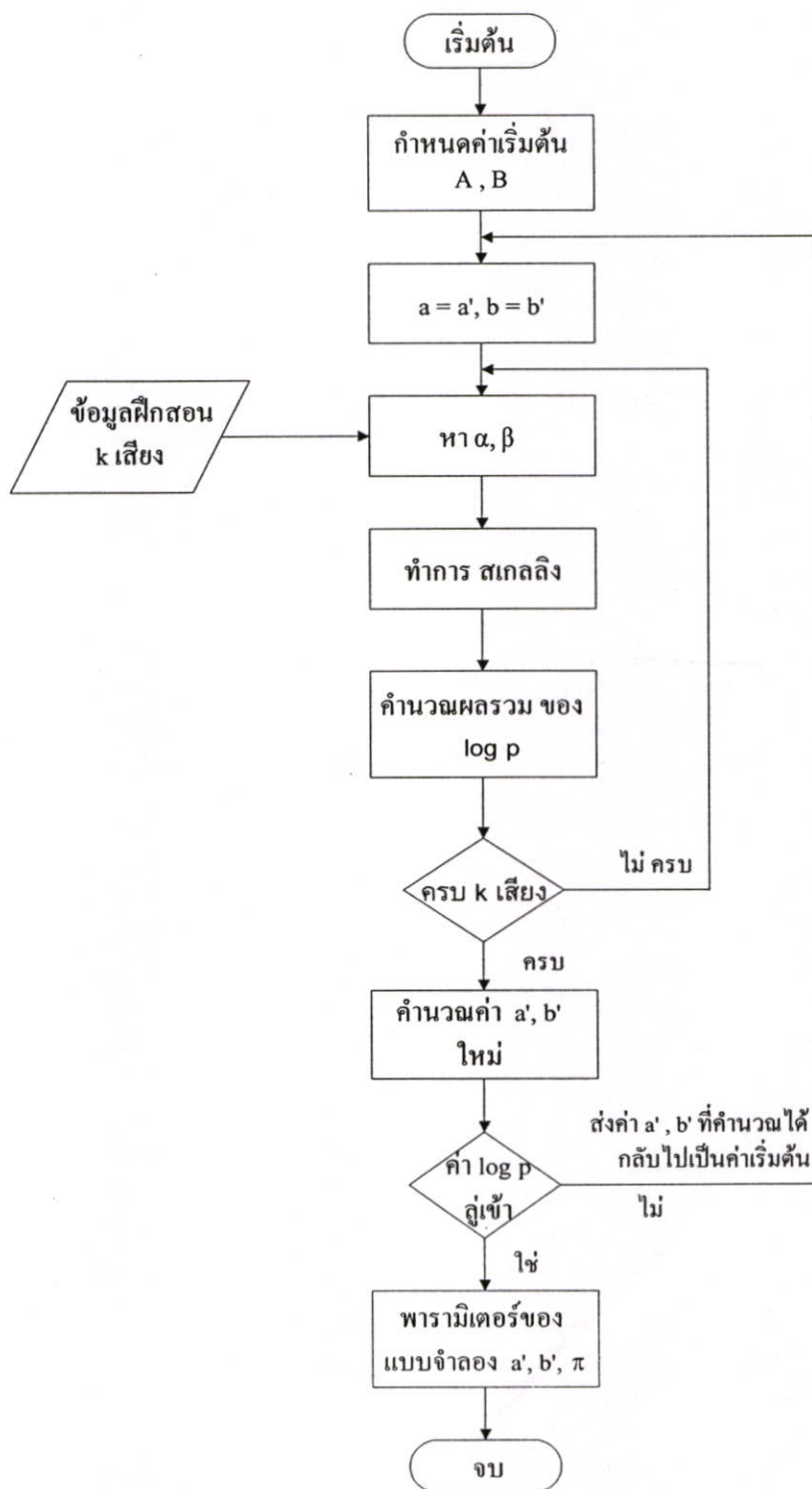
สำหรับการสร้างแบบอ้างอิง และแบบทดสอบเพื่อใช้ในการแยกหน่วยเสียงสระของตัวเลขภาษาไทยนั้นจะใช้การค้นหาแบบ K-Nearest Neighbor [8] [15] ค่า K ที่ใช้ในงานวิจัยนี้จะใช้ค่า K=5 เมื่อได้ค่าพารามิเตอร์ต่างๆ ที่ใช้ในการรู้จำเสียงสระของตัวเลขภาษาไทยแล้ว จึงนำค่าที่ได้มาสร้างแบบอ้างอิง และแบบทดสอบ โดยทำการทดลองแบบไม่ขึ้นกับผู้พูด (Speaker Independent Reference Model) ซึ่งการทดลองนี้จะใช้เสียงตัวอย่างจากผู้พูด เพศชาย 3 คน และเพศหญิง 3 คน จำนวน 3,000 เสียง โดยใช้ในการสร้างแบบอ้างอิง 1,500 เสียง และใช้ในการทดสอบ 1,500 เสียง

#### 3.3.2 การสร้างแบบอ้างอิง และแบบทดสอบในการแยกหน่วยเสียงวรรณยุกต์ของตัวเลข

ภาษาไทย

สำหรับการสร้างแบบอ้างอิง และแบบทดสอบในการแยกหน่วยเสียงวรรณยุกต์ของตัวเลขภาษาไทยนั้นจะใช้ HMM [9] [16] สำหรับการตัดสินใจ แบบ Left-right ซึ่งเป็นแบบจำลองที่เหมาะสมสำหรับการรู้จำรูปแบบของคำประเภทคำศัพท์เดี่ยว (Isolated Word) และเลือกใช้จำนวนสแตตเท่าๆกัน ทุกคำ ซึ่งจำนวนสแตตที่เลือกใช้สำหรับงานวิจัยนี้ เท่ากับ 6 โดยการสร้างโมเดลแบบอ้างอิงของเสียงวรรณยุกต์โท และวรรณยุกต์จัตวา จำนวน 2 โมเดล โดยในขั้นแรกจะต้องทำการกำหนดค่า A, B เริ่มต้น จากนั้น ทำการคำนวณหาค่า  $\alpha, \beta$  โดยการใช้การแก้ปัญหาที่ 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  $\lambda=(A',B',\pi)$  โดยการใช้การแก้ปัญหาที่ 3 ของ HMM จากนั้นทำการคำนวณซ้ำจนกว่าค่าผลรวมของ  $\log P$  ที่ได้ในแต่ละรอบ มีค่าลู่เข้า หรือไม่เปลี่ยนแปลงพารามิเตอร์ของแบบจำลอง  $\lambda'=(A',B',\pi)$  ค่าสุดท้ายจะเป็นแบบจำลองที่น่าจะทำให้เกิดลำดับของค่าปรากฏ O ที่ดีกว่า โดยรายละเอียดของขั้นตอนทฤษฎีการวิเคราะห์ต่างๆดังได้กล่าวมาแล้วก่อนหน้า ซึ่งเมื่อหาค่าพารามิเตอร์ของแต่ละระดับเสียงได้แล้วจะนำมาทำการสร้างแบบจำลองเสียงวรรณยุกต์ทั้ง 2 รูปแบบแสดงได้ดังรูปที่ 3.8 เพื่อใช้สำหรับการทดสอบ เมื่อได้ค่าพารามิเตอร์ต่างๆ ที่ใช้ในการรู้จำเสียงวรรณยุกต์ของตัวเลขภาษาไทยแล้ว จึงนำ

ค่าที่ได้มาสร้างแบบอ้างอิง และแบบทดสอบโดยทำการทดลองแบบไม่ขึ้นกับผู้พูด (Speaker Independent Reference Model) ซึ่งการทดลองนี้จะใช้เสียงตัวอย่างจากผู้พูด เพศชาย 3 คน และเพศหญิง 3 คน จำนวน 600 เสียง (เลข “3” และ “5” อย่างละ 300 เสียง) โดยใช้ในการสร้างแบบอ้างอิง 300 เสียง และใช้ในการทดสอบ 300 เสียง



รูปที่ 3.9 โฟลว์ชาร์ต การคำนวณหาค่าพารามิเตอร์ของแบบจำลองอ้างอิง

## บทที่ 4

### ผลการทดลอง

จากหลักการการรู้จำตัวเลขภาษาไทย ในบทที่ 3 การหาพารามิเตอร์ที่ใช้ในการแยกหน่วยเสียงสระ การสร้างแบบอ้างอิง และแบบทดสอบเสียงสระของตัวเลขภาษาไทย และการหาพารามิเตอร์ที่ใช้ในการแยกหน่วยเสียงวรรณยุกต์ การสร้างแบบอ้างอิง และแบบทดสอบเสียงวรรณยุกต์ของตัวเลขภาษาไทย ผ่านกระบวนการตัดสินใจ เพื่อใช้จำแนกเสียงตัวเลขภาษาไทย ดังนั้นผลการทดลองจะพิจารณาได้ตามขั้นตอน ดังนี้

#### 4.1 การทดลองแบบอ้างอิงเสียงสระของตัวเลขภาษาไทย

จากการทดลองแบบอ้างอิงเสียงสระ โดยทำการทดลองแบบไม่ขึ้นกับผู้พูด เพื่อใช้ในการแยกหน่วยเสียงสระของตัวเลขภาษาไทยนั้นจะใช้การค้นหาแบบ K-Nearest Neighborhood ที่ K=5 โดยใช้เสียงตัวอย่างจากผู้พูด เพศชาย 3 คน และเพศหญิง 3 คน จำนวน 1,500 เสียง (เสียงที่ใช้สำหรับการทดสอบ) โดยเป็นการทดสอบแสดงความถูกต้องในการรู้จำเสียงสระตัวเลขภาษาไทยแบบเฟรม และแบบเป็นคำตัวเลขภาษาไทย ดังตารางที่ 4.1 และ ตารางที่ 4.2

ตารางที่ 4.1 ผลการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด โดยทดสอบแบบเฟรม

| การทดสอบแบบเฟรม : ผู้ชาย 3 คน ผู้หญิง 3 คน ที่ KNN=5 |       |       |       |       |      |      |       |       |       |              |
|--|-------|-------|-------|-------|------|------|-------|-------|-------|--------------|
| เสียงสระ   | อี    | ออ    | อา    | อึ    | โอะ  | เอะ  | แอ    | เอา   | อุ    | %ความถูกต้อง |
| อี   | 15426 | 23    | 121   | 328   | 17   | 55   | 50    | 0     | 503   | 93.36        |
| ออ   | 7     | 19826 | 297   | 3     | 81   | 2    | 14    | 21    | 66    | 97.58        |
| อา   | 130   | 160   | 34116 | 14    | 13   | 0    | 94    | 1987  | 45    | 93.32        |
| อึ   | 117   | 20    | 22    | 18049 | 51   | 188  | 18    | 0     | 58    | 97.44        |
| โอะ  | 1     | 167   | 41    | 0     | 6742 | 0    | 60    | 0     | 297   | 92.26        |
| เอะ  | 148   | 12    | 1     | 341   | 17   | 8026 | 2     | 0     | 41    | 93.46        |
| แอ   | 47    | 0     | 156   | 11    | 0    | 81   | 14853 | 217   | 0     | 96.67        |
| เอา  | 10    | 157   | 3982  | 0     | 0    | 0    | 103   | 16260 | 1     | 79.27        |
| อุ   | 549   | 310   | 54    | 209   | 1294 | 68   | 71    | 0     | 16849 | 86.83        |
| % ค่าเฉลี่ย  |       |       |       |       |      |      |       |       |       | 92.24        |

ตารางที่ 4.2 ผลการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด โดยทดสอบแบบคำตัวเลขภาษาไทย

| การทดสอบแบบคำ : ผู้ชาย 3 คน ผู้หญิง 3 คน ที่ KNN=5 |     |     |          |     |     |     |     |     |     |              |        |
|--|-----|-----|----------|-----|-----|-----|-----|-----|-----|--------------|--------|
| ตัวเลข   | '1' | '2' | '3', '5' | '4' | '6' | '7' | '8' | '9' | '0' | %ความถูกต้อง |        |
| เสียงสระ   | อี  | ออ  | อา       | อี  | โอะ | เอะ | แอ  | เอา | อุ  |              |        |
| '1'  | อี  | 150 | 0        | 0   | 0   | 0   | 0   | 0   | 0   | 0            | 100.00 |
| '2'  | ออ  | 0   | 149      | 0   | 0   | 0   | 0   | 0   | 0   | 1            | 99.33  |
| '3'  | อา  | 0   | 0        | 135 | 0   | 0   | 0   | 0   | 15  | 0            | 90.00  |
| '4'  | อี  | 3   | 0        | 0   | 145 | 0   | 1   | 0   | 0   | 1            | 96.67  |
| '5'  | อา  | 0   | 0        | 145 | 0   | 0   | 0   | 0   | 5   | 0            | 96.67  |
| '6'  | โอะ | 0   | 0        | 0   | 0   | 136 | 0   | 0   | 0   | 14           | 90.67  |
| '7'  | เอะ | 0   | 0        | 0   | 1   | 0   | 147 | 2   | 0   | 0            | 98.00  |
| '8'  | แอ  | 0   | 0        | 0   | 0   | 1   | 0   | 149 | 0   | 0            | 99.33  |
| '9'  | เอา | 0   | 0        | 6   | 0   | 0   | 0   | 0   | 144 | 0            | 96.00  |
| '0'  | อุ  | 0   | 0        | 0   | 0   | 2   | 0   | 0   | 0   | 148          | 98.67  |
| % ค่าเฉลี่ย  |     |     |          |     |     |     |     |     |     |              | 96.53  |

จากตารางที่ 4.1 และตารางที่ 4.2 พบว่าค่าเปอร์เซ็นต์ความถูกต้องของสระจากตัวเลขภาษาไทยแบบเฟรมจะอยู่ที่ 92.24% และแบบเป็นคำตัวเลขภาษาไทยจะอยู่ที่ 96.53% ค่าความผิดพลาดที่เกิดขึ้นจากการทดสอบแบบเฟรมนั้นส่วนใหญ่จะเกิดขึ้นที่สระอุ และสระเอา ทั้งนี้สำหรับสระอุ เมื่อพิจารณาถึงความผิดพลาดที่เกิดขึ้นส่วนใหญ่จะตกอยู่ที่สระ “โอะ” ซึ่งจากบทที่ 2 เราจะพบว่าสระอุ และสระ “โอะ” เป็นกลุ่มสระเดียวกัน คือ กลุ่มสระหลัง ขณะที่ความผิดพลาดที่เกิดขึ้นกับสระ “เอา” นั้น เกิดขึ้นเนื่องจากสระ “เอา” เป็นสระอักษรที่ประกอบขึ้นจาก สระ “อะ” ร่วมกับตัวสะกด “ว” ดังนั้นความผิดพลาดที่เกิดขึ้นจึงไปตกอยู่ที่สระเอา ซึ่งเป็นสระกลุ่มเดียวกับสระ “อะ” แต่เป็นเสียงยาว เมื่อสังเกตจากค่าสเปกตรัมแอมพลิจูดและความเข้มสัญญาณในแต่ละความถี่ในกลุ่มสระที่เกิดความผิดพลาดนั้นจะพบว่าค่าใกล้เคียงกัน ส่วนค่าความผิดพลาดที่เกิดขึ้นกับสระต่างกลุ่มกันจะมีค่าน้อยกว่า

#### 4.2 การทดลองแบบอ้างอิงเสียงวรรณยุกต์ของตัวเลขภาษาไทย

จากการทดลองแบบอ้างอิงเสียงวรรณยุกต์จากตัวเลขภาษาไทย แบบที่ไม่ขึ้นกับผู้พูด เพื่อใช้ในการแยกหน่วยเสียงวรรณยุกต์ของตัวเลขภาษาไทยนั้น ผ่านกระบวนการรู้จำโดยวิธี HMM จำนวน 2 โมเดล โดยใช้เสียงตัวอย่างจากผู้พูด เพศชาย 3 คน และเพศหญิง 3 คน จำนวน 300 เสียง (เสียงที่ใช้สำหรับในการทดสอบ) สามารถแสดงผลของการทดสอบค่าเปอร์เซ็นต์ความถูกต้องตามตารางที่ 4.3 ดังนี้

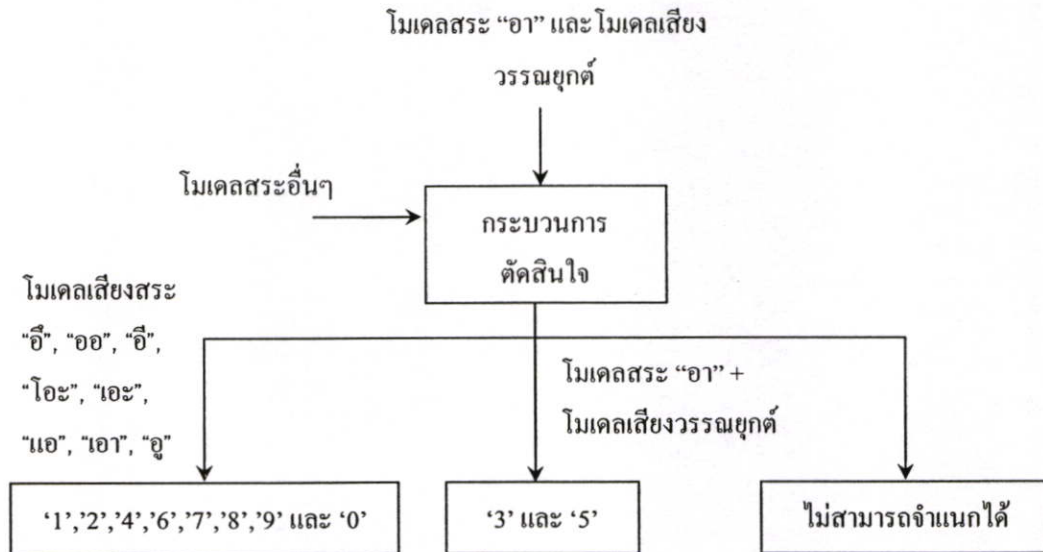
ตารางที่ 4.3 ผลการทดสอบแบบอ้างอิงเสียงวรรณยุกต์คำตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูด

| การรู้จำเสียงวรรณยุกต์ : HMM 2 โมเดล |                  |               |              |            |
|--------------------------------------|------------------|---------------|--------------|------------|
| เสียงวรรณยุกต์                       | เสียงจัตวา ('3') | เสียงโท ('5') | %ความถูกต้อง | %ค่าเฉลี่ย |
| เสียงจัตวา ('3')                     | 136              | 14            | 90.67        | 91.67      |
| เสียงโท ('5')                        | 11               | 139           | 92.67        |            |

จากตารางที่ 4.1 พบว่าค่าเปอร์เซ็นต์ความถูกต้องของวรรณยุกต์จากตัวเลขภาษาไทย ของเสียงจัตวา จะอยู่ที่ 90.67% และเสียงโท จะอยู่ที่ 92.67% ค่าเฉลี่ยความผิดพลาดอยู่ที่ 91.67% ค่าความผิดพลาดที่เกิดขึ้นจากการทดสอบจะพบว่าเป็นจากความยาวของเสียงที่แตกต่างกันมากของแต่ละบุคคลที่ใช้ในการทดสอบ และการมีเสียง Unvoiced เข้ามาปะปน ทำให้ระดับการเปลี่ยนแปลงของคาบพิทซ์เกิดความคลาดเคลื่อน

#### 4.3 การจำแนกเสียงตัวเลขภาษาไทย

เมื่อได้ค่าการทดลองแบบอ้างอิงเสียงสระของตัวเลขภาษาไทย โดยการแทนคลื่นเสียงของตัวเลขภาษาไทยด้วยเวกเตอร์คุณสมบัติขนาด 18 มิติของความเข้มแถบวิกฤติ (CBI) เพื่อจำแนกเสียงสระด้วยวิธี KNN โดยการสร้างแบบอ้างอิงเสียงสระ จำนวน 9 โมเดล และค่าการทดลองแบบอ้างอิงเสียงวรรณยุกต์ของตัวเลขภาษาไทย โดยการวิเคราะห์คาบพิทซ์เพื่อนำมาหาค่าความถี่มูลฐานเพื่อให้ได้ค่าควอนไทซ์พิทซ์ ผ่านกระบวนการรู้จำโดยใช้วิธี HMM โดยการสร้างแบบอ้างอิงเสียงสระ จำนวน 2 โมเดล แล้ว นำมาผ่านกระบวนการการตัดสินใจ รูปที่ 4.1 เพื่อจำแนกตัวเลขภาษาไทยได้ดังนี้



รูปที่ 4.1 กระบวนการตัดสินใจการรู้จำตัวเลขภาษาไทย

1. ถ้าผลการทดสอบแบบอ้างอิงการแยกหน่วยเสียงสระโดยวิธี KNN ที่ได้ อยู่ที่โมเดลเสียงสระ “อี”, “ออ”, “อิ”, “โอะ”, “เอะ”, “เออ”, “เอา” และ “อุ” ให้นำแบบทดสอบไปตัดสินใจว่าเป็นเสียงตัวเลขภาษาไทย ‘1’, ‘2’, ‘4’, ‘6’, ‘7’, ‘8’, ‘9’ และ ‘0’ ได้ตามลำดับ
2. ถ้าผลการทดสอบแบบอ้างอิงการแยกหน่วยเสียงสระโดยวิธี KNN ที่ได้ อยู่ที่โมเดลเสียงสระ “อา” ให้นำแบบทดสอบไปทดสอบกับแบบอ้างอิงการแยกหน่วยเสียงวรรณยุกต์ โดยวิธี HMM ถ้า
  - 2.1 ผลการทดสอบแบบอ้างอิงการแยกหน่วยเสียงวรรณยุกต์ โดยวิธี HMM ที่ได้ อยู่ที่โมเดลวรรณยุกต์จัตวา ให้นำแบบทดสอบไปตัดสินใจเสียงตัวเลขภาษาไทยเป็นเลข “3”
  - 2.2 ผลการทดสอบแบบอ้างอิงการแยกหน่วยเสียงวรรณยุกต์ โดยวิธี HMM ที่ได้ อยู่ที่โมเดลวรรณยุกต์โท ให้นำแบบทดสอบไปตัดสินใจเสียงตัวเลขภาษาไทยเป็นเลข “5”
3. ถ้าผลการทดสอบแบบอ้างอิงไม่เป็นไปตามข้อ 1 และข้อ 2 จะไม่สามารถจำแนกได้

เมื่อนำผลการทดลองแบบอ้างอิงการแยกหน่วยเสียงสระโดยวิธี KNN และผลการทดลองแบบอ้างอิงการแยกหน่วยเสียงวรรณยุกต์โดยวิธี HMM ที่ได้ มาผ่านกระบวนการตัดสินใจ ดังรูป 4.1 เพื่อใช้ในการตัดสินใจการรู้จำตัวเลขภาษาไทย ซึ่งสามารถแสดงผลของการทดสอบค่าเปอร์เซ็นต์ความถูกต้องตามตารางที่ 4.4 ดังนี้

ตารางที่ 4.4 ผลการทดลองแบบอ้างอิงเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลองอ้างอิงเสียงสระ และเสียงวรรณยุกต์ แบบไม่ขึ้นกับผู้พูด

| ตัวเลขภาษาไทย | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9' | '0' | %ความถูกต้อง |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------------|
| '1'           | 150 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 100.00       |
| '2'           | 0   | 149 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 99.33        |
| '3'           | 0   | 0   | 111 | 0   | 24  | 0   | 0   | 0   | 15  | 0   | 74.00        |
| '4'           | 3   | 0   | 0   | 145 | 0   | 0   | 1   | 0   | 0   | 1   | 96.67        |
| '5'           | 0   | 0   | 7   | 0   | 138 | 0   | 0   | 0   | 5   | 0   | 92.00        |
| '6'           | 0   | 0   | 0   | 0   | 0   | 136 | 0   | 0   | 0   | 14  | 90.67        |
| '7'           | 0   | 0   | 0   | 1   | 0   | 0   | 147 | 2   | 0   | 0   | 98.00        |
| '8'           | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 149 | 0   | 0   | 99.33        |
| '9'           | 0   | 0   | 6   | 0   | 0   | 0   | 0   | 0   | 144 | 0   | 96.00        |
| '0'           | 0   | 0   | 0   | 0   | 0   | 2   | 0   | 0   | 0   | 148 | 98.67        |
| % ค่าเฉลี่ย   |     |     |     |     |     |     |     |     |     |     | 94.47        |

จากตารางที่ 4.4 จะเห็นว่าค่าเปอร์เซ็นต์ความถูกต้องของเสียงตัวเลขภาษาไทย ภายหลังจากการผ่านกระบวนการการตัดสัทสัญญาณ มีค่าประมาณ 94.5% ซึ่งค่าความผิดพลาดส่วนใหญ่จะอยู่ที่เลข “3” ซึ่งเป็นเลขที่มีเสียงสระซ้ำกับเลข “5” จึงต้องนำเสียงวรรณยุกต์เข้ามาใช้ในการตัดสัทสัญญาณเพื่อจำแนกเลข “3” และ “5” ออกจากกัน ซึ่งจากการทดลองแบบอ้างอิงเสียงวรรณยุกต์ของตัวเลขภาษาไทยจะพบว่า เปอร์เซ็นต์ความถูกต้องของเสียงวรรณยุกต์โท มีค่ามากกว่าจึงทำให้เปอร์เซ็นต์ความถูกต้องของเลข “5” มีค่ามากกว่าเลข “3”

#### 4.4 การเปรียบเทียบผลการทดลองการจำแนกเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลองอ้างอิงเสียงสระ และเสียงวรรณยุกต์ กับวิธีค่าสัมประสิทธิ์เซปตรัม (MFCC)

วิธีค่าสัมประสิทธิ์เซปตรัม (Mel Frequency Cepstrum Coefficients) หรือ MFCC เป็นค่าลักษณะสำคัญที่นิยมใช้กันมากทั้งในระบบรู้จำผู้พูด และเสียงพูด ซึ่งสามารถคำนวณได้จากการแปลงโคไซน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform) ของค่าลอการิทึม (Logarithm) ของสเปกตรัมของสัญญาณเสียงแต่ส่วนย่อย สเปกตรัมของสัญญาณเสียงสามารถหาได้โดยการแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Fast Fourier Transformation) ขั้นตอนดังกล่าวอยู่บนแนวคิดที่ว่า สเปกตรัมของสัญญาณเสียงกำเนิดจาก 2 ส่วนประกอบ คือ เอนVELOPE ของสเปกตรัม (Spectrum Envelope) และโครงสร้างรายละเอียดของสเปกตรัม (Spectral Fine Structure) ทั้งสองส่วนสามารถแยกกันได้โดยการใส่ค่าลอการิทึม สัมประสิทธิ์เซปตรัมจึงเป็นการแทนสัญญาณในส่วนของเอนVELOPE ของสเปกตรัมเท่านั้น หลเมื่อทำการผ่านสเปกตรัมของสัญญาณเสียงเข้าไปในกลุ่มของตัวกรอง (Filter Bank) ซึ่งกระจายอยู่บนสเกลความถี่ตามสเกลเมล (Mel Scale) [8] ที่ออกแบบให้เหมาะสมกับการรับฟังของการได้ยินของ

มนุษย์ ค่าพลังงานของสเปกตรัมของเสียงที่ได้จากตัวกรองแต่ละตัวจะถูกนำมาคำนวณค่าสัมประสิทธิ์เซปตรัมหรือ MFCC

การทดลองการจำแนกเสียงตัวเลขภาษาไทยโดยใช้วิธีค่าสัมประสิทธิ์เซปตรัม (MFCC) จะใช้ข้อมูลเสียงชุดเดียวกันกับการจำแนกเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลองอ้างอิงเสียงสระ และเสียงวรรณยุกต์ ในขั้นตอนแรกจะทำการสร้างแบบอ้างอิงของเสียงตัวเลขภาษาไทยด้วยวิธี MFCC ขนาด 15 อันดับ [17] โดยใช้เสียงตัวอย่างจากผู้พูด เพศชาย 3 คน และเพศหญิง 3 คน จำนวน 1,500 เสียง และทำการทดลองแบบอ้างอิงเสียงตัวเลขภาษาไทยโดยทำการทดลองแบบไม่ขึ้นกับผู้พูด เพื่อใช้ในการแยกเสียงตัวเลขภาษาไทยด้วยวิธี MFCC และใช้การค้นหาแบบ K-NN ที่  $K=5$  โดยใช้เสียงตัวอย่างจากผู้พูด เพศชาย 3 คน และเพศหญิง 3 คน จำนวน 1,500 เสียง (เสียงที่ใช้สำหรับในการทดสอบ) สัญญาณเสียงตัวเลขภาษาไทยจะถูกแบ่งสัญญาณเสียงออกเป็นเฟรม ขนาด 27.2 วินาที ( $N=300$ ) และใช้ระยะในการเลื่อนเฟรมประมาณ 9 วินาที ( $M=100$ ) ผลการเปรียบเทียบผลการทดลองแบบอ้างอิงเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลองอ้างอิงเสียงสระ และเสียงวรรณยุกต์ แบบไม่ขึ้นกับผู้พูด กับวิธีค่าสัมประสิทธิ์เซปตรัม (MFCC) ดังตารางที่ 4.5

ตารางที่ 4.5 การเปรียบเทียบผลการทดลองแบบอ้างอิงเสียงตัวเลขภาษาไทยโดยใช้ผลการทดลองอ้างอิงเสียงสระ และเสียงวรรณยุกต์ แบบไม่ขึ้นกับผู้พูด กับวิธีค่าสัมประสิทธิ์เซปตรัม (MFCC)

| ตัวเลขภาษาไทย      | % ความถูกต้องของเสียงตัวเลขภาษาไทย   |                     |
|--------------------|--|---------------------|
|                    | การวิเคราะห์เสียงสระ และเสียงวรรณยุกต์ด้วยวิธี<br>ความเข้มแถบวิกฤติ และควอนไทซ์พิทช์ | MFCC<br>ที่ KNN = 5 |
| '1'                | 100.00   | 84.20               |
| '2'                | 99.33  | 97.48               |
| '3'                | 74.00  | 92.96               |
| '4'                | 96.67  | 85.67               |
| '5'                | 92.00  | 91.04               |
| '6'                | 90.67  | 61.15               |
| '7'                | 98.00  | 89.21               |
| '8'                | 99.33  | 67.71               |
| '9'                | 96.00  | 96.76               |
| '0'                | 98.67  | 97.46               |
| <b>% ค่าเฉลี่ย</b> | <b>94.47</b>   | <b>86.36</b>        |

จากตารางที่ 4.5 จะพบว่าค่าเปอร์เซ็นต์ความถูกต้องเฉลี่ยของเสียงตัวเลขภาษาไทยที่ได้จากวิธี MFCC โดยใช้การแยกแยะด้วยวิธี KNN ที่ K=5 มีค่าประมาณ 86.4% เมื่อเทียบกับด้วยวิธีการวิเคราะห์เสียงสระ และเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤติ และควอนไทซ์พิทช์ซึ่งมีค่าความถูกต้องเฉลี่ยประมาณ 94.5% ค่าความผิดพลาดส่วนใหญ่ที่เกิดขึ้นจากการรู้จำเสียงตัวเลขภาษาไทยด้วยวิธี MFCC จะอยู่ที่เลข “6” และเลข “8” ขณะที่ค่าความผิดพลาดที่เกิดขึ้นจากวิธีการวิเคราะห์เสียงสระ และเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤติ และ ควอนไทซ์พิทช์จะอยู่ที่เลข “3” ซึ่งจากผลการทดลองจะเห็นได้ว่าค่าเฉลี่ยความถูกต้องของการรู้จำเสียงตัวเลขภาษาไทยด้วยวิธีการวิเคราะห์เสียงสระ และเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤติ และควอนไทซ์พิทช์เป็นวิธีที่ให้ค่าความถูกต้องมากกว่าวิธี MFCC ถึงประมาณ 8%

## บทที่ 5

# สรุปผลงานวิจัยและข้อเสนอแนะ

### 5.1 สรุปผลการทดลอง

วิทยานิพนธ์เล่มนี้ได้ทำการศึกษาการรู้จำเสียงตัวเลขภาษาไทย โดยใช้รูปแบบของเสียงสระ (9 เสียง) โดยการแทนคลื่นเสียงของตัวเลขภาษาไทยด้วยเวกเตอร์คุณสมบัติขนาด 18 มิติของความเข้มแถบวิกฤติ (CBI) เพื่อจำแนกเสียงสระด้วยวิธี KNN และเสียงวรรณยุกต์ (2 เสียง) โดยการวิเคราะห์คาบพิทช์เพื่อนำมาหาค่าความถี่มูลฐานเพื่อให้ได้ค่าควอนไทซ์พิทช์ ผ่านกระบวนการรู้จำโดยใช้วิธี HMM ที่ประกอบขึ้นเป็นคำตัวเลขภาษาไทย และนำผลลัพธ์ของการจำแนกเสียงสระและเสียงวรรณยุกต์ที่ได้มาผ่านกระบวนการตัดสินใจ เพื่อจำแนกเสียงตัวเลขภาษาไทย ผลที่ได้จากการวิจัยสามารถสรุปได้ดังนี้

1. จากการทดสอบการรู้จำเสียงสระของตัวเลขภาษาไทย จำนวน 9 เสียง ได้แก่ เสียงสระ “อี”, “ออ”, “อิ”, “โอะ”, “เอะ”, “เอ”, “เอา” และ “อู” แบบไม่ขึ้นกับผู้พูด โดยการแทนคลื่นเสียงของตัวเลขภาษาไทยด้วยเวกเตอร์คุณสมบัติขนาด 18 มิติของความเข้มแถบวิกฤติ เพื่อจำแนกเสียงสระด้วยวิธี K-Nearest Neighbor นั้น

- สำหรับการทดสอบแบบแยกเฟรม จะให้ผลของการรู้จำของเฟรม มีค่าประมาณ 92%

- สำหรับการทดสอบแบบคำตัวเลขภาษาไทย จะให้ผลของการรู้จำของคำ มีค่าประมาณ 96%

2. จากการทดสอบการรู้จำเสียงวรรณยุกต์ของตัวเลขภาษาไทย ของเสียงวรรณยุกต์จัตวา และวรรณยุกต์โท แบบไม่ขึ้นกับผู้พูด โดยการวิเคราะห์คาบพิทช์เพื่อนำมาหาค่าความถี่มูลฐานเพื่อให้ได้ค่าควอนไทซ์พิทช์ ผ่านกระบวนการรู้จำโดยใช้วิธี HMM นั้น

- สำหรับการทดสอบแบบคำตัวเลขภาษาไทย (เลข “3” และเลข “5”) จะให้ผลของการรู้จำของคำ มีค่าประมาณ 92%

3. จากผลของการทดสอบการรู้จำเสียงสระ และการรู้จำเสียงวรรณยุกต์ของตัวเลขภาษาไทย ผ่านกระบวนการตัดสินใจเพื่อรู้จำตัวเลขภาษาไทย

- สำหรับการทดสอบแบบคำตัวเลขภาษาไทย จะให้ผลของการรู้จำของคำ มีค่าประมาณ 95%

และเมื่อนำผลการทดลองที่ได้การรู้จำเสียงตัวเลขภาษาไทยโดยการวิเคราะห์เสียงสระและเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤติ และควอนไทซ์พิทช์ไปเปรียบเทียบกับวิธี MFCC ขนาด 15 อันดับ โดยใช้การแยกแยะแบบ KNN ที่  $K=5$  โดยใช้ข้อมูลเสียงชุดเดียวกันกับการจำแนก

เสียงตัวเลขภาษาไทยด้วยวิธีการวิเคราะห์เสียงสระ และเสียงวรรณยุกต์ พบว่าการจำแนกเสียงตัวเลขภาษาไทยด้วยวิธีการวิเคราะห์เสียงสระ และเสียงวรรณยุกต์ด้วยวิธีความเข้มแถบวิกฤติ และควอนไทล์พิทช์ เป็นวิธีที่ให้ค่าความถูกต้องมากกว่าวิธี MFCC ถึง 8%

## 5.2 อภิปรายผลความผิดพลาด

จากผลการทดลองในบทที่ 4 จะเห็นว่าความผิดพลาดที่เกิดขึ้นกับการรู้จำเสียงสระนั้น ส่วนใหญ่เกิดจากกลุ่มสระเดียวกันเนื่องจากกลุ่มสระเดียวกันสระในกลุ่มเดียวกันมีลักษณะของแอลพีซีสเปกตรัมและค่าความเข้มสัญญาณในแต่ละความถี่วิกฤติที่ใกล้เคียงกันจึงทำให้การรู้จำเกิดความผิดพลาดขึ้นได้ ขณะที่ความผิดพลาดที่เกิดขึ้นกับการรู้จำเสียงวรรณยุกต์จะเกิดจากความยาวของเสียง หรือจำนวนเฟรมที่แตกต่างกันมากของคำตัวเลขภาษาไทยของแต่ละบุคคลที่ใช้ในการทดสอบ และการมีสัญญาณ Unvoiced ของเสียงพยัญชนะเข้ามาปะปน ทำให้ระดับการเปลี่ยนแปลงของคาบพิทช์มีความคลาดเคลื่อนได้ง่าย ส่วนความผิดพลาดที่เกิดจากการรู้จำตัวเลขภาษาไทยนั้น ส่วนใหญ่จะเกิดจากเลข “3” ซึ่งเป็นผลมาจากความผิดพลาดที่เกิดขึ้นจากการรู้จำเสียงวรรณยุกต์นั่นเอง

## 5.3 ข้อเสนอแนะ

1. การรู้จำเสียงสระของตัวเลขภาษาไทย โปรแกรมที่ใช้ในการคำนวณยังใช้เวลานานเนื่องจากข้อมูลที่ใช้มีจำนวนมาก
2. เสียงที่อัดจากไมโครโฟนมีสัญญาณรบกวนจากภายนอก ทำให้เสียงที่นำมาวิเคราะห์มีลักษณะผิดเพี้ยนไปบ้าง การมีขั้นตอนในการลดสัญญาณรบกวนเป็นขั้นตอนแรก ซึ่งเป็นส่วนที่สำคัญสำหรับการรู้จำเสียง จึงควรทำขั้นตอนนั้นให้สามารถได้เสียงที่มีความสมบูรณ์เพียงพอที่จะนำไปวิเคราะห์ต่อไป

## 5.4 การประยุกต์ใช้งาน

1. สามารถนำไปประยุกต์การรู้จำตัวเลขภาษาไทยแบบประโยค เช่น หมายเลขโทรศัพท์, หมายเลขบัตรประชาชน และหมายเลขบัตรเดบิตเงินเป็นต้น
2. สามารถนำไปประยุกต์ใช้งานที่พัฒนาทางด้านการควบคุมสั่งงานด้วยเสียงได้

## เอกสารอ้างอิง

- [1] ชันวา ศรีประมง. “การวิเคราะห์เสียงพูดภาษาไทยในแกนความถี่ฮาร์โมนิก”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2537.
- [2] ณิชุกร ทับทอง. “การรู้จำคำพูดภาษาไทย โดยใช้ลักษณะบ่งความต่างของหน่วยเสียง”, วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2538.
- [3] ทศเวท วีระวัฒน์. “การรู้จำเสียงคำไทยเฉพาะบุคคล” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2541.
- [4] วิไลวรรณ ขนิษฐานันท์, ภาษาและภาษาศาสตร์, พิมพ์ครั้งที่ 5. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์. 2533.
- [5] สุนทร อรอินทร์ และ อัฐ เครือฟัก. “การประมวลเสียงพูดโดยการประมาณเชิงเส้น”, วิทยานิพนธ์ สาขาวิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2538.
- [6] อภิรักษ์ ธรรมทวีธิกุล, สัทวิทยา: การวิเคราะห์ระบบเสียงในภาษา, พิมพ์ครั้งที่ 1. : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์. 2547.
- [7] กำชัย ทองหล่อ, หลักภาษาไทย, พิมพ์ครั้งที่ 2. : สำนักพิมพ์อมรการพิมพ์. 2547.
- [8] นงนุช สุขตั้งมั่น, “การรู้จำเสียงสระเสียงเดี่ยวและสระเสียงประสมในภาษาไทยโดยการใส่สเปกตรัมแอลพีซี บนสเกลบาร์ก”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2547.
- [9] จิตรลดา จารุมิศรี. “การออกแบบจำลองในการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทยโดยใช้เทคนิคการควอนไทซ์พิตซ์ และ Hidden Markov Modeling”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542.
- [10] L.R. Rabiner, B.H. Juang, **Fundamentals of Speech Recognition**, New Jersey : Prentice Hall, Inc. 1993.

- [11] ฌัญฐา จิตติวารงกุล. “ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2541.
- [12] L.R. Rabiner and R.W. Schafer, **Digital Processing of Speech Signals**, New Jersey : Prentice Hall, Inc. 1978.
- [13] E. Zwicker, H. Fastl, **Psychoacoustics: Facts and Models Second Edition**, Springer, 1999.
- [14] D.M. Howard, J. Angus, **Acoustic and Psychoacoustic**, Focal Press, 1996.
- [15] วรา คงคาวิฑูรย์. “การรู้จำเสียงสระเสียงเดี่ยวในภาษาไทยโดยใช้ทรานส์เฟอร์ฟังก์ชันบนสเกลบาร์ก”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542.
- [16] ไกรสิน ส่วงวัฒนา, อธิชัช อรุณศรีแสงชัย และ จิตรลดา จารุมิศรี. 2541. “แบบจำลองเสียงวรรณยุกต์สำหรับภาษาไทย โดยใช้เทคนิคคอนโวลูชันและ Hidden Markov Modeling”. หน้าที่ 110-118. ใน งานประชุมทางวิชาการทางวิทยาการคอมพิวเตอร์แห่งชาติ 2541. กรุงเทพฯ : คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
- [17] ชัย วุฒิวิวัฒน์ชัย, สุทัศน์ แซ่ตั้ง และวารินทร์ อัจฉริยะกุลพร. 2543. “ความก้าวหน้าของการพัฒนาระบบระบุผู้พูดภาษาไทย.” วารสารวิชาการเนคเทค, 2(7) : 24-35.

## ภาคผนวก ก.

### ผลงานวิจัยที่ได้รับการตีพิมพ์

1. ถักดี จันทร์เกษ, ไกรสิน ส่วงวัฒนา “Numeric Identification for Thai Spoken Language by Recognizing Vowels and Tones Using Critical Band Intensity and Quantized Pitch”, วิศวกรรมสารลาดกระบัง, ปีที่ 24, ฉบับที่ 1 เดือนมีนาคม 2550.



ISSN 0125-1724

# ลาดกระบัง

ลาดกระบังวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

## LADKRABANG ENGINEERING JOURNAL

Volume 24 Number 1

March 2007

|     |   |    |
|-----|---|----|
| 1.  | Estimation of Surface Tension for Polymer Blends<br><i>N. Kerdtronglang P. Piyamanoch N. Udomchapanich and S. Areevat</i>   | 1  |
| 2.  | Economical On-Site Efficiency Estimation Technique of Subway Tunnel Ventilation Fan Motor<br><i>A. Wongsupphaphol T. Pumpak and C. Chai-Uthai</i>                           | 7  |
| 3.  | Studies of Three-Phase Induction Motor Under Unbalanced Voltage Supply Conditions Using Dynamic Models<br><i>E. Kinnaras and A. Kunaokorn</i>                               | 13 |
| 4.  | A Comparison of Dynamic Models Used for A Grid-Connected Self-Excited Induction Generators<br><i>S. Jaikhan and A. Kunaokorn</i>  | 19 |
| 5.  | Analysis and Design of Static Variable AC Voltage Source<br><i>N. Pichaitin and V. Pichajarnroen</i>  | 25 |
| 6.  | Design and Analysis of An Electronic Ballast for A 250 W High Pressure Sodium Lamp Using Internal Ignition<br><i>P. Pongthai and V. Kinnaras</i>                            | 31 |
| 7.  | A Low Cost Controller for Induction Heating<br><i>Y. Chaisongkram and V. Kinnaras</i>   | 37 |
| 8.  | Genetic Algorithms Based Optimal Topology for Backpropagation Neural Network<br><i>K. Tkeaw and O. Pinnong</i>  | 43 |
| 9.  | Minimum Active-only Devices Current-Mode Universal Filter<br><i>S. Suriyethumwong P. Prommee M. Kungngern and K. Dejhan</i>   | 49 |
| 10. | A New Translinear Loops-Based Vector Normalization Circuit<br><i>S. Kinsukon and K. Dejhan</i>  | 55 |
| 11. | Numeric Identification for Thai Spoken Language by Recognizing Vowels and Tones Using Critical Band Intensity and Quantized Pitch<br><i>P. Chantraket and K. Songwetana</i> | 61 |



ISSN 0125-1724

# ลาดกระบัง

ลาดกระบังวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

## LADKRABANG ENGINEERING JOURNAL

ปีที่ 24 ฉบับที่ 1

มีนาคม 2550

|     |   |    |
|-----|---|----|
| 1.  | การประมาณค่าแรงดึงผิวของพอลิเมอร์ผสม<br><i>นิรุท บึงทองหลาง พงษ์ประภาพร ปิ่นไม่เมา ณัฐพร ชุตนะชัยภานิช สุทัศน์ อารีรัตน์</i>  | 1  |
| 2.  | วิธีการประเมินประสิทธิภาพการประเมินประสิทธิภาพของเครื่องจักรกลระบบระบายอากาศในอุโมงค์รถไฟฟ้าใต้ดิน<br><i>อรัชย์ นังสุกุลเดช สมานวิวัฒน์ อุมีภาศ ชัยวุฒิ ชัยกุลชัย</i>               | 7  |
| 3.  | การศึกษาการเคลื่อนที่ของยานพาหนะในระบบการขนส่งมวลชนภายใต้เงื่อนไขการเคลื่อนที่แบบไม่เป็นเส้นตรง<br><i>พรวิมล</i>  | 13 |
| 4.  | การเปรียบเทียบแบบจำลองทางกลศาสตร์ของระบบกำลังไฟฟ้าในระบบที่มีการควบคุมด้วยตนเองชนิดเชิงเส้น<br><i>เชษฐวัฒน์ กิเลสทัต อานันท์วัฒน์ ศุภกนก</i>  | 19 |
| 5.  | การวิเคราะห์และออกแบบแหล่งจ่ายแรงดันเชิงไม่แปรผันค่าได้แบบเบสติก<br><i>สิริวิวัฒน์ ไชยฐาน อานันท์วัฒน์ ศุภกนก</i>   | 25 |
| 6.  | การออกแบบและวิเคราะห์ระบบอิเล็กทรอนิกส์การเคลื่อนที่สำหรับมอเตอร์ขับเคลื่อนความเร็ว 250 วัตต์<br><i>พิชญ์พิชญ์ กฤษณะกมลแบบนภา</i>   | 31 |
| 7.  | ชุดควบคุมเชิงประจักษ์สำหรับระบบการให้ความร้อนด้วยวิธีอินฟราเรดแบบทางไฟฟ้า<br><i>ชยภัท ไชยสงคราม วิจิตร กิเลสทัต</i>   | 37 |
| 8.  | การใช้วิธีการทางคณิตศาสตร์ในการหาค่าที่น้อยที่สุดของฟังก์ชันค่าเหมาะสำหรับระบบควบคุมแบบเบสติก<br><i>คำความนิตย์พาศา</i>   | 43 |
| 9.  | กฎการปรับเทียบเสียง<br><i>จงการทองคำปาน พงษ์ประภาพร ปิ่นไม่เมา ณัฐพร ชุตนะชัยภานิช สุทัศน์ อารีรัตน์ อรัชย์ นังสุกุลเดช สมานวิวัฒน์ อุมีภาศ ชัยวุฒิ ชัยกุลชัย</i>                   | 49 |
| 10. | วงจรกรองความถี่แบบฟิลเตอร์เชิงอนุพันธ์<br><i>พรวิมล กิเลสทัต ณัฐพร ชุตนะชัยภานิช</i>  | 55 |
| 11. | Numeric Identification for Thai Spoken Language by Recognizing Vowels and Tones Using Critical Band Intensity and Quantized Pitch<br><i>Paikeew Chantraket and Kasin Songwetana</i> | 61 |

# Numeric Identification for Thai Spoken Language by Recognizing Vowels and Tones Using Critical Band Intensity and Quantized Pitch

Pakdee Chantraket and Kraisin Songwatana

Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.

Email: pakdee2@hotmail.com, kraisin@telecom.kmitl.ac.th

## Abstract

This paper presents a method for numeric recognition for Thai spoken language by classifying 9 vowels and 2 tones via K-Nearest Neighborhood (KNN) and Hidden Markov Modeling (HMM). Speech signals of Thai numerals are considered in 2 stages. In the first stage, speech signal is represented by frame of 18-dimension feature vector of Critical Band Intensity (CBI) and classified into Thai vowel via KNN. In the second stage, only speech signals of Thai numerals have a model as /-aa-/ is represented by quantized pitch which derived from pitch analysis and classified into tones via HMM. The output of vowel and tone, are used to classify for numerals.

## 1. Introduction

The using of Thai numerals is important in many applications such as voice-dialing telephone, automatic data entry, PIN system, etc. The words or syllables of numerals consist of consonants, tones and vowels that can be recognized by system of consonants recognition, tones recognition and vowels recognition. This paper presents a method of Thai numerals identification by recognizing vowels and tones. The Thai numerals consist of 10 numbers, 0-9, which constitute different pair of vowels and tones as shown in Table 1.

Table 1. The vowels and tones constitute Thai numerals.

| Numeric | Vowel  | Tone    |
|---------|--------|---------|
| 1       | /-v-/  | Low     |
| 2       | /-@@-/ | Rising  |
| 3       | /-aa-/ | Rising  |
| 4       | /-ii-/ | Low     |
| 5       | /-aa-/ | Falling |
| 6       | /-o-/  | Low     |
| 7       | /-e-/  | Low     |
| 8       | /-xx-/ | Low     |
| 9       | /-ay/  | Falling |
| 0       | /-uu-/ | Rising  |

The vowels recognition in Thai numeric is carried out classifying frames of transfer function of vocal tract on Bark Scale and using the 18-dimension feature vector of Critical Band Intensity to represent each frame. Each frame classified by KNN technique into vowel. As for the tones recognition in Thai numeric, the pitch or the fundamental frequency is calculated from each frame of signal using autocorrelation with center clipping method. After which, the sequence of the pitch is preprocessed using median filter to filter out unnecessary discontinuities, followed by quantization of data sequences in term of raising, decreasing or remain-the-same state in the sequence. The sequence of quantized data is classified by Hidden Markov Models into tone. The output of recognitions, vowels recognition and tones recognition, are used to classify for numeric identification.

## 2. Analysis of numeric and recognition model

From Table 1, the Thai numerals consist of 10 numbers which have 9 unique vowels and 3 unique tones. The numbers of '3' and '5' have the same vowel as /-aa-/ but their tones are different, rising and falling, respectively. Firstly, we consider the process of vowel recognition to classify the 9 unique vowels. After that, only the speech signal having vowel recognition as /-aa-/ is sent to 2 models of tone recognition via HMM to identify the tone and hence the number '3' or

'5'. On the other hand, the speech signal having vowel recognition via KNN as the other models are sent to identify the number ('1', '2', '4', '6', '7', '8', '9' and '0'). The block diagram of decision rule for recognition of Thai numerals is as followed in Figure 1.

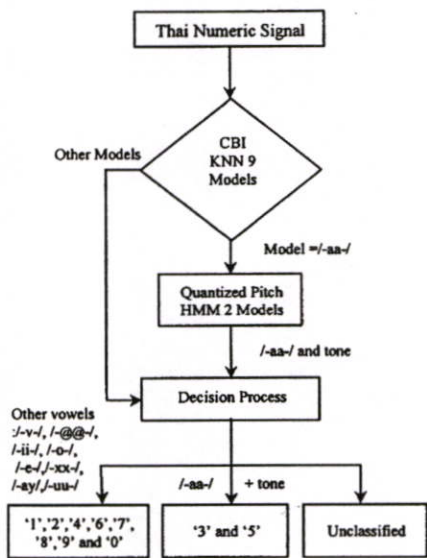


Figure 1. The decision rule for recognition of spoken numerals

2.1 Obtaining the vowel recognition from the Thai numeric signal

The vowel recognition has 3 processes. The block diagram is shown in Figure 2. [1] [2]

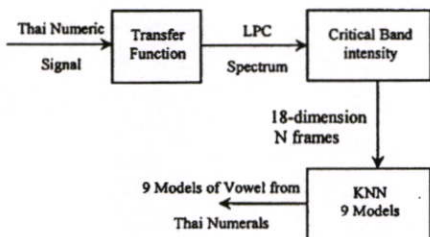


Figure 2. Step for analysis and classify the vowels of Thai numerals.

To evaluate the propose of recognition of vowels for the Thai numeric signal which are digitized at sampling rate 11.025 kHz and 8 bits quantization. Three hundred samples of signal are used to calculate the transfer function each frame of vocal tract from LPC coefficients [3] and each frame is a shifted from the previous by 100 samples to overlap the information. From each frame, an 18-dimension vector, with elements representing the critical band intensity, is then used for in the recognition process. Flowchart of this process is shown in Figure 3.

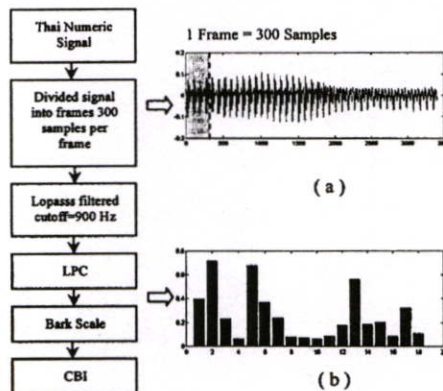


Figure 3. The flowchart for finding Critical Band Intensity (CBI): Figures (a-b) are Thai numeric signal as '1' and output of CBI.

2.2 K-Nearest Neighbor decision for vowel recognition of the frame of Thai numerals

The K-Nearest Neighbor algorithm calculates distances between set of 18-dimensional feature vector of Critical Band Intensities from N training utterance  $\vec{I}_m = [(\vec{I}_1, \vec{I}_2, \dots, \vec{I}_N)]^T$  belonging one of the 9 classes of vowels of Thai numerals, and unknown feature vector of 18 dimensions,  $I_p = [I_1, I_2, \dots, I_{18}]$ ; that is

$$|\vec{I}_p - \vec{I}_M| = \sqrt{\sum_{g=1}^K [(\vec{I}_p)_g - (\vec{I}_M)_g]^2} \quad (1)$$

when  $(\vec{I}_p)_g$  is the utterance in vector  $\vec{I}_p$ ,  $\vec{I}_p$  is the target vector, and  $(\vec{I}_M)_g$  is the nearest neighbor of  $\vec{I}_M$

K vectors with the least distance from the unknown vector are selected. The maximum likelihood decision process is then used to classify the unknown vector in the 9 models (vowels of Thai numerals)

Once of each frame of the Thai numerals word is classified into vowel by KNN method above, it is possible to classify the Thai numerals into vowel, by choosing the maximum likelihood of the vowel classified for each frame. We have performed speaker dependent experiments 1,500 words of Thai numerals (150 sets) from 3 males and 3 females subjects by using 12-order LPC and at K=5. They are used as training sets. The rest 1,500 words from the same 3 males and 3 females are used as test sets.

The vowel recognition of frames of Thai numerals results in average accuracy of 92.24% as shown in Table 2 and the vowel recognition classifying by words result in average accuracy of 96.53% is shown in Table 3.

Table 2. The vowel recognition using frames of Thai numerals as the test set.

| Frame Test : 3 males and 3 females KNN-5 |       |        |       |       |       |       |       |        |        |                 |
|--|-------|--------|-------|-------|-------|-------|-------|--------|--------|-----------------|
| Vowel                                    | l-v-l | l-00-l | l-a-l | l-i-l | l-e-l | l-o-l | l-u-l | l-ay-l | l-ua-l | %Total Accuracy |
| l-v-l                                    | 16426 | 23     | 121   | 328   | 17    | 55    | 50    | 0      | 303    | 93.36           |
| l-00-l                                   | 7     | 19826  | 297   | 3     | 81    | 2     | 14    | 21     | 66     | 97.38           |
| l-a-l                                    | 130   | 160    | 34136 | 14    | 13    | 0     | 94    | 1987   | 45     | 93.32           |
| l-i-l                                    | 117   | 20     | 22    | 18649 | 51    | 188   | 18    | 0      | 58     | 97.44           |
| l-e-l                                    | 1     | 167    | 41    | 0     | 6742  | 0     | 60    | 0      | 297    | 92.36           |
| l-o-l                                    | 148   | 12     | 1     | 341   | 17    | 8826  | 2     | 0      | 41     | 93.46           |
| l-u-l                                    | 47    | 0      | 156   | 11    | 0     | 81    | 14863 | 217    | 0      | 96.67           |
| l-ay-l                                   | 10    | 157    | 3982  | 0     | 0     | 0     | 103   | 16268  | 1      | 79.27           |
| l-ua-l                                   | 549   | 310    | 54    | 209   | 1294  | 68    | 71    | 0      | 16449  | 86.83           |
| % Average Accuracy                       |       |        |       |       |       |       |       |        |        | 92.24           |

Table 3. The vowel recognition using word of Thai numerals as the test set.

| Word Test : 3 males and 3 females KNN-5 |        |        |       |       |       |       |       |        |        |                 |
|---|--------|--------|-------|-------|-------|-------|-------|--------|--------|-----------------|
| Numerals                                | 1'     | ๒      | ๓     | ๔     | ๕     | ๖     | ๗     | ๘      | ๙      | %Total Accuracy |
| Vowel                                   | l-v-l  | l-00-l | l-a-l | l-i-l | l-e-l | l-o-l | l-u-l | l-ay-l | l-ua-l | Accuracy        |
| 1'                                      | l-v-l  | 180    | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 100.00          |
| ๒                                       | l-00-l | 0      | 149   | 0     | 0     | 0     | 0     | 0      | 0      | 99.33           |
| ๓                                       | l-a-l  | 0      | 0     | 136   | 0     | 0     | 0     | 13     | 0      | 90.00           |
| ๔                                       | l-i-l  | 3      | 0     | 0     | 146   | 0     | 1     | 0      | 0      | 96.67           |
| ๕                                       | l-e-l  | 0      | 0     | 0     | 146   | 0     | 0     | 5      | 0      | 96.67           |
| ๖                                       | l-o-l  | 0      | 0     | 0     | 0     | 136   | 0     | 0      | 14     | 90.67           |
| ๗                                       | l-u-l  | 0      | 0     | 0     | 0     | 1     | 0     | 147    | 2      | 98.00           |
| ๘                                       | l-ay-l | 0      | 0     | 0     | 0     | 0     | 1     | 149    | 0      | 99.33           |
| ๙                                       | l-ua-l | 0      | 0     | 6     | 0     | 0     | 0     | 0      | 144    | 96.00           |
| ๑                                       | l-ua-l | 0      | 0     | 0     | 0     | 2     | 0     | 0      | 148    | 98.67           |
| % Average Accuracy                      |        |        |       |       |       |       |       |        |        | 96.83           |

2.3 Obtaining the tone recognition from the Thai numeric signal

There are 4 steps in the analysis of tone as shown in Figure 4. [4]

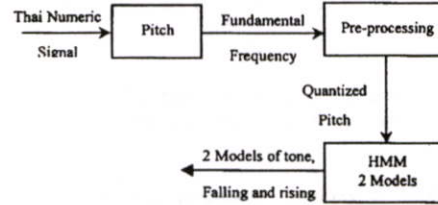


Figure 4. Steps for analysis and construction of tone recognition engine for Thai numerals

The sequence of Thai numeric signal is first segmented into frame (300 samples per frame) with a shifted from the previous by 100 samples. Each frame is first low-passed filtered to a bandwidth of 900 Hz to reduce the damping oscillation effect of the vocal tract response. The low-passed signal is then center-clipped to locate the signal's peak. The clipping level is set at a fixed percentage, 80%, of the smaller maximum absolute values between the first and the last 100 samples of the speech segment. The auto-correlation of the resultant signal is calculated to find the pitch for each frame [3]. The equation for auto-correlation is

$$R(k) = \sum_{n=0}^{N-1-k} y(n)y(n+k) \tag{3}$$

where  $y(n)$  is the nth sample of the center clipped signal,

$k$  is the auto-correlation time scale from 0 to 250,

$n$  is the sample index from 1 to N, and N is the number of sample is each frame.

From the equation is shown that the auto-correlation is a function of the periodic signal with period P. The local maximum of the auto-correlation function occurs at  $k=0, P, 2P$  and so on. The highest auto-correlation peak  $R(0)$  is located at  $k=0$  and the next highest peak  $R(P)$  is located at  $k=P$ . Hence, the fundamental frequency  $F_0$  equals to  $F_s/P$ , where  $F_s$  is the sampling frequency 11.025 kHz and P is the

quantized sequences from the same 3 males and 3 females are used as the test sets.

The tone recognition of the Thai numerals results in average accuracy of 91.67% as shown in Table 4.

| Tone          | Rising ('3') | Falling ('5') | %Accuracy | %Total Average |
|---------------|--------------|---------------|-----------|----------------|
| Rising ('3')  | 136          | 14            | 90.67     | 91.67          |
| Falling ('5') | 11           | 139           | 92.67     |                |

Table 4. The tone recognition of the Thai numerals.

According to the result of vowel recognition and the result of tone recognition, after the decision process we have the percentage of recognition accuracy for the Thai numerals when tested with the new data 1,500 words from 3 males and 3 females are used as the test sets. The recognition of Thai numerals results in average accuracy of 94.47% as shown in Table 4.

Table 4. The recognition accuracy for the Thai numerals.

| Thai Numerals Recognition |     |     |     |     |     |     |     |     |     |     |                 |
|---------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----------------|
| Numerals                  | 1'  | 2'  | 3'  | 4'  | 5'  | 6'  | 7'  | 8'  | 9'  | 0'  | %Total Accuracy |
| 1'                        | 168 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 100.00          |
| 2'                        | 0   | 149 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 99.33           |
| 3'                        | 0   | 0   | 111 | 0   | 24  | 0   | 0   | 0   | 15  | 0   | 74.00           |
| 4'                        | 3   | 0   | 0   | 148 | 0   | 0   | 1   | 0   | 0   | 1   | 96.67           |
| 5'                        | 0   | 0   | 7   | 0   | 138 | 0   | 0   | 0   | 5   | 0   | 92.00           |
| 6'                        | 0   | 0   | 0   | 0   | 0   | 136 | 0   | 0   | 0   | 14  | 90.67           |
| 7'                        | 0   | 0   | 0   | 1   | 0   | 0   | 147 | 2   | 0   | 0   | 98.00           |
| 8'                        | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 149 | 0   | 0   | 99.33           |
| 9'                        | 0   | 0   | 6   | 0   | 0   | 0   | 0   | 0   | 144 | 0   | 96.00           |
| 0'                        | 0   | 0   | 0   | 0   | 0   | 2   | 0   | 0   | 0   | 148 | 98.67           |
| %Average Accuracy         |     |     |     |     |     |     |     |     |     |     | 94.47           |

### 3. Conclusion

The method of classifying vowels and tones of Thai numeric speech using KNN technique and HMM technique can be used to recognize numbers in Thai spoken languages. The result of vowel recognition of each frames of word at 92.24% of the maximum likelihood vowels of frame and the result of tone recognition of each words of Thai numeric speech is and average of

91.67% accuracy. Hence, we can be classified into the Thai numerals with 94.47% accuracy.

### 4. Reference

- [1] K. Songwatana and N. Suktangman "Mixed and Unmixed vowels recognition in Thai spoken language by classifying frames of The Transfer Function on Bark Scale using KNN technique", vol.1, The 1<sup>st</sup> KMITL International Conference Integration of Science and Technology for Sustainable Development, pp. 251-254, 2004.
- [2] N. Suktangman, K. Khanthavivone and Kraisin Songwatana Optimizing Vowel Recognition in Thai Spoken Language Using Reduced LPC Spectrum and Reduced Feature Set of Critical Band Intensities, ISCIT 2006, pp W3A-5-W3A-9, 2006.
- [3] Lawrence Rabiner, Biing-Hwang Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993.
- [4] K. Songwatana, I. Arungsrisangchai and C. Charumit "An Implementation of Speaker independent tone Recognition Model For Thai Language Using Hidden Markov Model Technique on quantized Pitch Sequence", vol.1. ROVPIA '99, pp. 211-217, 1999.

## ประวัติผู้เขียน

|                                |  |
|--------------------------------|--|
| ชื่อ - สกุล                    | นายภักดี จันทรเกษ  |
| วัน เดือน ปีเกิด               | วันที่ 1 กันยายน พ.ศ. 2518 ที่ กรุงเทพมหานคร   |
| ที่อยู่ปัจจุบัน                | 119/886 หมู่ 5 ถนนสายไหม แขวงสายไหม เขตสายไหม<br>กรุงเทพมหานคร 10220   |
| ประวัติการศึกษา                | สำเร็จการศึกษาระดับปริญญาตรีวิศวกรรมศาสตรบัณฑิต สาขา<br>วิศวกรรมไฟฟ้า จากมหาวิทยาลัยธรรมศาสตร์ เมื่อปี พ.ศ. 2542 |
| ประสบการณ์การทำงาน<br>ปัจจุบัน | ตำแหน่งนักวิทยาศาสตร์ 6ว<br>สำนักฝนหลวงและการบินเกษตร<br>สำนักงานปลัดกระทรวงเกษตรและสหกรณ์                       |