

การรู้จำเสียงพยัญชนะต้นควบสระในภาษาไทยโดยวิธีสเปกตรัม LPC บนสเกล
บาร์กและ HIDDEN MARKOV MODELING

RECOGNITION OF CONSONANT VOWEL COMBINATION FOR THAI
SPOKEN LANGUAGE USING LPC SPECTRUM ON THE BARK SCALE AND
HIDDEN MARKOV MODELING

สุกิจ ปานแก้ว
SUKIT PANKAEW

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมโทรคมนาคม
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2550

**RECOGNITION OF CONSONANT VOWEL COMBINATION FOR THAI
SPOKEN LANGUAGE USING LPC SPECTRUM ON THE BARK SCALE AND
HIDDEN MARKOV MODELING**

SUKIT PANKAEW

**THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS 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	การรู้จำเสียงพยัญชนะต้นควบสระในภาษาไทยโดยวิธีสเปกตรัม LPC บนสเกลบาร์กและ HIDDEN MARKOV MODELING
ชื่อนักศึกษา	นายสุกิจ ป่านแก้ว
รหัสประจำตัว	45061091
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมโทรคมนาคม
พ.ศ.	2550
อาจารย์ที่ปรึกษา	รศ.ดร. ไกรสิน ส่งวัฒนา

บทคัดย่อ

งานวิจัยนี้เป็นการนำเสนอวิธีการรู้จำเสียงพยัญชนะต้นควบสระในภาษาไทย ซึ่งจำแนกเสียงพยัญชนะต้นทั้งหมด 21 เสียง โดยพิจารณารูปแบบการเปลี่ยนแปลงจากเสียงพยัญชนะต้นไปเป็นเสียงสระที่เป็นไปได้ในภาษาไทยทั้งหมด 189 รูปแบบ คลื่นเสียงในช่วงเวลาของการเปลี่ยนแปลงจากเสียงพยัญชนะต้นไปเป็นเสียงสระจะถูกแทนด้วยสเปกตรัม LPC โดยหาได้จากสัมประสิทธิ์ Linear Predictive Coding (LPC) แล้วทำการแปลงสเปกตรัม LPC ให้อยู่บนบาร์กสเกลเพื่อคำนวณความเข้มสัญญาณในแต่ละความถี่วิกฤต

จากค่าความเข้มสัญญาณในแต่ละความถี่วิกฤตที่ได้ จะนำไปผ่านกระบวนการรู้จำโดยใช้วิธี Hidden Markov Modeling เพื่อจำแนกเสียงพยัญชนะ สำหรับเสียงสระนั้นเราได้ใช้วิธีการ K-Nearest Neighbor ในการจำแนกด้วย

Thesis Title	Recognition of consonant vowel combination for Thai spoken language using LPC spectrum on the Bark scale and Hidden Markov Modeling
Student	Mr. Sukit Pankaew
Student ID.	45061091
Degree	Master of Engineering
Program	Telecommunication Engineering
Year	2007
Thesis Advisor	Assoc. Prof. Dr. Krisin Songwattana

ABSTRACT

This research presents the recognition method of consonant vowel (CV) combination for Thai spoken language for all 9 basic vowels and 21 first consonants. The main process considers 189 possible transition from consonant to vowel forms in Thai language. Speech signal during the transition period will be represented by spectrum LPC from the coefficients of Linear Predictive Coding (LPC) and calculate the critical band intensities on the Bark scale.

The critical band intensities on the Bark scale will be used in recognition process. In this research Hidden Markov Modeling is used to classify the first Thai consonant, while K-Nearest Neighbor method is used for identifying the vowels.

กิตติกรรมประกาศ

ขอขอบคุณ รศ.ดร. ไกรสิน ส่วงวัฒนา ซึ่งเป็นอาจารย์ผู้ควบคุมวิทยานิพนธ์ ที่ให้คำแนะนำ คำปรึกษา และความช่วยเหลือต่าง ๆ ที่สนับสนุนการทำวิทยานิพนธ์ฉบับนี้ จนสำเร็จได้ด้วยดี

ขอขอบคุณเพื่อนร่วมทีมวิจัยทั้ง คุณ คำ จันทร์วิวอน และ คุณนงนุช สุขตั้งมั่น ที่แนะนำ และให้ความช่วยเหลือต่าง ๆ ในการปรับปรุงงานวิจัยอย่างต่อเนื่อง

ขอขอบคุณเพื่อนร่วมงาน จากบริษัท บีเอฟเคที (ประเทศไทย) จำกัด ที่คอยให้การสนับสนุนใน ทุก ๆ ด้าน

ขอขอบคุณเพื่อน ๆ ทุกคนที่เป็นกำลังใจ และสนับสนุนตลอดมา

และสุดท้ายขอขอบคุณ พ่อ แม่ และครอบครัวทุกคน ที่เป็นแรงผลักดัน และเป็นกำลังใจสำคัญ ในการทำวิทยานิพนธ์ฉบับนี้

สุกิจ ปานแก้ว

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	2
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	2
1.3 สมมติฐานของการศึกษา.....	2
1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย.....	2
1.5 ขอบเขตการวิจัย.....	3
1.6 ขั้นตอนการศึกษา.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 กล่าวนำ.....	4
2.2 ทฤษฎีภาษาศาสตร์.....	4
2.2.1 อวัยวะที่เกี่ยวข้องกับเสียงพูด.....	4
2.2.2 การออกเสียงในภาษา (Articulatory Phonetics)	6
2.2.3 เสียงพยัญชนะและเสียงสระ.....	7
2.3 ระบบเสียงในภาษาไทย.....	8
2.3.1 เสียงพยัญชนะ.....	8
2.3.2 เสียงสระ.....	10
2.3.3 เสียงวรรณยุกต์.....	11
2.4 ลักษณะพยางค์และคำในภาษาไทย.....	11
2.5 การประมวลผลสัญญาณและกระบวนการรู้จำเสียงพูด.....	14
2.6 แบบจำลอง Linear Predictive Coding สำหรับการรู้จำเสียงพูด.....	15
2.6.1 แบบจำลอง Linear Predictive Coding.....	15
2.6.2 สมการการวิเคราะห์ LPC.....	17

สารบัญ (ต่อ)

	หน้า
2.6.3 การหาค่าสัมประสิทธิ์การประมาณ โดยวิธีอัตรสหสัมพันธ์.....	19
2.6.4 การประมวลผล LPC สำหรับการรู้จำเสียง.....	22
2.7 แถบความถี่วิกฤต และความเข้มสัญญาณในแถบความถี่วิกฤต.....	25
2.7.1 การหาความกว้างของแถบความถี่วิกฤต.....	26
2.7.2 สเกลอัตราแถบความถี่วิกฤต.....	30
2.7.3 ความเข้มแถบความถี่วิกฤต.....	32
2.8 กระบวนการตัดสินใจแบบ K-Nearest Neighbor Rule.....	33
2.9 กระบวนการตัดสินใจแบบ Hidden Markov Modeling.....	35
บทที่ 3 การรู้จำเสียงพยางค์ขณะต้นควบสระ.....	42
3.1 หลักการรู้จำเสียงพยางค์ขณะต้นควบสระ.....	42
3.2 ฐานข้อมูลเสียงพูด.....	43
3.3 เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียง.....	44
3.4 การหาค่าพารามิเตอร์สำหรับกระบวนการรู้จำ.....	44
3.5 กระบวนการรู้จำเสียงพยางค์ขณะต้นควบสระ.....	50
3.5.1 ขั้นตอนของกระบวนการรู้จำเสียงพยางค์ขณะต้นควบสระ.....	50
3.5.2 กระบวนการตัดสินใจสำหรับการรู้จำเสียงพยางค์ขณะต้นควบสระ.....	52
3.6 ผลการทดลอง.....	53
บทที่ 4 สรุปและอภิปรายผล.....	65
4.1 สรุปผลการทดลอง.....	65
4.2 อภิปรายผล และความผิดพลาด.....	65
4.3 ข้อเสนอแนะ.....	66
4.4 การประยุกต์ใช้งาน.....	66
บรรณานุกรม.....	67
ประวัติผู้เขียน.....	68

สารบัญตาราง

ตารางที่	หน้า
2.1 เสียงพยัญชนะในภาษาไทย.....	9
2.2 เสียงสระในภาษาไทย.....	10
2.3 พยางค์ในลักษณะ CV ^T C ที่เป็นไปได้ทั้งหมด.....	13
2.4 พยางค์ในลักษณะ CCV ^T C ที่เป็นไปได้ทั้งหมด.....	14
2.5 แสดงความถี่ศูนย์กลาง ความกว้าง และขอบเขตของแถบความถี่วิกฤต.....	31
3.1 เสียงพยัญชนะต้นควบสระในภาษาไทย.....	42
3.2 ผลการจำแนกเสียงสระ.....	53
3.3 ความถูกต้องของเสียงพยัญชนะต้นควบสระแต่ละกลุ่ม.....	54
3.4 ผลการจำแนกเสียงพยัญชนะ.....	55
3.5 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อะ” (/a-/).....	56
3.6 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อิ” (/i-/).....	57
3.7 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อี” (/v-/).....	58
3.8 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อุ” (/u-/).....	59
3.9 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เอะ” (/e-/).....	60
3.10 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “แอะ” (/x-/).....	61
3.11 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เออะ” (/q-/).....	62
3.12 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “โอะ” (/o-/).....	63
3.13 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เอะ” (/@-/).....	64

สารบัญรูป

รูปที่	หน้า
2.1 บริเวณต่าง ๆ ของอวัยวะในช่องปาก.....	5
2.2 เสียงวรรณยุกต์ในภาษาไทย.....	11
2.3 ขั้นตอนในกระบวนการรู้จำเสียงพูด.....	14
2.4 แบบจำลอง LPC สำหรับเสียงพูด.....	16
2.5 การสังเคราะห์เสียงพูดของมนุษย์ตามแบบจำลอง LPC.....	17
2.6 แผนภาพแสดงการประมวลผล LPC สำหรับการรู้จำเสียงพูด.....	22
2.7 การแบ่งช่วงของสัญญาณเสียง (Frame Blocking)	23
2.8 ระดับของสัญญาณทดสอบที่ถูกบดบังโดยสัญญาณรบกวนที่มีความถี่ต่าง ๆ.....	25
2.9 ความสัมพันธ์ระหว่างระดับอ้างอิง และจำนวนสัญญาณทดสอบ.....	26
2.10 ความสัมพันธ์ระหว่างระดับอ้างอิง และจำนวนสัญญาณทดสอบที่มีสัญญาณรบกวน.....	28
2.11 ระดับอ้างอิงของสัญญาณรบกวนความถี่แคบที่ถูกบดบังโดยสัญญาณทดสอบ.....	29
2.12 ระดับอ้างอิงของสัญญาณทดสอบที่ถูกบดบังโดยสัญญาณรบกวนความถี่แคบ.....	29
2.13 ความกว้างของแถบความถี่วิกฤตเมื่อเทียบกับความถี่ของสัญญาณทดสอบ.....	30
2.14 ความสัมพันธ์ระหว่างลำดับของแถบความถี่วิกฤตกับความถี่ของสัญญาณทดสอบ.....	30
2.15 ตัวอย่างของพื้นที่ Voronoi Tessellation.....	35
2.16 พารามิเตอร์ของแบบจำลอง Hidden Markov Model ที่มี 3 สถานะ.....	36
2.17 แผนภาพแสดงทิศทางการคำนวณ.....	39
3.1 ขั้นตอนในกระบวนการรู้จำเสียงพยางค์และสระ.....	43
3.2 สัญญาณเสียงและเส้นแบ่งหน่วยเสียงของประโยค “สมการที่หนึ่ง”	45
3.3 สัญญาณเสียงของพยางค์ “หนึ่ง” และส่วนที่นำไปใช้ในกระบวนการรู้จำ.....	45
3.4 ส่วนของสัญญาณเสียงที่ใช้จำแนกเสียงพยางค์และสระ.....	46
3.5 การตอบสนองของวงจรกรองสัญญาณ.....	46
3.6 ขั้นตอนของการหาค่าสเปกตรัม LPC.....	48
3.7 ความเข้มของสัญญาณในแต่ละความถี่วิกฤต.....	48
3.8 ขั้นตอนของกระบวนการรู้จำเสียงพยางค์และสระ.....	51

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การรู้จำเสียงพูด (Speech Recognition) ได้มีการพัฒนาอย่างต่อเนื่องเพื่อขยายความสามารถและความสะดวกในการติดต่อระหว่างอุปกรณ์คอมพิวเตอร์และมนุษย์ โดยทดแทนการสั่งการอุปกรณ์ดังกล่าวผ่านทางแป้นพิมพ์เพียงอย่างเดียว รวมทั้งมีการประยุกต์ใช้งานการรู้จำเสียงพูดกับอุปกรณ์สื่อสารสมัยใหม่ การรู้จำเสียงพูดที่มีการศึกษากันอย่างต่อเนื่องนี้มีอยู่ด้วยกัน 2 ลักษณะ คือการรู้จำเสียงพูดโดยขึ้นอยู่กับผู้พูด (Speaker Dependent Speech Recognition) ซึ่งผู้ใช้งานจำเป็นต้องบันทึกเสียงของตนเพื่อสร้างฐานข้อมูลเสียงสำหรับการสร้างรูปแบบเพื่อใช้เปรียบเทียบในกระบวนการรู้จำ และการรู้จำเสียงพูดโดยไม่ขึ้นอยู่กับผู้พูด (Speaker Independent Speech Recognition) ซึ่งกระบวนการรู้จำแบบนี้จะใช้ฐานข้อมูลเสียงจากผู้พูดกลุ่มหนึ่งแล้วนำไปสร้างรูปแบบอ้างอิงเพื่อใช้ในกระบวนการรู้จำเสียงของผู้พูดทั่วไป การรู้จำแบบนี้มีความยุ่งยากและซับซ้อนมากกว่าแบบแรก แต่ก็มีประโยชน์ในแง่ของการประยุกต์ใช้ที่ไม่จำเป็นต้องมีการสร้างฐานข้อมูลเสียงเฉพาะสำหรับผู้ใช้งานแต่ละคน

นอกจากการแบ่งการรู้จำเสียงพูดตามลักษณะที่ขึ้นอยู่กับผู้พูดแล้ว การรู้จำเสียงพูดยังสามารถแบ่งได้ตามรูปแบบของเสียงที่ใช้ในกระบวนการรู้จำ ซึ่งประกอบด้วย

1. การรู้จำเสียงโดยเปรียบเทียบเสียงพูดทั้งหมดซึ่งมีทั้งระบบการรู้จำคำพูดเดี่ยว (Isolated Word Recognition) และการรู้จำคำพูดต่อเนื่อง (Continuous Word Recognition) การรู้จำในลักษณะนี้จะใช้ฐานข้อมูลเสียงของคำ หรือกลุ่มคำต่อเนื่อง ในการสร้างรูปแบบอ้างอิงเพื่อใช้ในการเปรียบเทียบ ข้อดีของการรู้จำในลักษณะนี้คือสะดวก และไม่ยุ่งยากในการพัฒนาระบบ แต่รูปแบบอ้างอิงของเสียงที่ได้จะมีจำกัด และหากต้องการเพิ่มคำศัพท์ใหม่ก็จำเป็นต้องสร้างรูปแบบอ้างอิงใหม่เพื่อรองรับคำศัพท์ดังกล่าว และหากมีคำศัพท์มากเกินไปก็จำเป็นต้องใช้หน่วยประมวลผลที่มีประสิทธิภาพสูงเพื่อเปรียบเทียบเสียงพูดกับคำศัพท์ทั้งหมด ระบบนี้จึงเหมาะกับการใช้งานที่มีจำนวนคำศัพท์จำกัด

2. การรู้จำเสียงโดยการแยกหน่วยเสียง (Phonetic Recognition) แล้วพิจารณาแต่ละหน่วยเสียงเพื่อสร้างคำศัพท์จากหน่วยเสียงที่ได้ ซึ่งการรู้จำในลักษณะนี้จะต้องใช้ความรู้ทางภาษาศาสตร์มาประยุกต์ โดยจะต้องทำการศึกษารูปแบบที่เป็นไปได้ของแต่ละหน่วยเสียง แล้วสร้างกระบวนการรู้จำของแต่ละหน่วยเสียง ก่อนจะนำหน่วยเสียงเหล่านั้นมาหาคำศัพท์ที่เป็นไปได้ จะเห็นได้ว่ากระบวนการรู้จำในลักษณะนี้มีความยุ่งยากซับซ้อน และต้องพัฒนาตามรูปแบบหน่วย

เสียงของแต่ละภาษา แต่ก็มีข้อดีที่สามารถนำไปใช้ได้โดยไม่ถูกจำกัดด้วยจำนวนของคำศัพท์ สามารถใช้กับการรู้จำที่มีคำศัพท์จำนวนมากได้

สำหรับภาษาไทยนั้นประกอบด้วยหน่วยเสียง 3 หน่วยเสียง คือ เสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์ ซึ่งการรู้จำของแต่ละหน่วยเสียงก็จะมีลักษณะเฉพาะตัวของแต่ละหน่วยเสียง

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษารูปแบบของหน่วยเสียงพูดในภาษาไทย
2. เพื่อศึกษาและพัฒนาระบบการรู้จำเสียงพยัญชนะต้นควบสระของเสียงพูดภาษาไทย
3. เพื่อเป็นพื้นฐานในการพัฒนาระบบการรู้จำเสียงพูดในภาษาไทย โดยใช้การแยกหน่วยเสียง สำหรับการรู้จำเสียงพูดที่มีคำศัพท์จำนวนมาก

1.3 สมมติฐานของการศึกษา

ระบบเสียงในภาษาไทยประกอบด้วยหน่วยเสียง 3 หน่วยเสียง ได้แก่ เสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์ ซึ่งแต่ละหน่วยเสียงจะมีลักษณะที่ต่างกัน การรู้จำหน่วยเสียงแต่ละหน่วยเสียงจึงแตกต่างกันด้วย เนื่องจากเสียงพยัญชนะนั้นเกิดขึ้นในช่วงเวลาสั้น ๆ และมีจำนวนรูปแบบที่เป็นไปได้ทั้งหมดหลายรูปแบบ จึงทำให้กระบวนการรู้จำมีความยุ่งยากซับซ้อน และมีข้อผิดพลาดสูง ทำให้ไม่สามารถใช้กระบวนการรู้จำที่ใช้ค่าเวกเตอร์คุณสมบัติของแต่ละช่วงเวลามาพิจารณาเพียงอย่างเดียวได้ งานวิจัยนี้จึงทำการศึกษารูปแบบของการเปลี่ยนแปลงจากเสียงพยัญชนะไปเป็นเสียงสระ หรือเสียงพยัญชนะต้นควบสระ โดยมีรูปแบบที่เป็นไปได้ทั้งหมด 189 คู่เสียง เพื่อใช้เป็นกลุ่มของตัวอย่างสำหรับการจำแนก และนำไปใช้ในการรู้จำเสียงพยัญชนะต่อไป

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย

รูปแบบของหน่วยเสียงที่ประกอบขึ้นเป็นหน่วยเสียงในระบบภาษาไทยนั้นมีอยู่ด้วยกัน 4 รูปแบบตามลักษณะการรวมกันของหน่วยเสียงพยัญชนะ (C) หน่วยเสียงสระ (V) และหน่วยเสียงวรรณยุกต์ (T) ได้แก่ CV^T , CCV^T , CV^TC และ CCV^TC จะพบว่าสำหรับเสียงพยัญชนะต้นนั้นจะมีทั้งที่เป็นเสียงพยัญชนะเดี่ยวที่มีเสียงพยัญชนะเสียงเดียว (C) และเสียงพยัญชนะคู่ (CC) หรือเสียงควบกล้ำซึ่งประกอบด้วยเสียงพยัญชนะสองเสียง ในขณะที่เสียงพยัญชนะท้ายหรือตัวสะกดจะมีรูปแบบในลักษณะพยัญชนะเดี่ยวเท่านั้น

เสียงพยัญชนะต้นเดี่ยวนั้นมีทั้งหมด 21 เสียง ในขณะที่เสียงพยัญชนะควบกล้ำซึ่งเกิดจากเสียงพยัญชนะต้นเดี่ยว 2 เสียง มีทั้งหมด 12 เสียง ส่วนเสียงพยัญชนะท้ายนั้นมีด้วยกันทั้งหมด 9 เสียง สำหรับเสียงสระในภาษาไทยแบ่งได้เป็น 3 กลุ่ม คือ สระเสียงสั้นจำนวน 9 เสียง สระเสียง

ยาว 9 เสียง และสระเสียงผสม 6 เสียง ซึ่งทั้งสระเสียงยาว และสระเสียงผสมนั้น เกิดจากสระเสียงสั้นทั้งสิ้น เมื่อพิจารณาการเปลี่ยนแปลงของเสียงพยัญชนะไปเป็นเสียงสระในช่วงต้นของคำ เสียงพยัญชนะต้นควบสระจึงเกิดจากสระเสียงสั้นเท่านั้น ทำให้มีเสียงพยัญชนะต้นควบสระที่เป็นไปได้ทั้งหมด 189 คู่เสียง

1.5 ขอบเขตการวิจัย

งานวิจัยนี้เป็นการศึกษาการรู้จำเสียงพยัญชนะต้นควบสระในภาษาไทยในส่วนของพยัญชนะต้นเดี่ยว โดยรูปแบบในการรู้จำที่เป็นไปได้อาจเกิดจากจำนวนพยัญชนะต้นเดี่ยวทั้งหมด 21 เสียง กับสระเสียงสั้นจำนวน 9 เสียง ดังนั้นรูปแบบในการรู้จำทั้งหมดจึงเท่ากับ 189 รูปแบบ

1.6 ขั้นตอนการศึกษา

1. ศึกษากระบวนการรู้จำเสียงพูดที่ได้มีการศึกษามาแล้ว
2. ศึกษาทฤษฎีทางภาษาศาสตร์ และสัทศาสตร์ของภาษาไทย
3. ศึกษากระบวนการและอัลกอริทึมที่จำเป็นสำหรับการวิเคราะห์ และการรู้จำเสียงพูด
4. เตรียมฐานข้อมูลเสียงเพื่อใช้ในกระบวนการรู้จำเสียงพูด
5. วิเคราะห์ และทดสอบกระบวนการรู้จำเสียงพูดที่พัฒนาขึ้น
6. สรุปผลการวิจัย และเสนอแนวทางในการวิจัยต่อไป

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 กล่าวนำ

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยต่าง ๆ ที่เกี่ยวข้องกับงานวิจัย ซึ่งประกอบไปด้วย ส่วนของทฤษฎีทางภาษาศาสตร์ ได้แก่ ลักษณะเสียงพูดของภาษาไทย หน่วยเสียงของภาษาไทย และการพิจารณาเสียงพูดจากรูปสัญญาณคลื่นเสียง ซึ่งได้แก่ การแทนเสียงพูดด้วยสเปกตรัม LPC บนสเกลบาร์ก และกระบวนการรู้จำเสียงพูด

2.2 ทฤษฎีภาษาศาสตร์

ทฤษฎีภาษาศาสตร์ที่เกี่ยวข้องกับการรู้จำเสียงพูดนั้น คือการศึกษาเรื่องเสียงในภาษา โดยพิจารณาลักษณะการทำงานของอวัยวะต่าง ๆ ที่ใช้ในการเปล่งเสียง เพื่อใช้ในการจำแนกความแตกต่างของเสียงที่เกิดจากการเปล่งเสียงในแต่ละลักษณะ

2.2.1 อวัยวะที่เกี่ยวข้องกับเสียงพูด

อวัยวะที่เกี่ยวข้องกับการออกเสียงนั้นแบ่งได้เป็น 3 กลุ่ม คือ ปอดและหลอดลม กล่องเสียง และเส้นเสียง และอวัยวะที่อยู่เหนือเส้นเสียง ซึ่งอวัยวะแต่ละกลุ่มมีการทำงาน ดังนี้

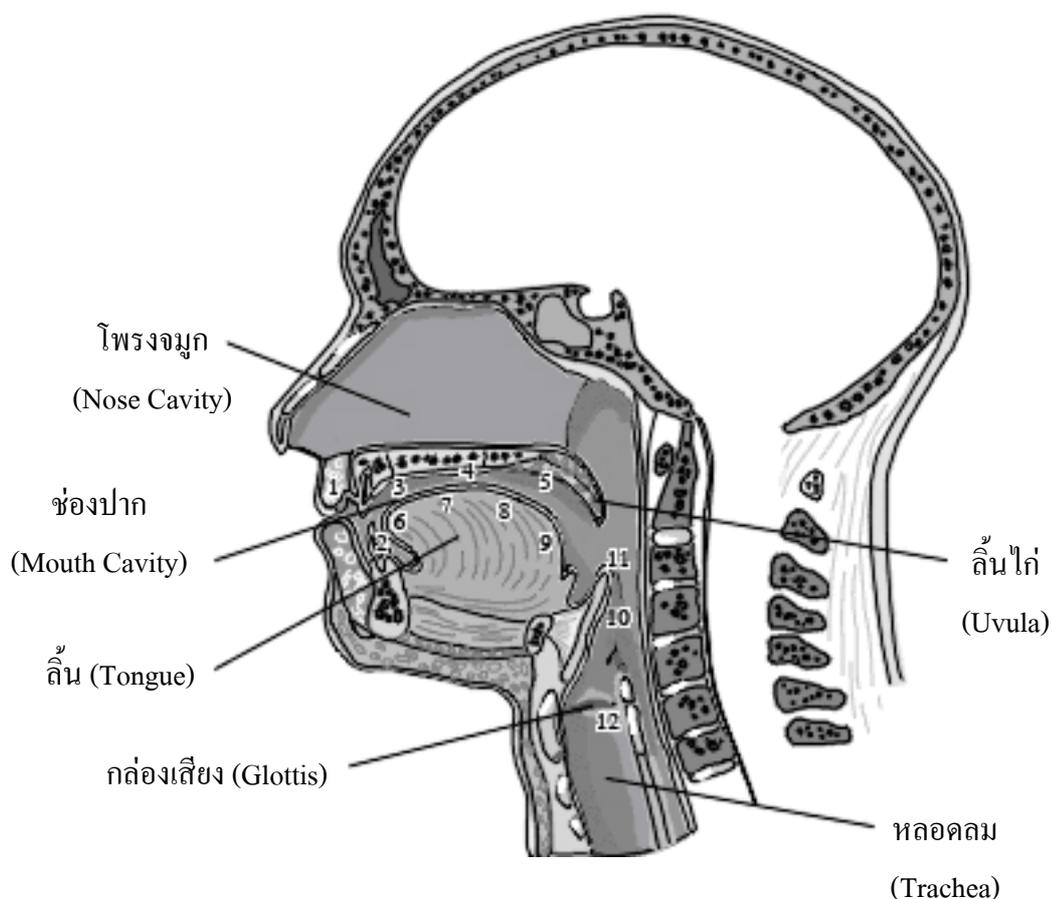
ปอดและหลอดลม เป็นอวัยวะเริ่มต้นของการออกเสียง โดยอากาศที่ใช้ในการออกเสียงเกิดจากการลดตัวลงของกระบังลมซึ่งเป็นแผ่นกล้ามเนื้อรูปโค้งได้ปอด เมื่อกระบังลมลดตัวลงจะทำให้ขนาดในช่องอกเพิ่มขึ้น อากาศจะไหลจากภายนอกร่างกายเข้าไปในปอดเป็นลมหายใจเข้า เมื่อสุดลมหายใจเข้าแล้ว กระบังลมจะโค้งตัวขึ้นทำให้ขนาดของช่องอกลดลง ทำให้ความดันอากาศภายในช่องอกสูงกว่าความดันภายนอก อากาศก็จะไหลจากปอดกลายเป็นลมหายใจออก ลมหายใจออกนี้จะผ่านจากปอดทางหลอดลม ไปยังกล่องเสียงและอวัยวะที่อยู่เหนือกล่องเสียง ซึ่งอวัยวะเหล่านี้จะทำหน้าที่ดัดแปลงเสียงที่ผ่านเส้นเสียงเป็นเสียงพูดต่าง ๆ

กล่องเสียงและเส้นเสียง กล่องเสียงประกอบด้วย กระดูก กระดูกอ่อน และกล้ามเนื้อหลายชิ้น ส่วนประกอบที่สำคัญที่สุดของกล่องเสียง คือ เส้นเสียง ซึ่งมีอยู่ 2 เส้น มีลักษณะเป็นแผ่นประกอบด้วยเส้นเอ็นและกล้ามเนื้อ และพาดตัวตามแนวนอนอยู่ตรงกลางของกล่องเสียง เส้นเสียงจะทำหน้าที่ปิดกั้นไม่ให้อาหารพลัดลงไปหลอดลม และเปลี่ยนอากาศจากปอดให้เป็นคลื่นเสียง

อวัยวะที่อยู่เหนือเส้นเสียง อากาศจากปอดที่ผ่านเส้นเสียงจะผ่านช่องคอ และออกสู่ภายนอกทางช่องปากหรือช่องจมูก อวัยวะที่อยู่เหนือเส้นเสียงสามารถแบ่งออกได้เป็น 2 กลุ่ม คือ

อวัยวะที่ไม่เคลื่อนที่ในขณะที่ออกเสียง (Passive articulator) หรือที่เรียกว่า ฐานที่เกิดเสียง ซึ่งอยู่ทางด้านบนของช่องปาก ได้แก่ ริมฝีปากบน ฟันบน ปุ่มเหงือก เพดานแข็ง เพดานอ่อน และ ลิ้นไก่

อวัยวะที่เคลื่อนที่ขณะออกเสียง (Active Articulator) หรือที่เรียกว่า กรณ ซึ่งอยู่ทางด้านล่างของช่องปาก ได้แก่ ริมฝีปากล่าง ฟันล่าง ปลายลิ้น ลิ้นส่วนปลาย ลิ้นส่วนหน้า และลิ้นส่วนหลัง



รูปที่ 2.1 บริเวณต่างๆ ของอวัยวะในช่องปาก

- | | |
|--------------------------------------|--|
| 1. ริมฝีปาก (Lips) | 7. ลิ้นส่วนปลาย (The lamina of the tongue) |
| 2. ฟัน (Teeth) | 8. ลิ้นส่วนหน้า (The front of the tongue) |
| 3. ปุ่มเหงือก (Alveolar ridge) | 9. ลิ้นส่วนหลัง (The back of the tongue) |
| 4. เพดานแข็ง (Hard palate) | 10. แผ่นเนื้อปากหลอดลม (Epiglottis) |
| 5. เพดานอ่อน (Soft palate) | 11. กรวยคอ (Pharynx) |
| 6. ปลายลิ้น (The apex of the tongue) | 12. เส้นเสียง (Vocal cords) |

เสียงที่เกิดจากอวัยวะต่าง ๆ เหล่านี้ มี 3 ลักษณะ คือ

1. หากเพดานอ่อนและลิ้นไก่สามารถเคลื่อนไปติดกับผนังคอ ซึ่งทำให้อากาศจากปอดผ่านไปทางช่องจมูกไม่ได้ เสียงที่ได้จะเป็นเสียงที่เกิดจากช่องปากเท่านั้น
2. หากเพดานอ่อนและลิ้นไก่อยู่ในตำแหน่งปกติ และมีการปิดกั้นอากาศในช่องปาก ทำให้อากาศจากปอดจะผ่านออกทางช่องจมูกเท่านั้น เสียงที่เกิดขึ้นจะเป็นเสียงนาสิก (Nasal Sound)
3. หากเพดานอ่อนและลิ้นไก่อยู่ในตำแหน่งปกติ และไม่มีการปิดกั้นอากาศในช่องปาก อากาศจะผ่านออกได้ทั้งทางช่องปาก และช่องจมูก เสียงที่เกิดขึ้นจะเป็นเสียงขึ้นจมูก (Nasalized Sound)

2.2.2 การออกเสียงในภาษา (Articulatory Phonetics)

ในการศึกษาการออกเสียงในภาษานั้น จะต้องพิจารณาวิธีการออกเสียง และอวัยวะต่าง ๆ ที่ใช้ในการออกเสียง ซึ่งสามารถพิจารณาได้ 4 ลักษณะดังนี้

1. การทำงานของเส้นเสียง ลักษณะการทำงานของเส้นเสียงที่เกี่ยวข้องกับการออกเสียงในภาษา มี 2 ลักษณะ คือ

- เสียงโฆษะ หรือ เสียงก้อง (Voiced sound) เป็นเสียงที่เกิดจากเส้นเสียงวางตัวชิดกัน แต่ประกบกันไม่สนิท อากาศจากปอดที่ผ่านเส้นเสียง จะทำให้เส้นเสียงสั่นเกิดเป็นเสียงก้อง
- เสียงอโฆษะ หรือ เสียงไม่ก้อง (Unvoiced sound) เกิดจากเส้นเสียงวางตัวอยู่ห่างกัน อากาศจึงผ่านเส้นเสียงได้อย่างอิสระและไม่ทำให้เส้นเสียงเคลื่อนไหว

2. ฐานกรณ์ของเสียง (Place of articulation) คือบริเวณที่มีการตัดแปลงอากาศจากปอดเป็นเสียงต่าง ๆ ซึ่งเสียงในภาษาไทยนั้นเกิดจากฐานที่เกิดเสียง ดังนี้

- ฐานริมฝีปาก (Bilabial)
- ฐานริมฝีปากและฟัน (Labiodental)
- ฐานปุ่มเหงือก (Alveolar)
- ฐานเพดานแข็ง (Palatal)
- ฐานเพดานอ่อน (Velar)
- ฐานเส้นเสียง (Glottal)

3. ประเภทของเสียง (Manner of Articulation) แม้ว่าเสียงในภาษาที่เกิดจากฐานกรณ์เดียวกัน ก็สามารถสร้างเสียงที่แตกต่างกันได้ โดยใช้ลักษณะการออกเสียงที่ต่างกัน ซึ่งเสียงในภาษาไทยมีลักษณะการออกเสียง ดังนี้

- เสียงหยุด (Stop) เป็นเสียงที่เมื่อผ่านกล่องเสียงเข้ามาในช่องปากแล้วจะถูกกั้นที่ฐานกรณ์แห่งใดแห่งหนึ่ง เป็นช่วงระยะเวลาสั้น ๆ แล้วฐานกรณ์นั้นจะเปิดออก อากาศก็จะถูกปล่อยออกมา อากาศที่โดนกั้นแล้วถูกปล่อยออกมาจะเหมือนการระเบิด บางคนจึงเรียกเสียงแบบนี้

ว่า เสียงระเบิด (Plosive) เสียงหยุดยังสามารถแบ่งย่อยเป็น 2 ลักษณะคือ เสียงชนิด หรือ เสียงหยุดที่มีลมพ่นออกมา (Aspirated Sound) และเสียงสติด หรือเสียงหยุดที่ไม่มีลม (Unaspirated Sound)

- เสียงแทรก (Fricative) เป็นเสียงที่ฐานกรณ์ในช่องปากปิดกั้นอากาศไว้ไม่สนิทเหมือนเสียงหยุด ยังมีอากาศเล็ดลอดออกมาได้ ทำให้เกิดเสียงในขณะที่แทรกออกมา

- เสียงหยุดแทรก (Affricate) เป็นเสียงในช่องปากที่มีลักษณะเหมือนกับเสียงหยุดและตามด้วยเสียงแทรก

- เสียงนาสิก (Nasal) เป็นเสียงที่เมื่ออากาศผ่านกล่องเสียงแล้ว จะเข้าสู่ช่องจมูก โดยที่ช่องปากมีฐานกรณ์ปิดกั้นไว้ไม่ให้อากาศออกทางช่องปาก

- เสียงลอดข้าง (Lateral) เป็นเสียงที่อากาศในช่องปากออกสู่ภายนอกปากโดยผ่านข้าง ๆ ลิ้น และไม่ผ่านส่วนตรงกลางของลิ้น

- เสียงลิ้นรัว (Trill) เป็นเสียงที่เมื่ออากาศเข้ามาอยู่ในช่องปากแล้ว มีการกระดกปลายลิ้นรัวเพดานหลาย ๆ ครั้ง

- เสียงเลื่อน (Glide) เป็นเสียงที่ตำแหน่งของลิ้น เมื่อเริ่มต้นเสียงต่างไปจากตำแหน่งในตอนท้ายเสียง เสียงประเภทนี้บางครั้งเรียกว่า เสียงกึ่งสระ (Semivowel) เพราะตำแหน่งลิ้นและปากคล้ายสระ

4. ทิศทางเดินของอากาศ การพิจารณาการออกเสียงในภาษาจากทิศทางเดินของอากาศ แบ่งได้ 2 ลักษณะ คือ

- เสียงจากทิศทางลมออก (Egressive) คือเสียงทั่วไปในภาษา โดยลมที่ออกจากปอดจะเคลื่อนที่ผ่านอวัยวะออกเสียงต่าง ๆ ออกมาทางช่องปาก หรือช่องจมูก

- เสียงทางทิศทางลมเข้า (Ingressive) หรือเสียงที่ดูดลมจากภายนอกเข้าไปในปาก มีใช้บ้างแต่พบไม่มากนัก เช่น การสูดปากแสดงความเจ็บปวด

2.2.3 เสียงพยัญชนะและเสียงสระ

ทั้งเสียงพยัญชนะและเสียงสระล้วนมีจุดกำเนิดจากปอดที่ทำหน้าที่ขับอากาศให้ผ่านเส้นเสียงที่อยู่ด้านบนของกล่องเสียง อากาศที่ผ่านเส้นเสียงจะทำให้เกิดเสียงโห่หรือโห่ตามลักษณะตำแหน่งของเส้นเสียงทั้งสอง เมื่อเสียงผ่านออกมาก็จะมีจุดกั้นหรือฐานกรณ์ต่าง ๆ ในปากกั้นอากาศที่จะผ่านออกมาจากปาก และลักษณะของเสียงต่าง ๆ จะขึ้นอยู่กับลักษณะอากาศที่ออกจากปาก เสียงที่ถูกกั้นในช่องปากเช่นนี้เป็นลักษณะของเสียงพยัญชนะ แต่ถ้าเป็นเสียงสระ จะเป็นเสียงที่ไม่มีอะไรปิดกั้นทางเดินของอากาศในช่องปากเลย

สำหรับการอธิบายเสียงพยัญชนะนั้นจะต้องกล่าวถึง สภาพของเส้นเสียงที่ทำให้เกิดเสียงโห่หรือโห่ ฐานกรณ์ของเสียง และประเภทของเสียง ในขณะที่การอธิบายเสียงสระซึ่งไม่มีจุดปิดกั้นเสียงในช่องปาก จะใช้ลักษณะของช่องคอและช่องปากที่มีรูปร่างต่างกัน สิ่งที่ทำให้ช่อง

คอและช่องปากมีลักษณะต่างกัน นั่นคือ ลิ้นและริมฝีปาก ดังนั้นการอธิบายเสียงสระจึงต้องกล่าวถึงระดับลิ้นในปาก ส่วนของลิ้นที่ใช้ในการออกเสียง และลักษณะของริมฝีปาก

2.3 ระบบเสียงในภาษาไทย

ระบบเสียงในภาษาไทยนั้น ประกอบด้วย หน่วยเสียง (Phoneme) หรือเสียงที่นับว่าเป็นเสียงที่สำคัญใช้สื่อความหมายหรือทำให้คำต่างกัน 3 ประเภท ได้แก่ หน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์

2.3.1 เสียงพยัญชนะ

ตัวพยัญชนะไทย 44 ตัว ถ้าแบ่งตามฐานกรณ์ และประเภทของเสียงจะลดเหลือ 21 เสียง ดังตารางที่ 2.1 ซึ่งตัวพยัญชนะในภาษาไทยอาจเทียบกับตัวหน่วยเสียงได้ดังนี้

/k-/	ก	/ph-/	พ ภ ผ
/kh-/	ข ค ฃ	/f-/	ฝ ฟ
/ng-/	ง หง-	/m-/	ม หม-
/c-/	จ	/y-/	ย ญ หย- หญ- อย-
/ch-/	ช ฉ ฌ	/r-/	ร รร-
/d-/	ด ฎ	/l-/	ล พ ลล-
/t-/	ต ฏ	/w- /	ว หว-
/th-/	ท ฑ ฒ ฑ ฏ ฐ	/s-/	ส ศ ษ ซ ทร-
/n-/	น ณ หน-	/h-/	ห ฮ
/b-/	บ	/z-/	อ
/p-/	ป		

หน่วยเสียงทั้ง 21 เสียงสามารถใช้เป็นเสียงพยัญชนะต้นของคำได้ แต่จะมีหน่วยเสียง 9 เสียงที่สามารถใช้เป็นเสียงพยัญชนะท้ายหรือตัวสะกด ได้แก่

/-z/	อ - (ไม่มีตัวสะกด)	/-m/	ม หม
/-b/	บ ป พ ภ ฟ	/-n/	ณ น ล ร ญ พ
/-d/	ด ฎ ต ฏ ส ศ ษ ซ จ ช ฌ ท ฑ ฏ ฐ ฒ	/-ng/	ง
/-k/	ก ข ค ฃ	/-w/	ว
		/-y/	ย

นอกจากนี้เสียงพยัญชนะ 21 ตัวอาจควบกล้ำกันเป็นเสียงต่าง ๆ ได้ โดยในภาษาไทยมาตรฐานเสียงพยัญชนะที่เกิดได้ในตำแหน่งที่สองมี 3 เสียง คือ [-r-], [-l-] และ [-w-] ส่วนเสียงที่เกิดในตำแหน่งแรกของเสียงควบกล้ำมีเฉพาะเสียงประเภทเสียงหยุดอโฆษะ (Unvoiced stop)

เท่านั้น เสียงควบกล้ำจึงมี 12 เสียง ซึ่งเกิดในตำแหน่งต้นคำเท่านั้น ไม่เกิดเป็นตัวสะกดเลย เสียงพยัญชนะควบกล้ำมีดังนี้

/pr-/	ปร	/phr-/	พร พร
/pl-/	ปล	/phl-/	พล พล
/tr-/	ตร	/thr-/	ทร
/kr-/	กร	/khr-/	กร ขร
/kl-/	กล	/khl-/	ขล กล
/kw-/	กว	/khw-/	ขว กว

ตารางที่ 2.1 เสียงพยัญชนะในภาษาไทย

	ฐานริมฝีปาก (Bilabial)	ฐานริมฝีปากและฟัน (Labiodental)	ฐานปุ่มเหงือก (Alveolar)	ฐานเพดานแข็ง (Palatal)	ฐานเพดานอ่อน (Velar)	ฐานเส้นเสียง (Glottal)
เสียงหยุดอโหชนะที่ไม่มีลม (Unaspirated unvoiced stop)	p		t	c	k	z
เสียงหยุดอโหชนะที่มีลม (Aspirated unvoiced stop)	ph		th	ch	kh	
เสียงหยุดอโหชนะ (Voiced stop)	b		d			
เสียงแทรกอโหชนะ (Unvoiced fricative)		f	s			h
เสียงนาสิก (Nasal)	m		n		ng	
เสียงเลื่อน (Glide)	w			y		
เสียงลิ้นร้ว (Trill)			r			
เสียงลอคข้าง (Lateral)			l			

2.3.2 เสียงสระ

เสียงสระในภาษาไทยนั้น เริ่มต้นจากสระเสียงสั้น ซึ่งมีทั้งหมด 9 เสียงคือ

/-a-/	อะ	/-x-/	แอะ
/-i-/	อิ	/-q-/	เออะ
/-v-/	อี	/-o-/	โอะ
/-u-/	อุ	/-@-/	เออะ
/-e-/	เอะ		

และสระเหล่านี้มีทั้งเสียงสั้นและเสียงยาว ความยาวของเสียงสระนั้นมีความสำคัญ เพราะทำให้ความหมายของคำต่างกันได้ ความยาวทำให้จำนวนสระในภาษาไทยเพิ่มขึ้นจากเดิมอีก 9 เสียงคือ

/-aa-/	อา	/-xx-/	แอ
/-ii-/	อี	/-qq-/	เออ
/-vv-/	เอี	/-oo-/	โอ
/-uu-/	อุ	/-@@-/	เออ
/-ee-/	เอ		

ตารางที่ 2.2 เสียงสระในภาษาไทย

		หน้า (Front)	กลาง (Central)	หลัง (Back)
สูง (High)		/-i-/ , /-ii-/	/-v-/ , /-vv-/	/-u-/ , /-uu-/
กลาง (Mid)	สูง	/-e-/ , /-ee-/	/-q-/ , /-qq-/	/-o-/ , /-oo-/
	ต่ำ	/-x-/ , /-xx-/		
ต่ำ (Low)		/-a-/ , /-aa-/		/-@-/ , /-@@-/

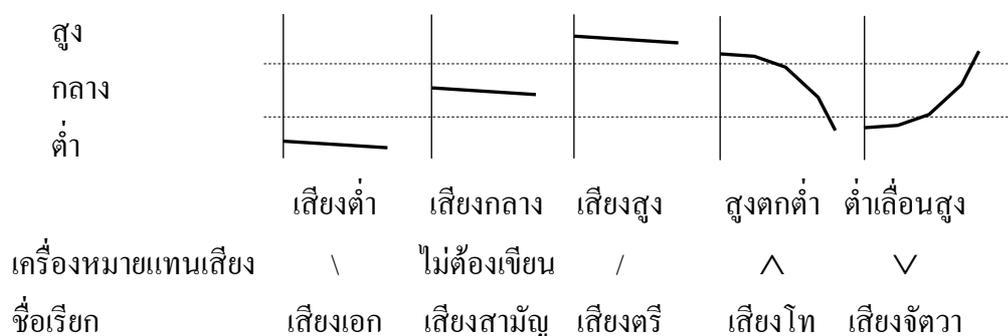
นอกจากสระเดี่ยว (Monophthong) แล้ว ภาษาไทยยังมีสระผสม (Diphthong) ซึ่งมีทั้งที่เป็นเสียงสั้นและเสียงยาวอีก 6 เสียงคือ

/-ie-/	เอียะ	/-iie-/	เอีย
/-va-/	เอือะ	/-vva-/	เอือ
/-ua-/	อัวะ	/-uua-/	อิว

ดังนั้นภาษาไทยจึงมีเสียงสระทั้งหมด 24 เสียง และนอกจากนี้ในภาษาไทยยังมีสระที่เรียกว่าสระอักษร ได้แก่ อ่า ไอ โอ เอา ซึ่งแม้จะมีรูปเฉพาะ แต่เมื่อพิจารณาเสียงแล้วจะเหมือนกับสระปกติที่รวมกับพยัญชนะท้ายนั่นเอง เช่น /-am/, /-ay/, /-aw/

2.3.3 เสียงวรรณยุกต์

เสียงวรรณยุกต์ คือ ระดับเสียงสูงต่ำในคำหนึ่งคำ คำที่มีเสียงพยัญชนะและเสียงสระเหมือนกัน แต่มีเสียงวรรณยุกต์ที่ต่างกัน ก็จะมีความหมายต่างกันด้วย วรรณยุกต์จึงเป็นเสียงที่มีความสำคัญในภาษาไทย จึงนับเป็นหน่วยเสียงด้วยเช่นกัน ภาษาไทยมีเสียงวรรณยุกต์ต่างกัน ไป 5 เสียง เสียงเหล่านี้อาจเขียนเป็นเส้นเสียงที่แสดงระดับสูงต่ำหรือขึ้นลงได้ดังนี้



รูปที่ 2.2 เสียงวรรณยุกต์ในภาษาไทย

2.4 ลักษณะพยางค์และคำในภาษาไทย

พยางค์ (Syllable) คือ เสียงที่เปล่งออกมาแต่ละครั้ง จึงประกอบด้วยเสียงสระเพียงเสียงเดียว ซึ่งอาจมีความหมายหรือไม่มีความหมายก็ได้ ส่วนคำ (Word) นั้น อาจหมายถึงพยางค์เดียวหรือพยางค์ที่รวมกันหลาย ๆ พยางค์ก็ได้ แต่คำนั้นจะต้องมีความหมาย หน่วยเสียงในภาษาไทยที่รวมกันเป็นพยางค์นั้นประกอบด้วย เสียงพยัญชนะ (Consonant: C) เสียงสระ (Vowel: V) และเสียงวรรณยุกต์ (Tone: T) โดยเสียงพยัญชนะต้นนั้นอาจจะเป็นพยัญชนะเพียงเสียงเดียว หรือพยัญชนะควบกล้ำก็ได้ ดังนั้นรูปแบบ ของพยางค์ในภาษาไทยที่เป็นไปได้มีดังนี้

1. CV^T (พยัญชนะต้น + สระ + วรรณยุกต์)
2. CCV^T (พยัญชนะควบกล้ำ + สระ + วรรณยุกต์)
3. CV^TC (พยัญชนะต้น + สระ + วรรณยุกต์ + พยัญชนะท้าย)
4. CCV^TC (พยัญชนะควบกล้ำ + สระ + วรรณยุกต์ + พยัญชนะท้าย)

พยางค์แบบ CV^T เสียงพยัญชนะต้นที่เป็นไปได้ทั้งหมดมี 21 เสียง รวมกับเสียงสระ 24 เสียง ซึ่งจะแบ่งเป็นสระเสียงสั้น 12 เสียงซึ่งผันวรรณยุกต์ได้ 4 เสียง และสระเสียงยาว 12 เสียงผัน

วรรณยุกต์ได้ 5 เสียง เพราะฉะนั้นจะมีพยางค์ที่เป็นไปได้ไม่เกิน $2,268 ((21 \times 12 \times 4) + (21 \times 12 \times 5))$ เสียง

พยางค์แบบ CCV^T เสียงพยัญชนะควบกล้ำที่เป็นไปได้ทั้งหมดมี 12 เสียง รวมกับเสียงสระ 24 เสียง ซึ่งจะแบ่งเป็นสระเสียงสั้น 12 เสียงซึ่งผันวรรณยุกต์ได้ 4 เสียง และสระเสียงยาว 12 เสียงผันวรรณยุกต์ได้ 5 เสียง เพราะฉะนั้นจะมีพยางค์ที่เป็นไปได้ไม่เกิน $1,296 ((12 \times 12 \times 4) + (12 \times 12 \times 5))$ เสียง

พยางค์แบบ CV^TC การพิจารณาพยางค์ในลักษณะ CV^TC ที่เป็นไปได้ทั้งหมดนั้นต้องแยกพิจารณาคำแต่ละลักษณะดังนี้

คำเป็น คือคำที่ลงท้ายด้วยเสียง /-m/, /-n/, /-ng/, /-y/, /-w/ ประกอบด้วย

- พยัญชนะต้น 21 เสียงรวมกับเสียงสระ 16 เสียง (ยกเว้น /-e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง (ยกเว้น /-y/, /-w/) ผันวรรณยุกต์ได้ 5 เสียง มีพยางค์ที่เป็นไปได้ทั้งหมด 5,040 เสียง

- พยัญชนะต้น 21 เสียงรวมกับเสียงสระ 8 เสียง (/ -e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง (ยกเว้น /-y/, /-w/) ในรูปสามัญ มีพยางค์ที่เป็นไปได้ทั้งหมด 504 เสียง (สระกลุ่มนี้มีแต่รูปสามัญเท่านั้น)

- พยัญชนะต้น 21 เสียงรวมกับเสียงสระ 18 เสียง (ยกเว้น /-e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-) และลงท้ายด้วยเสียงสะกด 1 เสียง (/ -y/ รวมกับ /-w/) ผันวรรณยุกต์ได้ 5 เสียง มีพยางค์ที่เป็นไปได้ทั้งหมด 1,890 เสียง (เสียง /-y/ และ /-w/ มีไม่ครบทุกสระ)

- พยัญชนะต้น 21 เสียงรวมกับเสียงสระ 7 เสียง (/ -e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-) และลงท้ายด้วยเสียงสะกด 1 เสียง (/ -y/ รวมกับ /-w/) ในรูปสามัญ มีพยางค์ที่เป็นไปได้ทั้งหมด 147 เสียง

คำตาย คือคำที่ลงท้ายด้วยเสียง /-k/, /-d/, /-b/ ประกอบด้วย

- พยัญชนะต้น 21 เสียงรวมกับเสียงสระ 16 เสียง (ยกเว้น /-e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง ผันวรรณยุกต์ได้ 4 เสียง มีพยางค์ที่เป็นไปได้ทั้งหมด 4,032 เสียง

- พยัญชนะต้น 21 เสียงรวมกับเสียงสระ 8 เสียง (/ -e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง ในรูปสามัญ มีพยางค์ที่เป็นไปได้ทั้งหมด 504 เสียง

ดังนั้นพยางค์ที่เป็นไปได้จะมีทั้งหมด $5,040 + 504 + 1,890 + 147 + 4,032 + 504 = 12,117$ เสียง

ตารางที่ 2.3 พยางค์ในลักษณะ CV^TC ที่เป็นไปได้ทั้งหมด

		พยัญชนะ	สระ	ตัวสะกด	วรรณยุกต์	พยางค์ที่เป็นไปได้
CV ^T C	คำเป็น	21	16	3	5	5,040
		21	8	3	1	504
		21	18	1	5	1,890
		21	7	1	1	147
	คำตาย	21	16	3	4	4,032
		21	8	3	1	504
รวม						12,117

พยางค์แบบ CCV^TC การพิจารณาพยางค์ในลักษณะ CCV^TC ที่เป็นไปได้ทั้งหมด ต้องแยกพิจารณาคำแต่ละลักษณะดังนี้

คำเป็น คือคำที่ลงท้ายด้วยเสียง /-m/, /-n/, /-ng/, /-y/, /-w/ ประกอบด้วย

- พยัญชนะต้น 12 เสียงรวมกับเสียงสระ 16 เสียง (ยกเว้น /-e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง (ยกเว้น /-y/, /-w/) ผันวรรณยุกต์ได้ 5 เสียง มีพยางค์ที่เป็นไปได้ทั้งหมด 2,880 เสียง

- พยัญชนะต้น 12 เสียงรวมกับเสียงสระ 8 เสียง (/ -e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง (ยกเว้น /-y/, /-w/) ในรูปสามัญ มีพยางค์ที่เป็นไปได้ทั้งหมด 288 เสียง

- พยัญชนะต้น 12 เสียงรวมกับเสียงสระ 18 เสียง (ยกเว้น /-e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-) และลงท้ายด้วยเสียงสะกด 1 เสียง (/ -y/ รวมกับ /-w/) ผันวรรณยุกต์ได้ 5 เสียง มีพยางค์ที่เป็นไปได้ทั้งหมด 1,080 เสียง

- พยัญชนะต้น 12 เสียงรวมกับเสียงสระ 7 เสียง (/ -e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-) และลงท้ายด้วยเสียงสะกด 1 เสียง (/ -y/ รวมกับ /-w/) ในรูปสามัญ มีพยางค์ที่เป็นไปได้ทั้งหมด 84 เสียง

คำตาย คือคำที่ลงท้ายด้วยเสียง /-k/, /-d/, /-b/ ประกอบด้วย

- พยัญชนะต้น 12 เสียงรวมกับเสียงสระ 16 เสียง (ยกเว้น /-e-, /-x-, /-q-, /-o-, /-@-, /-ie-, /-va-, /-ua-) และลงท้ายด้วยเสียงสะกด 3 เสียง ผันวรรณยุกต์ได้ 4 เสียง มีพยางค์ที่เป็นไปได้ทั้งหมด 2,304 เสียง

- พยัญชนะต้น 12 เสียงรวมกับเสียงสระ 8 เสียง (/e-/ , /-x-/ , /-q-/ , /-o-/ , /-@-/ , /-ie-/ , /-va-/ , /-ua-/) และลงท้ายด้วยเสียงสะกด 3 เสียง ในรูปสามัญ มีพยางค์ที่เป็นไปได้ทั้งหมด 288 เสียง

ดังนั้นพยางค์ที่เป็นไปได้จะมีทั้งหมด $2,880 + 288 + 1,080 + 84 + 2,304 + 288 = 6,924$ เสียง

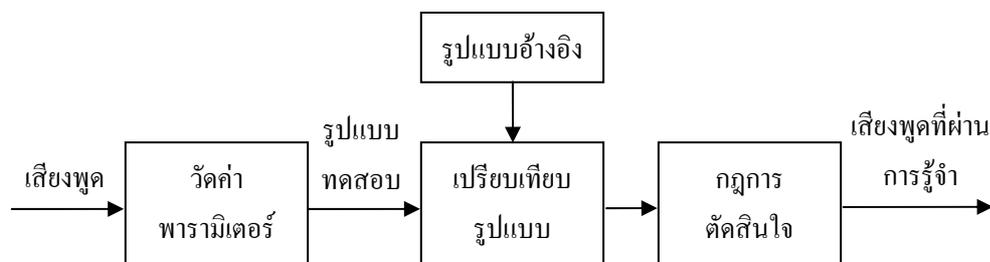
ตารางที่ 2.4 พยางค์ในลักษณะ CCV^TC ที่เป็นไปได้ทั้งหมด

		พยัญชนะ	สระ	ตัวสะกด	วรรณยุกต์	พยางค์ที่เป็นไปได้	
CCV ^T C	คำเป็น	12	16	3	5	2,880	
		12	8	3	1	288	
		12	18	1	5	1,080	
		12	7	1	1	84	
	คำตาย	12	16	3	4	2,304	
		12	8	3	1	288	
	รวม						6,924

2.5 การประมวลผลสัญญาณและกระบวนการรู้จำเสียงพูด

กระบวนการรู้จำเสียงนั้นถูกพัฒนามาจากทฤษฎีต่าง ๆ มากมาย เช่น ทฤษฎีทางภาษาศาสตร์ สถิติ ทฤษฎีทางการสื่อสาร จากรูปที่ 2.3 ซึ่งแสดงระบบการรู้จำเสียงทั่วไป จะเห็นได้ว่าจะต้องประกอบด้วยขั้นตอนพื้นฐาน 3 ขั้นตอน คือ

1. การวัดค่าพารามิเตอร์
2. การเปรียบเทียบรูปแบบ
3. การตัดสินใจ



รูปที่ 2.3 ขั้นตอนในกระบวนการรู้จำเสียงพูด

ในขั้นตอนของการวัดค่าพารามิเตอร์นั้นนับว่าเป็นขั้นตอนที่สำคัญ การเลือกค่าพารามิเตอร์ที่มีประสิทธิภาพ จะส่งผลให้การเปรียบเทียบรูปแบบ และการตัดสินใจมีความถูกต้องสูงด้วย การวัดค่าพารามิเตอร์นั้นจะต้องอาศัยการประมวลผลสัญญาณ (Signal Processing) ซึ่งจะแปลงลักษณะของรูปแบบคลื่นเสียง (Waveform) ไปเป็นค่าพารามิเตอร์ที่ใช้เป็นตัวแทนของเสียงพูด และสามารถนำไปวิเคราะห์ในขั้นตอนต่อไป

ค่าพารามิเตอร์ที่ใช้แทนเสียงพูดนั้นมีหลายชนิด เช่น ค่าพลังงาน (Energy) ของคลื่นเสียง ค่าอัตราการตัดค่าศูนย์ (Zero crossing rate) และค่าการห่อหุ้มสเปกตรัม (Spectrum envelope) โดยเฉพาะค่าสเปกตรัมนั้น นับว่าเป็นพารามิเตอร์หลักที่ใช้ในระบบการรู้จำ ซึ่งมีอยู่หลายลักษณะ ในงานวิจัยนี้ได้ศึกษาวิธีการวิเคราะห์สเปกตรัมจากแบบจำลองสเปกตรัมที่เรียกว่า Linear Predictive Coding (LPC) แบบจำลองการวิเคราะห์ LPC นั้นสามารถแสดงดังรูปที่ 2.4 ซึ่งจะวิเคราะห์สเปกตรัมของเสียงทีละช่วงเวลา (Speech frame) และผลลัพธ์ที่ได้ในแต่ละช่วงเวลา คือ เวกเตอร์ของค่าสัมประสิทธิ์ หรือค่าพารามิเตอร์ LPC ซึ่งแสดงค่าสเปกตรัมของแบบจำลองทุกโพลที่ตรงกับสเปกตรัมของสัญญาณบนช่วงเวลานั้น

2.6 แบบจำลอง Linear Predictive Coding สำหรับการรู้จำเสียงพูด

แบบจำลอง Linear Predictive Coding (LPC) นั้นถูกใช้อย่างแพร่หลายในระบบการรู้จำเสียงเพื่อใช้หาค่าพารามิเตอร์ที่เป็นตัวแทนของเสียงพูด เนื่องจาก

1. แบบจำลอง LPC เป็นแบบจำลองที่ดีสำหรับสัญญาณเสียง โดยเฉพาะช่วงที่เป็นเสียงโหมะ และสามารถประมาณค่าการห่อหุ้มสเปกตรัมของทางเดินเสียงได้ดี แม้ว่าในช่วงเสียงโหมะจะมีประสิทธิภาพลดลงบ้าง แต่ผลที่ได้ก็สามารถนำไปใช้ในกระบวนการรู้จำเสียงได้
2. แบบจำลอง LPC นั้น สามารถวิเคราะห์สัญญาณเสียงในลักษณะแยกพิจารณาแหล่งกำเนิดและทางเดินของเสียงได้ จึงอาจใช้สร้างตัวแทนของลักษณะทางเดินของเสียงพูดได้
3. แบบจำลอง LPC นั้น ง่ายและตรงไปตรงมา สามารถนำไปใช้ได้กับอุปกรณ์และโปรแกรมประมวลผลต่าง ๆ และยังต้องการทรัพยากรในการคำนวณไม่มากนัก
4. แบบจำลอง LPC เหมาะสมกับระบบการรู้จำต่าง ๆ และให้ประสิทธิภาพสูงเมื่อเปรียบเทียบกับแบบจำลองอื่น ๆ

2.6.1 แบบจำลอง Linear Predictive Coding

แนวคิดเบื้องต้นของแบบจำลอง LPC คือสัญญาณเสียงที่เวลา n หรือ $s(n)$ สามารถหาค่าโดยประมาณจากผลรวมเชิงเส้นของค่าตัวอย่างของสัญญาณเสียงก่อนหน้า จำนวน p ตัวอย่าง

$$s(n) \approx a_1s(n-1) + a_2s(n-2) + \dots + a_p s(n-p) \quad (2.1)$$

โดยที่ค่าสัมประสิทธิ์ a_1, a_2, \dots, a_p จะเป็นค่าคงที่ตลอดช่วงของเสียงพูดที่ทำการวิเคราะห์ สมการ (2.1) สามารถเขียนใหม่โดยเพิ่มพจน์ของค่าการกระตุ้น (Excitation Term) ได้ดังนี้

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (2.2)$$

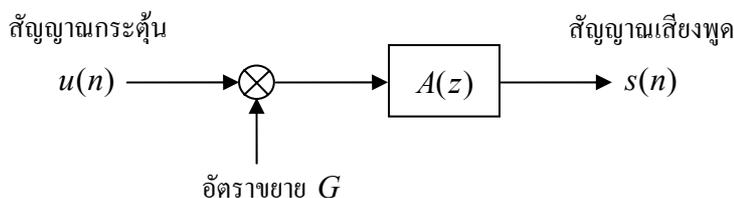
โดยที่ $u(n)$ คือค่าการกระตุ้นบรรทัดฐาน (Normalized Excitation) และ G คืออัตราขยายของการกระตุ้น จากสมการที่ (2.2) สามารถเขียนในโดเมน z ได้เป็น

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + GU(z) \quad (2.3)$$

และได้ค่าทรานส์เฟอร์ฟังก์ชัน

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (2.4)$$

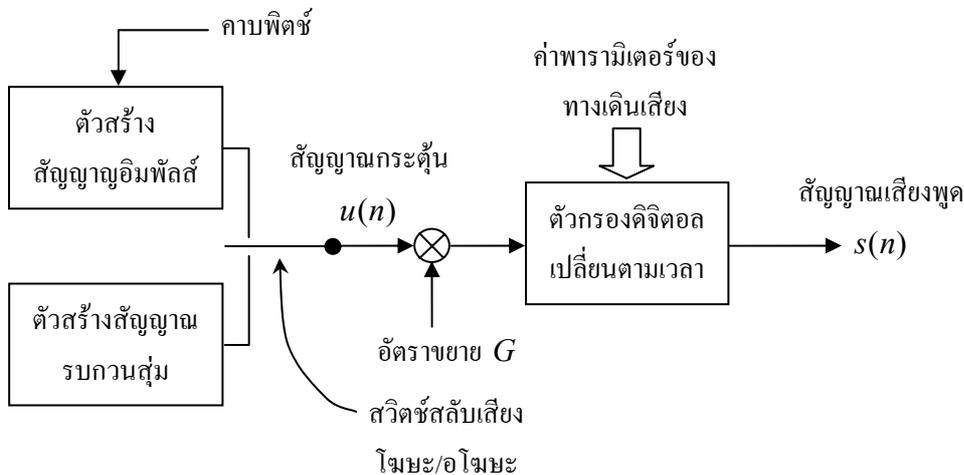
จากสมการ (2.3) สามารถอธิบายได้ดังรูปที่ 2.4 ว่าสัญญาณกระตุ้นบรรทัดฐาน $u(n)$ จะถูกปรับขนาดด้วยอัตราขยาย G และถูกนำเข้าสู่ระบบ $H(z) = 1/A(z)$ เพื่อสร้างสัญญาณเสียง $s(n)$



รูปที่ 2.4 แบบจำลอง LPC สำหรับเสียงพูด

ฟังก์ชันของค่าการกระตุ้นสำหรับเสียงพูดของมนุษย์ในความเป็นจริงนั้นอาจจะมีทั้งที่เป็นสัญญาณพัลส์กึ่งคาบ (Quasi-periodic Pulse) หรือเสียงโฆชะ และสัญญาณรบกวนสุ่ม (Random Noise) หรือเสียงโฆชะ ดังนั้นการสังเคราะห์เสียงพูดจึงสามารถแสดงได้ดังรูปที่ 2.5 ซึ่งมีส่วนของ

สวิตช์ที่ทำหน้าสลับแหล่งกำเนิดเสียงระหว่างเสียงโฆมะ และอโฆมะ เพื่อส่งเข้าไปยังตัวกรอง $H(z)$ ซึ่งถูกควบคุมคุณลักษณะโดยค่าพารามิเตอร์ของทางเดินของเสียง



รูปที่ 2.5 การสังเคราะห์เสียงพูดของมนุษย์ตามแบบจำลอง LPC

2.6.2 สมการการวิเคราะห์ LPC

จากรูปที่ 2.4 เราเขียนความสัมพันธ์ของ $s(n)$ และ $u(n)$ ได้เป็น

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2.5)$$

ถ้าเรากำหนดให้ค่าโดยประมาณของสัญญาณเสียง $\tilde{s}(n)$ นั้นได้จากผลรวมเชิงเส้นของค่าตัวอย่างของสัญญาณเสียงก่อนหน้า จะได้ว่า

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

เราจะได้อัปเดตค่าความผิดพลาดของการประมาณ (Prediction Error) $e(n)$ ซึ่งนิยามได้ว่า

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (2.7)$$

ซึ่งจะได้ค่าทรานส์เฟอร์ฟังก์ชันของความผิดพลาดเป็น

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p a_k z^{-k} \quad (2.8)$$

จะเห็นว่าค่าสัญญาณ $s(n)$ ซึ่งสร้างจากระบบเชิงเส้นตามรูปที่ 2.4 นั้นจะมีค่าความผิดพลาดของการประมาณ $e(n)$ เท่ากับค่าการกระตุ้นที่มีการปรับขนาดหรือ $Gu(n)$

สิ่งที่เราต้องการหาจากระบบเชิงเส้นดังกล่าวคือเซตของค่าสัมประสิทธิ์ของการประมาณ (Prediction Coefficient) หรือ $\{a_k\}$ จากสัญญาณเสียงโดยตรง ซึ่งสเปกตรัมคุณสมบัติของตัวกรองดิจิทัลในรูปที่ 2.5 นั้นตรงกับรูปคลื่นของสัญญาณเสียงในช่วงเวลาที่ทำการวิเคราะห์ เนื่องจากสเปกตรัมคุณสมบัติของสัญญาณเสียงนั้นเปลี่ยนแปลงตามเวลา ดังนั้นการหาค่าสัมประสิทธิ์ของการประมาณที่เวลา n นั้น จะต้องประมาณจากช่วงสั้น ๆ ของสัญญาณเสียงที่เวลานั้น การประมาณค่าสัญญาณเสียงนั้นจะต้องหาค่าสัมประสิทธิ์ของการประมาณในช่วงเวลาสั้น ๆ ที่ทำให้ค่าความผิดพลาดเฉลี่ยกำลังสอง (Mean-squared error) มีค่าน้อยที่สุด

การสร้างสมการในการหาค่าสัมประสิทธิ์ของการประมาณนั้น เริ่มจากการกำหนดช่วงสั้น ๆ ของสัญญาณเสียง และค่าผิดพลาดที่เวลา n ดังนี้

$$s_n(m) = s(n+m) \quad (2.9a)$$

$$e_n(m) = e(n+m) \quad (2.9b)$$

และค่าความผิดพลาดเฉลี่ยกำลังสองที่เวลา n จะเท่ากับ

$$E_n = \sum_m e_n^2(m) \quad (2.10)$$

จากนิยามของ $e_n(m)$ และ $s_n(m)$ จะได้ว่า

$$E_n = \sum_m \left[s_n(m) - \sum_{k=1}^p a_k s_n(m-k) \right]^2 \quad (2.11)$$

ในการหาค่าสัมประสิทธิ์การประมาณจากสมการที่ (2.11) นั้นจะต้องหาค่าอนุพันธ์ของ E_n เมื่อเทียบกับ a_k แต่ละตัว แล้วให้เท่ากับ 0

$$\frac{\partial E_n}{\partial a_k} = 0, k = 1, 2, \dots, p \quad (2.12)$$

กำหนดให้

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^p \hat{a}_k \sum_m s_n(m-i)s_n(m-k) \quad (2.13)$$

โดยที่ $\sum s_n(m-i)s_n(m-k)$ คือพจน์ของโควาเรียนซ์ (Covariance) ในช่วงสั้น ๆ ของ $s_n(m)$

$$\phi_n(i, k) = \sum_m s_n(m-i)s_n(m-k) \quad (2.14)$$

สมการ (2.14) สามารถเขียนในรูปย่อได้ว่า

$$\phi_n(i, 0) = \sum_{k=1}^p \hat{a}_k \phi_n(i, k) \quad (2.15)$$

ซึ่งอธิบายเขตของสมการ p สมการที่มีตัวแปร p ตัว และค่าความผิดพลาดเฉลี่ยกำลังสองที่น้อยที่สุด \hat{E}_n จะเท่ากับ

$$\begin{aligned} \hat{E}_n &= \sum_m s_n^2(m) - \sum_{k=1}^p \hat{a}_k \sum_m s_n(m)s_n(m-k) \\ &= \phi_n(0, 0) - \sum_{k=1}^p \hat{a}_k \phi_n(0, k) \end{aligned} \quad (2.16)$$

2.6.3 การหาค่าสัมประสิทธิ์การประมาณโดยวิธีอัตโนมัติ

การหาค่าสัมประสิทธิ์ของการประมาณจากสมการที่ (2.16) นั้นทำได้หลายวิธี แต่การวิจัยนี้จะใช้วิธีอัตโนมัติ (Autocorrelation) ซึ่งเป็นวิธีที่สะดวกและตรงไปตรงมา โดยจะกำหนดให้ช่วงของสัญญาณเสียง $s_n(m)$ ที่อยู่นอกช่วง $0 \leq m \leq N-1$ มีค่าเป็น 0 และให้มีค่าเท่ากับผลคูณของสัญญาณเสียง $s(m+n)$ กับวินโดว์ที่มีความยาวจำกัด $w(m)$ ซึ่งมีค่าเป็น 0 ภายนอกช่วง $0 \leq m \leq N-1$ ดังนั้นสัญญาณเสียงจะเขียนได้เป็น

$$s_n(m) = \begin{cases} s(m+n) \cdot w(m), & 0 \leq m \leq N-1 \\ 0, & \text{otherwise} \end{cases} \quad (2.17)$$

จากสมการที่ (2.17) ค่าความผิดพลาดของการประมาณสัญญาณ $e_n(m)$ ที่อยู่นอกช่วง $0 \leq m \leq N-1+p$ นั้นจะมีค่าเป็น 0 ด้วย เนื่องจากสัญญาณในบริเวณนี้ถูกกำหนดให้มีค่าเป็น 0 ในขณะที่ช่วง $m=0$ ถึง $m=p-1$ และ $m=N-1$ ถึง $m=N-1+p$ หรือบริเวณขอบของวินโดว์อาจจะมีค่าความผิดพลาดสูงเนื่องจากการประมาณที่ใช้ค่า 0 แทนสัญญาณก่อนหน้าในการคำนวณ ซึ่งค่าผิดพลาดดังกล่าวจะเกิดขึ้นอย่างมากกับเสียงโฆยะ แต่สำหรับเสียงโฆยะนั้นค่าผิดพลาดบริเวณขอบจะมีไม่มากนัก การกำหนดวินโดว์ในสมการที่ (2.17) จะมีผลทำให้ลดค่าความผิดพลาดของการประมาณสัญญาณบริเวณขอบได้

จากสมการ (2.17) จะได้ค่าความผิดพลาดในการประมาณเฉลี่ยกำลังสองเท่ากับ

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m) \quad (2.18)$$

และ $\phi_n(i, k)$ จะเขียนได้เป็น

$$\phi_n(i, k) = \sum_{m=0}^{N-1+p} s_n(m-i)s_n(m-k), \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix} \quad (2.19)$$

หรือ

$$\phi_n(i, k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k), \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix} \quad (2.20)$$

เนื่องจากสมการ (2.20) เป็นฟังก์ชันของ $i-k$ ดังนั้นฟังก์ชันโควาเรียนซ์ $\phi_n(i, k)$ จะลดรูปเป็นฟังก์ชันอัตตสหสัมพันธ์อย่างง่าย คือ

$$\phi_n(i, k) = r_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \quad (2.21)$$

เนื่องจากฟังก์ชันอัตตสหสัมพันธ์ที่ได้เป็นฟังก์ชันสมมาตร คือ $r_n(-k) = r_n(k)$ สมการ LPC จึงเขียนได้เป็น

$$\sum_{k=1}^p r_n(i-k)\hat{a}_k = r_n(i), 1 \leq i \leq p \quad (2.22)$$

และเขียนให้อยู่ในรูปเมตริกซ์ได้ว่า

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdots & r_n(p-3) \\ \vdots & \vdots & \vdots & & \vdots \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \cdots & r_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ \vdots \\ r_n(p) \end{bmatrix} \quad (2.23)$$

เมตริกซ์ขนาด $p \times p$ ของค่าอัตโนมัติสหสัมพันธ์จะเป็นเมตริกซ์ที่ค่าระหว่างแนวทแยงมุมสมมาตรกัน วิธีที่ใช้แก้สมการดังกล่าวเพื่อหาค่าสัมประสิทธิ์ของการประมาณ ได้แก่ วิธีของ Durbin ซึ่งมีวิธีการดังนี้

$$E^{(0)} = r(0) \quad (2.24)$$

$$k_i = \frac{r(i) - \sum_{j=1}^{L-1} \alpha_j^{(i-1)} r(i-j)}{E^{(i-1)}}, 1 \leq i \leq p \quad (2.25)$$

$$\alpha_i^{(i)} = k_i \quad (2.26)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad (2.27)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (2.28)$$

ผลรวมของสมการที่ (2.25) จะไม่ได้ใช้ในกรณีที่ $i = 1$ และสมการที่ (2.26-2.28) นั้นจะมีการคำนวณแบบวนซ้ำสำหรับ $i = 1, 2, \dots, p$ และผลลัพธ์สุดท้ายที่ได้คือ

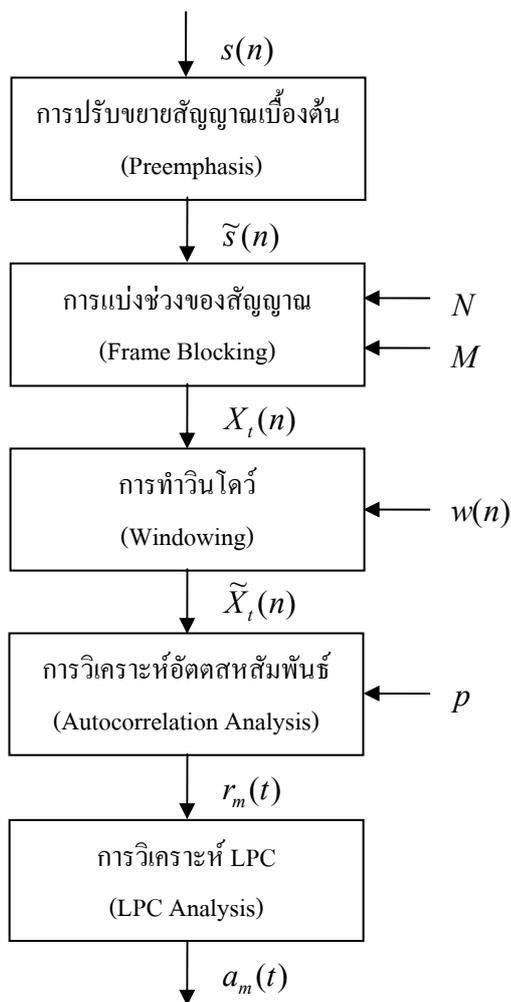
$$a_m = \text{LPC coefficients} = \alpha_m^{(p)}, 1 \leq m \leq p \quad (2.29)$$

$$k_m = \text{PARCOR coefficients} \quad (2.30)$$

$$g_m = \text{log area ratio coefficients} = \log \left(\frac{1 - k_m}{1 + k_m} \right) \quad (2.31)$$

2.6.4 การประมวลผล LPC สำหรับการรู้จำเสียง

การประมวลผล LPC ที่มีการใช้งานในระบบรู้จำเสียงอย่างแพร่หลายนั้นประกอบด้วยขั้นตอนพื้นฐาน ซึ่งแสดงได้ดังรูปที่ 2.6 โดยในแต่ละขั้นตอนมีรายละเอียดดังนี้



รูปที่ 2.6 แผนภาพแสดงการประมวลผล LPC สำหรับการรู้จำเสียงพูด

1. การปรับขยายสัญญาณเบื้องต้น (Preemphasis) เสียงพูดที่อยู่ในรูปสัญญาณดิจิทัล $s(n)$ จะถูกส่งผ่านระบบตัวกรองอันดับต่ำ (Low-order Digital Filter) เพื่อให้สเปกตรัมของสัญญาณเรียบมากขึ้นและลดความซับซ้อนของข้อมูล ซึ่งตัวกรองดิจิทัลที่ใช้นั้นมีทั้งแบบคงที่และแบบปรับค่าได้ สำหรับตัวกรองแบบคงที่ที่นิยมใช้จะเป็นระบบอันดับหนึ่ง (Fixed First-order System)

$$H(z) = 1 - \tilde{a}z^{-1}, 0.9 \leq a \leq 1.0 \quad (2.32)$$

ซึ่งในกรณีนี้สัญญาณขาออก $\tilde{s}(n)$ ที่ได้จะสัมพันธ์กับสัญญาณขาเข้า $s(n)$ ดังนี้

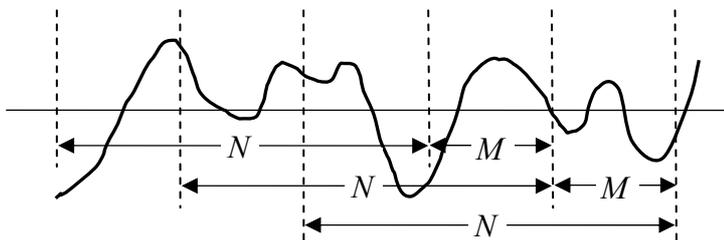
$$\tilde{s}(n) = s(n) - \tilde{a}s(n-1) \quad (2.33)$$

โดยค่า \tilde{a} ที่นิยมใช้มีค่าประมาณ 0.95 สำหรับตัวกรองแบบปรับค่าได้นั้นจะมีค่าทรานส์เฟอร์ฟังก์ชันเป็น

$$H(z) = 1 - \tilde{a}_n z^{-1} \quad (2.34)$$

ค่า \tilde{a}_n นั้นจะเปลี่ยนแปลงได้ตามเวลา n ตามเงื่อนไขที่กำหนด ตัวอย่างเงื่อนไขสำหรับ \tilde{a}_n ได้แก่ $\tilde{a}_n = r_n(1)/r_n(0)$

2. การแบ่งช่วงของสัญญาณ (Frame Blocking) สัญญาณที่ผ่านกระบวนการ Preemphasis $\tilde{s}(n)$ นั้นจะถูกแบ่งออกเป็นเฟรม เฟรมละ N ตัวอย่าง โดยเฟรมที่อยู่ติดกันจะมีอยู่ห่างกัน M ตัวอย่าง รูปที่ 2.7 นั้นอธิบายการแบ่งช่วงของสัญญาณในกรณีที่ $M = (1/3)N$ ซึ่งเฟรมแรกจะประกอบด้วยสัญญาณจำนวน N ตัวอย่าง เฟรมที่สองจะเริ่มหลังจากเฟรมแรกไป M ตัวอย่างโดยมีส่วนที่ซ้อนทับกันระหว่างเฟรมแรกและเฟรมที่สอง $N - M$ ตัวอย่าง ในทำนองเดียวกัน เฟรมที่สามจะเริ่มหลังจากเฟรมแรก $2M$ และมีส่วนที่ซ้อนทับกับเฟรมแรกอยู่ $N - 2M$ ตัวอย่าง จากตัวอย่างข้างต้น เมื่อ $M \leq N$ จะมีส่วนที่ซ้อนทับกันระหว่างเฟรมซึ่งทำให้การประมาณค่า LPC มีความสัมพันธ์ระหว่างเฟรมที่ติดกัน และจะยิ่งมีความละเอียดมากขึ้นเมื่อ $M \ll N$ ในทางกลับกัน ถ้า $M > N$ จะมีบางส่วนของสัญญาณที่ไม่ถูกประมวลผล ซึ่งทำให้ความสัมพันธ์ระหว่างเฟรมในการประมาณค่า LPC จะมีส่วนของสัญญาณรบกวนรวมเข้ามาด้วย ค่าของ N และ M ที่ใช้กันทั่วไปคือ 300 และ 100 ตัวอย่างตามลำดับ สำหรับสัญญาณเสียงที่มีอัตราการสุ่มตัวอย่าง 6.67 kHz หรือประมาณ 45 ms ต่อเฟรม และแต่ละเฟรมห่างกัน 15 ms



รูปที่ 2.7 การแบ่งช่วงของสัญญาณเสียง (Frame Blocking)

3. การทำวินโดว์ (Windowing) สัญญาณเสียงแต่ละเฟรมที่แบ่งออกมา จะผ่านกระบวนการทำวินโดว์ เพื่อลดปัญหาความไม่ต่อเนื่องของสัญญาณบริเวณจุดเริ่มต้นและตอนท้ายของเฟรม โดยการลดค่าของสัญญาณให้เป็น 0 ที่บริเวณตอนต้นและตอนท้ายของแต่ละเฟรม ถ้ากำหนดให้วินโดว์เท่ากับ $w(n), 0 \leq n \leq N-1$ จะได้สัญญาณเฟรม l^{th} ที่ผ่านการทำวินโดว์เป็น

$$\tilde{x}_l(n) = x_l(n)w(n), 0 \leq n \leq N-1 \quad (2.35)$$

วินโดว์ที่นิยมใช้ในการประมาณค่า LPC สำหรับวิธีอัตตสหสัมพันธ์ คือวินโดว์แบบแฮมมิง (Hamming) ซึ่งอยู่ในรูป

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (2.36)$$

4. การวิเคราะห์อัตตสหสัมพันธ์ (Autocorrelation Analysis) สัญญาณแต่ละเฟรมที่ผ่านการทำวินโดว์ เมื่อผ่านกระบวนการวิเคราะห์อัตตสหสัมพันธ์ จะได้

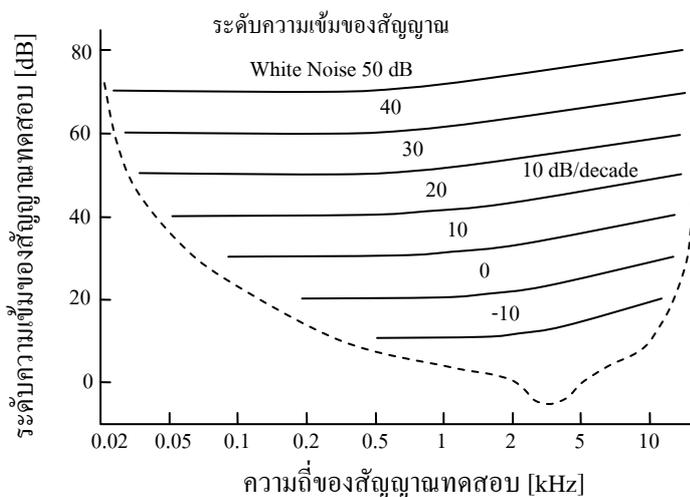
$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (2.37)$$

ซึ่งค่าสูงสุดของอัตตสหสัมพันธ์ p จะเป็นอันดับของการวิเคราะห์ LPC โดยทั่วไปค่า p จะมีค่าระหว่าง 8 ถึง 16 นอกจากนี้การวิเคราะห์อัตตสหสัมพันธ์ยังได้ค่าพลังงานของเฟรม l^{th} ซึ่งเท่ากับค่าอัตตสหสัมพันธ์ที่ 0 หรือ $R_l(0)$ ซึ่งสามารถนำไปใช้ในกระบวนการตรวจหาเสียงพูดได้

5. การวิเคราะห์ LPC (LPC Analysis) เป็นกระบวนการเปลี่ยนค่าอัตตสหสัมพันธ์ทั้ง $p+1$ ค่าไปเป็นเซตของค่าพารามิเตอร์ LPC ซึ่งได้แก่ ค่าสัมประสิทธิ์ LPC ค่าสัมประสิทธิ์การสะท้อน หรือ สัมประสิทธิ์ PARCOR และค่าสัมประสิทธิ์อัตราพื้นที่ลอการิทึม (Log area ratio Coefficient) ซึ่งการแปลงค่าอัตตสหสัมพันธ์ไปเป็นค่าสัมประสิทธิ์ LPC นั้น จะใช้วิธีของ Durbin ซึ่งอธิบายไว้ก่อนหน้านี

2.7 แถบความถี่วิกฤต และความเข้มสัญญาณในแถบความถี่วิกฤต

จากสมมติฐานของ Fletcher ที่บอกว่าส่วนของสัญญาณรบกวนที่มีผลในการบดบัง (Masking) สัญญาณทดสอบ (Test tone) จะเป็นส่วนที่มีสเปกตรัมใกล้เคียงกับสัญญาณทดสอบ และการบดบังจะเกิดขึ้นเมื่อกำลังของสัญญาณทดสอบ และสัญญาณรบกวนที่มีสเปกตรัมใกล้เคียงกับสัญญาณนั้น มีค่าเท่ากัน ส่วนของสัญญาณรบกวนที่มีสเปกตรัมอยู่นอกบริเวณดังกล่าว จะไม่มีผลต่อการบดบังสัญญาณ โดยแถบของความถี่บริเวณดังกล่าวนี้จะมีแบนด์วิธที่สร้างกำลังทางเสียงเดียวกับสัญญาณทดสอบ และสัญญาณรบกวนที่อยู่ในแถบนี้ เมื่อสัญญาณทดสอบถูกบดบัง



รูปที่ 2.8 ระดับของสัญญาณทดสอบที่ถูกบดบังโดยสัญญาณรบกวนที่มีความถี่ต่าง ๆ

การบดบังสัญญาณทดสอบโดยสัญญาณรบกวน White Noise ซึ่งเป็นสัญญาณรบกวนความถี่กว้าง (Broadband Noise) ที่ระดับความเข้มสัญญาณต่าง ๆ ซึ่งแสดงได้ดังรูปที่ 2.8 เส้นประในรูปคือระดับสัญญาณที่หูของมนุษย์เริ่มที่จะได้ยินเมื่อไม่มีระดับสัญญาณรบกวน หรือ Threshold in Quiet เช่น สัญญาณที่ความถี่ 2 kHz หากไม่มีสัญญาณรบกวนใด ๆ มนุษย์จะเริ่มได้ยินเสียงเมื่อสัญญาณนั้นมีความเข้มประมาณ 0 dB (10^{-12} w/m²) ขึ้นไป ส่วนเส้นทึบนั้นจะเป็นระดับที่มนุษย์เริ่มได้ยินเสียงในกรณีที่มีสัญญาณรบกวนซึ่งมีระดับความเข้มต่าง ๆ หรือ Masking Threshold เช่น สัญญาณทดสอบความถี่ 2 kHz และสัญญาณรบกวนที่มีความเข้ม 20 dB จะเริ่มได้ยินเมื่อสัญญาณนั้นมีความเข้มมากกว่า 40 dB จะเห็นว่าสัญญาณรบกวนนั้นจะมีระดับการบดบังสัญญาณทดสอบที่ขึ้นกับความถี่ แม้ว่าสัญญาณรบกวนที่เป็นสัญญาณ White Noise นั้นจะไม่ขึ้นกับความถี่ก็ตาม อย่างไรก็ตามระดับการบดบังของสัญญาณรบกวนที่ขึ้นกับความถี่นี้จะเกิดขึ้นกับความถี่ตั้งแต่ 500 Hz ขึ้นไปและจะมีอัตราการเปลี่ยนแปลงประมาณ 10 dB/decade

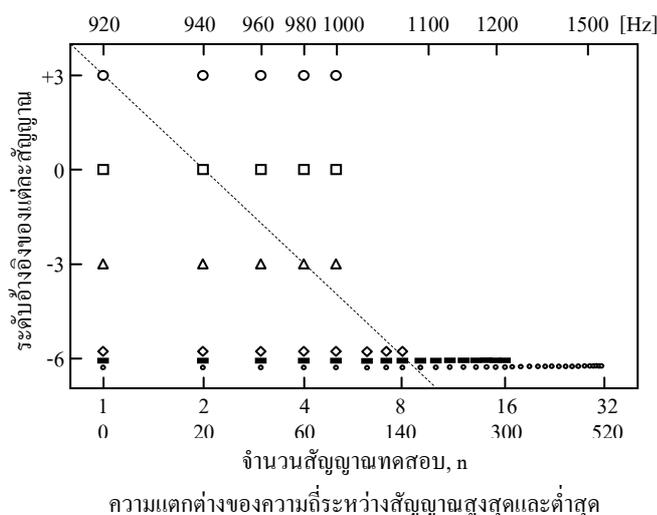
ที่ระดับความถี่ต่ำกว่า 500 Hz ระดับการบดบังของสัญญาณรบกวนจะไม่ขึ้นกับความถี่แถบความถี่วิกฤตในช่วงนี้จึงมีความกว้างเท่า ๆ กัน แต่ระดับความถี่ที่สูงกว่านี้ จะมีระดับการบดบังสัญญาณเพิ่มขึ้นประมาณ 10 dB/decade ดังนั้นความกว้างของแถบความถี่วิกฤตในช่วงนี้จึงเพิ่มขึ้นในลักษณะเดียวกัน

จากสมมติฐานของ Fletcher แสดงว่ามนุษย์จะเริ่มได้ยินเสียงสัญญาณทดสอบเมื่อสัญญาณรบกวนที่มีแถบความถี่ใกล้เคียงกับแถบความถี่วิกฤตในช่วงความถี่ศูนย์กลางของสัญญาณทดสอบ และสัญญาณรบกวนนั้นมีกำลังเท่ากับสัญญาณทดสอบ ความกว้างของแถบความถี่วิกฤตที่ความถี่ต่ำกว่า 500 Hz ความเข้มของสัญญาณทดสอบที่ถูกบดบังจะสูงกว่าระดับความเข้มของสัญญาณรบกวนอยู่ประมาณ 17 dB ซึ่งจะหาความกว้างของแถบความถี่วิกฤต ได้ประมาณ 50 Hz

แต่ในความเป็นจริง เราสามารถได้ยินสัญญาณทดสอบที่มีความเข้มเพียงครึ่งหนึ่งหรือหนึ่งในสี่ ของกำลังของสัญญาณรบกวนในแถบความถี่ที่ใกล้เคียงกับความถี่ของสัญญาณทดสอบ ดังนั้นความกว้างของความถี่วิกฤตในย่านความถี่ต่ำกว่า 500 Hz จะมีความกว้างประมาณ 100 Hz ส่วนในย่านความถี่ที่สูงกว่า 500 Hz จะมีความกว้างของแถบความถี่วิกฤตประมาณ 20 เปอร์เซ็นต์ของความถี่ศูนย์กลางนั้น

2.7.1 การหาความกว้างของแถบความถี่วิกฤต

การหาความกว้างของแถบความถี่วิกฤต จะใช้การวัดระดับอ้างอิงโดยตรง โดยเปลี่ยนทั้งความกว้างแถบความถี่ และค่าที่เกี่ยวข้องโดยตรงกับความกว้างแถบความถี่ ในกรณีนี้ระดับของสัญญาณจะเป็นฟังก์ชันของจำนวนสัญญาณทดสอบ ซึ่งแต่ละเสียงจะมีค่าแอมพลิจูดเท่ากันซึ่งใช้ประมาณค่าความกว้างของแถบความถี่วิกฤตในช่วงความถี่ 1 kHz

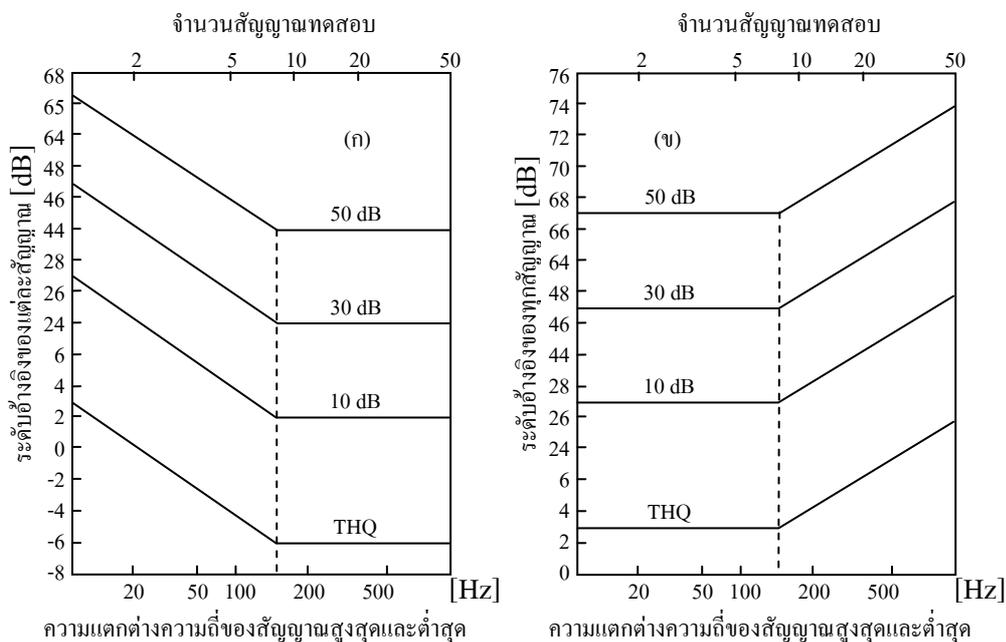


รูปที่ 2.9 ความสัมพันธ์ระหว่างระดับอ้างอิง และจำนวนสัญญาณทดสอบ

รูปที่ 2.9 แสดงระดับอ้างอิงของสัญญาณซึ่งเป็นฟังก์ชันของจำนวนของสัญญาณ และความแตกต่างของความถี่ระหว่างสัญญาณสูงสุดและสัญญาณต่ำสุด จำนวนของสัญญาณทดสอบที่ระดับอ้างอิงต่าง ๆ จะแสดงด้วยสัญลักษณ์ที่แตกต่างกัน มีการกำหนดให้ความถี่ของแต่ละสัญญาณทดสอบมีค่าความแตกต่างคงที่ โดยเท่ากับ 20 Hz ระดับอ้างอิงที่วัดจากสัญญาณทดสอบที่ความถี่ 920 Hz จะเกิดขึ้นที่ระดับความดันเสียง +3 dB เมื่อเพิ่มสัญญาณทดสอบที่ระดับเดียวกันอีกเสียงซึ่งมีความถี่ 940 Hz (เพิ่มขึ้น 20 Hz) แล้ววัดระดับอ้างอิง จะได้ระดับอ้างอิงของสัญญาณเสียงทั้งสองที่ระดับ 0 dB สำหรับสัญญาณเสียงที่มีจำนวนเสียง 4 และ 8 เสียงจะมีระดับอ้างอิงอยู่ที่ -3 dB และ -6 dB ตามลำดับ ซึ่งจะเห็นว่าระดับอ้างอิงของสัญญาณเสียงจะลดลงเมื่อจำนวนของสัญญาณเสียงเพิ่มขึ้น แต่เมื่อเพิ่มจำนวนเสียงไปถึงจุด ๆ หนึ่ง ระดับอ้างอิงจะไม่ลดลงอีก ดังจะเห็นได้จากรูปที่จำนวนเสียง 16 และ 32 เสียงนั้นระดับอ้างอิงจะไม่ต่างจากระดับอ้างอิงของสัญญาณเสียงที่มีจำนวนเสียง 8 เสียงมากนัก จะสังเกตได้ว่าระดับอ้างอิงจะลดลง 3 dB เมื่อมีการเพิ่มจำนวนสัญญาณทดสอบเป็น 2 เท่า หมายความว่าที่ระดับอ้างอิงใด ๆ ระดับแรงดันเสียงทุกเสียง จะค่อนข้างคงที่และไม่ขึ้นกับจำนวนของเสียง แต่จะใช้ได้กับสัญญาณเสียงที่มีจำนวนเสียงไม่เกิน 8 เสียงเท่านั้น ถ้ามีการเพิ่มจำนวนเสียงมากกว่านี้ระดับอ้างอิงจะไม่ลดลงตามจำนวนของสัญญาณเสียง ระดับอ้างอิงของการได้ยินของมนุษย์สามารถประมาณได้ด้วยความเข้มของสัญญาณเสียงทั้งหมด โดยที่สัญญาณเสียงเหล่านั้นอยู่ในช่วงแถบความถี่วิกฤต ส่วนที่อยู่นอกเหนือแถบความถี่นี้จะไม่มีผลต่อระดับอ้างอิงของการได้ยิน เช่น มนุษย์จะได้ยินเสียงที่ความถี่ 920 Hz เมื่อมีระดับความเข้มตั้งแต่ +3 dB หรือสัญญาณเสียงใด ๆ ที่อยู่ในช่วงแถบความถี่วิกฤตที่มีความถี่ศูนย์กลางที่ 920 Hz และมีความเข้มรวมกันมากกว่า +3 dB จากการวัดระดับอ้างอิงการคำนวณค่าความกว้างของแถบความถี่วิกฤต จะคำนวณจากจำนวนของเสียง และระยะห่างระหว่างเสียงแต่ละเสียง จากรูปจะได้ความกว้างของแถบความถี่วิกฤต เท่ากับ $(9-1)*20 \text{ Hz} = 180 \text{ Hz}$ สำหรับสัญญาณเสียงที่มีความถี่ศูนย์กลาง 920 Hz

การหาค่าความกว้างของแถบความถี่วิกฤตโดยไม่มีสัญญาณรบกวนนั้น จะได้ค่าที่ถูกต้องเมื่ออยู่ในย่านความถี่ที่ระดับอ้างอิงไม่ขึ้นกับความถี่ ซึ่งอยู่ในช่วง 500 Hz ถึง 2 kHz เท่านั้น แต่อย่างไรก็ตาม เนื่องจากสัญญาณรบกวนที่บดบังสัญญาณทดสอบที่มีลักษณะคงรูป (Uniform Masking Noise) นั้นจะทำให้ระดับอ้างอิงไม่ขึ้นกับความถี่ได้ ดังนั้นหากทำการวัดระดับอ้างอิงโดยมีสัญญาณรบกวนในลักษณะนี้จะทำให้สามารถวัดค่าระดับอ้างอิง ตลอดช่วงความถี่ที่มนุษย์สามารถได้ยินได้ ซึ่งผลของการวัดค่าระดับอ้างอิงโดยมีสัญญาณรบกวนสามารถแสดงได้ดังรูปที่ 2.10 โดยจำนวนสัญญาณเสียงที่มีระดับอ้างอิงเท่ากันจะเป็นค่าที่แสดงในแกนด้านบน ส่วนความกว้างของแถบสัญญาณที่ได้จากการเปลี่ยนแปลงจำนวนสัญญาณเสียงจะแสดงอยู่ในแกนด้านล่าง จะเห็นว่าระดับอ้างอิงของแต่ละเสียงจะลดลงในช่วงที่ความแตกต่างของความถี่มีค่าน้อยกว่าค่า ๆ

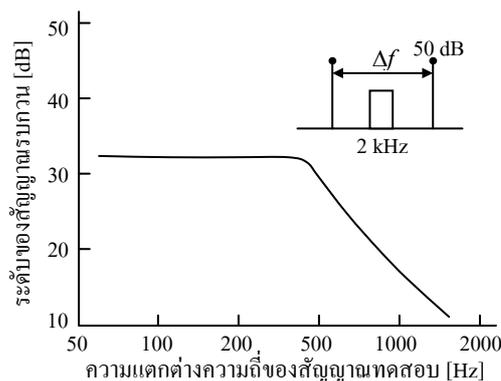
หนึ่ง และในช่วงที่ความแตกต่างของความถี่มีค่ามากกว่าค่านี้ระดับอ้างอิงจะคงที่ ในรูปที่ 2.10 (ก) จะแสดงระดับอ้างอิงของเสียงทดสอบแต่ละเสียง ในขณะที่รูป 2.10 (ข) จะแสดงระดับอ้างอิงโดยเปรียบเทียบกับระดับแรงดันเสียงรวมของทุกเสียง ซึ่งจะเห็นว่าระดับแรงดันเสียงโดยรวมนั้น จะคงที่ที่ระดับอ้างอิงในบริเวณที่ความแตกต่างของความถี่น้อยกว่าความกว้างของแถบความถี่วิกฤต และระดับแรงดันโดยรวมจะเพิ่มขึ้นเมื่อความแตกต่างของความถี่มากกว่าความกว้างของแถบความถี่วิกฤต ซึ่งแสดงให้เห็นว่าองค์ประกอบของเสียงในส่วนที่เกินจากแถบความถี่วิกฤต จะไม่มีผลต่อการวัดระดับอ้างอิง หรือระดับการบังคับของสัญญาณ



รูปที่ 2.10 ความสัมพันธ์ระหว่างระดับอ้างอิง และจำนวนสัญญาณทดสอบที่มีสัญญาณรบกวน

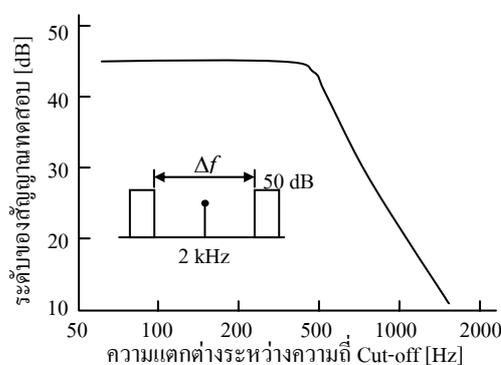
นอกจากนี้ในการหาค่าความกว้างของแถบความถี่วิกฤต อาจจะใช้สัญญาณเสียงสองความถี่ที่มีระดับสัญญาณเท่ากันทำหน้าที่บังคับสัญญาณ และให้สัญญาณรบกวนที่มีความถี่ในช่วงแคบเป็นสัญญาณทดสอบ โดยทำการวัดระดับอ้างอิงของสัญญาณทดสอบที่ถูกบังคับเทียบกับความแตกต่างของความถี่ของสัญญาณเสียงทั้งสอง ซึ่งมีสัญญาณทดสอบอยู่ที่ความถี่กึ่งกลาง ดังนั้นช่วงความถี่ของสัญญาณทดสอบจึงต้องแคบกว่าความกว้างของแถบความถี่วิกฤต ในรูปที่ 2.11 ภาพเล็กนั้นจะแสดงสัญญาณเสียงทั้งสองที่ใช้ในการบังคับสัญญาณ และสัญญาณรบกวนความถี่แคบที่ใช้เป็นสัญญาณทดสอบในแกนความถี่ ส่วนภาพใหญ่จะเป็นระดับอ้างอิงที่ถูกบังคับด้วยสัญญาณเสียง 2 เสียง ที่มีระดับ 50 dB และสัญญาณทดสอบมีความถี่ศูนย์กลางที่ 2 kHz โดยระดับอ้างอิงของสัญญาณรบกวนที่ถูกบังคับโดยสัญญาณเสียงทั้งสองนั้น จะแสดงเปรียบเทียบกับความแตกต่างของ

ความถี่ของสัญญาณเสียงทั้งสอง เมื่อความแตกต่างของความถี่ของสัญญาณเสียงมีค่าน้อย ระดับอ้างอิงจะไม่ขึ้นกับความแตกต่างของความถี่ แต่เมื่อความแตกต่างของความถี่เกินระดับหนึ่ง ระดับอ้างอิงจะลดลง ซึ่งจุดนั้นจะเป็นความกว้างของแถบความถี่วิกฤต เมื่อทำการวัดโดยเปลี่ยนความถี่ศูนย์กลางไปเรื่อย ๆ จะได้ความสัมพันธ์ระหว่างความกว้างของแถบความถี่วิกฤต กับความถี่ศูนย์กลาง



รูปที่ 2.11 ระดับอ้างอิงของสัญญาณรบกวนความถี่แคบที่ถูกบังคับโดยสัญญาณทดสอบ

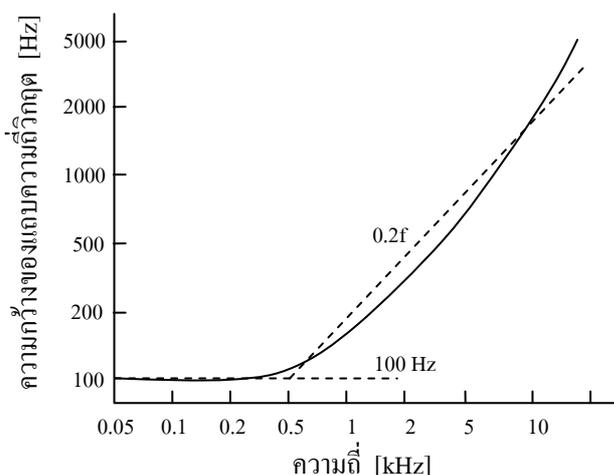
การหาค่าความกว้างของแถบความถี่วิกฤต ยังอาจใช้สัญญาณรบกวนความถี่แคบเป็นสัญญาณบังคับ แล้วใช้สัญญาณเสียงเป็นสัญญาณทดสอบ โดยให้ความถี่ที่แตกต่างกันของสัญญาณรบกวนความถี่แคบทั้งสอง เป็นความแตกต่างของความถี่ ซึ่งจะเปลี่ยนค่าไปเรื่อย ๆ เพื่อทำการวัดระดับอ้างอิง โดยใช้สัญญาณเสียงที่มีความถี่ 2 kHz เป็นสัญญาณทดสอบ และสัญญาณรบกวนทั้งสองสัญญาณที่ใช้ จะมีระดับ 50 dB โดยมีความกว้างของแถบความถี่ที่ 200 Hz ซึ่งแสดงดังรูปที่ 2.12



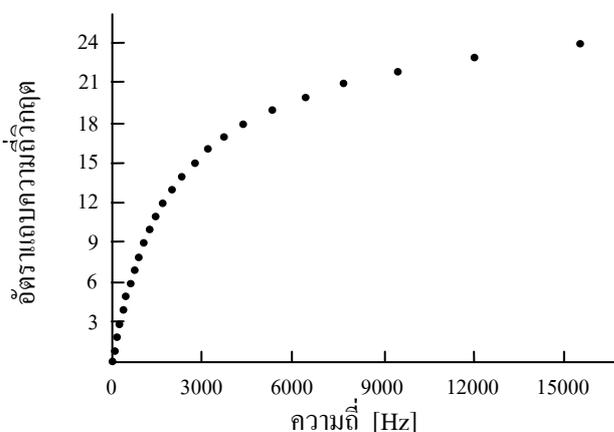
รูปที่ 2.12 ระดับอ้างอิงของสัญญาณทดสอบที่ถูกบังคับโดยสัญญาณรบกวนความถี่แคบ

2.7.2 สเตลอัตราแถบความถี่วิกฤต

จากหัวข้อที่ผ่านมาได้แสดงวิธีการต่าง ๆ ในการประมาณค่าความกว้างของแถบความถี่วิกฤต ซึ่งแถบความถี่วิกฤตแรกที่อยู่ในช่วงความถี่ต่ำสุดที่มนุษย์สามารถได้ยินเสียง จะอยู่ที่ประมาณ 20-100 Hz แต่เพื่อความสะดวกจึงรวมช่วงความถี่ 0-20 Hz ที่มนุษย์ไม่สามารถได้ยินเสียงเข้าไว้ด้วย และกำหนดให้แถบความถี่แรกของความถี่วิกฤตมีค่าเท่ากับ 0-100 Hz ดังรูปที่ 2.13 ซึ่งเป็นข้อมูลของแถบความถี่วิกฤตที่ได้จากการหาค่าเฉลี่ยของวิธีต่าง ๆ จะเห็นว่าที่ความถี่ต่ำกว่า 500 Hz นั้นแถบความถี่วิกฤตจะมีความกว้าง 100 Hz ส่วนความถี่ที่สูงกว่านั้นจะมีความกว้างประมาณ 20 เปอร์เซ็นต์ของความถี่ศูนย์กลาง ค่าความกว้างของแถบความถี่วิกฤต รวมทั้งขอบเขตบนและขอบเขตล่างที่ได้จากการทดลองสามารถแสดงได้ ดังตารางที่ 2.5



รูปที่ 2.13 ความกว้างของแถบความถี่วิกฤตเมื่อเทียบกับความถี่ของสัญญาณทดสอบ



รูปที่ 2.14 ความสัมพันธ์ระหว่างลำดับของแถบความถี่วิกฤตกับความถี่ของสัญญาณทดสอบ

ตารางที่ 2.5 แสดงความถี่ศูนย์กลาง ความกว้าง และขอบเขตของแถบความถี่วิกฤต

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

ความหมายของแถบความถี่วิกฤตนั้นมีความสำคัญต่อการอธิบายลักษณะในการได้ยินของมนุษย์ โดยใช้เป็นแบบจำลองของเสียงในรูปแบบของสเกลอัตราแถบความถี่วิกฤต ซึ่งอ้างอิงจาก

หลักความเป็นจริงเกี่ยวกับระบบการรับฟังเสียงของคน ที่แบ่งการวิเคราะห์เสียงออกเป็นส่วนต่าง ๆ ตามแถบความถี่วิกฤต และหากยึดเอาแถบความถี่วิกฤตแรกเป็นจุดอ้างอิง แล้วทำการคำนวณแถบความถี่วิกฤตที่ความถี่ต่อ ๆ ไป จะเห็นว่าจากความถี่ 0-16 kHz จะสามารถแบ่งแถบความถี่วิกฤตออกเป็น 24 แถบ โดยอันดับของแถบความถี่วิกฤตแต่ละแถบนั้น จะเรียกว่า “บาร์ก” (Bark) ในรูปที่ 2.14 จะแสดงความสัมพันธ์ระหว่างสเกลอัตราแถบความถี่วิกฤต และความถี่ ตลอดช่วงความถี่เสียงที่มนุษย์สามารถได้ยิน ซึ่งมีความสัมพันธ์กันตามสมการที่ 2.38

$$\frac{z}{Bark} = 13 \arctan\left(\frac{0.76f}{kHz}\right) + 3.5 \arctan\left(\frac{f}{7.5kHz}\right)^2 \quad (2.38)$$

และ

$$\frac{\Delta f_G}{Hz} = 25 + 75 \left[1 + 1.4 \left(\frac{f}{kHz} \right)^2 \right]^{0.69} \quad (2.39)$$

2.7.3 ความเข้มแถบความถี่วิกฤต

ความสามารถในการรับรู้เสียงในย่านความถี่ต่าง ๆ ของระบบการได้ยินของมนุษย์นั้น สามารถอธิบายได้ด้วย ความเข้มของสัญญาณเสียงที่คำนวณโดยแบ่งเป็นย่านความถี่ ตามแถบความถี่วิกฤต ซึ่งเรียกว่า ความเข้มแถบความถี่วิกฤต (Critical Band Intensity) โดยค่าความเข้มของแถบความถี่วิกฤต สามารถคำนวณได้จาก

$$I_G(f) = \int_{f-0.5\Delta f_G(f)}^{f+0.5\Delta f_G(f)} \frac{dI}{df} df \quad (2.40)$$

และสามารถเขียนได้เป็น

$$I_G(z) = \int_{z-0.5Bark}^{z+0.5Bark} \frac{dI}{dz} dz \quad (2.41)$$

สำหรับการแสดงค่าในสเกลลอการิทึม นั้น เราใช้ค่า I_0 เป็นค่าอ้างอิง ดังนั้นระดับแถบความถี่วิกฤต (Critical Band Level) จะหาได้จาก

$$L_G = 10 \cdot \log \frac{I_G}{I_0} dB \quad (2.42)$$

จากสมการจะเห็นว่าความเข้มแถบความถี่วิกฤตเป็นค่าความเข้มส่วนหนึ่งของเสียงในช่วงความถี่ของแถบความถี่วิกฤต และการเปลี่ยนจากสเกลความถี่เป็นสเกลอัตราแถบความถี่วิกฤตเป็น

การเปลี่ยนจากขนาดที่ขึ้นกับความถี่เป็นขนาดในช่วงความถี่เดียวคือ 1 บาร์ก ซึ่งไม่ขึ้นกับช่วงความถี่ใด ๆ

2.8 กระบวนการตัดสินใจแบบ K-Nearest Neighbor Rule

ภายหลังจากกระบวนการวัดค่าพารามิเตอร์แล้ว พารามิเตอร์ที่ได้จะถูกนำไปทดสอบ โดยเปรียบเทียบกับพารามิเตอร์อ้างอิง ก่อนจะเข้าสู่กระบวนการตัดสินใจเพื่อจำแนกค่าพารามิเตอร์ที่อยู่ในรูปของเวกเตอร์คุณสมบัติ (Feature Vector)

ขั้นตอนของการตัดสินใจนั้นสามารถทำได้หลายวิธีขึ้นอยู่กับลักษณะของเวกเตอร์คุณสมบัติที่มี กระบวนการตัดสินใจแบบ K-Nearest Neighbor Rule นั้นนับเป็นวิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลาย กระบวนการตัดสินใจแบบนี้จะต้องมีเวกเตอร์อ้างอิง (Training Vector) ทั้งหมด N ตัว และต้องรู้ว่าเวกเตอร์เหล่านี้เป็นสมาชิกของกลุ่ม (Class) $\omega_i, i = 1, 2, \dots, M$ ใด เมื่อมีเวกเตอร์คุณสมบัติที่ต้องการทดสอบ (Test Vector) x การจำแนกเวกเตอร์ทดสอบจะมีขั้นตอนดังนี้

1. หาค่าระยะทางระหว่างเวกเตอร์ x กับเวกเตอร์อ้างอิงแต่ละตัว จำนวน N ตัว
 2. หาเวกเตอร์อ้างอิงจำนวน k ตัว ซึ่งมีระยะทางระหว่างเวกเตอร์อ้างอิงกับเวกเตอร์ทดสอบสั้นที่สุด จากเวกเตอร์อ้างอิงนี้จะได้กลุ่มของเวกเตอร์อ้างอิงเหล่านั้นด้วย โดยค่า k ที่ใช้มักจะเป็นเลขคี่ เพื่อป้องกันปัญหาในการตัดสินใจภายหลัง และไม่ควรเป็นจำนวนเท่าของจำนวนกลุ่มทั้งหมด M
 3. จากเวกเตอร์อ้างอิงที่ได้จำนวน k ตัว จะต้องหาจำนวนเวกเตอร์ที่อยู่ในกลุ่ม ω_i จำนวน k_i ตัว โดยที่ $\sum_i k_i = k$
 4. จะได้ว่าเวกเตอร์ทดสอบ x นั้นเป็นสมาชิกของกลุ่ม ω_i ที่มีจำนวน k_i สูงที่สุด
- ค่าระยะทางระหว่าง เวกเตอร์ทดสอบกับเวกเตอร์อ้างอิงแต่ละตัวนั้น อาจจะใช้ค่าระยะทางแบบ Euclidean ซึ่งเท่ากับ

$$\text{Euclidean Distanced: } d_e = |x - \tilde{x}_i| \quad (2.43)$$

หรือใช้ค่าระยะทางแบบ Mahalanobis

$$\text{Mahalanobis Distance: } d_e = \sqrt{(x - \tilde{x}_i)^T (x - \tilde{x}_i)} \quad (2.44)$$

โดยที่ \tilde{x}_i คือเวกเตอร์อ้างอิงแต่ละตัว

ในกรณีที่ $k = 1$ ซึ่งเป็นกระบวนการตัดสินใจที่เรียกว่า Nearest Neighbor Rule เวกเตอร์ทดสอบ x จะถูกจัดให้อยู่ในกลุ่มของเวกเตอร์อ้างอิงที่ใกล้ที่สุด การตัดสินใจแบบนี้จะได้ผลที่ดีเมื่อมีจำนวนเวกเตอร์อ้างอิงมากพอ ซึ่งหาก $N \rightarrow \infty$ แล้ว ความน่าจะเป็นของความผิดพลาดในการจำแนกของการตัดสินใจตามกฎ NN หรือ P_{NN} นั้นจะมีค่าอยู่ระหว่าง

$$P_B \leq P_{NN} \leq P_B \left(2 - \frac{M}{M-1} P_B \right) \leq 2P_B \quad (2.45)$$

โดยที่ P_B คือค่าต่ำสุดของค่าผิดพลาด Bayesian จะเห็นว่าค่าผิดพลาดสูงสุดของการตัดสินใจตามกฎ NN คือ 2 เท่าของ P_B ในขณะที่การตัดสินใจตามกฎ kNN นั้นจะให้ผลที่ดีกว่า เช่น เมื่อ $k = 2$ จะได้ว่าค่าผิดพลาดของแบบ kNN นั้นจะมีค่าระหว่าง

$$P_B \leq P_{kNN} \leq P_B + \frac{1}{\sqrt{ke}} \quad \text{หรือ} \quad P_B \leq P_{kNN} \leq P_B + \sqrt{\frac{2P_{NN}}{k}} \quad (2.46)$$

จะเห็นว่าเมื่อ $k \rightarrow \infty$ แล้ว การตัดสินใจตามกฎ kNN นั้นจะได้ค่าที่ดีที่สุด นอกจากนี้ถ้าค่าผิดพลาด Bayesian มีค่าน้อย ๆ จะได้ว่า

$$P_{NN} \approx 2P_B \quad (2.47)$$

$$P_{3NN} \approx 2P_B + 3P_B^2 \quad (2.48)$$

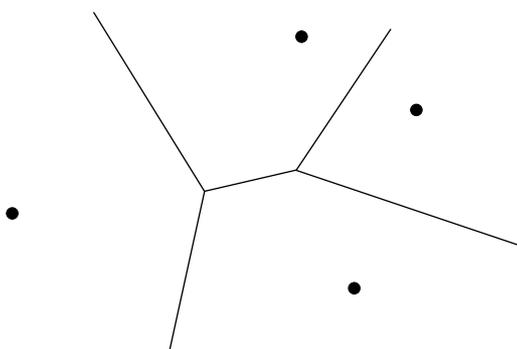
ดังนั้นเมื่อ $N \rightarrow \infty$ มีค่ามากและค่าผิดพลาด Bayesian มีค่าน้อยการจำแนกแบบ 3NN จะได้ผลใกล้เคียงกับการจำแนกแบบ Bayesian เช่น ถ้าความผิดพลาด Bayesian เท่ากับ 1.0 % แล้วความผิดพลาดของ 3NN จะเท่ากับ 1.03%

สมมติว่า N มีค่ามาก รูปทรงที่มีจุดศูนย์กลางเป็น x จะมีรัศมีที่มีความต่อเนื่อง โดยมีตัวอย่างที่มีระยะห่างน้อยที่สุด k ตัวอย่างรอบ ๆ จุดศูนย์กลางนี้ ตัวอย่างที่อยู่รอบ ๆ เหล่านี้จะอยู่ใกล้กับตัวอย่าง x มากเมื่อเทียบกับตัวอย่างอ้างอิงทั้งหมด และถ้า k มีค่ามาก ตัวอย่างที่อยู่รอบ ๆ ตัวอย่างทดสอบ x ก็จะเป็นตัวอย่างที่เป็นสมาชิกของกลุ่มที่มีความน่าจะเป็นของเงื่อนไขสูงที่สุด

สำหรับกรณีที่มีการใช้การตัดสินใจแบบ Nearest Neighbor โดยที่ $k = 1$ เวกเตอร์คุณสมบัติอ้างอิง $x_i, i = 1, 2, \dots, N$ จะสร้างส่วนแบ่งของพื้นที่ที่เป็น l มิติ หรือ R_l จำนวน N พื้นที่ซึ่งแต่ละพื้นที่จะนิยามว่า

$$R_i = \{x : d(x, x_i) < d(x, x_j), i \neq j\} \quad (2.49)$$

ซึ่งพื้นที่ R_i นี้จะประกอบไปด้วยจุดทุกจุดในพื้นที่ที่ใกล้กับ x_i มากกว่าจุดอื่น ๆ ของตัวอย่างอ้างอิง เมื่อเทียบตามระยะทาง d ซึ่งส่วนแบ่งของพื้นที่ที่คุณสมบัตินี้เรียกว่า Voronoi Tessellation ตัวอย่างของพื้นที่ Voronoi Tessellation ในกรณีที่ $l = 2$ สามารถแสดงได้ดังรูปที่ 2.15



รูปที่ 2.15 ตัวอย่างของพื้นที่ Voronoi Tessellation

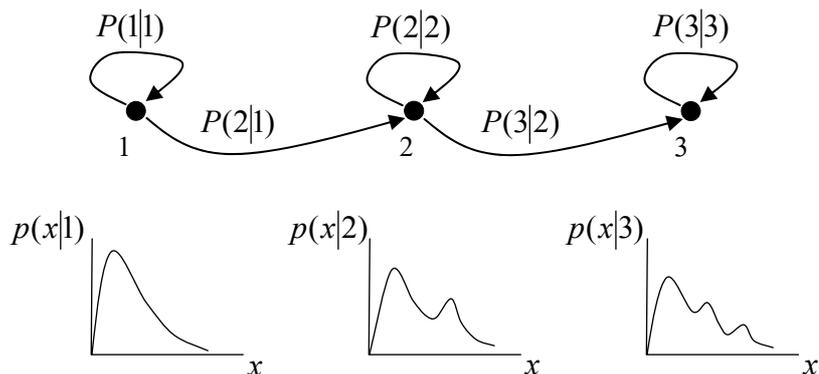
แม้ว่าเทคนิค (k)NN นี้จะให้ผลของการตัดสินใจได้ค่อนข้างดี แต่เทคนิคนี้ก็มีความสำคัญคือความยุ่งยากซับซ้อนในการหาตัวอย่างที่ใกล้จุด x ที่สุดจากตัวอย่างอ้างอิงทั้งหมด N ตัวอย่างซึ่งจำนวนครั้งในการคำนวณจะเป็นสัดส่วนโดยตรงกับค่า kN และหากมิติของเวกเตอร์คุณสมบัติมีค่ามากก็จะยิ่งทำให้ปัญหานี้มีความซับซ้อนยิ่งขึ้นด้วย

2.9 กระบวนการตัดสินใจแบบ Hidden Markov Modeling

กระบวนการตัดสินใจแบบ Hidden Markov Modeling เป็นการสร้างแบบจำลองชนิดหนึ่งซึ่งเหมาะสมกับลำดับที่ไม่คงที่และคาดการณ์ล่วงหน้าไม่ได้ (Non-stationary Stochastic Sequence) โดย HMM จะสร้างแบบจำลองของลำดับการสังเกต (Observation) ของกระบวนการที่คงที่บางส่วน (Piecewise Stationary Process) จึงสามารถนำมาใช้ในกระบวนการรู้จำเสียงพูด โดยสร้างแบบจำลองของการเปล่งเสียง ซึ่งอาจจะเป็นคำ บางส่วนของคำ หรือแม้กระทั่งประโยคของคำพูดก็ได้

แบบจำลอง HMM เป็นกลไกของลำดับที่คาดการณ์ล่วงหน้าไม่ได้ที่มีจำนวนสถานะจำกัด (Stochastic Finite State Automaton) ซึ่งสร้างแถวของการสังเกตที่อยู่ในรูปของเวกเตอร์ของการสังเกต x_1, x_2, \dots, x_N แบบจำลอง HMM จึงประกอบด้วยจำนวนสถานะ K สถานะ ซึ่งเป็นผลมา

จากการเปลี่ยนสถานะที่อยู่ติดกันจากสถานะ i ไปยังสถานะ j รูปที่ 2.16 แสดงแผนภาพของ HMM ที่มี 3 สถานะโดยลูกศรจะแสดงทิศทางของการเปลี่ยนสถานะ ซึ่งแบบจำลองนี้จะให้ข้อมูลของการเปลี่ยนแปลงระหว่างสถานะ ($P(i|j), i, j = 1, 2, 3$) และข้อมูลสถิติคงที่ของแต่ละสถานะ ($p(x|i), i = 1, 2, 3$) แบบจำลอง HMM ในลักษณะนี้เป็นแบบซ้ายไปขวา (Left to Right) ซึ่งไม่มีการเปลี่ยนสถานะไปยังสถานะที่ต่ำกว่า



รูปที่ 2.16 พารามิเตอร์ของแบบจำลอง Hidden Markov Model ที่มี 3 สถานะ

หากกำหนดให้มีรูปแบบอ้างอิงที่รู้ค่า M รูปแบบ กระบวนการรู้จำนั้นจะต้องจำแนกเวกเตอร์ทดสอบที่ไม่ทราบค่า โดยเปรียบเทียบว่าเวกเตอร์นี้มีรูปแบบที่เข้ากับรูปแบบอ้างอิงใดมากที่สุด โดยรูปแบบอ้างอิงนี้จะอยู่ในรูปของแบบจำลอง HMM ซึ่งแต่ละรูปแบบของข้อมูลอ้างอิงทั้งหมดจะประกอบด้วยพารามิเตอร์ต่าง ๆ ดังนี้

- จำนวนสถานะของแต่ละรูปแบบ $K_s, s = 1, 2, \dots, M$
- ความหนาแน่นของความน่าจะเป็น (Probability Density) $p(x|j), j = 1, 2, \dots, K_s$ ซึ่งแสดงการกระจายตัวของการสังเกตที่ออกจากสถานะ j

- ความน่าจะเป็นของการเปลี่ยนสถานะ $p(i|j), i, j = 1, 2, \dots, K_s$ ระหว่างสถานะต่าง ๆ
- ความน่าจะเป็นของสถานะเริ่มต้น $p(i), i = 1, 2, \dots, K_s$

กระบวนการรู้จำแบบ HMM นั้นประกอบด้วย 2 ส่วน ส่วนของกระบวนการรู้จำ (Recognition Process) ซึ่งจะเปรียบเทียบและหาว่ารูปแบบทดสอบจะเข้ากับรูปแบบอ้างอิงใดได้มากที่สุด และกระบวนการสร้างรูปแบบอ้างอิง (Training Process) เพื่อหาค่าพารามิเตอร์ต่าง ๆ ของแบบจำลอง HMM ดังที่กล่าวมาแล้ว

1. กระบวนการรู้จำ ในขั้นตอนนี้จะเป็นการหาว่าลำดับของการสังเกต $X : x_1, x_2, \dots, x_N$ ทั้งหมด N ตัวอย่างซึ่งเป็นผลของเวกเตอร์ที่ไม่ทราบรูปแบบ ควรจะถูกจัดให้อยู่กลุ่มใดจาก

แบบจำลอง HMM ที่ใช้อ้างอิงทั้งหมด M รูปแบบ โดยรูปแบบ S^* ที่ตรงกับเวกเตอร์นั้นจะหาได้จาก

$$S^* = \arg \max_S P(S|X) \quad (2.50)$$

และสำหรับแบบจำลองอ้างอิงที่มีความน่าจะเป็นเท่ากัน จะได้ว่า

$$S^* = \arg \max_S p(X|S) \quad (2.51)$$

เพื่อความสะดวกเราจะกำหนดให้ $S = \{P(i|j), p(x|i), P(i), K_s\}$ แทนเซตของค่าพารามิเตอร์ที่อธิบายแบบจำลอง HMM แต่ละตัว ซึ่งจะมีการเปลี่ยนสถานะที่อยู่ติดกันที่เป็นไปได้ Ω_i มากกว่า 1 ค่า และการเปลี่ยนสถานะนี้มีความน่าจะเป็นของการเกิด $P(\Omega_i|S)$ ดังนั้นจะได้ว่า

$$p(X|S) = \sum_i p(x, \Omega_i|S) = \sum_i p(x|\Omega_i, S)P(\Omega_i|S) \quad (2.52)$$

ในการหาค่า $p(X|S)$ ที่มากที่สุดจากแบบจำลองที่เป็นไปได้ทั้งหมดจะต้องทำการคำนวณสมการที่ 2.52 กับแบบจำลองแต่ละตัว ซึ่งมีทั้งหมด M ตัวโดยใช้อัลกอริทึมของ Viterbi ซึ่งจะกำหนดค่าความหนาแน่นของความน่าจะเป็นของเหตุการณ์ที่ต่อเนื่องกัน หรือ $\alpha(i_k)$ โดยมี (1) เส้นทางที่สถานะ $i_k (i_k \in \{1, 2, \dots, K_s\})$ ที่ขั้นตอน k (2) การสังเกต x_1, x_2, \dots, x_{k-1} ที่มีการเปลี่ยนแปลงจนถึงสถานะก่อนหน้านี้ และ (3) การสังเกต x_k ที่ถูกส่งมาจากสถานะ i_k ที่ขั้นตอน k จากการกำหนดค่า $\alpha(i_k)$ จะได้ความสัมพันธ์แบบวนซ้ำดังนี้

$$\begin{aligned} \alpha(i_{k+1}) &\equiv p(x_1, x_2, \dots, x_{k+1}, i_{k+1}|S) \\ &= \sum_{i_k} \alpha(i_k) P(i_{k+1}|i_k) p(x_{k+1}|i_{k+1}), k = 1, 2, \dots, N-1 \end{aligned} \quad (2.53)$$

โดยที่

$$\alpha(i_1) = P(i_1) p(x_1|i_1) \quad (2.54)$$

ผลคูณของ $P(i_{k+1}|i_k)p(x_{k+1}|i_{k+1})$ ในสมการที่ 2.53 จะได้ข้อมูลเฉพาะของการเปลี่ยนสถานะสุดท้าย และ $\alpha(i_k)$ จะเป็นข้อมูลสะสมของเส้นทางจนถึงขั้นตอน k จะเห็นว่าในการกำหนด $\alpha(i_k)$ นั้นจะไม่ขึ้นกับลำดับการสังเกต x_{k+1}, \dots, x_N

กำหนดให้ $\beta(i_k)$ เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็นจากการสังเกต x_{k+1}, \dots, x_N ที่เกิดขึ้นที่ขั้นตอน $k+1, \dots, N$ โดยให้ที่ขั้นตอน k เส้นทางของสถานะจะอยู่ที่ i_k จะได้สมการวนซ้ำของ $\beta(i_k)$ เป็น

$$\begin{aligned}\beta(i_k) &\equiv p(x_{k+1}, x_{k+2}, \dots, x_N | i_k, S) \\ &= \sum_{i_{k+1}} \beta(i_{k+1}) P(i_{k+1} | i_k) p(x_{k+1} | i_{k+1}), k = N-1, N-2, \dots, 1\end{aligned}\quad (2.55)$$

โดยที่

$$\beta(i_N) = 1, i_N \in \{1, 2, \dots, K_s\} \quad (2.56)$$

ดังนั้นความหนาแน่นของความน่าจะเป็นของเหตุการณ์ที่ต่อเนื่องกันที่มีสถานะ i_k ที่ขั้นตอน k และมีลำดับการสังเกต x_1, \dots, x_N จะได้จาก

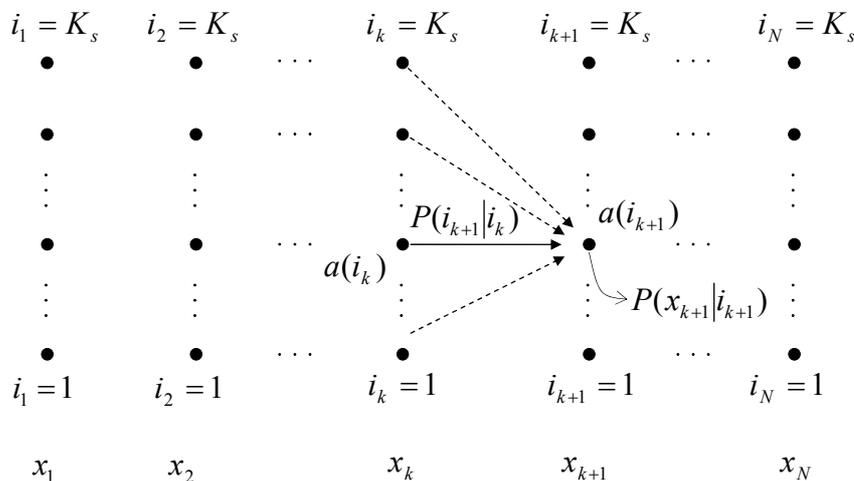
$$\begin{aligned}\gamma(i_k) &\equiv p(x_1, x_2, \dots, x_N, i_k | S) \\ &= p(x_1, x_2, \dots, x_k, i_k | S) p(x_{k+1}, x_{k+2}, \dots, x_N | i_k, S) \\ &= \alpha(i_k) \beta(i_k)\end{aligned}\quad (2.57)$$

จากจุดเริ่มต้นของการคำนวณที่ต้องการหาค่า $p(X|S)$ ที่มากที่สุด จากสมการที่ 2.52 จะสามารถเขียนได้เป็น

$$p(X|S) = \sum_{i_N=1}^{K_s} \alpha(i_N) \quad (2.58)$$

จากการคำนวณของสมการที่ 2.58 จะต้องคำนวณค่า $\alpha(i_k), k=1, 2, \dots, N$ ทุกค่า ซึ่งสามารถแสดงได้จากแผนภาพในรูปที่ 2.17 แต่ละจุดซึ่งอยู่ในขั้นตอน k และมีสถานะ $i_k, i_k = 1, 2, \dots, K_s$ จะมีการคำนวณ $\alpha(i_k)$ ตามสมการที่ 2.53 ดังนั้นจำนวนการคำนวณทั้งหมดจะเท่ากับ NK_s^2 ในกระบวนการตัดสินใจจะต้องคำนวณเพื่อหาค่า $p(X|S)$ ตามสมการที่ 2.58

ของแต่ละแบบจำลอง และแถวของรูปแบบการสังเกตที่ไม่ทราบกลุ่มจะถูกจำแนกให้อยู่ในกลุ่มของแบบจำลองอ้างอิง S ที่มีค่า $p(X|S)$ สูงที่สุด



รูปที่ 2.17 แผนภาพแสดงทิศทางการคำนวณ

2. กระบวนการสร้างรูปแบบอ้างอิง เริ่มต้นโดยกำหนดให้ลำดับของการสังเกตของเวกเตอร์อ้างอิง $x_k, k = 1, 2, \dots, N$ นั้น ประกอบด้วยเวกเตอร์ที่แควนไตซ์แล้ว ซึ่งแต่ละเวกเตอร์ของการสังเกตจะมีค่าเพียงค่าเดียวจากค่าซึ่งแตกต่างกันที่เป็นไปได้ทั้งหมด L ค่า ในพื้นที่ขนาด L มิติ ดังนั้นการสังเกตจะอธิบายได้ด้วยเลขจำนวนเต็ม $r, r = 1, 2, \dots, L$ ก่อนที่จะเริ่มทำการคำนวณ จะต้องมีนิยามค่าต่าง ๆ ดังนี้

- $\xi_k(i, j, X|S) \equiv$ ความน่าจะเป็นของเหตุการณ์ที่ต่อเนื่อง โดยมี (1) เส้นทางที่ผ่านสถานะ i ที่ขั้นตอน k (2) เส้นทางที่ผ่านสถานะ j ที่ขั้นตอนต่อมา $k + 1$ และ (3) แบบจำลองที่สร้างลำดับที่เป็นไปได้ของการสังเกต X จากค่าพารามิเตอร์ของแบบจำลอง S

- $\gamma_k(i, X|S) \equiv$ ความน่าจะเป็นของเหตุการณ์ที่ต่อเนื่อง โดยมี (1) เส้นทางที่ผ่านสถานะ i ที่ขั้นตอน k และ (2) แบบจำลองที่สร้างลำดับที่เป็นไปได้ของการสังเกต X จากค่าพารามิเตอร์ของแบบจำลอง S

จากการนิยามข้างต้น จะได้ว่า

$$\xi_k(i, j) \equiv \xi_k(i, j|X, S) = \frac{\xi_k(i, j, X|S)}{P(X|S)} \tag{2.59}$$

และจากสมการที่ 2.53 และ 2.55 จะเขียนสมการที่ 2.59 เป็น

$$\xi_k(i, j) = \frac{\alpha(i_k = i)P(j|i)P(x_{k+1}|j)\beta(i_{k+1} = j)}{P(X|S)} \quad (2.60)$$

โดยที่ $\alpha(i_k = i)$ จะคิดจากเส้นทางที่ผ่านมาโดยสิ้นสุดที่ขั้นตอน k ที่สถานะ i ส่วน $\beta(i_{k+1} = j)$ คิดจากส่วนข้างหน้าของเส้นทางที่ขั้นตอน $k+1$ ที่สถานะ j และผลคูณของ $P(j|i)P(x_{k+1}|j)$ จะคิดจากข้อมูลที่ขั้นตอน k นอกจากนี้ยังมีข้อมูลอื่น ๆ ที่น่าสนใจซึ่งได้แก่ความน่าจะเป็นของการอยู่ที่ขั้นตอน k ที่สถานะ i ตามแบบจำลองและลำดับการสังเกตที่ได้ ซึ่งเท่ากับ

$$\gamma_k(i) \equiv \gamma_k(i|X, S) = \frac{\alpha(i_k = i)\beta(i_k = i)}{P(X|S)} \quad (2.61)$$

สำหรับการคำนวณไปข้างหน้าจะพบว่า

- $\sum_{k=1}^N \gamma_k(i)$ คือจำนวนครั้งที่คาดว่าสถานะ i จะเกิดขึ้นจากแบบจำลอง S และลำดับการสังเกต X ที่ได้ เมื่อตัวชี้ด้านบนของผลบวกเท่ากับ คำนี้นี้จะเป็นจำนวนครั้งที่คาดว่าจะมีการเปลี่ยนสถานะจากสถานะ i
- $\sum_{k=1}^{N-1} \xi_k(i, j)$ คือจำนวนครั้งที่คาดว่าจะมีการเปลี่ยนสถานะจากสถานะ i ไปเป็นสถานะ j จากแบบจำลอง S และลำดับการสังเกต X ที่ได้

จากนิยามข้างต้น จะสามารถนำไปสร้างสมการของการประมาณค่าพารามิเตอร์แบบจำลองที่ไม่ทราบค่าได้

$$\bar{P}(j|i) = \frac{\sum_{k=1}^{N-1} \xi_k(i, j)}{\sum_{k=1}^{N-1} \gamma_k(i)} \quad (2.62)$$

$$\bar{P}_x(r|i) = \frac{\sum_{(k=1 \text{ and } x \rightarrow r)}^N \gamma_k(i)}{\sum_{k=1}^N \gamma_k(i)} \quad (2.63)$$

$$\bar{P}(i) = \gamma_1(i) \quad (2.64)$$

การคำนวณของสมการที่ 2.63 จะรวมเฉพาะค่า $\gamma_k(i)$ ที่สอดคล้องกับการสังเกต x_k ที่ใช้ค่าที่ไม่ต่อเนื่อง r^{th} สำหรับอัลกอริทึมที่จะต้องทำซ้ำ ๆ นั้น จะมีขั้นตอนดังนี้

1. กำหนดเงื่อนไขเริ่มต้นสำหรับปริมาณที่ไม่ทราบค่า แล้วคำนวณค่า $P(X|S)$

2. ค่าพารามิเตอร์ของแบบจำลองที่ได้จากการประมาณในแต่ละครั้ง จะต้องนำไปประมาณค่าแบบจำลอง \bar{S} ใหม่ โดยใช้สมการที่ 2.62 ถึง 2.64
3. คำนวนค่า $P(X|\bar{S})$ ถ้า $P(X|\bar{S}) - P(X|S) > \varepsilon$ โดยกำหนดให้ $S = \bar{S}$ แล้วกลับไปประมาณค่าพารามิเตอร์ในขั้นตอนที่ 2 หรือหยุดคำนวณถ้า $P(X|\bar{S}) - P(X|S) \leq \varepsilon$

บทที่ 3

การรู้จำเสียงพยัญชนะต้นควบสระ

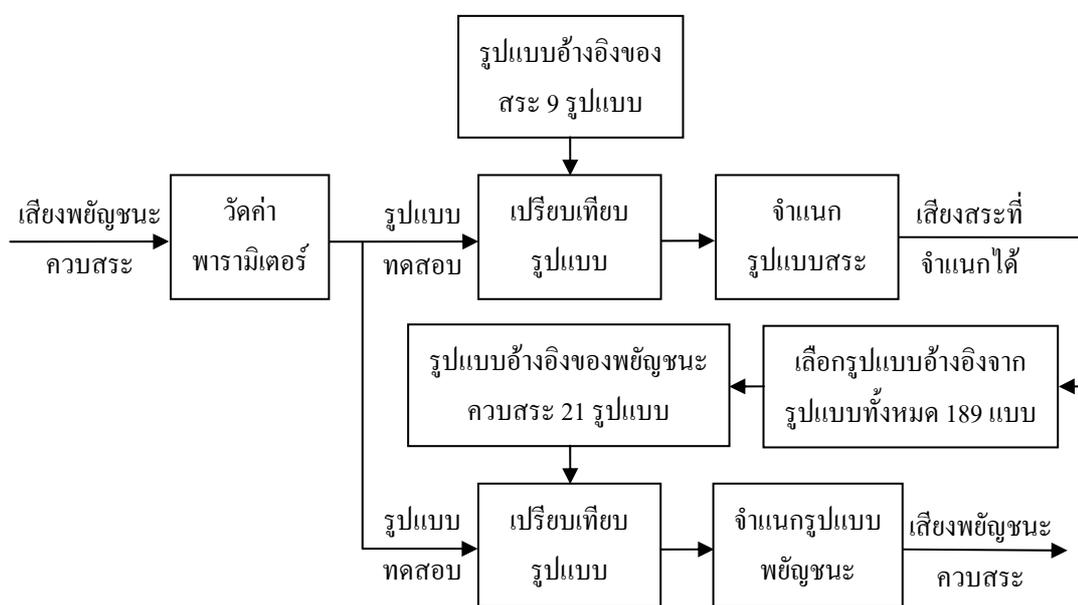
3.1 หลักการรู้จำเสียงพยัญชนะต้นควบสระ

เสียงพยัญชนะต้นควบสระสำหรับงานวิจัยนี้เกิดจากการรวมกันของเสียงพยัญชนะต้นเดี่ยวทั้งหมด 21 เสียง กับสระเสียงสั้นทั้งหมด 9 เสียง ดังนั้นจำนวนของรูปแบบของกระบวนการรู้จำเสียงพยัญชนะต้นควบสระจึงมีทั้งหมด 189 เสียง ซึ่งแสดงได้ดังตารางที่ 3.1

ตารางที่ 3.1 เสียงพยัญชนะต้นควบสระในภาษาไทย

สระ \ พยัญชนะ	อะ /a-/	อิ /i-/	อี /v-/	อุ /u-/	เอะ /e-/	แอะ /x-/	เออะ /q-/	โอะ /o-/	เอะ /@-/
ก /k-/	/k-a-/	/k-i-/	/k-v-/	/k-u-/	/k-e-/	/k-x-/	/k-q-/	/k-o-/	/k-@-/
ข /kh-/	/kh-a-/	/kh-i-/	/kh-v-/	/kh-u-/	/kh-e-/	/kh-x-/	/kh-q-/	/kh-o-/	/kh-@-/
ง /ng-/	/ng-a-/	/ng-i-/	/ng-v-/	/ng-u-/	/ng-e-/	/ng-x-/	/ng-q-/	/ng-o-/	/ng-@-/
จ /c-/	/c-a-/	/c-i-/	/c-v-/	/c-u-/	/c-e-/	/c-x-/	/c-q-/	/c-o-/	/c-@-/
ช /ch-/	/ch-a-/	/ch-i-/	/ch-v-/	/ch-u-/	/ch-e-/	/ch-x-/	/ch-q-/	/ch-o-/	/ch-@-/
ด /d-/	/d-a-/	/d-i-/	/d-v-/	/d-u-/	/d-e-/	/d-x-/	/d-q-/	/d-o-/	/d-@-/
ต /t-/	/t-a-/	/t-i-/	/t-v-/	/t-u-/	/t-e-/	/t-x-/	/t-q-/	/t-o-/	/t-@-/
ท /th-/	/th-a-/	/th-i-/	/th-v-/	/th-u-/	/th-e-/	/th-x-/	/th-q-/	/th-o-/	/th-@-/
น /n-/	/n-a-/	/n-i-/	/n-v-/	/n-u-/	/n-e-/	/n-x-/	/n-q-/	/n-o-/	/n-@-/
บ /b-/	/b-a-/	/b-i-/	/b-v-/	/b-u-/	/b-e-/	/b-x-/	/b-q-/	/b-o-/	/b-@-/
ป /p-/	/p-a-/	/p-i-/	/p-v-/	/p-u-/	/p-e-/	/p-x-/	/p-q-/	/p-o-/	/p-@-/
พ /ph-/	/ph-a-/	/ph-i-/	/ph-v-/	/ph-u-/	/ph-e-/	/ph-x-/	/ph-q-/	/ph-o-/	/ph-@-/
ฟ /f-/	/f-a-/	/f-i-/	/f-v-/	/f-u-/	/f-e-/	/f-x-/	/f-q-/	/f-o-/	/f-@-/
ม /m-/	/m-a-/	/m-i-/	/m-v-/	/m-u-/	/m-e-/	/m-x-/	/m-q-/	/m-o-/	/m-@-/
ย /y-/	/y-a-/	/y-i-/	/y-v-/	/y-u-/	/y-e-/	/y-x-/	/y-q-/	/y-o-/	/y-@-/
ร /r-/	/r-a-/	/r-i-/	/r-v-/	/r-u-/	/r-e-/	/r-x-/	/r-q-/	/r-o-/	/r-@-/
ล /l-/	/l-a-/	/l-i-/	/l-v-/	/l-u-/	/l-e-/	/l-x-/	/l-q-/	/l-o-/	/l-@-/
ว /w-/	/w-a-/	/w-i-/	/w-v-/	/w-u-/	/w-e-/	/w-x-/	/w-q-/	/w-o-/	/w-@-/
ส /s-/	/s-a-/	/s-i-/	/s-v-/	/s-u-/	/s-e-/	/s-x-/	/s-q-/	/s-o-/	/s-@-/
ห /h-/	/h-a-/	/h-i-/	/h-v-/	/h-u-/	/h-e-/	/h-x-/	/h-q-/	/h-o-/	/h-@-/
อ /z-/	/z-a-/	/z-i-/	/z-v-/	/z-u-/	/z-e-/	/z-x-/	/z-q-/	/z-o-/	/z-@-/

กระบวนการรู้จำเสียงพยัญชนะต้นควบสระนั้น มีขั้นตอนหลักเหมือนกับระบบการรู้จำเสียงพูดทั่วไป คือ การหาค่าพารามิเตอร์ เปรียบเทียบพารามิเตอร์ดังกล่าวกับรูปแบบอ้างอิง และ จำแนกตัวอย่างทดสอบตามรูปแบบอ้างอิงที่เหมาะสม แต่เนื่องจากรูปแบบของพยัญชนะต้นควบสระที่ต้องจำแนกนั้นมีจำนวนมาก ในขั้นตอนของกระบวนการรู้จำจึงจำเป็นต้องแบ่งเป็น 2 ขั้นตอนย่อย คือการรู้จำเสียงสระจากรูปแบบทั้งหมด 9 รูปแบบ แล้วนำสระที่ได้ไปใช้ในการเลือกรูปแบบของพยัญชนะ ทำให้รูปแบบของพยัญชนะที่ต้องทำการจำแนกจะเหลือเพียง 21 รูปแบบ ขั้นตอนของกระบวนการรู้จำเสียงพยัญชนะต้นควบสระ แสดงได้ดังรูปที่ 3.1



รูปที่ 3.1 ขั้นตอนในกระบวนการรู้จำเสียงพยัญชนะควบสระ

3.2 ฐานข้อมูลเสียงพูด

ฐานข้อมูลเสียงพูดที่ใช้ในการวิจัยนี้ จะใช้ฐานข้อมูล LOTUS Corpus (Large vOcabulary Thai continUous Speech recognition Corpus) ซึ่งพัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) เพื่อใช้ในการวิจัยและพัฒนาระบบรู้จำเสียงพูดต่อเนื่อง (Large Vocabulary Continuous Speech Recognition: LVCSR) สำหรับภาษาไทย ซึ่งฐานข้อมูลนี้ครอบคลุมคำศัพท์ภาษาไทย จำนวนไม่ต่ำกว่า 5,000 คำ จากฐานข้อมูลบทความข่าวหรือบทความทั่วไป

ฐานข้อมูล LOTUS Corpus นี้บันทึกเสียงผู้พูดทั้งหมด 248 คน เป็นชาย 124 คน และหญิง 124 คน โดยใช้ไมโครโฟน 2 ชนิดคือ ไมโครโฟน Cross-talk ที่มีคุณภาพสูงและไมโครโฟน

Unidirectional ที่มีคุณภาพปานกลาง และทำการบันทึกเสียงใน 2 สภาพแวดล้อม คือ สภาพแวดล้อมแบบห้องเงียบ และสภาพแวดล้อมแบบสำนักงาน

เสียงพูดจะถูกบันทึกจากไมโครโฟน SONY F-720 Dynamic Microphone แล้วบันทึกลงเครื่องคอมพิวเตอร์ด้วยโปรแกรม Cool Edit 2000 โดยมีอัตราการสุ่มตัวอย่าง 16 kHz ที่ความละเอียด 16 บิต แบบ Stereo Windows PCM โดยใช้ High Quality ที่ 999 และใช้ Pre/Post Filter

ฐานข้อมูล LOTUS Corpus นี้สามารถดาวน์โหลดไฟล์สัญญาณเสียงในรูปแบบ .wav และไฟล์กำกับฐานข้อมูลได้จากเว็บไซต์ <http://www.nectec.or.th/rdi/lotus>

งานวิจัยนี้ได้เลือกตัวอย่างเสียงจากฐานข้อมูลเสียง LOTUS Corpus เพื่อใช้ทดสอบเสียงพยัญชนะต้นควบสระ 189 กลุ่ม โดยแต่ละกลุ่มจะใช้ตัวอย่างเสียงกลุ่มละไม่เกิน 80 เสียง จากตัวอย่างเสียงของผู้พูดที่เป็นชาย 12 คน และเป็นหญิง 12 คน และเลือกเสียงที่บันทึกด้วยไมโครโฟนแบบ Crosstalk ที่บันทึกในสภาพแวดล้อมแบบห้องเงียบ และสภาพแวดล้อมแบบสำนักงาน

เนื่องจากฐานข้อมูลเสียง LOTUS Corpus เป็นฐานข้อมูลเพื่อการพัฒนาการรู้จำเสียงทั่วไป ซึ่งมีตัวอย่างของเสียงตามจำนวนคำศัพท์ที่ใช้ทั่วไปในชีวิตประจำวัน จึงไม่ได้ครอบคลุมเสียงพยัญชนะต้นควบสระทุกเสียง บางเสียงอาจจะมีตัวอย่างเสียงไม่มากนัก โดยเฉพาะเสียง /c-q-, /ng-e-, /ng-i-, /ng-v- และ /w-v- ซึ่งพบได้น้อยในคำศัพท์ภาษาไทย จึงไม่มีตัวอย่างเสียงดังกล่าวในฐานข้อมูล และไม่ได้อยู่รวมอยู่ในการศึกษาของงานวิจัยนี้ ดังนั้นในงานวิจัยนี้จึงมีกลุ่มของเสียงพยัญชนะต้นควบสระทั้งหมด 184 กลุ่ม นอกจากนี้เพื่อลดความซับซ้อนของปัญหาในการวิเคราะห์เสียงสระ ในงานวิจัยยังเลือกใช้เสียงพยางค์ที่มีหน่วยเสียงสระเป็นเสียงสระเดี่ยวเท่านั้น

3.3 เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียง

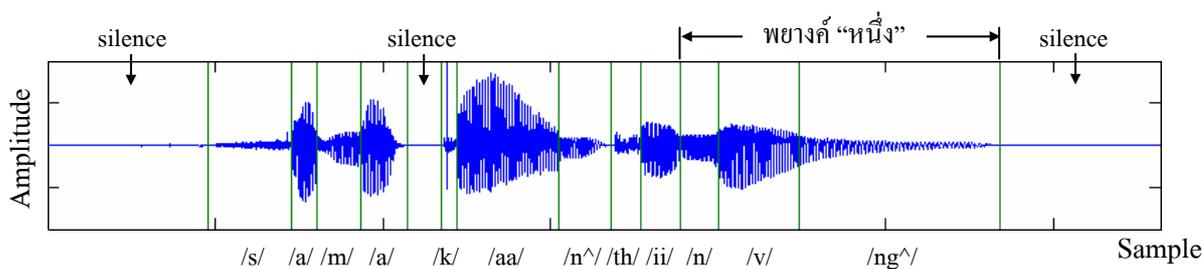
1. คอมพิวเตอร์ส่วนบุคคล หน่วยประมวลผล Intel® Pentium® III Mobile CPU 1133 MHz หน่วยความจำ (RAM) 256 MB และใช้ระบบปฏิบัติการ Microsoft Windows XP Professional เวอร์ชัน 2002 Service Pack 2
2. โปรแกรมเพื่อการคำนวณทางคณิตศาสตร์ MATLAB® ของบริษัท The MathWorks, Inc. เวอร์ชัน 6.1.0.450 Release 12.1

3.4 การหาค่าพารามิเตอร์สำหรับกระบวนการรู้จำ

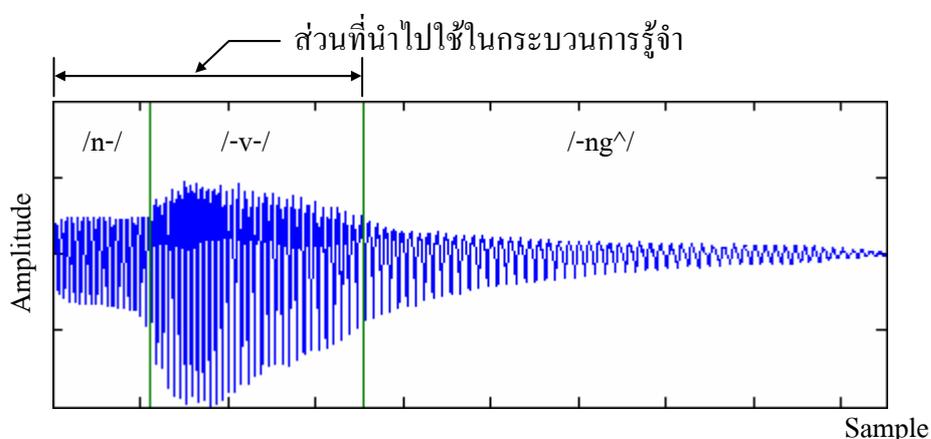
ฐานข้อมูลเสียง LOTUS Corpus ที่ใช้ในงานวิจัยนี้มีการจัดเก็บโดยแบ่งเป็นไฟล์ของประโยคต่าง ๆ ในรูปแบบ .wav แต่ละไฟล์จะมีไฟล์ข้อความที่ระบุหน่วยเสียง และค่าของจุดแบ่ง

ของแต่ละหน่วยเสียงของไฟล์นั้นด้วย รูปที่ 3.2 แสดงตัวอย่างไฟล์เสียงของประโยค “สมการที่หนึ่ง” ซึ่งระบุตำแหน่งจะหน่วยเสียงต่าง ๆ ในประโยค

จากไฟล์เสียงดังกล่าวจะต้องทำการตัดแบ่งประโยคออกเป็นพยางค์ เพื่อนำไปใช้เป็นฐานข้อมูลสำหรับการรู้จำเสียงพยางค์ต้นควบสระ โดยจะต้องแบ่งตามกลุ่มของเสียงทั้งหมด 184 กลุ่ม ตัวอย่างของสัญญาณเสียงของพยางค์ “หนึ่ง” ซึ่งประกอบด้วยหน่วยเสียง /n-/ , /-v-/ และ /-ng^/ ดังที่แสดงตามรูปที่ 3.3 นั้นจะถูกตัดส่วนของพยางค์ที่แยกออก แล้วจัดอยู่ในกลุ่มของเสียงพยางค์ต้นควบสระ /n-v-/

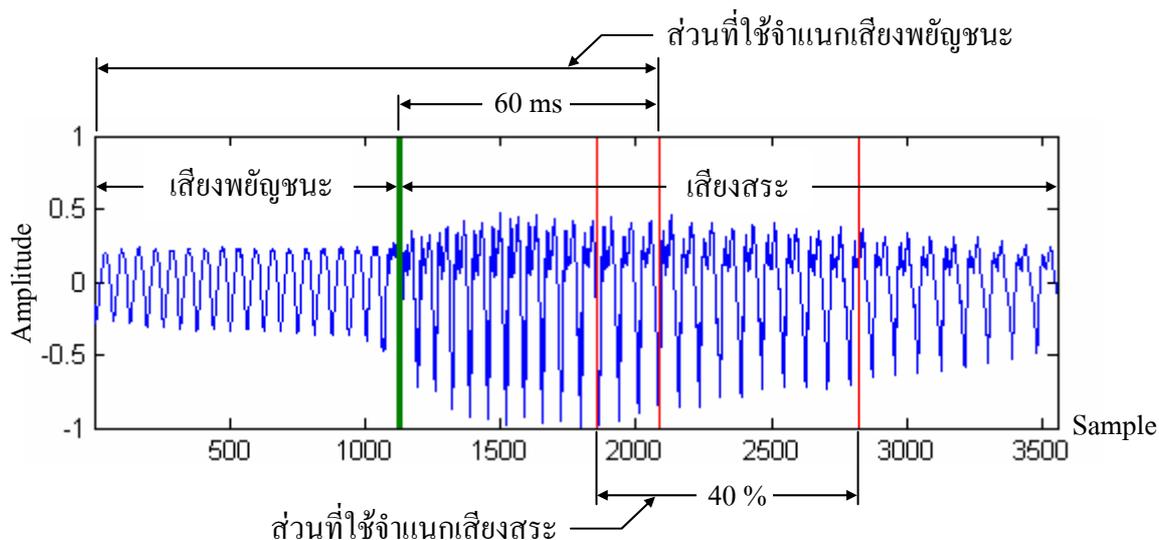


รูปที่ 3.2 สัญญาณเสียงและเส้นแบ่งหน่วยเสียงของประโยค “สมการที่หนึ่ง”



รูปที่ 3.3 สัญญาณเสียงของพยางค์ “หนึ่ง” และส่วนที่นำไปใช้ในกระบวนการรู้จำ

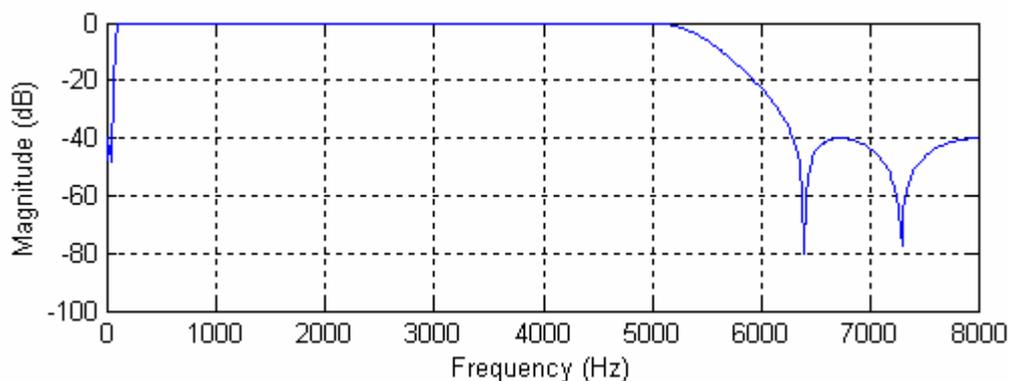
สัญญาณเสียงที่นำมาใช้ในแต่ละพยางค์จะถูกแยกออกเป็นสองส่วน คือส่วนที่ใช้สำหรับการจำแนกเสียงสระ และส่วนที่ใช้สำหรับจำแนกเสียงพยางค์ ซึ่งแสดงได้ดังรูปที่ 3.4 โดยส่วนที่ใช้สำหรับจำแนกสระนั้นจะใช้สัญญาณเสียงช่วงกลางของเสียงสระประมาณ 40 เปอร์เซ็นต์ของเสียงสระทั้งหมด และส่วนที่ใช้สำหรับการจำแนกเสียงพยางค์จะใช้ส่วนของเสียงพยางค์ทั้งหมดรวมกับเสียงสระอีก 60 ms



รูปที่ 3.4 ส่วนของสัญญาณเสียงที่ใช้จำแนกเสียงพยัญชนะและเสียงสระ

ค่าพารามิเตอร์ที่เลือกใช้ในระบบการรู้จำเสียงพยัญชนะควบสระสำหรับงานวิจัยนี้ ใช้ค่าสเปกตรัม LPC บนสเกลบาร์ก ซึ่งเป็นที่นิยมในการรู้จำเสียงพูด ขั้นตอนของการหาค่าสเปกตรัม LPC บนสเกลบาร์ก มีดังนี้

1. การกรองสัญญาณ (Filtering) สัญญาณเสียงพูดที่บันทึกไว้นั้นจะถูกกรองด้วยวงจรกรองสัญญาณผ่านความถี่ เพื่อขจัดสัญญาณของไฟฟ้ากระแสตรง เสียงฮัมที่ความถี่ต่ำ และสัญญาณรบกวนที่อาจเกิดขึ้นกับสัญญาณเสียงที่บันทึกได้ โดยในงานวิจัยนี้ใช้วงจรกรองคิจิตอลอันดับ 4 แบบ Elliptic ที่มีความถี่ Cutoff ที่ 100 – 5000 Hz โดยตอบสนองความถี่ที่ขนาด -0.1 dB ในช่วงผ่านความถี่ และมีการลดทอนสัญญาณที่ 40 dB การตอบสนองความถี่ของวงจรกรองสัญญาณที่ใช้แสดงได้ดังรูปที่ 3.5

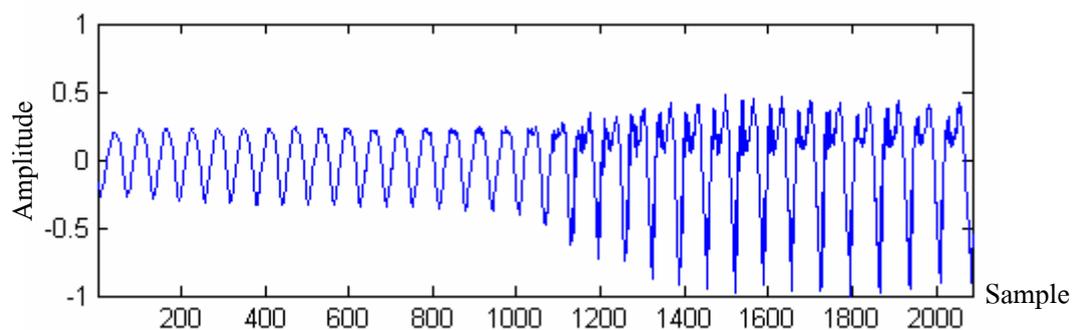


รูปที่ 3.5 การตอบสนองของวงจรกรองสัญญาณ

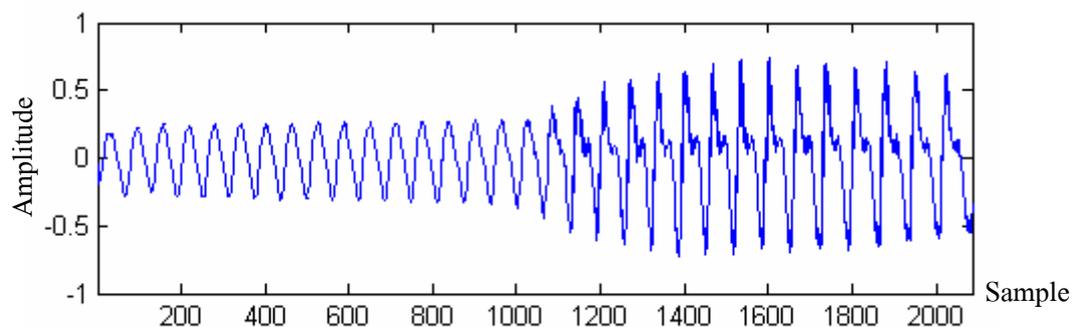
2. การหาค่าสเปกตรัม LPC ในการหาค่าสเปกตรัม LPC ของสัญญาณเสียงจะเริ่มจากการผ่านกระบวนการปรับขยายสัญญาณเบื้องต้น (Preemphasis) เพื่อให้อัตราส่วนของสัญญาณเสียงต่อสัญญาณรบกวนมีค่าคงที่มากที่สุด ตลอดช่วงความถี่ โดยเน้นความถี่สูงให้มีขนาดสูงขึ้น สัญญาณเสียงที่ผ่านกระบวนการปรับขยายสัญญาณเบื้องต้นจะถูกแบ่งออกเป็นเฟรม โดยงานวิจัยนี้กำหนดให้แต่ละเฟรมมีความยาว 20 ms การวิเคราะห์สัญญาณเสียงจะทำการวิเคราะห์ทีละเฟรม โดยแต่ละเฟรมจะเลื่อนไปครั้งละ 5 ms สัญญาณเสียงแต่ละเฟรมที่ทำการวิเคราะห์จะเริ่มการวิเคราะห์ด้วยการทำวินโดว์เพื่อลดปัญหาความไม่ต่อเนื่องของสัญญาณที่ขอบเฟรม และไม่ทำให้สเปกตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไป ซึ่งวินโดว์ที่ใช้ในงานวิจัยนี้จะเป็นวินโดว์แบบแฮมมิง (Hamming) ซึ่งอยู่ในรูป

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (3.1)$$

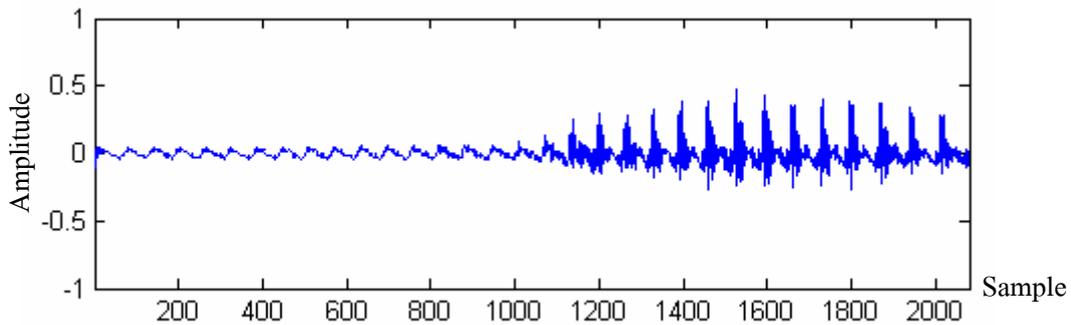
สัญญาณที่ผ่านการทำวินโดว์แล้ว จะถูกนำมาหาค่าสัมประสิทธิ์ LPC ซึ่งงานวิจัยนี้ใช้ LPC อันดับ 14 เพื่อแทนสัญญาณเสียงพูด และการหาค่าสัมประสิทธิ์ LPC นั้นจะใช้วิธีอัตโนมัติสัมพันธ์ (Autocorrelation) แล้วจึงนำค่าสัมประสิทธิ์ที่ได้ไปคำนวณค่าสเปกตรัม LPC ต่อไป รูปที่ 3.6 นั้นจะแสดงขั้นตอนของการหาค่าสเปกตรัม LPC



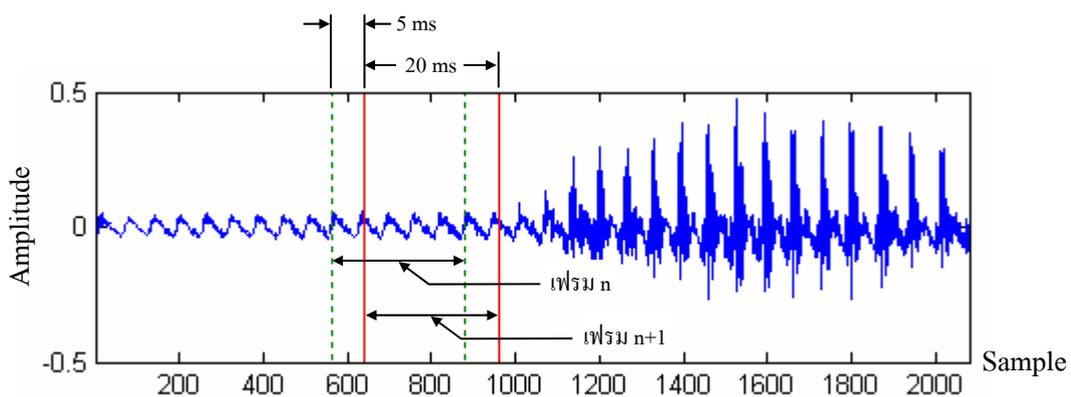
(ก) สัญญาณเสียงพูดที่จะนำมาหาค่าสเปกตรัม LPC



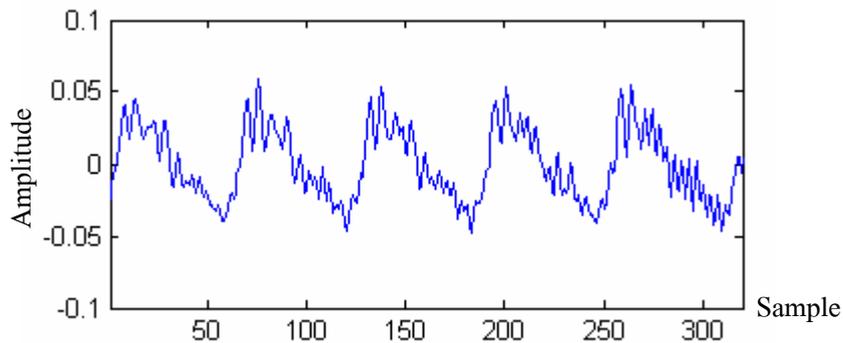
(ข) สัญญาณเสียงพูดที่ผ่านวงจรกรองสัญญาณผ่านความถี่



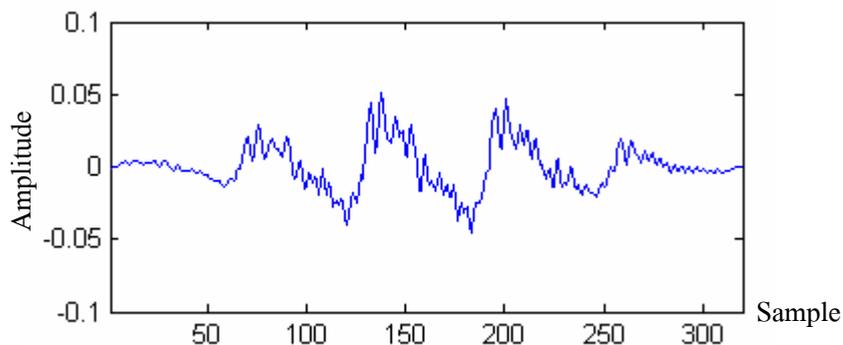
(ค) สัญญาณเสียงพูดที่ผ่านการปรับขยายสัญญาณเบื้องต้น



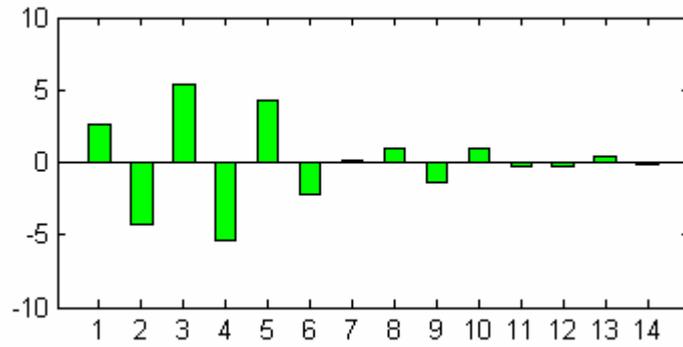
(ง) การแบ่งเฟรมของสัญญาณเสียงเพื่อนำไปหาค่าสเปกตรัม LPC



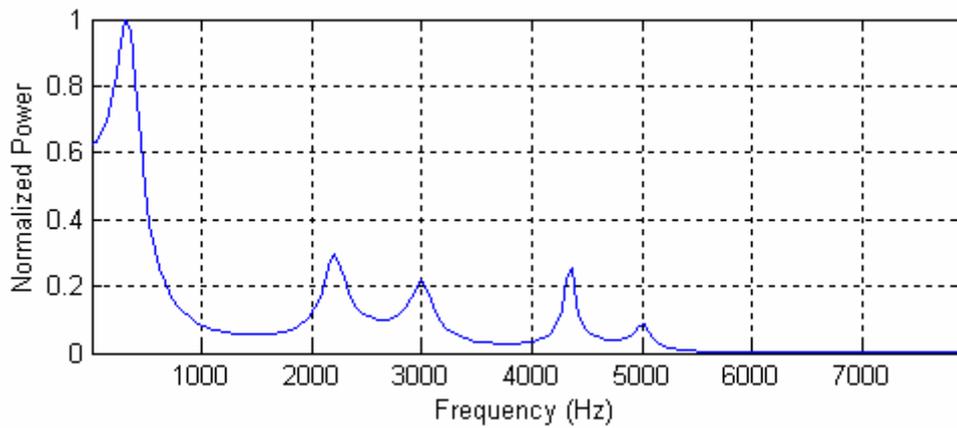
(จ) เฟรมของสัญญาณเสียงที่นำไปวิเคราะห์



(ฉ) เฟรมของสัญญาณเสียงที่ผ่านการทำวินโดว์

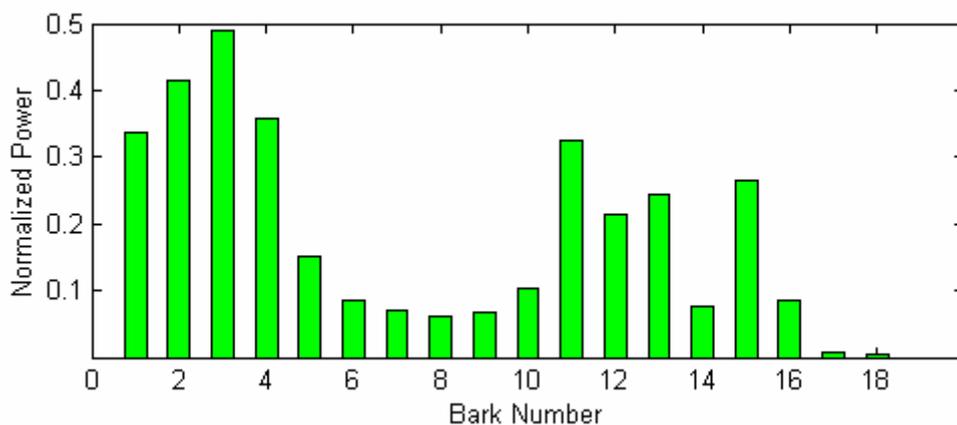


(ข) ค่าสัมประสิทธิ์ LPC อันดับ 14 ของเฟรมที่ทำการวิเคราะห์



(ข) ค่าสเปกตรัม LPC ของเฟรมที่ทำการวิเคราะห์

รูปที่ 3.6 ขั้นตอนของการหาค่าสเปกตรัม LPC



รูปที่ 3.7 ความเข้มของสัญญาณในแต่ละความถี่วิกฤต

3. การแปลงค่าสเปกตรัม LPC จากสเกลความถี่ไปเป็นสเกลบาร์ก หรือสเกลอัตราความถี่วิกฤต (Critical Band Rate Scale) ซึ่งแสดงการตอบสนองความถี่ที่มนุษย์สามารถได้ยิน โดยจะต้อง

คำนวณหาความเข้มของสัญญาณในแต่ละความถี่วิกฤต (Critical Band Intensity) และค่าความเข้มของสัญญาณบนสเกลบาร์กนี่จะนำไปใช้เป็นเวกเตอร์คุณสมบัติของแต่ละเฟรมสำหรับกระบวนการรู้จำต่อไป ความเข้มของสัญญาณที่ได้จากสเปกตรัม LPC ที่ได้ในแต่ละความถี่วิกฤต สามารถแสดงได้ดังรูปที่ 3.7

เนื่องจากสัญญาณเสียงที่ใช้วิเคราะห์นั้นผ่านตัวกรองสัญญาณ ทำให้ความถี่สูงสุดที่เป็นไปได้คือ 5 kHz ซึ่งอยู่ในช่วงของบาร์กที่ 18 ดังนั้นเวกเตอร์คุณสมบัติที่ได้จึงมีขนาด 18 มิติ นอกจากนี้ในงานวิจัยยังได้เสนอให้ใช้ค่าลอการิทึมของความเข้มสัญญาณ ในกระบวนการรู้จำ เนื่องจากค่าดังกล่าวจะให้ผลของการรู้จำที่มีความถูกต้องสูงกว่า ค่าความเข้มของสัญญาณที่ไม่อยู่ในรูปของค่าลอการิทึม

3.5 กระบวนการรู้จำเสียงพยัญชนะต้นควบสระ

กระบวนการรู้จำเสียงพยัญชนะต้นควบสระของงานวิจัยนี้ ประกอบด้วยขั้นตอนต่าง ๆ ทั้งการสร้างรูปแบบอ้างอิง และการทดสอบสัญญาณเสียง รวมทั้งกระบวนการตัดสินใจที่งานวิจัยเลือกใช้ ซึ่งมีรายละเอียดดังนี้

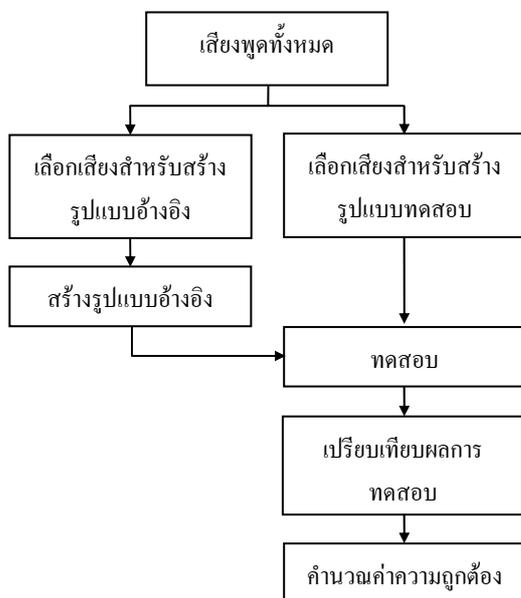
3.5.1 ขั้นตอนของกระบวนการรู้จำเสียงพยัญชนะต้นควบสระ

กระบวนการรู้จำเสียงพยัญชนะต้นควบสระของงานวิจัยนี้ จะเริ่มจากการแบ่งเสียงพูดทั้งหมดจำนวน 9,083 พยางค์ ออกเป็น 2 กลุ่ม คือเสียงสำหรับสร้างรูปแบบอ้างอิง จำนวน 4,647 พยางค์ และเสียงสำหรับสร้างรูปแบบทดสอบ จำนวน 4,436 พยางค์ โดยหลังจากสร้างรูปแบบอ้างอิงแล้ว จะนำเสียงของรูปแบบทดสอบซึ่งทราบกลุ่มของเสียงพยัญชนะต้นควบสระ มาทำการจำแนกกลุ่ม แล้วจึงนำผลลัพธ์ที่ได้ไปเปรียบเทียบกับกลุ่มของรูปแบบทดสอบนั้น เพื่อนำไปคำนวณเปอร์เซ็นต์ความถูกต้องของระบบการรู้จำเสียงพูดของงานวิจัย รูปที่ 3.8 (ก) จะอธิบายกระบวนการรู้จำเสียงพูดของงานวิจัยนี้

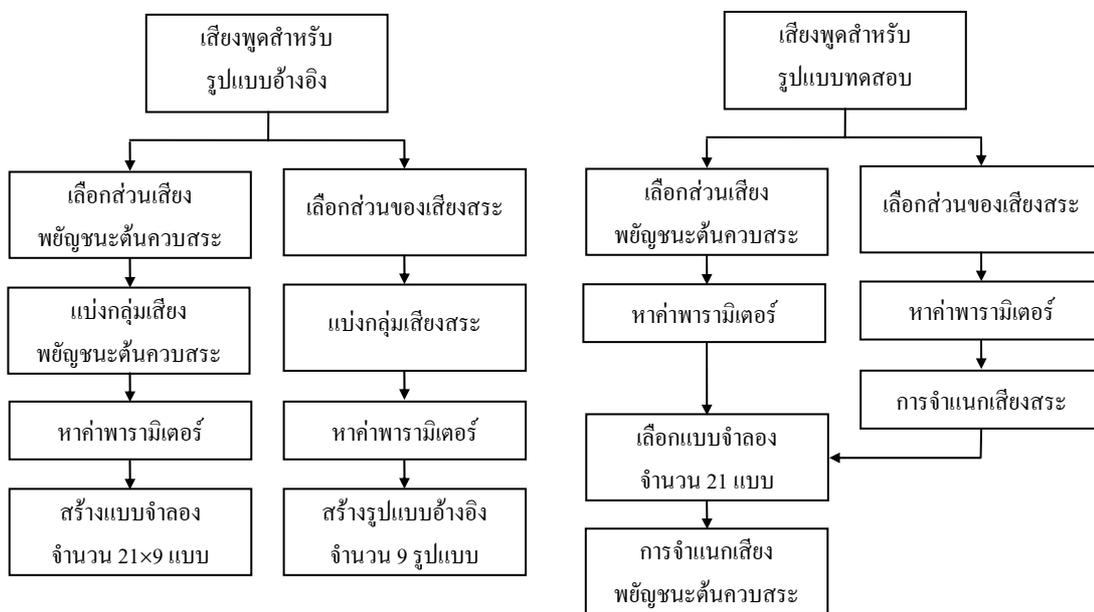
จากภาพรวมของกระบวนการรู้จำเสียงพยัญชนะต้นควบสระนั้น สามารถพิจารณาขั้นตอนย่อยของกระบวนการออกเป็นส่วนสำคัญ 2 ส่วน ซึ่งแสดงได้ดังรูปที่ 3.8 (ข) และ (ค) ได้แก่ ส่วนของการสร้างรูปแบบอ้างอิง และส่วนของการทดสอบรูปแบบ ซึ่งมีรายละเอียดดังนี้

การสร้างรูปแบบอ้างอิง เสียงพูดแต่ละพยางค์จะถูกแบ่งสัญญาณออกเป็น 2 ส่วนสำหรับการจำแนกเสียงพยัญชนะต้นควบสระ และส่วนสำหรับการจำแนกสระ ทั้งสองส่วนนี้จะถูกจัดให้อยู่ในกลุ่มของเสียงนั้น ๆ โดยจะมีเสียงพยัญชนะต้นควบสระทั้งหมด 184 กลุ่ม (ยกเว้น /c-q/, /ng-e-, /ng-i-, /ng-v- และ /w-v-) และเสียงสระอีก 9 กลุ่ม เสียงทั้งหมดจะนำไปหาค่าพารามิเตอร์สำหรับกระบวนการจำแนก ในรูปของเวกเตอร์คุณสมบัติ ก่อนที่จะนำไปสร้างแบบจำลองและรูปแบบอ้างอิง โดยเวกเตอร์คุณสมบัติของเสียงสำหรับการจำแนกพยัญชนะต้นควบสระ จะนำไป

สร้างแบบจำลองอ้างอิง และเวกเตอร์คุณสมบัติของเสียงสำหรับการจำแนกสระ จะใช้สร้างรูปแบบอ้างอิงสำหรับสระแต่ละกลุ่ม ดังนั้นในขั้นตอนนี้จะได้แบบจำลองอ้างอิงสำหรับเสียงพยัญชนะต้นควบสระ 184 แบบจำลอง และรูปแบบอ้างอิงสำหรับเสียงสระอีก 9 กลุ่ม



(ก) ขั้นตอนหลักของการรู้จำเสียงพยัญชนะต้นควบสระ



(ข) การสร้างรูปแบบอ้างอิง

(ค) การรู้จำรูปแบบทดสอบ

รูปที่ 3.8 ขั้นตอนของกระบวนการรู้จำเสียงพยัญชนะต้นควบสระ

การทดสอบรูปแบบ มีการเริ่มต้นในลักษณะเดียวกับการสร้างรูปแบบอ้างอิง ก็จะต้องแบ่งสัญญาณเสียงเพื่อหาเวกเตอร์คุณสมบัติของส่วนที่เป็นเสียงพยัญชนะต้นควบสระ และส่วนที่เป็นสระ สำหรับเวกเตอร์คุณสมบัติของส่วนที่เป็นสระจะถูกนำไปจำแนกเสียงก่อน เพื่อระบุว่ารูปแบบทดสอบมีเสียงสระอยู่ในกลุ่มได้ เมื่อได้เสียงสระแล้วจะนำไปเลือกแบบจำลองอ้างอิงของเสียงพยัญชนะต้นควบสระที่มีเสียงสระตรงกับที่จำแนกได้ ทำให้ในขั้นตอนี้จะมีแบบจำลองอ้างอิงที่เป็นไปได้ทั้งหมดลดลงจาก 184 แบบจำลอง เหลือเพียง 21 รูปแบบ และแบบจำลองอ้างอิงที่ได้นี้จะถูกใช้เปรียบเทียบกับเวกเตอร์คุณสมบัติของพยัญชนะต้นควบสระที่ต้องการทดสอบเพื่อจำแนกกลุ่มของสัญญาณทดสอบในขั้นตอนต่อไป

3.5.2 กระบวนการตัดสินใจสำหรับการรู้จำเสียงพยัญชนะต้นควบสระ

ในกระบวนการรู้จำเสียงพยัญชนะต้นควบสระของงานวิจัยนี้ ประกอบด้วยการจำแนกเสียงสระ และการจำแนกเสียงพยัญชนะ ซึ่งการจำแนกทั้งสองขั้นตอนนั้นมีความแตกต่างกันตามลักษณะของหน่วยเสียง โดยเสียงสระนั้นมีลักษณะเป็นเสียงโหมะที่ค่อนข้างคงที่ตลอดช่วงของหน่วยเสียง การวิจัยนี้จึงใช้กระบวนการตัดสินใจแบบ K-Nearest Neighbor Rule ในขณะที่เสียงพยัญชนะต้นควบสระจะมีการเปลี่ยนแปลงสเปกตรัมตลอดช่วงของหน่วยเสียง งานวิจัยนี้จึงใช้กระบวนการตัดสินใจแบบ Hidden Markov Modeling

กระบวนการตัดสินใจแบบ K-Nearest Neighbor Rule ที่ใช้ในการจำแนกเสียงสระนั้น จะต้องคำนวณหาระยะทางระหว่างเวกเตอร์ทดสอบของแต่ละเฟรมกับเวกเตอร์อ้างอิงทั้งหมด โดยงานวิจัยนี้เลือกใช้ระยะทางแบบ Mahalanobis เพื่อหาว่าเวกเตอร์อ้างอิง K เวกเตอร์ที่อยู่ใกล้เวกเตอร์ทดสอบมากที่สุดจัดอยู่ในกลุ่มของสระใดบ้าง และนำเวกเตอร์ K เวกเตอร์ของทุก ๆ เฟรมมารวมกัน เพื่อหาว่าเวกเตอร์ที่ได้จากทุก ๆ เฟรมนั้นจัดอยู่ในกลุ่มสระใดมากที่สุด ซึ่งกลุ่มสระนั้นคือผลลัพธ์ของการจำแนกเสียงสระ สำหรับงานวิจัยได้ทดสอบเปรียบเทียบประสิทธิภาพของการจำแนกโดยใช้ค่า K ต่าง ๆ พบว่าที่ค่า $K = 2$ จะได้เปอร์เซ็นต์ของความถูกต้องสูงที่สุด

กระบวนการตัดสินใจแบบ Hidden Markov Modeling ที่ใช้ในการจำแนกเสียงพยัญชนะต้นควบสระนั้น ในงานวิจัยนี้จะใช้แบบจำลอง HMM ชนิดซ้ายไปขวา ที่มีจำนวนสถานะ 6 สถานะ ซึ่งเหมาะที่จะใช้อธิบายการเปลี่ยนแปลงของสัญญาณเสียงในช่วงดังกล่าว ในขั้นตอนของการสร้างแบบจำลองอ้างอิงนั้น จะมีการสร้างแบบจำลองของกลุ่มเสียงพยัญชนะต้นควบสระทั้งหมด 184 กลุ่ม ตามจำนวนเสียงที่เป็นไปได้ทั้งหมด แต่ในขั้นตอนของการจำแนกเสียงจะใช้เสียงสระที่ได้จากการจำแนกสระในขั้นตอนแรก เลือกแบบจำลองที่จะใช้ในการรู้จำ โดยเลือกแบบจำลองของกลุ่มที่มีเสียงสระตรงกับเสียงสระที่จำแนกได้เท่านั้น ดังนั้นแบบจำลองที่ใช้สำหรับเปรียบเทียบกับรูปแบบทดสอบจึงลดลงเหลือ 21 แบบจำลอง การแบ่งกระบวนการรู้จำเสียงพยัญชนะต้นควบสระ

ออกเป็น 2 ขั้นตอนย่อยเช่นนี้ จะทำให้ความผิดพลาดในการจำแนก และเวลาในการประมวลผลเพื่อเปรียบเทียบรูปแบบลดลงได้อย่างมาก

3.6 ผลการทดลอง

การทดสอบระบบการรู้จำเสียงพยัญชนะต้นควบสระนั้น จะแยกพิจารณาความถูกต้องของการจำแนกเสียงสระ และการจำแนกเสียงพยัญชนะต้นควบสระ ซึ่งได้เปอร์เซ็นต์ความถูกต้องเท่ากับ 91.95 % และ 89.38 % ตามลำดับ โดยความถูกต้องของการจำแนกเสียงพยัญชนะต้นควบสระนั้นจะพิจารณากรณีที่การจำแนกสระในขั้นตอนแรกถูกต้องเท่านั้น เมื่อพิจารณาความถูกต้องรวมของระบบการรู้จำ จะได้เปอร์เซ็นต์ความถูกต้องรวมเท่ากับ 82.19 % ตารางที่ 3.2 – 3.13 จะแสดงรายละเอียดของผลการทดสอบการจำแนกเสียงสระ และเสียงพยัญชนะต้นควบสระ

ตารางที่ 3.2 ผลการจำแนกเสียงสระ

ผลกรจำแนก สระ	q	a	x	e	i	v	u	o	@	% ความ ถูกต้อง
q	259	-	-	-	-	1	-	-	1	99.23%
a	22	472	65	13	5	6	-	-	10	79.60%
x	-	13	431	6	-	-	-	-	1	95.57%
e	5	-	9	456	3	1	-	-	-	96.20%
i	1	-	-	17	511	15	-	1	-	93.76%
v	9	-	-	10	11	383	-	-	-	92.74%
u	9	-	-	-	6	34	444	16	-	87.23%
o	31	-	-	-	-	5	8	569	5	92.07%
@	6	4	3	-	-	-	-	5	554	96.85%
ค่าเฉลี่ย										91.95%

ตารางที่ 3.3 ความถูกต้องของเสียงพยัญชนะต้นควบสระแต่ละกลุ่ม

สระ พยัญชนะ	a	e	i	o	q	u	v	x	@	
z	100%	84.6%	57.6%	95.0%	100%	94.1%	93.1%	80.0%	64.1%	81.25%
ph	78.6%	80.0%	96.7%	84.6%	94.1%	95.0%	87.5%	80.0%	89.2%	88.01%
kh	87.5%	100%	100%	68.6%	100%	100%	85.0%	75.0%	95.0%	88.41%
th	90.5%	80.8%	97.3%	96.8%	100%	97.5%	69.4%	82.8%	92.6%	88.76%
ch	100%	97.2%	38.5%	95.5%	100%	100%	95.5%	100%	100%	92.36%
c	100%	90.0%	65.0%	100%	-	92.3%	96.0%	100%	92.3%	90.41%
s	100%	97.4%	100%	95.0%	95.0%	91.7%	88.9%	100%	85.0%	94.08%
f	96.6%	100%	100%	100%	100%	100%	94.1%	100%	100%	98.04%
h	100%	88.9%	100%	90.6%	100%	100%	100%	65.2%	100%	88.80%
r	100%	95.5%	73.1%	83.3%	100%	85.3%	100%	100%	92.3%	91.27%
l	100%	97.4%	52.6%	83.8%	100%	100%	100%	87.2%	50.0%	84.52%
d	77.3%	100%	89.7%	87.5%	100%	96.9%	100%	75.0%	75.0%	86.88%
y	75.0%	100%	93.5%	100%	100%	97.5%	100%	66.7%	100%	92.59%
n	100%	100%	77.5%	84.6%	100%	52.9%	100%	82.8%	90.0%	87.10%
m	90.9%	100%	71.8%	96.2%	100%	97.1%	100%	72.7%	96.7%	90.00%
w	100%	84.0%	87.5%	85.3%	100%	100%	-	78.9%	100%	89.16%
b	92.6%	100%	73.7%	94.4%	100%	100%	100%	87.5%	96.8%	91.89%
p	100%	94.9%	69.2%	100%	100%	100%	100%	69.2%	100%	92.42%
t	81.5%	71.4%	93.8%	84.2%	95.0%	100%	100%	86.8%	90.0%	88.57%
k	88.9%	92.0%	73.3%	92.0%	84.2%	100%	100%	80.6%	86.1%	86.73%
ng	97.1%	-	-	100%	100%	100%	-	100%	100%	98.72%
	79.60%	96.20%	93.76%	92.07%	99.23%	87.23%	92.74%	95.57%	96.85%	

ตารางที่ 3.4 ผลการจำแนกเสียงพยัญชนะ

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	169	3	3	1	-	-	-	-	3	4	-	1	9	-	1	-	-	8	4	2	-	81.25%
ph	3	235	13	11	-	-	-	-	4	-	-	1	-	-	-	-	-	-	-	-	-	88.01%
kh	-	16	183	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	88.41%
th	-	9	7	221	5	-	-	-	-	2	-	-	-	-	-	-	-	-	-	5	-	88.76%
ch	-	-	-	2	133	4	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	92.36%
c	3	-	4	-	3	132	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	90.41%
s	-	-	-	5	9	2	270	1	-	-	-	-	-	-	-	-	-	-	-	-	-	94.08%
f	-	1	-	-	-	-	-	100	-	1	-	-	-	-	-	-	-	-	-	-	-	98.04%
h	1	5	2	6	-	-	-	-	111	-	-	-	-	-	-	-	-	-	-	-	-	88.80%
r	-	-	-	-	-	-	-	-	-	230	16	-	3	-	-	-	-	-	-	3	-	91.27%
l	1	-	-	1	-	-	-	-	4	11	202	1	-	4	1	1	1	3	9	-	-	84.52%
d	5	-	-	-	-	-	-	-	-	2	11	192	-	8	-	-	3	-	-	-	-	86.88%
y	2	-	-	-	-	-	-	-	-	-	4	3	150	3	-	-	-	-	-	-	-	92.59%
n	-	-	-	-	-	-	-	-	4	2	4	19	-	216	1	-	1	-	1	-	-	87.10%
m	-	-	-	-	-	-	-	-	-	-	1	-	1	2	198	9	8	-	-	-	1	90.00%
w	-	-	-	-	-	-	-	-	8	1	-	1	-	1	3	148	3	-	-	1	-	89.16%
b	-	1	-	-	-	-	-	-	-	-	4	-	-	1	4	5	170	-	-	-	-	91.89%
p	-	-	-	-	-	-	-	-	-	1	-	-	4	-	-	-	-	122	5	-	-	92.42%
t	11	1	-	3	-	-	-	-	-	2	1	-	-	1	-	3	-	6	217	-	-	88.57%
k	8	5	2	1	-	-	-	-	3	-	-	-	-	-	-	-	-	3	4	170	-	86.73%
ng	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	77	98.72%
ค่าเฉลี่ย																					89.38%	

ตารางที่ 3.5 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อะ” (/a-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
ph	-	11	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	78.57%
kh	-	-	7	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	87.50%
th	-	-	1	19	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	90.48%
ch	-	-	-	-	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
c	-	-	-	-	-	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
s	-	-	-	-	-	-	21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
f	-	-	-	-	-	-	-	28	-	1	-	-	-	-	-	-	-	-	-	-	-	96.55%
h	-	-	-	-	-	-	-	-	15	-	-	-	-	-	-	-	-	-	-	-	-	100%
r	-	-	-	-	-	-	-	-	-	23	-	-	-	-	-	-	-	-	-	-	-	100%
l	-	-	-	-	-	-	-	-	-	-	25	-	-	-	-	-	-	-	-	-	-	100%
d	-	-	-	-	-	-	-	-	-	-	5	17	-	-	-	-	-	-	-	-	-	77.27%
y	-	-	-	-	-	-	-	-	-	-	3	3	18	-	-	-	-	-	-	-	-	75.00%
n	-	-	-	-	-	-	-	-	-	-	-	-	-	26	-	-	-	-	-	-	-	100%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	1	-	-	-	1	90.91%
w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	26	-	-	-	-	-	100%
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	25	-	-	-	-	92.59%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	-	-	-	100%
t	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	4	22	-	-	81.48%
k	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24	-	88.89%
ng	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	34	97.14%
ค่าเฉลี่ย																					93.43%	

ตารางที่ 3.6 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อิ” (/i-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	19	-	-	-	-	-	-	-	3	-	-	1	9	-	-	-	-	-	-	1	-	57.58%
ph	1	29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	96.67%
kh	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
th	-	1	-	36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	97.30%
ch	-	-	-	-	5	4	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	38.46%
c	3	-	4	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	65.00%
s	-	-	-	-	-	-	40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
f	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	100%
r	-	-	-	-	-	-	-	-	-	19	3	-	3	-	-	-	-	-	1	-	-	73.08%
l	-	-	-	1	-	-	-	-	-	4	10	-	-	4	-	-	-	-	-	-	-	52.63%
d	1	-	-	-	-	-	-	-	-	-	-	35	-	-	-	-	3	-	-	-	-	89.74%
y	2	-	-	-	-	-	-	-	-	-	-	-	29	-	-	-	-	-	-	-	-	93.55%
n	-	-	-	-	-	-	-	-	-	-	-	8	-	31	1	-	-	-	-	-	-	77.50%
m	-	-	-	-	-	-	-	-	-	-	-	-	1	2	28	8	-	-	-	-	-	71.79%
w	-	-	-	-	-	-	-	-	-	-	-	1	-	1	3	35	-	-	-	-	-	87.50%
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	1	14	-	-	-	-	73.68%
p	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	9	-	-	-	69.23%
t	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1	30	-	-	93.75%
k	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	-	73.33%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ค่าเฉลี่ย																					81.80%	

ตารางที่ 3.7 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อี” (/v-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	27	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	93.10%
ph	-	35	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	87.50%
kh	-	5	34	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	85.00%
th	-	6	5	25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	69.44%
ch	-	-	-	-	21	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95.45%
c	-	-	-	-	1	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	96.00%
s	-	-	-	-	4	-	32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	88.89%
f	-	1	-	-	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	94.12%
h	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	100%
r	-	-	-	-	-	-	-	-	-	40	-	-	-	-	-	-	-	-	-	-	-	100%
l	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-	100%
d	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-	-	100%
y	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	100%
n	-	-	-	-	-	-	-	-	-	-	-	-	-	34	-	-	-	-	-	-	-	100%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	-	-	-	-	-	-	100%
w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	100%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	-	-	-	100%
t	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	100%
k	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	100%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ค่าเฉลี่ย																					91.91%	

ตารางที่ 3.8 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “อุ” (-u-)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	16	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	94.12%
ph	1	38	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95.00%
kh	-	-	27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
th	-	1	-	39	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	97.50%
ch	-	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
c	-	-	-	-	-	12	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	92.31%
s	-	-	-	3	-	-	33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	91.67%
f	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	-	-	-	-	-	-	-	-	10	-	-	-	-	-	-	-	-	-	-	-	-	100%
r	-	-	-	-	-	-	-	-	-	29	5	-	-	-	-	-	-	-	-	-	-	85.29%
l	-	-	-	-	-	-	-	-	-	-	25	-	-	-	-	-	-	-	-	-	-	100%
d	-	-	-	-	-	-	-	-	-	1	-	31	-	-	-	-	-	-	-	-	-	96.88%
y	-	-	-	-	-	-	-	-	-	-	1	-	39	-	-	-	-	-	-	-	-	97.50%
n	-	-	-	-	-	-	-	-	-	-	4	4	-	9	-	-	-	-	-	-	-	52.94%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33	1	-	-	-	-	-	97.06%
w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	100%
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	-	-	-	-	100%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	100%
t	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	-	-	100%
k	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	-	100%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	100%
ค่าเฉลี่ย																					94.59%	

ตารางที่ 3.9 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เอะ” (/e-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	33	2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	84.62%
ph	1	16	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	80.00%
kh	-	-	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
th	-	-	-	21	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	80.77%
ch	-	-	-	1	35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	97.22%
c	-	-	-	-	2	18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	90.00%
s	-	-	-	-	-	-	37	1	-	-	-	-	-	-	-	-	-	-	-	-	-	97.37%
f	-	-	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	1	-	1	1	-	-	-	-	24	-	-	-	-	-	-	-	-	-	-	-	-	88.89%
r	-	-	-	-	-	-	-	-	-	21	1	-	-	-	-	-	-	-	-	-	-	95.45%
l	-	-	-	-	-	-	-	-	-	-	38	-	-	-	-	-	-	-	1	-	-	97.44%
d	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	100%
y	-	-	-	-	-	-	-	-	-	-	-	-	19	-	-	-	-	-	-	-	-	100%
n	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	100%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	19	-	-	-	-	-	-	100%
w	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	21	-	-	-	-	-	84.00%
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	100%
p	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	37	1	-	-	94.87%
t	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	3	-	-	10	-	-	71.43%
k	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	23	-	92.00%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ค่าเฉลี่ย																					92.11%	

ตารางที่ 3.10 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เอะ” (/x-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	4	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	80.00%
ph	-	24	-	5	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	80.00%
kh	-	4	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	75.00%
th	-	1	-	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	82.76%
ch	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
c	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
s	-	-	-	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
f	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	-	3	-	5	-	-	-	-	15	-	-	-	-	-	-	-	-	-	-	-	-	65.22%
r	-	-	-	-	-	-	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-	100%
l	-	-	-	-	-	-	-	-	4	-	34	-	-	-	-	-	1	-	-	-	-	87.18%
d	4	-	-	-	-	-	-	-	-	-	2	30	-	4	-	-	-	-	-	-	-	75.00%
y	-	-	-	-	-	-	-	-	-	-	-	-	6	3	-	-	-	-	-	-	-	66.67%
n	-	-	-	-	-	-	-	-	4	-	-	-	-	24	-	-	1	-	-	-	-	82.76%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	-	6	-	-	-	-	72.73%
w	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	15	3	-	-	-	-	78.95%
b	-	-	-	-	-	-	-	-	-	-	4	-	-	1	-	-	35	-	-	-	-	87.50%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	4	-	-	69.23%
t	1	1	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33	-	-	86.84%
k	-	4	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	29	-	80.56%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	100%
ค่าเฉลี่ย																					81.90%	

ตารางที่ 3.11 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เออะ” (-q-)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง	
z	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
ph	-	16	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	94.12%
kh	-	-	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
th	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
ch	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
s	-	-	-	-	1	-	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95.00%
f	-	-	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
r	-	-	-	-	-	-	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-	-	100%
l	-	-	-	-	-	-	-	-	-	-	17	-	-	-	-	-	-	-	-	-	-	-	100%
d	-	-	-	-	-	-	-	-	-	-	-	21	-	-	-	-	-	-	-	-	-	-	100%
y	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	100%
n	-	-	-	-	-	-	-	-	-	-	-	-	-	17	-	-	-	-	-	-	-	-	100%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	-	-	-	-	-	-	-	100%
w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	100%
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	100%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	100%
t	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	38	-	-	-	95.00%
k	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	16	-	-	84.21%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	-	100%
ค่าเฉลี่ย																						97.30%	

ตารางที่ 3.12 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “โอะ” (/o-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง	
z	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	95.00%
ph	-	33	5	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	84.62%
kh	-	7	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	68.57%
th	-	-	-	30	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	96.77%
ch	-	-	-	1	21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95.45%
c	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
s	-	-	-	2	-	-	38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95.00%
f	-	-	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	-	2	1	-	-	-	-	-	29	-	-	-	-	-	-	-	-	-	-	-	-	-	90.63%
r	-	-	-	-	-	-	-	-	-	30	5	-	-	-	-	-	-	-	-	1	-	-	83.33%
l	-	-	-	-	-	-	-	-	-	-	31	-	-	-	-	1	-	-	-	5	-	-	83.78%
d	-	-	-	-	-	-	-	-	-	-	4	28	-	-	-	-	-	-	-	-	-	-	87.50%
y	-	-	-	-	-	-	-	-	-	-	-	-	19	-	-	-	-	-	-	-	-	-	100%
n	-	-	-	-	-	-	-	-	-	2	-	3	-	33	-	-	-	-	-	1	-	-	84.62%
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25	-	1	-	-	-	-	-	96.15%
w	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	29	-	-	-	1	-	-	85.29%
b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	34	-	-	-	-	-	94.44%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	100%
t	5	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	32	-	-	84.21%
k	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	23	-	92.00%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	100%
ค่าเฉลี่ย																						88.93%	

ตารางที่ 3.13 ผลการจำแนกเสียงพยัญชนะต้นควบสระ สำหรับสระ “เอะ” (/@-/)

ผลการจำแนก พยัญชนะ	z	ph	kh	th	ch	c	s	f	h	r	l	d	y	n	m	w	b	p	t	k	ng	% ความ ถูกต้อง
z	25	-	-	-	-	-	-	-	-	2	-	-	-	-	1	-	-	7	4	-	-	64.10%
ph	-	33	1	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	89.19%
kh	-	-	38	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	95.00%
th	-	-	1	25	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	92.59%
ch	-	-	-	-	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
c	-	-	-	-	-	36	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	92.31%
s	-	-	-	-	4	2	34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	85.00%
f	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	100%
h	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-	-	-	-	-	100%
r	-	-	-	-	-	-	-	-	-	36	2	-	-	-	-	-	-	-	1	-	-	92.31%
l	1	-	-	-	-	-	-	-	-	7	16	1	-	-	1	-	-	3	3	-	-	50.00%
d	-	-	-	-	-	-	-	-	-	1	-	15	-	4	-	-	-	-	-	-	-	75.00%
y	-	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-	100%
n	-	-	-	-	-	-	-	-	-	-	-	4	-	36	-	-	-	-	-	-	-	90.00%
m	-	-	-	-	-	-	-	-	-	-	1	-	-	-	29	-	-	-	-	-	-	96.67%
w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	100%
b	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	30	-	-	-	-	96.77%
p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-	-	100%
t	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	36	-	-	90.00%
k	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	3	1	31	-	86.11%
ng	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	100%
ค่าเฉลี่ย																					87.36%	

บทที่ 4

สรุปและอภิปรายผล

4.1 สรุปผลการทดลอง

งานวิจัยนี้ได้ศึกษากระบวนการรู้จำเสียงพยางค์ระดับความสระ โดยใช้สเปกตรัม LPC บนสเกลบาร์กเป็นเวกเตอร์คุณสมบัติ โดยแบ่งกระบวนการรู้จำออกเป็น 2 ขั้นตอนย่อย คือการจำแนกเสียงสระ และการจำแนกเสียงพยางค์ระดับความสระ ซึ่งการจำแนกสระนั้นใช้กระบวนการตัดสินใจแบบ K-Nearest Neighbor Rule ส่วนการจำแนกเสียงพยางค์ระดับความสระนั้นใช้กระบวนการตัดสินใจแบบ Hidden Markov Modeling โดยงานวิจัยนี้ใช้ตัวอย่างเสียงทั้งหมด 9,083 เสียง แบ่งเป็นเสียงสำหรับการสร้างรูปแบบอ้างอิง จำนวน 4,647 เสียง และเสียงสำหรับทดสอบ จำนวน 4,436 เสียง ซึ่งผลของการวิจัยสรุปได้ดังนี้

- การจำแนกเสียงสระ จากเสียงตัวอย่างทดสอบทั้งหมด 4,436 เสียง มีเสียงที่จำแนกได้ถูกต้อง 4,079 เสียง คิดเป็นเปอร์เซ็นต์ความถูกต้อง เท่ากับ 91.95 %
- การจำแนกเสียงพยางค์ระดับความสระ จากเสียงตัวอย่างที่จำแนกเสียงสระได้ถูกต้องจำนวน 4,079 เสียง สามารถจำแนกเสียงพยางค์ระดับความสระได้ถูกต้อง 3,646 เสียง คิดเป็นเปอร์เซ็นต์ความถูกต้อง เท่ากับ 89.38 %
- ระบบการรู้จำเสียงพยางค์ระดับความสระ มีเปอร์เซ็นต์ความถูกต้องรวม เท่ากับ 82.19 %

4.2 อภิปรายผล และความผิดพลาด

จากผลการทดลองในแต่ละขั้นตอนของกระบวนการรู้จำ จะพบว่า

- การจำแนกเสียงสระ จากการทดสอบสระแต่ละเสียงมีเปอร์เซ็นต์ความถูกต้องแตกต่างกัน โดยสระ “เออะ” (/q/) มีความถูกต้องสูงสุดถึง 99.23 % เสียงสระ “เอะ” (/@/), “เอะ” (/e/) และ “เอะ” (/x/) มีความถูกต้องมากกว่า 95 % ในขณะที่เสียงสระ “อิ” (/i/), “อี” (/v/), “โอะ” (/o/) และ “อุ” (/u/) มีเปอร์เซ็นต์ความถูกต้องต่ำกว่า 90 % สำหรับ สระ “อะ” /a/ มีเปอร์เซ็นต์ความถูกต้องต่ำที่สุด คือ 79.60 % เนื่องจากช่วงเวลาของเสียงนี้สั้นกว่าสระอื่น ๆ เมื่อทำการจำแนกโดยใช้ K-Nearest Neighbor Rule จึงมีจำนวนเฟรมที่ใช้ในการตัดสินใจน้อย และเกิดความผิดพลาดได้ง่าย นอกจากนี้เสียงสระต่าง ๆ ยังสามารถแบ่งตามลักษณะของรูปปากได้เป็น 3 กลุ่ม คือ กลุ่มของสระ /a/ /e/ และ /x/ กลุ่มของสระ /u/ /o/ และ /@/ และกลุ่มของสระ /i/ และ /v/
- การจำแนกเสียงพยางค์ จากการทดสอบจะพบว่าพยางค์ที่ออกเสียงได้ยาก หรือมีช่วงเวลาของเสียงพยางค์ยาว เช่น เสียง “ง” (/ng/), “ฟ” (/f/) และ “ส” (/s/) จะมีความถูกต้องสูง ในขณะที่เสียง “อ” (/z/) ซึ่งมีลักษณะคล้ายเสียงสระ และมีช่วงเวลาของเสียงพยางค์สั้น จะมี

เปอร์เซ็นต์ความถูกต้องต่ำ นอกจากนี้ยังพบว่าการทดสอบการรู้จำเสียงพยางค์ต้นควบสระนั้น มีกลุ่มของเสียงพยางค์ต้นควบสระ 77 คู่ที่มีเปอร์เซ็นต์ความถูกต้อง 100 % ในขณะที่อีก 47 กลุ่ม และ 33 กลุ่มจะมีความถูกต้องมากกว่า 90 % และ 80 % ตามลำดับ

4.3 ข้อเสนอแนะ

1. ระบบการรู้จำเสียงพยางค์ต้นควบสระที่งานวิจัยนี้นำเสนอ ซึ่งแยกกระบวนการรู้จำออกเป็น 2 ขั้นตอนย่อย สามารถปรับปรุงประสิทธิภาพได้โดยการพัฒนาทั้งระบบการจำแนกเสียงสระ หรือการจำแนกเสียงพยางค์ต้นควบสระ

2. การจำแนกเสียงสระของงานวิจัยนี้เลือกใช้การตัดสินใจแบบ K-Nearest Neighbor Rule ซึ่งมีเวลาในการประมวลผลขึ้นอยู่กับจำนวนของเวกเตอร์คุณสมบัติที่ใช้เป็นรูปแบบอ้างอิง การลดจำนวนเวกเตอร์คุณสมบัติดังกล่าวโดยเลือกเฉพาะเวกเตอร์คุณสมบัติที่เหมาะสม จะช่วยให้เวลาในการประมวลผลลดลง โดยไม่ทำให้ประสิทธิภาพของการจำแนกลดลงได้

4.4 การประยุกต์ใช้งาน

1. จากงานวิจัย สามารถนำหลักการของกระบวนการรู้จำนี้ ไปประยุกต์ใช้งานกับการรู้จำเสียงพยางค์ต้นควบกล้ำในภาษาไทยได้

2. สามารถนำหลักการรู้จำเสียงพยางค์ต้นควบสระของงานวิจัยนี้ ไปประยุกต์ใช้งานกับการรู้จำเสียงพยางค์ท้ายในภาษาไทยได้

3. สามารถนำหลักการรู้จำเสียงพยางค์ต้นควบสระของงานวิจัยนี้ ไปพัฒนาระบบการรู้จำเสียงพูดภาษาไทยที่มีคำศัพท์จำนวนมากได้

บรรณานุกรม

- [1] กำชัย ทองหล่อ. **หลักภาษาไทย**. พิมพ์ครั้งที่ 5. กรุงเทพมหานคร : รวมสาส์น. 2519.
- [2] นงนุช สุขตั้งมั่น. “การรู้จำเสียงสระเสียงเดี่ยวและสระเสียงประสมในภาษาไทยโดยการใช้สเปกตรัมแอลพีซี บนสเกลบาร์ก.” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2547.
- [3] วรา คงกาวิฑูร, นริศ บุญศักดิ์เฉลิม และไกรสิน ส่งวัฒนา. “การรู้จำหน่วยเสียงสระเดี่ยวสำหรับภาษาไทยโดยการใช้ทรานส์เฟอร์ฟังก์ชันของอวัยวะกำทอนเสียงบนสเกลบาร์ก.” **NECTEC Technical Journal**, Vol. 3, No. 10, March-June, 2001. หน้า 1-6.
- [4] วิไลวรรณ ขนิษฐานันท์. **ภาษาและภาษาศาสตร์**. กรุงเทพมหานคร : มหาวิทยาลัยธรรมศาสตร์. 2533.
- [5] Bishnu S. Atal and Lawrence Rabiner. “A Pattern Recognition Approach to Voice-Unvoiced-Silence Classification with Applications to Speech Recognition.” **IEEE Transactions on acoustic, speech, and signal processing**, Vol. ASSP-24, No. 3, June, 1976. pp.201-212.
- [6] Eberhard Zwicker and Hugo Fastl. **Psychoacoustics : Facts and Models**. 2nd ED. Heidelberg : Springer. 1999.
- [7] Klaus Fellbaum. “**Model of the Human Speech Production**.” Online. Available : <http://www.kt.tu-cottbus.de/speech-analysis/tech.html>. 2007.
- [8] Lawrence Rabiner and Biing-Hwang Juang. **Fundamentals of Speech Recognition**. New Jersey : Prentice-Hall, Inc. 1993.
- [9] Patcharikra Cotsomrong, Treepop Sunpetchniyom, Sawit Kasuriya, Nattanun Thatphithakkul, and Chai Wutiw WATCHAI. “LOTUS : Large vOcabulary Thai continUous Speech Recognition Corpus.” **NSTDA Annual Conference S&T in Thailand: Towards the Molecular Economy (NAC2005)**, March, 2005.
- [10] Pradit Mittrapiyanuruk, Chatchawarn Hansakunbuntheung, Virongrong Tesprasit, Virach Sornlertlumvanich. “Issues in Thai Text-to-Speech Synthesis: The NECTEC Approach.” **NECTEC Technical Journal**, Vol. 2, No.7, March-June, 2000. pp.36-47
- [11] Sergios Theodoridis and Konstantinos Koutroumbas. **Pattern Recognition**. San Diego : Academic Press. 2003.

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายสุกิจ ป่านแก้ว
วัน เดือน ปีเกิด	30 มกราคม พ.ศ. 2519
ที่อยู่	82 ซ. รังสิต-ปทุมธานี 14 ต. ประชาธิปัตย์ อ. ธัญบุรี จ. ปทุมธานี 12130
ประวัติการศึกษา	2541 วิศวกรรมศาสตร์บัณฑิต สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธรรมศาสตร์