การแยกสระเสียงสั้น และสระเสียงยาวในภาษาไทย โดยใช้การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง

SHORT AND LONG VOWELS CLASSIFYING IN THAI SPOKEN USING 2nd ORDER POLYNOMIAL CURVE FITTING ON THE VOICE ENERGY FUNCTION

กัญน์ สิทธิประเสริฐ
KAN SITTIPRASERT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2547

ISBN 974-15-1216-3

การแยกสระเสียงสั้น และสระเสียงยาวในภาษาไทย โดยใช้การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง

SHORT AND LONG VOWELS CLASSIFYING IN THAI SPOKEN USING 2nd ORDER POLYNOMIAL CURVE FITTING ON THE VOICE ENERGY FUNCTION

กัญน์ สิทธิประเสริฐ
KAN SITTIPRASERT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2547

ISBN 974-15-1216-3

SHORT AND LONG VOWELS CLASSIFYING IN THAI SPOKEN USING 2nd ORDER POLYNOMIAL CURVE FITTING ON THE VOICE ENERGY FUNCTION

KAN SITTIPRASERT

THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING

SCHOOL OF GRADUATE STUDIES

KINGMONGKUT' INSTITUTE OF TECHNOLOGY LADKRABANG

2004

ISBN 974-15-1216-3

COPYRIGHT 2004

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

หัวข้อวิทยานิพนธ์	การแยกสระเสียงสั้น และสระเสียงยาวในภาษาไทย โดยใช้การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง
นักศึกษา	นายกัญน์ สิทธิประเสริฐ
รหัสประจำตัว	42061186
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้า
พ.ศ.	2547
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร. ไกรสิน ส่งวัฒนา

บทคัดย่อ

วิทยานิพนธ์นี้เสนอ วิธีการแยกสระเสียงสั้น และ สระเสียงยาวในภาษาไทย โดยใช้ลักษณะของพลังงานเสียง เป็นตัวบ่งชี้ โดยนำสัญญาณเสียงแต่ละเสียงมาคำนวณหาค่าพลังงานของเสียงในรูปแบบต่างๆ และนำลักษณะของพลังงานเสียง มาประดิษฐ์ฟังก์ชันพหุนามอันดับสอง เพื่อนำค่าสัมประสิทธิ์ ของฟังก์ชั่นพหุนามไปใช้ในการแยกสระเสียงสั้นและเสียงยาว

สระในภาษาไทยประกอบด้วยสระเดี่ยว เสียงสั้นจำนวน 9 เสียง(อะ อิ อึ อุ โอะ เอะ แอะ เออะ เอาะ) สระเดี่ยว เสียงยาว จำนวน 9 เสียง (อา อี อือ อู โอ เอ แอ เออ ออ) สระผสมเสียงสั้นจำนวน 3 เสียง (เอียะ เอือะ อัวะ) และสระผสมเสียงยาวจำนวน 3 เสียง (เอีย เอือ อัว) โดยใช้ข้อมูลเสียงจากเพศชาย 5 คน และ เพศหญิง 5 คน รวมทั้งสิ้น 5,040 เสียง

Thesis Title	Short and Long Vowels Classifying in Thai Spoken Using 2nd Order Polynomial Curve Fitting on the Voice Energy Function
Student	Mr. Kan Sittiprasert
Student ID	42061186
Degree	Master Student
Programme	Master of Engineering in Electrical Engineering
Year	2004
Thesis Advisor	Assoc.Prof.Dr. Kraisin Songwatana

ABSTRACT

This thesis presents method short and long vowels classifying in Thai spoken language by using voice energy feature. The variation in energy of vowels is calculated and the feature of energy is fitted with 2nd order polynomial regression function. The coefficient of polynomial regression is used to classify the short and long vowels.

The vowels in Thai spoken language consist of 9 unmixed short vowels (a, i,ω, u, o, e, ε, γ, ɔ̀) , 9 unmixed long vowels (aa, ii, ωω, uu, oo, ee, εε, γγ, ɔ̀ɔ̀) , 3 mixed short vowels (ia, ωa, ua) and 3 mixed long vowels (iia, ωωa, uua). The voice data samples consist of 5,040 words from 5 male and 5 female.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี ด้วยความกรุณาของ รศ.ดร. ไกรสิน ส่งวัฒนา อาจารย์ที่ปรึกษา ที่กรุณาแนะนำ ให้คำปรึกษาในการวิจัยครั้งนี้ และให้ความเมตตาแก่ศิษย์ด้วยดีเสมอมา ผู้เขียนรู้สึกซาบซึ้งในความกรุณาเป็นอย่างยิ่ง และขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆ ให้คำแนะนำ คำปรึกษา ที่ดีมาโดยตลอดระยะเวลาที่ศึกษาอยู่ให้แก่ศิษย์

ขอขอบคุณ พี่ๆ เพื่อนๆ และน้องๆ ทุกท่านทั้งที่ห้องวิจัย ที่ทำงาน และที่เรียนด้วยกัน สำหรับความเป็นมิตร และความช่วยเหลือต่างๆ ที่มีให้กันระหว่างศึกษาอยู่

ขอขอบคุณเจ้าของเสียงทุกๆ ท่าน ที่ได้เสียสละเวลามาบันทึกเสียงให้

สุดท้ายขอกราบขอบพระคุณ บิดา มารดา และครอบครัว ที่ให้ทุกสิ่งทุกอย่าง กับลูก

ประโยชน์อันถึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้เขียนขอมอบแด่ผู้มีพระคุณทุกท่าน

ขอขอบคุณด้วยใจจริง

กัญน์ สิทธิประเสริฐ

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ.....III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 สมมุติฐานของการศึกษา 3

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย 3

1.5 ขอบเขตการวิจัย 3

1.6 ขั้นตอนการศึกษา 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 กล่าวนำ 4

2.2 ทฤษฎีภาษาศาสตร์ 4

2.2.1 อวัยวะออกเสียง 4

2.2.2 กระบวนการผลิตเสียงพูด 6

2.2.3 เสียงพูดของมนุษย์ 7

2.3 หน่วยเสียงสำคัญในภาษาไทย 9

2.3.1 หน่วยเสียงสระ 9

2.3:2 หน่วยเสียงพยัญชนะ 11

2.3.3 หน่วยเสียงวรรณยุกต์ 13

2.4 ลักษณะ พยางค์ และ คำในภาษาไทย 14

2.5 หลักการประมวลผลในเชิงเวลา 14

สารบัญ (ต่อ)

หน้า

2.5.1 แบบจำลองระบบกำเนิดเสียงพูดเชิงเลข 16
2.5.2 พารามิเตอร์ของระบบกำเนิดเสียงพูด 16
2.6 การวิเคราะห์สัญญาณเสียงพูดเบื้องต้น 17
2.6.1 กรรมวิธีการปรับบรรทัดฐานแอมปลิจูด 17
2.6.2 กรรมวิธีการเน้นล่วงหน้า 17
2.6.3 การแบ่งช่วงสัญญาณ 18
2.6.4 การวินโดว์ 19
2.7 พลังงานของสัญญาณเสียงพูด 20
2.8 การประมาณค่าในช่วงด้วยเส้นโค้ง 22
2.8.1 การประมาณค่าในช่วงเชิงเส้น 23
2.8.2 การประมาณค่าในช่วงกำลังสอง 23
2.9 ระเบียบวิธีการถดถอยแบบกำลังสองน้อยที่สุด 25
2.9.1 การถดถอยแบบเชิงเส้น 25
2.9.2 การประยุกต์การถดถอยเชิงเส้นกับข้อมูลไม่เป็นเชิงเส้น 28
2.9.3 การถดถอยพหุนาม 31
2.10การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชั่นพลังงานเสียง 34

บทที่ 3 วิธีการดำเนินการวิจัย 35
3.1 การสร้างรายการคำทดสอบ 35
3.2 การคัดเลือกผู้บอกภาษา 35
3.3 การเก็บข้อมูล 35
3.4 เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียง 35
3.5 ข้อจำกัดในการดำเนินการวิจัย 36

สารบัญ (ต่อ)

หน้า

บทที่ 4 การทดลองและผลการทดลอง 37

4.1 กล่าวนำ 37

4.2 การกำหนดขอบเขตของพยางค์ 37

4.3 การเตรียมข้อมูลเสียง 40

4.4 การหาค่าพลังงานเสียง 42

4.5 การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชั่นพลังงานเสียง 44

4.6 การถดถอยแบบพหุนามของพลังงานเสียง 46

4.7 การทดสอบอัตราการแยกแยะสระเสียงสั้น และ สระเสียงยาว 49

บทที่ 5 สรุปผล และข้อเสนอแนะ 52

5.1 การทดลอง 52

5.2 ข้อสังเกต และ ปัญหาที่พบในการทดลอง 53

5.3 ข้อเสนอแนะ 53

บรรณานุกรม 54

ภาคผนวก ก. ผลงานวิจัยที่ได้รับการตีพิมพ์ 56

ประวัติผู้เขียน 57

สารบัญตาราง

ตารางที่ หน้า

2.1 เสียงสระเดี่ยวในภาษาไทย 10

2.2 เสียงพยัญชนะ ในภาษาไทย 12

4.1 การแจกแจงสัมประสิทธิ์การถดถอยพหุนามอันดับสอง 50

4.2 อัตราการแยกแยะ (แบบอ้างอิง) 51

4.3 อัตราการแยกแยะ (แบบทดสอบ) 51

สารบัญรูป

รูปที่ หน้า

1.1 ส่วนประกอบของคำในภาษาไทยโดยแยกตามลักษณะของหน่วยเสียง....2

2.1 อวัยวะในระบบการพูดของมนุษย์....6

2.2 แผนภาพระบบเสียงพูดของมนุษย์....7

2.3 ตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง....10

2.4 การเกิดเสียงสระผสมในภาษาไทย....11

2.5 องค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์....14

2.6 บล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น....15

2.7 แบบจำลองของระบบกำเนิดเสียงพูดแบบดิจิตอล....16

2.8 ขนาดสเปคตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส....18

2.9 การแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์....18

2.10 ส่วนของสัญญาณที่ตัดมาวิเคราะห์....19

2.11 วินโดว์แบบสี่เหลี่ยม(ก) ในโดเมนเวลา (ข) ในโดเมนความถี่....20

2.12 วินโดว์แบบแฮมมิ่ง (a) ในโดเมนเวลา(b) ในโดเมนความถี่....20

2.13 การประมาณค่าในช่วงโดยการใช้ฟังก์ชันเส้นตรง....23

2.14 การประมาณค่าในช่วงโดยใช้ฟังก์ชันกำลังสอง....23

2.15 การถดถอยแบบเชิงเส้นโดยการประดิษฐ์ฟังก์ชันเส้นตรงจากชุดของ ข้อมูลที่กำหนดมาให้....26

2.16 การแปลงสมการกำลังไปสู่สมการเชิงเส้นโดยใช้ลอการิทึมเพื่อใช้ วิธีถดถอยแบบเชิงเส้น....28

2.17 การแปลงสมการเอกซ์โพเนนเชียลไปสู่สมการเชิงเส้นโดยใช้ลอการิทึม เพื่อใช้วิธีถดถอยแบบเชิงเส้น....29

2.18 การแปลงสมการอัตราการเพิ่มสู่จุดอิ่มตัวไปสู่สมการเชิงเส้นเพื่อใช้ วิธีถดถอยแบบเชิงเส้น....30

2.19 การถดถอยแบบพหุนามโดยการประดิษฐ์ฟังก์ชันพหุนามจากชุดของข้อมูล....31

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.1 ขั้นตอนการแยกสระเสียงสั้นและสระเสียงยาว....37

4.2 การตัดเสียงโดยใช้ค่าพลังงานและความถี่มูลฐาน....39

4.3 สัญญาณเสียง "ขอ"....40

4.4 สัญญาณเสียง "ขอ" ที่ผ่านการปรับบรรทัดฐาน Amplitude....40

4.5 สัญญาณเสียง "ขะ"....41

4.6 สัญญาณเสียง "ขะ" ที่ถูกแบ่งเฟรมที่ 1ถึงเฟรมที่ 3....41

4.7 สัญญาณเสียง "ขะ" ในเฟรมที่ 1 ถึง 3 ที่ผ่านการ preemphasis....41

4.8 สัญญาณเสียง "ขะ" ในเฟรมที่ 1 ถึง 3 ที่ผ่านฟังก์ชั่นวินโดว์แบบแฮมมิ่ง....42

4.9 ลักษณะพลังงานเสียง....42

4.10 ลักษณะพลังงานเสียงแบบ Teager Energy ของเสียงสระต่างๆ....44

4.11 ลักษณะพลังงานที่ผ่านการ Normalize....46

4.12 ฟังก์ชั่นพหุนามของเสียงสระ....49

4.13 การแจกแจงค่าสัมประสิทธิ์การถดถอยพหุนามอันดับสอง (a_2)....49

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

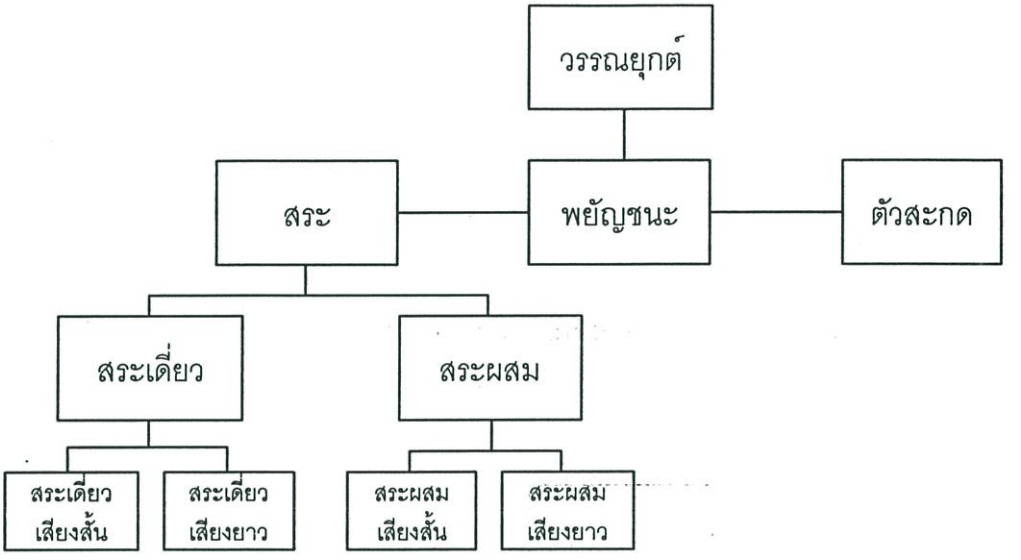
ในปัจจุบันนี้ เทคโนโลยีทางด้านคอมพิวเตอร์ได้เข้ามามีบทบาทในชีวิตประจำวันของมนุษย์เป็นอย่างมาก ซึ่งโดยปกติการติดต่อระหว่างคอมพิวเตอร์กับมนุษย์จะทำได้โดย เครื่องคอมพิวเตอร์จะรับคำสั่งเข้าทางแป้นพิมพ์ และจะแสดงผลออกทางจอภาพหรือเครื่องพิมพ์ เพราะฉะนั้นการติดต่อกับคอมพิวเตอร์อย่างมีประสิทธิภาพ ต้องอาศัยความชำนาญและการฝึกฝน การที่จะเพิ่มความสามารถของมนุษย์ในการติดต่อกับเครื่อง(Man-Machine Communication) จึงเป็นที่มาของการพัฒนาการติดต่อกับคอมพิวเตอร์โดยใช้เสียงพูดแทน ในปัจจุบันการรู้จำเสียงพูด (Speech Recognition) ถูกนำมาประยุกต์ใช้งานกันอย่างกว้างขวางโดยเฉพาะในระบบสื่อสารโทรคมนาคม จุดมุ่งหมายหลักของการรู้จำเสียงพูดก็คือการเพิ่มพูนความสามารถของอุปกรณ์ต่างๆในการรับรู้และสื่อสารโต้ตอบกับมนุษย์ได้ เพื่อเพิ่มทางเลือกในการควบคุมสั่งการอุปกรณ์เครื่องมือต่างๆ โดยเฉพาะเครื่องคอมพิวเตอร์ ซึ่งการใช้เสียงพูดควบคุมสั่งการนี้ถือได้ว่าเป็นวิธีการที่เป็นธรรมชาติมากที่สุดของมนุษย์

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

1. พิจารณาทั้งหน่วยภาษาที่เปล่งเสียงออกมาทั้งชุดมีทั้งระบบการรู้จำคำพูดเดี่ยว (Isolated word Recognition) และระบบรู้จำคำพูดต่อเนื่อง(Continuous word Recognition) ซึ่งข้อดีของระบบเหล่านี้คือง่าย เนื่องจากมีการหลีกเลี่ยงผลกระทบอันเนื่องจากฐานรากของเสียงภายในคำหรือกลุ่มคำนั้น แต่ข้อเสียคือสามารถรู้จำคำได้จำนวนคำที่จำกัด เนื่องจากต้องใช้เนื้อที่จำนวนมากในการจัดเก็บแบบจำลองอ้างอิง และต้องใช้เวลาในการคำนวณเพื่อเปรียบเทียบมากตามจำนวนของแบบจำลองอ้างอิงที่มีอยู่

2. พิจารณาโดยการแยกแยะรายละเอียดของหน่วยเสียง(Phonetic Recognition) วิธีนี้จะพิจารณาลักษณะของหน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ ดัง

แสดงในรูปที่ 1.1 โดยจะใช้หน่วยเสียงย่อยเหล่านี้เป็นหลักในการรู้จำเสียงพูด ซึ่งวิธีนี้เหมาะสำหรับการพัฒนาไปสู่ระบบการรู้จำคำจำนวนมาก



รูปที่ 1.1 ส่วนประกอบของคำในภาษาไทยโดยแยกตามลักษณะของหน่วยเสียง

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาลักษณะความแตกต่างของสระเสียงสั้นและสระเสียงยาว
2. เพื่อศึกษาวิธีการแยกแยะสระเสียงสั้น และ สระเสียงยาว
3. เพื่อศึกษาวิธีการหาพลังงานของเสียง ที่เหมาะสมที่จะนำมาใช้ในการแยกสระเสียงสั้น และ สระเสียงยาว
4. เพื่อหาตัวแปร หรือค่า พารามิเตอร์ ใดๆ ที่จะใช้เป็นตัวอ้างอิง ของความต่างทางพลังงานเสียงสระเสียงสั้น และ สระเสียงยาว
5. เพื่อเป็นองค์ประกอบในการพัฒนาไปสู่ระบบการรู้จำคำพูดภาษาไทยทั้งระบบ

1.3 สมมติฐานของการศึกษา

การศึกษาวิธีการรู้จำเสียงสระ ยังไม่สามารถระบุสระเสียงสั้น และ สระเสียงยาวได้ เนื่องจากความแตกต่างของสระเสียง เสียงสั้น และ เสียงยาวที่เห็นได้ชัดที่สุดคือ ระยะเวลาในการออกเสียง แต่ความแตกต่างนี้ไม่ทำให้สามารถแยกแยะเสียงสั้นและยาวระหว่างบุคคลได้

โดยลักษณะความแตกต่างของสระเสียงสั้น และ สระเสียงเสียงยาว อีกประการหนึ่ง คือ ลักษณะการหยุดของเสียงแบบทันทีทันใดใน สระเสียงเสียงสั้น และ การหยุดของเสียงแบบค่อยๆ ลดระดับเสียงลงอย่างช้าๆ ในสระเสียง

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย

เสียงสระในภาษาไทยมีทั้งหมด 12 คู่เสียง(สั้น,ยาว) ประกอบด้วยเสียงสระเดี่ยว 9 คู่เสียง (สั้นยาว) คือ (อะ อา) (อิ อี) (อึ อือ) (อุ อู) (โอะ โอ) (เอาะ ออ) (เอะ เอ) (แอะ แอ) (เออะ เออ) และ สระผสม 3 คู่เสียง (สั้น ยาว) คือ (เอียะ เอีย) (เอือะ เอือ) (อัวะ อัว)

การแยกแยะสระเสียงสั้นและสระเสียงยาว โดยจะใช้พลังงานเสียงเป็นตัวบ่งชี้ โดยที่พลังงานเสียงจะเป็นตัวบอกถึงลักษณะการเปล่งเสียงตลอดคำว่ามีลักษณะเป็นอย่างไร ซึ่งจะทำให้เห็นถึงช่วงของการหยุด หรือ ช่วงการจบคำ ว่ามีลักษณะเป็นอย่างไร และ หาค่าพารามิเตอร์ ที่ช่วยบ่งชี้ความต่างดังกล่าว เพื่อให้สามารถแยกแยะสระเสียงสั้น และ สระเสียงยาว ได้

1.5 ขอบเขตการวิจัย

งานวิจัยนี้ได้เลือกศึกษาเสียงสระในภาษาไทยโดยแบ่งออกเป็นเสียงสระเดี่ยว(สั้น ยาว) และเสียงสระผสม(สั้น ยาว) ที่ผันตามเสียงพยัญชนะต้น 21 รูปเสียง แต่ไม่ผันตามวรรณยุกต์ และพยัญชนะสะกด

1.6 ขั้นตอนการศึกษา

1. ศึกษาระบบการรู้จำภาษาในลักษณะต่างๆ ที่ได้มีการศึกษามาแล้ว
2. ศึกษาทฤษฎีทางด้านภาษาศาสตร์ และอัลกอลิธึมต่างๆที่ใช้ในการวิเคราะห์เสียงพูด
3. กำหนดขอบเขตการวิจัย
4. บันทึกข้อมูลเสียง โดยใช้ผู้พูดทั้งหมด 10 คน เพศชาย 5 คน เพศหญิง 5 คน
5. เก็บข้อมูล วิเคราะห์เสียง และทดสอบความถูกต้อง
6. สรุปผลการวิจัย และเสนอแนะแนวทางในการวิจัยต่อไป

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 กล่าวนำ

ในบทนี้จะกล่าวถึงทฤษฎีต่างๆที่เกี่ยวข้องกับงานวิจัยนี้ โดยจะเริ่มจากทฤษฎีทางภาษาศาสตร์ ซึ่งจะกล่าวถึงเรื่องทฤษฎีการสร้างเสียงพูดประกอบด้วย หน่วยเสียงสระ หน่วยเสียงพยัญชนะ หน่วยเสียงวรรณยุกต์ และลักษณะคำ และพยางค์ ในภาษาไทย , การเตรียมข้อมูลเสียง , การหาค่าพลังงานเสียง , การประมาณค่าพหุนาม และการถดถอยพหุนาม

2.2 ทฤษฎีภาษาศาสตร์

มนุษย์สร้างเสียงพูดขึ้นมาได้โดยมนุษย์จะมีเส้นเอ็นในลำคอที่เรียกว่าเส้นเสียง (Vocal Cords) ซึ่งอยู่ในช่องหลอดลม และจากการสั่นของเส้นเสียงจะทำให้เกิดคลื่นเสียงความถี่ต่างๆ ผ่านเข้าสู่ลำคอ ผ่านจากลำคอเข้าสู่ปากหรือช่องจมูกออกไปภายนอกซึ่งขนาดและรูปร่างของอวัยวะภายในช่องปากนี้จะเป็นสิ่งกำหนดว่าคลื่นเสียงความถี่ไหนจะยอมให้ปรากฏออกมาให้ได้ยิน (Passed Through) หรือคลื่นเสียงความถี่ไหนที่จะถูกดูดซับไว้ไม่ยอมให้ปรากฏออกมา (Absorbed) ปรากฏการณ์ที่ทำให้เกิดเสียงพูดออกมาได้นี้ เนื่องจากรูปร่างของช่องคอและช่องปากของมนุษย์ที่สามารถเปลี่ยนแปลงไปมาได้อย่างอิสระ ทำให้ผู้พูดมีความสามารถในการกำหนดคลื่นเสียงระดับต่างๆ เหล่านี้นั้นเองที่ทำให้เกิดการแปลงเป็นเสียงพูดออกมา

2.2.1 อวัยวะออกเสียง (The Organs of Speech)

ธรรมชาติได้กำหนดตำแหน่งของอวัยวะที่ใช้ในการทำให้เกิดเสียงของมนุษย์เอาไว้ นั้นคือริฝีปาก ฟัน เพดาน และ ลิ้น ที่อยู่ตามลำดับตลอดช่องของกล้ามเนื้อช่องคอ เพื่อทำให้สามารถเปลี่ยนขนาดและรูปร่างได้ตามต้องการ ดังรูปที่ 2.1 เราจะเปล่งคำพูดออกมาโดยการเปล่งออกมาเป็นลำดับขั้นของความถี่คลื่นเสียงที่ได้คัดเลือกไว้ออกมา นั้นคือ จะมีการเปลี่ยนแปลงตำแหน่งของช่องทางเดินเสียง (Vocal Tract) ของมนุษย์เป็นลำดับขั้นอย่างต่อเนื่อง

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

1) อวัยวะที่เป็นส่วนกระทำอาการ (Articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลักลมไปยังส่วนต่างๆ อวัยวะที่เป็นส่วนกระทำการที่สำคัญ คือลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

2) อวัยวะที่เป็นฐานที่เกิดเสียงต่างๆ (Point of articulator) หมายถึง ตำแหน่งหรือฐานที่เกิดของเสียงต่างๆ เช่น ริมฝีปาก ฟัน เพดานส่วนต่างๆ เป็นต้น

จากรูปที่ 2.1 อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูดมี ดังนี้

ริมฝีปาก (Lips) เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ยื่นออกมา หรือ ห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่างๆ นี้ ล้วนมีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น

ฟัน (Teeth) เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่าง หรือกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอผ่านได้ออกมาทำให้เกิดเป็นเสียงชนิดที่เรียกว่าเสียงเสียดแทรกที่เกิดที่ฟัน เป็นต้น

ปุ่มเหงือก (Gum ridge) คือส่วนที่อยู่ต่อจากฟันบน

เพดานแข็ง หรือเพดานปาก (Hard palate) หมายถึง เฉพาะเพดานที่โค้งเป็นกระดูกแข็งเท่านั้น

เพดานอ่าน (Soft palate) คือส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดกับหลังคอ นอกจากเวลาออกเสียงนาสิก เท่านั้นที่เพดานอ่อนจะลดระดับลงมา

ลิ้น (Tongue) เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งเป็น 3 ส่วน ด้วยกันตามหน้าที่ในการออกเสียง

ปลายลิ้น (Blade of the tongue) หมายถึงส่วนปลายซึ่งสามารถยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้ง่าย

หน้าลิ้น (Front of the tongue) หมายถึงลิ้นที่อยู่ตรงข้ามกับเพดานแข็ง ขณะที่วางลิ้นราบกับปากตอนที่ไม่ได้พูด

หลังลิ้น (Back of the tongue) หมายถึง ส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ขณะที่วางลิ้นราบกับปาก

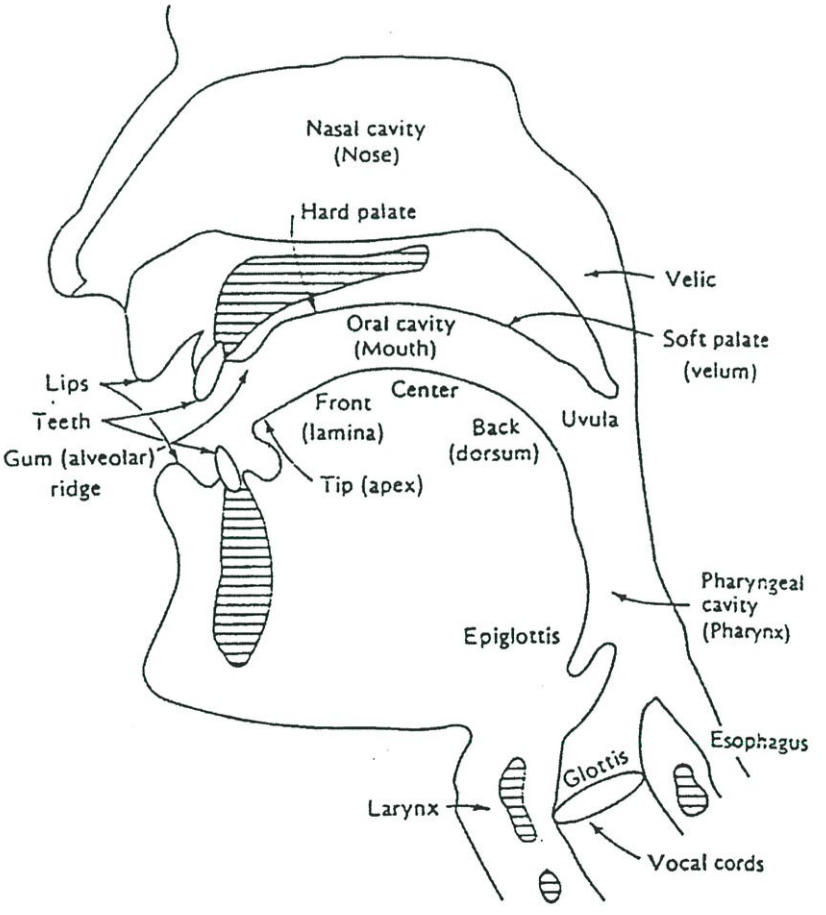
ช่องปาก (Oral cavity) ช่องปากนี้เป็นอวัยวะอีกส่วนหนึ่งที่มีความสำคัญในการพูด กล่าวคือช่องปากจะทำหน้าที่เป็นช่องกำทอน (Resonant chamber) ซึ่งมีบทบาทในการกำทอนเสียงที่เดินทางผ่านมาถึงบริเวณนี้ ทั้งนี้เพราะช่องปากสามารถจะเปลี่ยนแปลงเป็นช่องกำทอนรูปร่างต่างๆ กัน เนื่องจากรูปร่างของมันแปรผันไปตามการจัดท่าทางของลิ้น ริมฝีปากและขากรรไกร

โพรงจมูก (Nasal cavity) คุณสมบัติลักษณะของเสียงที่พูดที่เกิดขึ้นจะแปรผันไปตามการปิด – เปิดของช่องทางออกสู่โพรงจมูก ซึ่งเป็นผลจากการยกขึ้นหรือเลื่อนลงของเพดานอ่อน

ช่องคอ (Pharynx) หมายถึงโพรงคอซึ่งอยู่ถัดจากปากลงไปจนถึงเส้นเสียง

ลิ้นปิดหลอดลม (Epiglottis) เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่ อยู่ต่อจากโคนลิ้นลงไปในลำคอ มีหน้าที่ปิดช่องลมขณะรับประทานอาหาร และปิดช่องลมและจมูก

เส้นเสียง (Vocal cords) เป็นอวัยวะสำคัญที่เกิดของเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ 2 แผ่น ปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างระหว่างเส้นเสียง (glottis) เส้นเสียงทั้งสองสามารถจะดึงออกให้ห่างจากกัน หรือดึงเข้ามาประชิดกันก็ได้ เส้นเสียงเป็นส่วนทำให้เกิดเสียงพูดขึ้น



รูปที่ 2.1 แสดงอวัยวะในระบบการพูดของมนุษย์

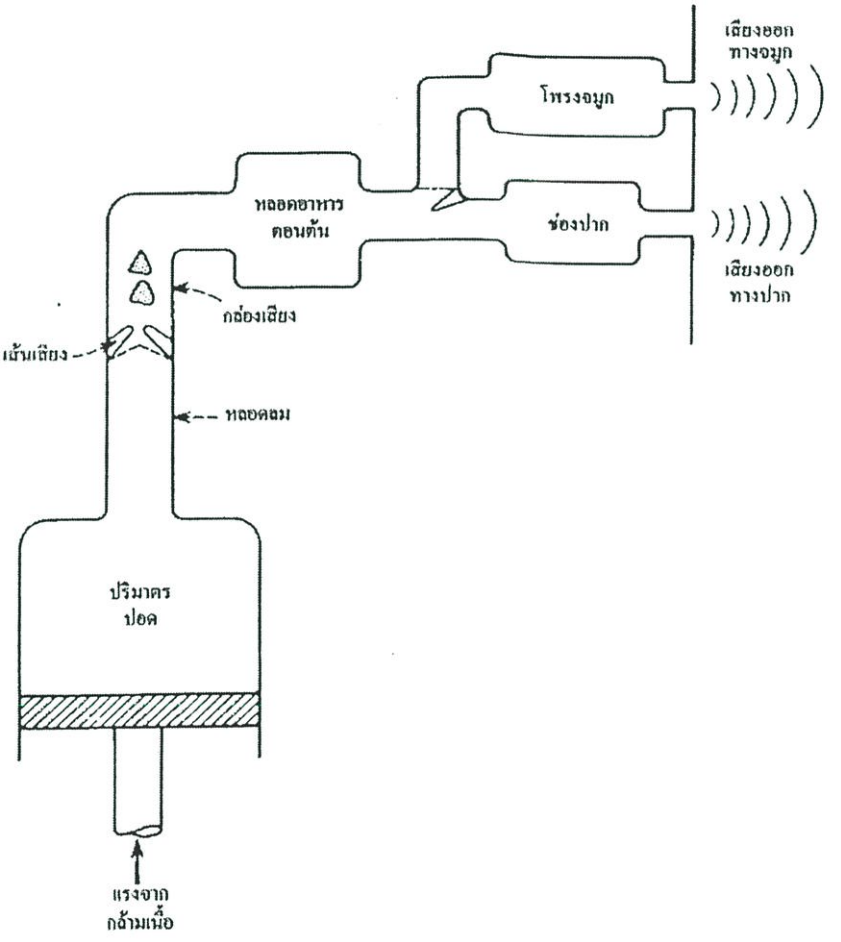
2.2.2 กระบวนการผลิตเสียงพูด

จากระบบเสียงพูดของมนุษย์ สามารถแสดงเป็นแผนภาพของระบบกำเนิดเสียงดังรูปที่ 2.2 ซึ่งเราสามารถจำแนกกลไกสร้างเสียงพูด ของมนุษย์ได้ 3 แบบ ดังนี้

- อากาศไหลจากปอดจะถูกมอดูเลท (Modulate) โดยการสั่นของเส้นเสียงทำให้เกิดคลื่นเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (Quasi_periodic puls – like excitation)

- อากาศไหลจากปอดถูกทำให้ปั่นป่วน ด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปากทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (Noise – like excitation)

- อากาศที่ไหลถูกกัก และเกิดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้อากาศที่มีแรงดันพุ่งออกไปอย่างรวดเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น (Transient excitation)



รูปที่ 2.2 แผนภาพระบบเสียงพูดของมนุษย์

2.2.3 เสียงพูดของมนุษย์

เสียงพูดเป็นคลื่นความยาว (Longitudinal wave) เกิดจากการสั่นของอนุภาคตัวกลาง นั้นคืออากาศ และทิศทางการสั่นของอนุภาคจะอยู่ในทิศเดียวกันกับทิศทางของการเคลื่อนที่ คลื่นเสียงเป็นคลื่นที่เปลี่ยนแปลงไปตามเวลา

เสียงพูดแบ่งได้เป็น 2 ชนิด ตามการกำเนิดเสียงหรือโหมด (Mode) การกระตุ้น คือ

2.2.3.1 เสียงวอยซ์ หรือเสียงโฆษะ (Voiced sound) เกิดจากการบังคับอากาศให้ผ่านสายเสียงทำให้มีการเปลี่ยนแปลงความตึงหย่อนของเส้นเสียง โดยเส้นเสียงจะสั่นและเกิดเป็นพัลส์ (Pulse) ของอากาศไปกระตุ้นอวัยวะกำทอนเกิดเป็นเสียงวอยซ์ ตัวอย่างเสียงวอยซ์ ได้แก่ เสียงสระ เสียงพยัญชนะ ที่ออกเสียงจากลำคอ (Voiced consonants)

2.2.3.2 เสียงอันวอยซ์หรือเสียงอโฆษะ (Voiceless หรือ Unvoiced sounds) เป็นเสียงที่เกิดในช่องปากหรือโพรงจมูกโดยอวัยวะภายในช่องปาก ริมฝีปาก ขวางการไหลของอากาศได้เป็นช่องเล็กๆ อากาศจึงไหลผ่านอย่างรวดเร็วและปั่นป่วนจนกระทั่งสร้างเป็นเสียงรบกวนช่วงความถี่กว้าง (Broad – spectrum noise) ตัวอย่างเสียงอันวอยซ์ ได้แก่ เสียงพยัญชนะที่ไม่ได้เกิดจากลำคอ (Voiceless consonants)

2.2.3.4 ความยาวของเสียง (Length)

การเกิดเสียงใดเสียงหนึ่งเปล่งออกมาได้นานเท่าใด เสียงพูดบางเสียงอาจจะเปล่งออกมาได้ติดต่อกันได้นาน เช่น เสียงสระ เสียงพยัญชนะ หรือเสียงพยัญชนะเสียดแทรก

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

2.2.3.5 ระดับเสียง สูง – ต่ำ (Pitch)

เสียงพูดจะมีระดับ สูงหรือต่ำ อยู่ที่ความถี่พื้นฐานของเสียง (Fundamental frequency) ถ้าความถี่ต่ำเสียงก็จะต่ำอวัยวะส่วนที่ทำให้เสียงมีระดับ สูง – ต่ำ คือ เส้นเสียง ดังนั้นระดับเสียงสูง – ต่ำ ก็คืออัตราการสั่นสะบัดของเส้นเสียงนั้นเอง

ในการพูดเสียงที่มีระดับสูง – ต่ำ ได้คือเสียงก้องเท่านั้น เพราะมีการสั่นสะเทือนของเส้นเสียง ที่ทำให้เกิดมีความถี่ระดับต่างๆ ได้ ในภาษาไทยระดับเสียงสูง – ต่ำ ของคำเราเรียกว่า "วรรณยุกต์"

2.2.3.6 ความดัง (Loudness)

ความดังขึ้นอยู่กับปริมาณของลม ที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่งๆ

2.2.3.7 การลงน้ำหนัก (Stress)

การออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นที่อยู่ข้างเคียง (เพื่อต้องการเรียกร้องความสนใจเป็นพิเศษ หรือแสดงอารมณ์ อย่างใดอย่างหนึ่ง)

2.2.3.8 ช่วงต่อของเสียง (Juncture)

ช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งเสียงอีกเสียงหนึ่ง ซึ่งเรียงกันมาเป็นลำดับเสียงที่ประกอบกันเข้าเป็นพยางค์ จะมีช่วงต่อของเสียงแนบสนิทจนไม่เห็นร่องรอย (Close juncture) แต่ถ้าเสียงปรากฏอยู่คนละพยางค์ หรือคนละคำจะมีช่วงต่อ "ห่าง" จนสังเกตได้ชัด

(Open juncture) ดังนั้นช่วงต่อของเสียง โดยเฉพาะช่วงต่อห่างจะมีความสำคัญมากในการแบ่งคำในภาษา

2.3 หน่วยเสียงสำคัญในภาษาไทย

"หน่วยเสียง" (Phoneme) เป็นหน่วยเล็กที่สุดของภาษา หน่วยดังกล่าวได้แก่เสียงสำคัญๆ ในภาษาใดภาษาหนึ่งซึ่งทำหน้าที่ให้ความหมายของคำที่ใช้ในภาษานั้น และทำให้ความหมายของคำนั้นๆ มีความหมายแตกต่างจากคำอื่นๆ หน่วยเสียงสำคัญในภาษาไทยมี 3 ประเภทใหญ่ๆ คือ เสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์ หน่วยเสียงทั้ง 3 นี่เองที่ประกอบกันเข้าเป็นคำที่ใช้ในภาษาไทย

เสียงพูดของมนุษย์ซึ่งมีความแตกต่างกันมากมายนั้น ถ้าเราพิจาณาอย่างถ้วนๆ จะพบว่าสามารถแบ่งออกเป็น 2 ประเภทใหญ่ คือ

1. เสียงเรียง (Segmental sound) เป็นหน่วยเสียงที่แยกออกจากเสียงอื่นได้โดยเด็ดขาด เพราะมันมีลักษณะเด่นเฉพาะตัว ในภาษาไทยได้แก่เสียงสระ และเสียงพยัญชนะ

2. เสียงซ้อน (Supra – segmental feature) เป็นเสียงที่ทำหน้าที่เป็นส่วนประกอบของเสียงอื่นเพราะไม่สามารถแยกเปล่งเสียงได้ตามลำพัง ในภาษาไทยได้แก่เสียงวรรณยุกต์ และทำนองเสียงเป็นต้น

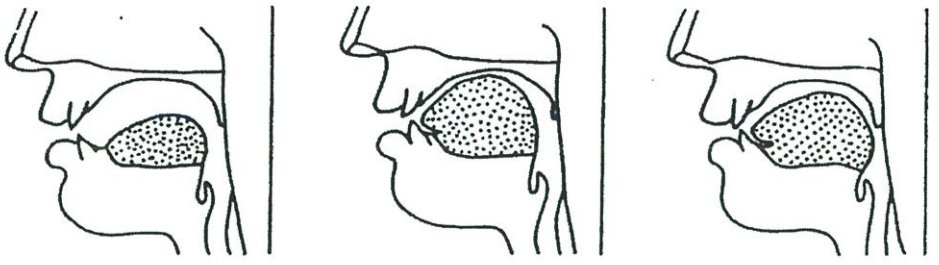
2.3.1 หน่วยเสียงสระ

2.3.1.1 ลักษณะของเสียงสระ

ลักษณะสำคัญของเสียงสระก็คือ " เป็นเสียงก้องที่เปล่งเสียงออกมาโดยให้ลมออกทางช่องปากโดยไม่ถูกลิ้นกักหรือขัดขวาง" ดังนั้นเวลาเราออกเสียงสระจะออกเสียงได้สะดวกและออกเสียงได้นาน ทั้งนี้เพราะคุณสมบัติของเสียงสระมีความดังเด่นกว่าเสียงอื่นๆ ที่เรียงอยู่ข้างเสมอ อวัยวะที่เกี่ยวข้องกับการออกเสียงสระได้แก่ลิ้น กับริมฝีปาก ถ้าลิ้นส่วนใดทำหน้าที่เพียงเพียงส่วนเดียว เสียงที่เกิดขึ้นก็จะมีเพียงเส้นเสียงเดียว เสียงเช่นนี้เรียกว่า "สระเดี่ยว" แต่ถ้าลิ้นส่วนอื่นทำหน้าที่ร่วมด้วยเสียงสระนั้นเรียกว่า "สระผสม"

ฉะนั้นการอธิบายเรื่องการออกเสียงสระจึงเป็นเรื่องที่ยากกว่าการออกเสียงพยัญชนะเพราะไม่มีจุดสัมผัสในปากที่เราสามารถใช้เป็นหลักในการอธิบายได้ การที่สระมีเสียงต่างๆ กันก็เพราะช่องคอ ช่องปากมีรูปร่างต่างกันในขณะที่อากาศกำลังผ่านออกจากปากไป สิ่งที่มีส่วนทำให้ช่องปากช่องคอมีรูปร่างต่างๆ กันออกไปก็คือ ลิ้น และริมฝีปาก เพื่อความสะดวกในการอธิบายเสียงเราจะแบ่งลิ้นออกเป็น 3 ส่วน คือ ส่วนหน้า (Front) ส่วนกลาง (Center) และส่วนหลัง (Back) ในขณะที่เราหุบปากอยู่นิ่งๆ ส่วนหน้าของลิ้นอยู่ใต้เพดานแข็ง (hard palate) ส่วนหลังอยู่ใต้

เพดานอ่อน (Velum) และส่วนกลางอยู่ระหว่างกลางของสองส่วนนี้ รูปต่อไปนี้แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง



[a] "_ะ" [i] "_ิ" [u] "_ุ"

รูปที่ 2.3 แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง

สำหรับภาษาไทยมีหน่วยเสียงสระทั้ง 24 หน่วยเสียง แยกออกเป็นสระเดี่ยว 18 หน่วยเสียง และสระผสม 6 หน่วยเสียง

<u>สระเดี่ยว</u>

เสียงสระเดี่ยว 18 หน่วยเสียง พิจารณาการเกิดเสียงได้เป็น 2 กรณีใหญ่ๆ คือ

1.การเกิดจากส่วนต่างๆ ของลิ้นหมายถึงลมผ่านส่วนหน้า ส่วนกลาง หรือส่วนหลังของลิ้น

2. การเกิดจากลมผ่านลิ้นในขณะที่ลิ้นอยู่ในระดับ สูง กลาง หรือ ต่ำ

ตารางที่ 2.1 เสียงสระเดี่ยวในภาษาไทย

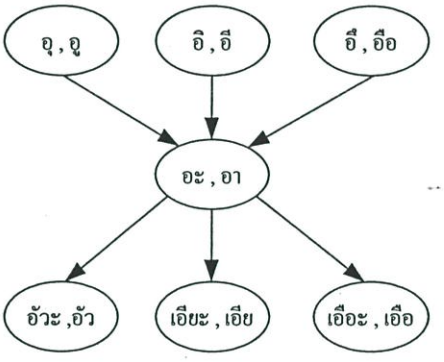
สระ	ระดับลิ้น	หน่วยเสียง	สัญลักษณ์
หน้า	สูง	อิ, อี	" i ", " i: "
	กลาง	เอะ, เอ	" e ", " e: "
	ต่ำ	แอะ, แอ	" ɛ ", " ɛ: "
กลาง	สูง	อึ, อือ	" ɯ ", " ɯ: "
	กลาง	เออะ, เออ	" ɔ ", " ɔ: "
	ต่ำ	อะ, อา	" a ", " a: "
หลัง	สูง	อุ, อู	" u ", " u: "
	กลาง	โอะ, โอ	" o ", " o: "
	ต่ำ	เอาะ, ออ	" ɤ ", " ɤ: "

นอกจากนี้ หน่วยเสียงสระเดี่ยว 18 หน่วย สามารถแบ่งตามความสั้น – ยาว ของการออกเสียงได้เป็น

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

<u>สระผสม</u>

เสียงผสม 6 หน่วยเสียง เกิดจากลมผ่านกระทบลิ้น 2 ส่วนคือ ส่วนบน และส่วนล่าง ซึ่งในขณะออกเสียงลิ้น จะอยู่ในระดับสูง แล้วลดลงต่ำโดยเสียงหลังเป็นสระ (อะ อา) เสมอ ดังแผนผัง ดังนี้



รูปที่ 2.4 แสดงการเกิดเสียงสระผสมในภาษาไทย

2.3.1.2 หน้าที่ของหน่วยเสียงสระในภาษาไทย

หน่วยเสียงสระในภาษาไทยทั้ง 24 หน่วยเสียงนี้ ทำหน้าที่เป็นแกนกลางของพยางค์ หรือคำกล่าวคือ คำทุกคำในภาษาไทยจะต้องมีเสียงสระอยู่ด้วย และเสียงสระในภาษาไทยจะสามารถเกิดกับเสียงพยัญชนะต้นได้ทุกเสียง และสามารถเกิดกับหน่วยเสียงวรรณยุกต์ได้ทุกหน่วย แต่ไม่สามารถเกิดกับหน่วยเสียงพยัญชนะสะกดได้ทุกหน่วย หน่วยเสียงสระที่ทำให้เกิดคำหรือพยางค์ใช้ได้มากที่สุดในภาษามักเป็นหน่วยเสียงยาว

2.3.2 **หน่วยเสียงพยัญชนะ**

เสียงพยัญชนะในภาษาไทยมีทั้งหมด 21 หน่วยเสียง (44 รูป) หน่วยเสียงพยัญชนะออกเสียงได้ไม่สะดวกเท่าหน่วยเสียง เพราะเวลาออกเสียงลมหายใจที่พุ่งออกมาจากหลอดลมจะถูกขัดขวางตามส่วนต่างๆ ของปาก เสียงพยัญชนะจึงออกเสียงให้ยาวนาน อย่างเสียงสระไม่ได้ และเสียงพยัญชนะก็ไม่ใช่เสียงก้องเสมอไป

2. เป็นพยัญชนะสะกดของพยางค์ ในตำแหน่งนี้เสียงพยัญชนะในภาษาไทยสามารถเกิดได้ 9 หน่วยเสียงคือ / ป / (แม่กบ) / ต / (แม่ กด) / ก / (แม่กด) / ม / (แม่กม) / ง / (แม่กง) / น / (แม่กน) / ว / (แม่เกอว) และ ไม่มีเสียงพยัญชนะ สะกด (แม่กา)

2.3.3 หน่วยเสียงวรรณยุกต์

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

ในภาษาไทยหน่วยเสียงวรรณยุกต์เป็นหน่วยเสียงสำคัญ มีความหมายต่างกัน ดังนั้นอาจกล่าวได้ว่าหน้าที่ ของหน่วยเสียงวรรณยุกต์ ก็คือ การทำให้เกิดคำขึ้นใช้ในภาษามากขึ้น และเป็นวิธีการสร้างคำขึ้นใช้เพิ่มในภาษาเป็นวิธีแรก ทั้งนี้เพราะถ้าเราเปลี่ยนเสียงวรรณยุกต์ ก็จะทำให้คำเกิดความหมายเพิ่มขึ้นใหม่ นั้นเอง

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

2.3.3.1 ลักษณะของเสียงวรรณยุกต์

สามารถแบ่งออกตามลักษณะระดับเสียงได้เป็น 2 กลุ่มใหญ่ คือ

1. กลุ่มวรรณยุกต์ระดับ (Level tone) มี 3 หน่วยเสียง คือ

1.1 หน่วยเสียงวรรณยุกต์ระดับ ต่ำ (Low tone) แทนด้วยสัญลักษณ์ / \ /

คือเสียงวรรณยุกต์เอกหน่วยเสียงนี้จะปรากฏในพยางค์ของภาษาไทย ได้ทุกแบบ

1.2 หน่วยเสียงวรรณยุกต์ระดับกลาง (Mid tone) ไม่มีสัญลักษณ์

คือเสียงวรรณยุกต์สามัญ หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ ที่มีตัวสะกดเป็นพยัญชนะกัก (พยางค์คำตาย)

1.3 หน่วยเสียงวรรณยุกต์ระดับสูง (High tone) แทนด้วยสัญลักษณ์ / ˊ /

คือเสียงวรรณยุกต์ตรี หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่ผสมด้วยสระเสียงยาว ซึ่งมีตัวสะกดเป็นเสียงกัก

2. กลุ่มวรรณยุกต์เปลี่ยนระดับ (Contour tone) มี 2 หน่วยเสียง คือ

2.1 หน่วยเสียงวรรณยุกต์เปลี่ยนตก (Falling tone) แทนด้วยสัญลักษณ์ / ˆ /

คือเสียงวรรณยุกต์ไทย หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มีสระเสียงสั้น และมีเสียงพยัญชนะสะกดเป็นพยัญชนะกัก

2.2 หน่วยเสียงวรรณยุกต์ เปลี่ยนขึ้น (Rising tone) แทนด้วยสัญลักษณ์ / ∨ /

คือเสียงวรรณยุกต์จัตวา หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มีเสียงพยัญชนะสะกดเป็นเสียงกักเลย

2.4 ลักษณะ พยางค์ และคำในภาษาไทย

ศ.ดร. กาญจนา นาคสกุล (2520 : 104) ได้ให้ความหมายของพยางค์ในระบบเสียงภาษาไทยว่า "พยางค์ หมายถึง จำนวนเสียงที่ดังเด่นซึ่งปรากฏในกลุ่มเสียงที่เรียงกันเป็นคำพูด ส่วนเสียงอื่นๆ ที่อยู่ข้างเคียงก็จะประกอบกันเข้าเป็นส่วนของพยางค์" เสียงที่ดังเด่นในกลุ่มเสียงก็คือเสียงสระ ซึ่งมีลักษณะประจำตัวก็คือเป็นเสียงก้อง ซึ่งดังเด่นกว่าเสียงอื่นๆ ดังนั้นเสียงสระจึงมักเป็นเสียงที่ทำให้เกิดพยางค์ ถ้ามีเสียงสระเด่นอยู่กี่เสียง พยางค์ก็จะมีจำนวนเท่านั้นด้วย

พยางค์ที่เปล่งออกมาครั้งหนึ่งๆ อาจมีความหมายหรือไม่ก็ได้ แต่เมื่อใดพยางค์ที่ประกอบขึ้นจากเสียงสระพยัญชนะ และวรรณยุกต์ เป็นอย่างน้อยที่สุด และกลุ่มเสียงเหล่านี้มีความหมายและสามารถปรากฏได้โดยลำพัง พยางค์นั้นๆ ก็กลายเป็นคำในภาษา

คำในภาษาไทยส่วนใหญ่จะเป็นคำพยางค์เดียว ซึ่งเป็นคำพื้นฐาน (Bass words) ของภาษา ภาษาไทยจึงจัดอยู่ในตระกูลภาษา คำโดด หรือ คำพยางค์เดียว (Monosyllabic language) หน่วย เสียงที่ประกอบกันเข้าเป็นพยางค์ จะต้องมีอย่างน้อย 3 หน่วยคือ หน่วยเสียงพยัญชนะต้น 1 หน่วย หน่วยเสียงสระ 1 หน่วย และหน่วยเสียงวรรณยุกต์ 1 หน่วย และมีหน่วยเสียงอย่างมากไม่เกิน 5 หน่วย คือ เพิ่มหน่วยเสียงพยัญชนะต้นที่เป็นเสียงควบกล้ำอีก 1 หน่วย และหน่วยเสียงพยัญชนะสะกด อีก 1 หน่วย โดยมีองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์แสดงได้ ดังรูปที่ 2.5

		วรรณยุกต์	
พยัญชนะต้น	(ควบ)	สระ	(พยัญชนะสะกด)

รูปที่ 2.5 แสดงองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์

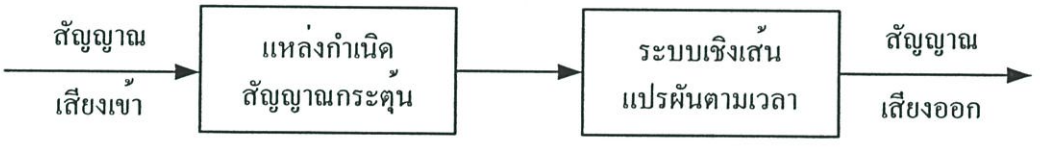
2.5 หลักการประมวลผลในเชิงเวลา

หลักการประมวลผลในเชิงเวลา คือการวิเคราะห์สัญญาณของเสียงพูดโดยตรงจากรูปคลื่นสัญญาณ (wave form) ซึ่งแตกต่างจากการประมวลผลเชิงความถี่ (Frequency – Domain) ตัวอย่างของการประมวลผลของสัญญาณเสียงพูดเชิงเวลา เช่น การหาค่าอัตราการตัดศูนย์(Zero Crossing Rate),การหาค่าพลังงาน (Energy) และค่าออโตคอรีเลชั่น (Autocorrelation) โดยการ

ประมวลผลหาค่าเหล่านี้ เป็นไปโดยง่าย มีขั้นตอนที่ไม่ซับซ้อน แต่ให้องค์ประกอบสำคัญของสัญญาณเสียง

สัญญาณเสียงพูด ที่มีอัตราการสุ่มตัวอย่างของสัญญาณเป็น 8000 ค่าใน 1 วินาที จะเห็นได้ชัดว่าคุณสมบัติของสัญญาณเสียงพูดจะมีลักษณะที่เปลี่ยนแปลงไปตามเวลาตัวอย่างเช่น การเปลี่ยนแปลงสัญญาณเสียงพูดระหว่างเสียงพูดที่เป็นเสียงก้อง หรือ เสียงโฆษะ (Voice) และเสียงไม่ก้อง หรือ เสียงอโฆษะ (unvoice) โดยที่เสียง จะมีการเปลี่ยนแปลงของความถี่มูลฐานของเสีย (fundamental frequency) ภายในช่วงที่เป็นโฆษะด้วย โดยการที่แสดงรูปแบบสัญญาณเป็นสามารถทำให้ง่าย ต่อการสังเกตถึงองค์ประกอบ และคุณลักษณะของเสียงนั้น เช่น ความเข้มของเสียง (Intensity), ชนิดของเสียง (เสียงก้องหรือเสียงไม่ก้อง) หรือ excitation mode , ความถี่มูลฐาน (pitch หรือ fundamental frequency) และอาจรวมถึงสัมประสิทธิ์ในอวัยวะกำเนิดเสียง เช่น ความถี่กำทอน (resonant หรือ formant frequency)

ในบทที่ผ่านมาเราได้ทราบการทำงานของระบบเสียงพูดของมนุษย์มาแล้ว ซึ่งเราสามารถแสดงบล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น ดังรูปที่ 2.6 จากรูปมีการแยกภาคแหล่งกำเนิดสัญญาณกระตุ้นออกจากส่วนกำทอนเสียง ซึ่งแทนด้วยระบบเชิงเส้นแปรผันตามเวลา



รูปที่ 2.6 บล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น

จากรูปที่ 2.6 แหล่งกำเนิดสัญญาณ กระตุ้นทำหน้าที่แทน การทำงานของปอดและกล่องเสียง ส่วนนี้จะผลิตขบวนพัลส์ที่มีคาบเวลาพิทช์ ขณะเปล่งเสียงโฆษะ(voice) และให้กำเนิดเสียงซึ่งคล้ายเสียงรบกวนขณะเปล่งเสียงอโฆษะ(unvoice)

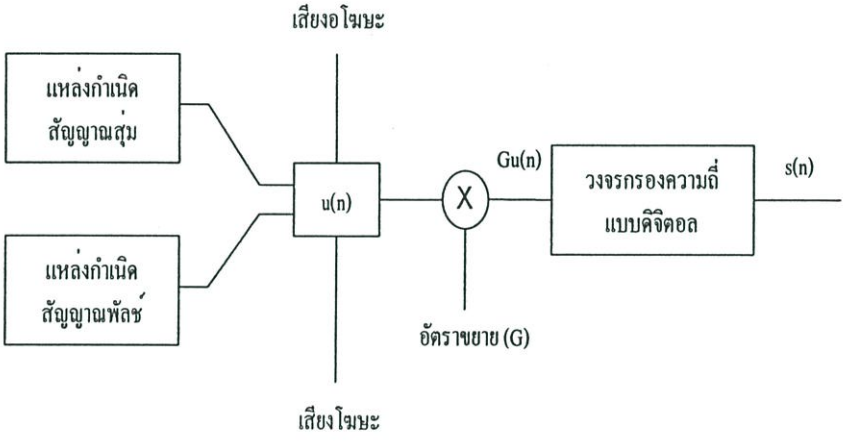
ส่วนที่สองเป็นท่อกำทอนเสียง จะแทนการทำงานของช่องปาก และโพรงจมูก ทำหน้าที่เสมือนตัวกรองสัญญาณ (filter) ที่ยอมให้ความถี่ฟอร์แมนท์ผ่านได้ ซึ่งสามารถแทนด้วยระบบเชิงเส้นแปรผันตามเวลา(time – varying linear system)

2.5.1 แบบจำลองระบบกำเนิดเสียงพูดเชิงเลข

จากแบบจำลองรูปที่ 2.6 ประกอบกับการใช้เทคนิคการประมวลผลสัญญาณเชิงเลข(digital signal processing) สามารถแสดงเป็นแบบจำลองระบบกำเนิดเสียงพูดเชิงเลขดังรูปที่ 2.7

2.5.1.1 แหล่งกำเนิดสัญญาณกระตุ้น แบ่งเป็น 2 ชนิด ตามประเภทเสียงโดยชนิดแรกเป็นแหล่งกำเนิดเสียงโฆษะ(voice) ทำหน้าที่ผลิตขบวนสัญญาณพัลส์ ด้วยคาบเวลาคงที่ และชนิดที่สองเป็นแหล่งกำเนิดสัญญาณเสียงรบกวนแบบสุ่ม ในกรณีที่กำเนิดเสียงอโฆษะ(unvoice)

2.5.1.2 แบบจำลองอวัยวะกำทอนเสียง เป็นวงจรกรองเชิงดิจิตอล ซึ่งเป็นชนิดที่มีแต่โพลทั้งหมด (all – pole digital filter) หรือ วงจรเชิงเลขแบบป้อนกลับ (recursive digital filter)



รูปที่ 2.7 แสดงแบบจำลองของระบบกำเนิดเสียงพูดแบบดิจิตอล

2.5.2 พารามิเตอร์ (Parameter) ของระบบกำเนิดเสียงพูด

การสร้างระบบเสียงพูดจำเป็นต้องวิเคราะห์หาค่าตัวแปรหรือพารามิเตอร์ที่ต้องใช้ตามแบบจำลองและระบบประมวลสัญญาณเสียง ซึ่งประกอบด้วย

2.5.2.1 ดัชนีหรือพารามิเตอร์บอกชนิดเสียงพูดว่าเป็นเสียงโฆษะ(voice)หรืออโฆษะ (unvoice)

2.5.2.2 คาบเวลาพิทช์ของขบวนพัลส์ในกรณีเสียงโฆษะ(voice)

2.5.2.3 อัตราการขยาย (gain) ของวงจรกรอง หรือค่ารากกำลังสองเฉลี่ย (rms value) ของเสียง

2.5.2.4 พารามิเตอร์ของแบบจำลองท่อกำทอนหรือสัมประสิทธิ์ของวงจรกรอง

2.5.2.5 อัตราการสุ่มตัวอย่างสัญญาณเสียงหรืออัตราเร็วข้อมูล

2.6 การวิเคราะห์สัญญาณเสียงพูดเบื้องต้น

โดยธรรมชาติของสัญญาณเสียงพูดจะไม่เสถียร และเปลี่ยนแปลงตามเวลา (Non-Stationary) ดังนั้น เมื่อต้องการนำสัญญาณเสียงพูดมาประมวลผลสัญญาณดิจิทัล (Digital Signal Processing) จึงจำเป็นต้องแบ่งสัญญาณพูดออกเป็นช่วงเวลาสั้น ๆ (Short Time) เพื่อให้สัญญาณเสียงมีความเสถียรและไม่เปลี่ยนแปลงเวลา (Stationary) จากนั้นจึงจะสามารถนำสัญญาณเสียงไปประมวลผลต่อไปได้ กรอบเสียงพูด (Speech Frame) ความยาวประมาณ 20-30 มิลลิวินาที ทำให้สัญญาณเสียงพูดในแต่ละกรอบเสียงพูดเป็นสัญญาณที่มีความเสถียรและไม่เปลี่ยนแปลงตามเวลา การเหลื่อมกรอบเสียงพูด (Frame Overlap) จะทำให้รอยต่อของลักษณะสำคัญของเสียงพูด จากกรอบเสียงพูดหนึ่งไปยังอีกกรอบเสียงพูดหนึ่งเรียบ (Smooth) ขึ้น กรรมวิธีการวิเคราะห์สัญญาณเสียงพูดเบื้องต้นประกอบด้วย[3]

2.6.1 กรรมวิธีการปรับบรรทัดฐานแอมปลิจูด (Amplitude Normalization)

กรรมวิธีการปรับบรรทัดฐานแอมปลิจูดของสัญญาณเสียงพูดเป็นการเพิ่มหรือลดขนาดของสัญญาณเสียงพูด เพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน จึงจำเป็นต้องปรับให้มีขนาดของสัญญาณเสียงพูดอยู่ในบรรทัดฐานเดียวกัน เพื่อง่ายต่อการวัดคุณลักษณะและเปรียบเทียบสัญญาณเสียง การปรับบรรทัดฐานแอมปลิจูดแสดงดังสมการที่ 2.1

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

เมื่อ $\bar{S}[i]$ คือสัญญาณเสียงที่ปรับบรรทัดฐานแอมปลิจูดแล้ว

$S[i]$ คือสัญญาณเสียงพูดดิจิทัล

N คือ จำนวนบิต(Bit) ที่ใช้แทนค่าของสัญญาณเสียงพูด

2.6.2 กรรมวิธีการเน้นล่วงหน้า (Preamphasis)

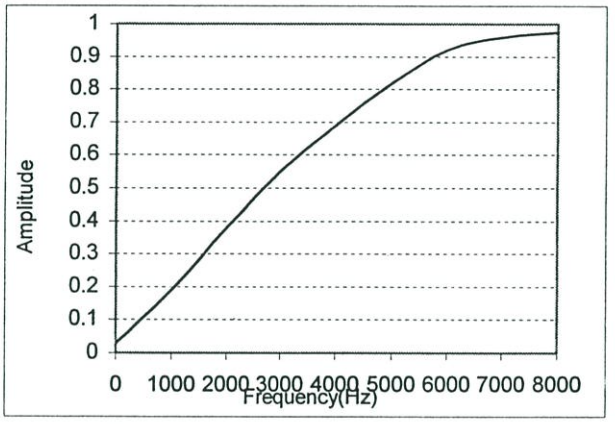
เนื่องจากสัญญาณเสียงพูดของมนุษย์ จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำเมื่อเทียบกับแถบความถี่ปฏิบัติงาน (Bandwidth) ไม่เกิน 5 กิโลเฮิรตซ์ ดังนั้น เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal to Noise Ratio : SNR) มีค่าค่อนข้างคงที่ตลอดช่วงความถี่ปฏิบัติงานนี้ เราจึงต้องมีการพรีเอมฟาซิส โดยเน้นความถี่สูงให้มีขนาดสูงขึ้นนั่นคือ การพรีเอมฟาซิสก็คือการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (high pass filter) ซึ่งมักนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชันถ่ายโอนเป็น

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

เมื่อเทียบกับรูปที่ 2.8 เราจะได้ว่า

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

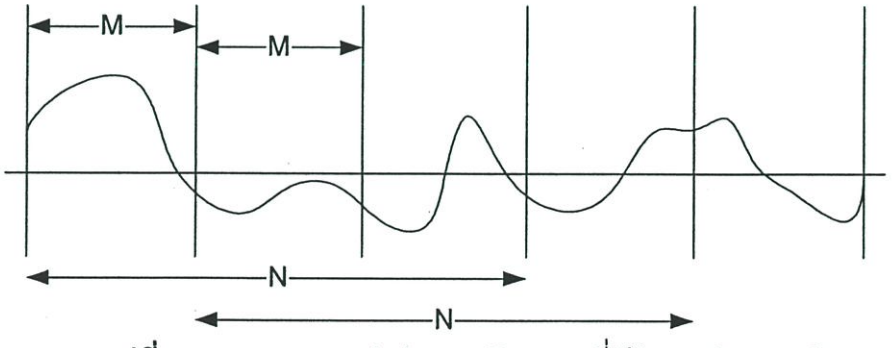
ยิ่งค่า a ใกล้ 1 เท่าใด ความถี่สูงก็จะยิ่งถูกขยายมากขึ้นเท่านั้น ค่า a ที่นิยมสำหรับใช้ในการหาพารามิเตอร์ของ LPC คือ 15/16= 0.9375 [3] เมื่อนำฟังก์ชันถ่ายโอนมาพล็อตกราฟของขนาดเทียบกับความถี่จะได้ดังรูปที่ 2.8



รูปที่ 2.8 แสดงขนาดสเปคตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส

2.6.3 **การแบ่งช่วงสัญญาณ** (Frame Blocking)

สัญญาณที่ผ่านการพรีเอมฟาซิสแล้ว $S[i]$ จะถูกตัดแบ่งออกเป็นช่วงๆ หรือ เฟรม ช่วงละ N ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ N ตัวอย่างสัญญาณ ดังรูปที่ 2.9



รูปที่ 2.9 แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์

โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ M ช่วงสัญญาณ จะเห็นได้ว่าถ้าค่า M โตกว่าค่า N ในการเลื่อนของช่วงในการวิเคราะห์จะทำให้บางสัญญาณไม่ถูกให้ในการวิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N ก็

จะทำให้ตัวอย่างสัญญาณ ทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่ก็จะทำให้การคำนวณช้าลง

2.6.4 **การวินโดว์** (Windowing)

หลักการกำหนดขนาดของวินโดว์

- วินโดว์จะต้องสั้นพอที่คุณสมบัติของเสียงที่เราสนใจจะวิเคราะห์นั้น ยังไม่เปลี่ยนแปลงในวินโดว์

- วินโดว์จะต้องยาวพอที่ จำนวนตัวอย่างสัญญาณในวินโดว์สามารถนำมาคำนวณหาคุณสมบัติที่ต้องการได้

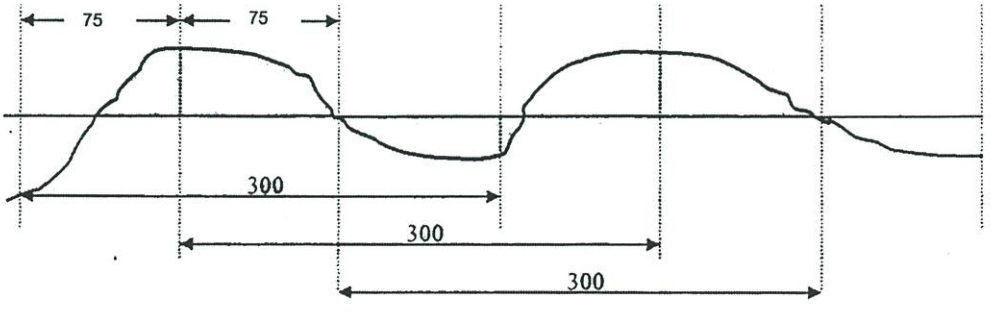
- วินโดว์ที่ติดกันไม่ควรจะสั้นจนกระโดดข้ามข้อมูลบางส่วนไป แต่ควรเลื่อนวินโดว์ให้น้อยกว่าขนาดของเฟรม

เนื่องจาก เราใช้ความถี่ในการสุ่มสัญญาณ 11.025 กิโลเฮิรตซ์ ในการวิเคราะห์นี้ เราเลือกใช้ค่า N=300 และ ค่า M=75 นั่นคือ ช่วงในการวิเคราะห์นี้ และระยะในการเลื่อนเฟรมประมาณ 6.803 มิลลิวินาที

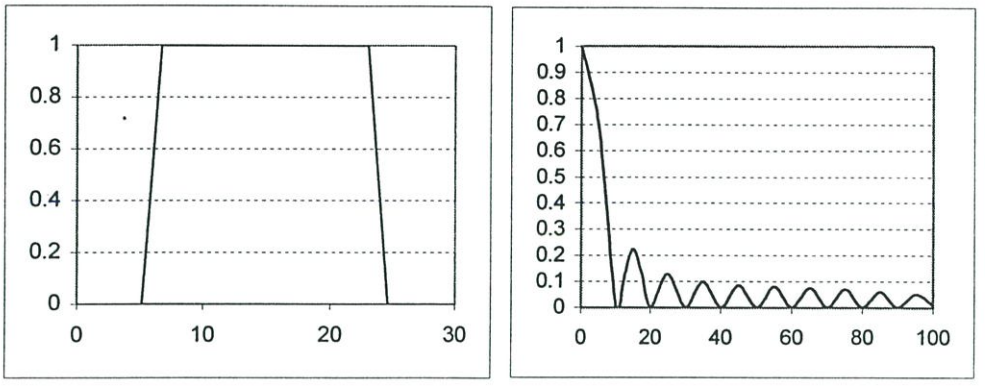
พิจารณาช่วงสัญญาณ N ตัวอย่างสัญญาณของช่วงใดๆ ที่ตัดมาวิเคราะห์ (รูปที่ 2.10) จะเห็นว่าที่ขอบของเฟรมที่ตัดมานี้มีความไม่ต่อเนื่องของสัญญาณ ถ้ามองในโดเมนความถี่ ก็จะมีความถี่สูงเกิดขึ้นดังนั้นเพื่อที่จะลดองค์ประกอบทางความถี่ที่สูงเหล่านี้ เราจะคูณด้วยฟังก์ชันวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปคตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก ในที่นี้จะให้ฟังก์ชันวินโดว์แฮมมิ่ง(Hamming window function) ซึ่งนิยามโดยสมการดังนี้คือ

$$w(n) = 0.54 - 0.46\cos(2\pi n/(N-1)) \quad \text{เมื่อ } n = 1,2,\ldots,N-1 \qquad (2.4)$$

รูปที่ 2.11 และ รูปที่ 2.12 แสดงองค์ประกอบทางเวลาและทางความถี่ของฟังก์ชันวินโดว์แบบสี่เหลี่ยม และแฮมมิ่ง ตามลำดับ จะเห็นว่าสเปคตรัมของวินโดว์แฮมมิ่งมีริพเพิล (ripple) น้อยกว่าของวินโดว์สี่เหลี่ยม

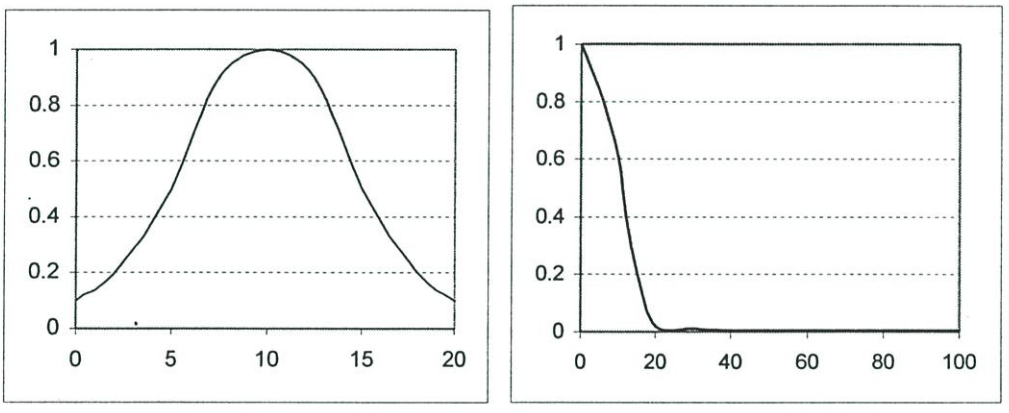


รูปที่ 2.10 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์



(a) (b)

รูปที่ 2.11 แสดงวินโดว์แบบสี่เหลี่ยม(ก) ในโดเมนเวลา (ข) ในโดเมนความถี่



(ก) (ข)

รูปที่ 2.12 แสดงวินโดว์แบบแฮมมิ่ง (a) ในโดเมนเวลา(b) ในโดเมนความถี่

ดังนั้นเราจะได้ว่า

$$x'(n) = x(n) \cdot w(n) \tag{2.5}$$

2.7 พลังงานของสัญญาณเสียงพูด

พลังงานของสัญญาณเสียงพูด ถูกนำมาใช้ในการวิเคราะห์เสียงพูด และเป็นคุณลักษณะที่นิยมนำมาใช้กันอย่างแพร่หลาย เนื่องจากเป็นวิธีที่คำนวณง่าย และรวดเร็ว พลังงานของสัญญาณเสียงพูดเป็นคุณสมบัติที่แสดงให้เห็นว่ามีสัญญาณเสียง (รวมทั้งสัญญาณรบกวน) เกิดขึ้น ณ เวลานั้นหรือไม่ การคำนวณหาค่าพลังงาน จะทำทีละกรอบเสียงพูด โดย E_n คือ สัญญาณเสียงพูดที่ i ในกรอบเสียงพูด n และในแต่ละกรอบเสียงพูดจะมีสัญญาณเสียงพูดจำนวน K ซึ่งวิธีการคำนวณหาค่าพลังงานของสัญญาณเสียงพูดมีดังนี้คือ

1) พลังงานสัมบูรณ์(Absolute Energy)

เป็นการหาผลรวมของสัญญาณเสียงพูดสัมบูรณ์ในแต่ละกรอบเสียงพูด ดังสมการที่ 2.6

$$En = \sum_{i=1}^{w} [S_n[i]] \tag{2.6}$$

2) พลังงานเฉลี่ย (Root Mean Square Energy)

เป็นการหาพลังงานของสัญญาณเสียงพูดจากรากที่สองของผลรวมกำลังสองเฉลี่ย ดังสมการที่ 2.7

$$En = \left[\frac{1}{W} \sum_{i=1}^{w} S_n^2[i] \right]^{1/2} \tag{2.7}$$

3) พลังงานกำลังสอง (Square Energy)

เป็นการวัดค่าพลังงานจากสัญญาณเสียงพูดยกกำลังสอง ทำให้ค่าพลังงานมีความไวต่อสัญญาณที่มีขนาดใหญ่ การหาค่าพลังงานกำลังสองแสดงดังสมการที่ 2.8

$$En = \left[\sum_{i=1}^{w} S_n^2[i] \right] \tag{2.8}$$

4) พลังงานของ Teager (Teager Energy)

เมื่อพูดถึงพลังงานของสัญญาณ มักจะหมายถึงค่าผลรวมเฉลี่ยกำลังสองของขนาดสัญญาณ เมื่ออยู่ในเชิงเวลา หรือ ทำการแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform ,DFT) แล้วยกกำลังขนาดของความถี่ของสัญญาณเสียงพูด ซึ่งเป็นการหาค่าพลังงานในเชิงความถี่ ตัวอย่างของสัญญาณความถี่ 10Hzและสัญญาณความถี่ 1000Hz ไม่ว่าจะหาค่าพลังงานในเชิงเวลา หรือ ในเชิงความถี่ พลังงานของสัญญาณ 1000 Hz จะมีค่ามากกว่าสัญญาณความถี่ 10 Hz เสมอ พลังงานของสัญญาณรายคาบ (Sinusoidal) เกิดจากการคูณกันของขนาดกำลังสองกับความถี่กำลังสอง สัญญาณเสียงมีลักษณะเป็นรายคาบในช่วงเวลาสั้นๆ ซึ่งได้กำหนดขึ้นเป็นกรอบเสียงพูด โดยมีสมการการเคลื่อนที่ของสัญญาณเสียงดังสมการที่ 2.9

$$S_i = A\cos(\Omega i + \phi) \tag{2.9}$$

โดยที่ S_i คือตัวอย่างของสัญญาณเสียงพูด

A คือขนาดของสัญญาณเสียงพูด

Ω คือ ความถี่ดิจิทัล $\Omega = 2\pi f / f_s$ ซึ่ง f คือความถี่แอนาลอก (Analog Frequency) และ f_s คืออัตราการสุ่มตัวอย่าง (Sampling Rate)

ϕ คือ เฟสเริ่มต้น

ดังที่ได้กล่าวแล้วว่าพลังงานของสัญญาณรายคาบ จะเท่ากับขนาดของสัญญาณกำลังสองคูณกับความถี่กำลังสอง ในสมการที่ 10 แสดงให้เห็นว่าพลังงานของสัญญาณสามารถหาได้จากสัญญาณตัวเอง ลบด้วยผลคูณของสัญญาณข้างหน้ากับสัญญาณข้างหลัง

$$S^2[i] - S[i+1]S[i-1] = A^2\Omega^2 \qquad (2.10)$$

ค่าพลังงานที่ได้เป็นเพียงค่าพลังของสัญญาณที่เวลาใดๆเราเรียกว่าพลังงานชั่วขณะ (Instantaneous Energy , *Ei*)และพลังงานของแต่ละกรอบเสียงพูดสามารถหาได้จากสมการที่ 2.10

2.8 การประมาณค่าในช่วงด้วยเส้นโค้ง

การทดลองส่วนใหญ่ในทางวิศวกรรมและวิทยาศาสตร์ มักจะให้ผลลัพธ์ในลักษณะของข้อมูลตามตำแหน่งต่างๆ ซึ่งเป็นข้อมูลเพียงตามบางจุดบางตำแหน่งเท่านั้น มีบ่อยครั้งที่ต้องการทราบข้อมูลเหล่านี้ที่ตำแหน่งอื่นๆ ซึ่งต่างไปจากตำแหน่งที่ทดลองมาได้ ซึ่งความเข้าใจในการประมาณค่าในช่วง (interpolation) และ การประมาณค่านอกช่วง (extrapolation) จะมีความสำคัญที่สามารถนำไปสู่การคำนวณเพื่อหาค่าที่เที่ยงตรงที่อยู่ตามตำแหน่งต่างๆ นอกเหนือจากตำแหน่งที่ได้จากการทดลอง

การประมาณค่าในช่วงที่สำคัญ ๆ มี 3 วิธีการคือ การประมาณค่าในช่วงโดย

1 การใช้ผลต่างจากการแบ่งย่อยของนิวตัน (Newton's divided differences)

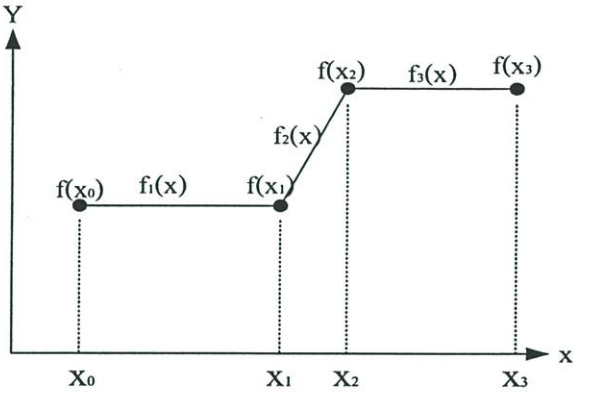
2 การใช้ฟังก์ชันพหุนามของลากรองจ์ (Lagrange polynomials)

3 การประมาณค่าในช่วงโดยเส้นโค้ง (spline interpolation)

วิธีการประมาณค่าในช่วง โดยการใช้ผลต่างจากการแบ่งย่อยของนิวตันและโดยการใช้ฟังก์ชันพหุนามของลากรองจ์ ก่อให้เกิดฟังก์ชันพหุนามที่ลากผ่านทุกข้อมูลที่กำหนดมาให้ ซึ่งจะมีการกระจายที่เกิดการเบี่ยงเบนไปจากลักษณะของการกระจายที่ควรจะเป็น ซึ่งการประมาณค่าในช่วงด้วยเส้นโค้งจะเป็นวิธีการปรับปรุงที่สร้างขึ้นเพื่อแก้ปัญหาดังกล่าว โดยการประดิษฐ์ฟังก์ชันเส้นโค้ง (spline function) ที่เราจะใช้เชื่อมต่อระหว่างข้อมูลทีละ 2 ข้อมูล ผลลัพธ์ที่เกิดขึ้นจากการใช้ฟังก์ชันเส้นโค้งต่างๆ เหล่านี้มาเชื่อมต่อกัน จะก่อให้เกิดลักษณะของการกระจายในภาพรวมที่ใกล้ความจริงมากขึ้น

2.8.1 การประมาณค่าในช่วงเชิงเส้น

การสร้างฟังก์ชันโดยเชื่อมต่อข้อมูลด้วยการใช้เส้นตรง (linear spline) จัดว่าเป็นวิธีที่ทำได้สะดวกที่สุด รูปที่ 2.13 แสดงการเชื่อมต่อข้อมูล 4 ข้อมูลด้วยฟังก์ชันเส้นตรง 3 ฟังก์ชัน



รูปที่ 2.13 การประมาณค่าในช่วงโดยการใช้ฟังก์ชันเส้นตรง

$$f_1(x) = f(x_0) + m_1(x - x_0) \qquad x_0 \le x \le x_1 \qquad (2.11)$$

$$f_2(x) = f(x_1) + m_2(x - x_1) \qquad x_1 \le x \le x_2 \qquad (2.12)$$

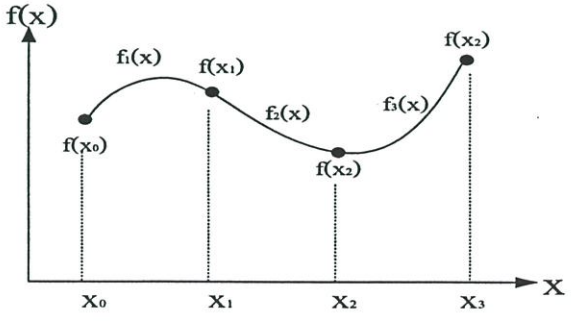
$$f_3(x) = f(x_2) + m_3(x - x_2) \qquad x_2 \le x \le x_3 \qquad (2.13)$$

โดย

$$m_i = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \qquad i = 1,2,3 \qquad (2.14)$$

2.8.2 การประมาณค่าในช่วงกำลังสอง

การสร้างฟังก์ชันเพื่อเชื่อมต่อข้อมูลด้วยการใช้เส้นโค้งกำลังสอง(quadratic spline) สามารถทำได้โดยไม่ยากนัก รูปที่ 2.14 แสดงข้อมูล 4 ข้อมูลที่ถูกเชื่อมต่อด้วยฟังก์ชันกำลังสอง 3 ฟังก์ชัน



รูปที่ 2.14 การประมาณค่าในช่วงโดยใช้ฟังก์ชันกำลังสอง

ฟังก์ชันกำลังสองที่จำเป็นต้องหาซึ่งเชื่อมต่อข้อมูลเหล่านี้ คือ

$$f_1(x) = a_1x^2 + b_1x + c_1 \qquad x_0 \le x \le x_1 \qquad (2.15)$$

$$f_2(x) = a_2x^2 + b_2x + c_1 \qquad x_0 \le x \le x_1 \qquad (2.16)$$

$$f_3(x) = a_3x^2 + b_3x + c_3 \qquad x_0 \le x \le x_1 \qquad (2.17)$$

โดย $a_i, b_i, c_i + c3, i = 1,2,3$ เป็นค่าคงตัวที่ไม่รู้ค่า ซึ่งในที่นี้มีทั้งหมด 9 ค่า และสามารถคำนวณหาได้จากเงื่อนไขดังต่อไปนี้

(a) ที่จุดต่อภายในใดๆ ฟังก์ชันทั้ง 2 ด้านที่เชื่อมเข้าหากันที่จุดต่อภายในจุดใดจุดหนึ่งจำเป็นต้องมีค่าเท่ากันที่ตำแหน่งของจุดต่อนั้น เช่น ที่จุดต่อ x_1 เงื่อนไขที่เกิดขึ้นนี้คือ

$$f_1(x_1) = a_1x_1^2 + b_1x_1 + c_1 = f(x_1) \qquad (2.18)$$

และ $$f_2(x_1) = a_2x_1^2 + b_2x_1 + c_2 = f(x_1) \qquad (2.19)$$

ในกรณีของรูปที่ 2.14 ที่มีจุดต่อภายใน 2 จุด ดังนั้นจึงเกิดเงื่อนไขเช่นนี้ 4 เงื่อนไข

(b) ฟังก์ชันแรกต้องผ่านข้อมูลแรกที่ x_0 นั่นคือ

$$f_1(x_0) = a_1x_0^2 + b_1x_0 + c_1 = f(x_0) \qquad (2.20)$$

และฟังก์ชันท้ายต้องผ่านข้อมูลท้ายที่ x_3 นั่นคือ

$$f_3(x_3) = a_3x_3^2 + b_3x_3 + c_3 = f(x_3) \qquad (2.21)$$

นั่นคือ ก่อให้เกิดเงื่อนไขเพิ่มขึ้นอีก 2 เงื่อนไข

(c) ที่จุดต่อภายในใดๆ ค่าความชันของฟังก์ชันทั้ง 2 ด้านของจุดต่อต้องมีค่าเท่ากันที่ตำแหน่งของจุดต่อนั้น เช่นที่จุดต่อ x_1 เงื่อนไขที่เกิดขึ้น คือ

$$f_1'(x_1) = f_2'(x_1) \qquad (2.22)$$

ดังเช่นในกรณีของรูป 2.14 จะมีเงื่อนไขเช่นนี้อยู่ 2 เงื่อนไข

ดังนั้นจำนวนเงื่อนไขที่เกิดขึ้นทั้งหมดในกรณีของรูปที่ 2.14 คือ 4+2+2= 8 เงื่อนไข แต่เนื่องจากเรามีค่าคงตัวที่ไม่รู้ค่าทั้งหมด 9 ค่า ซึ่งเกินกว่าเงื่อนไขที่มีอยู่ 1 ค่า ดังนั้นเพื่อให้แก้ปัญหานี้ได้ เราอาจกำหนดค่าสัมประสิทธิ์ a_1 ให้มีค่าเท่ากับศูนย์ ซึ่งมีความหมายเช่นเดียวกันกับการกำหนดให้ฟังก์ชัน $f_1(x)$ แรกนั้นมีลักษณะการกระจายแบบเชิงเส้นตรงในรูปแบบเช่นเดียวกันกับดังที่อธิบายในหัวข้อ 2.8.1 นั่นคือ ในขั้นตอนท้ายสุด ตัวไม่รู้ค่าที่เหลืออยู่ 8 ค่านี้ สามารถคำนวณได้จากระบบสมการที่ประกอบด้วย 8 สมการย่อยจากเงื่อนไขทั้งหมดดังกล่าว ในรูปแบบดังนี้

$$\begin{bmatrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix}_{(8\times 8)} \begin{bmatrix} b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} & \\ & \\ & \\ & \\ & \\ & \\ & \\ & \end{bmatrix}_{8\times 8} \tag{2.23}$$

สำหรับปัญหาโดยทั่วไปที่ประกอบด้วยข้อมูลทั้งหมด $n+1$ ข้อมูล เราจำเป็นต้องประดิษฐ์ฟังก์ชันนั่นคือ $f_1(x), f_2(x), \ldots, f_n(x)$ และเนื่องจากแต่ละฟังก์ชันกำลังสองต่างประกอบด้วยค่าคงตัวทั้งหมด $3n$ ค่า ซึ่งจะให้เงื่อนไขรวมทั้งสิ้น $2(n-1)+2+(n+1)=3n-1$ เงื่อนไข ในทำนองเดียวกันหากเรากำหนดให้ฟังก์ชันแรกมีการกระจายแบบเชิงเส้นกล่าวคือ $a_1 = 0$ ดังนั้นปัญหาโดยทั่วไปนี้จึงประกอบด้วยค่าคงตัวที่ไม่รู้ค่า $3n-1$ ค่า ซึ่งสามารถคำนวณหาได้จากเงื่อนไขทั้งหมด $3n-1$ เงื่อนไขที่มีอยู่ดังกล่าว

2.9 ระเบียบวิธีการถดถอยแบบกำลังสองน้อยสุด (Least squares regression)

เป็นกระบวนการการกลั่นกรองข้อมูล อย่างเป็นขั้นเป็นตอนที่จะก่อให้เกิดผลลัพธ์ที่สามารถใช้เป็นมาตรฐานสำหรับข้อมูลที่กำหนดมาให้ชุดนั้น ซึ่งระเบียบวิธีการต่าง ๆ ประกอบด้วย

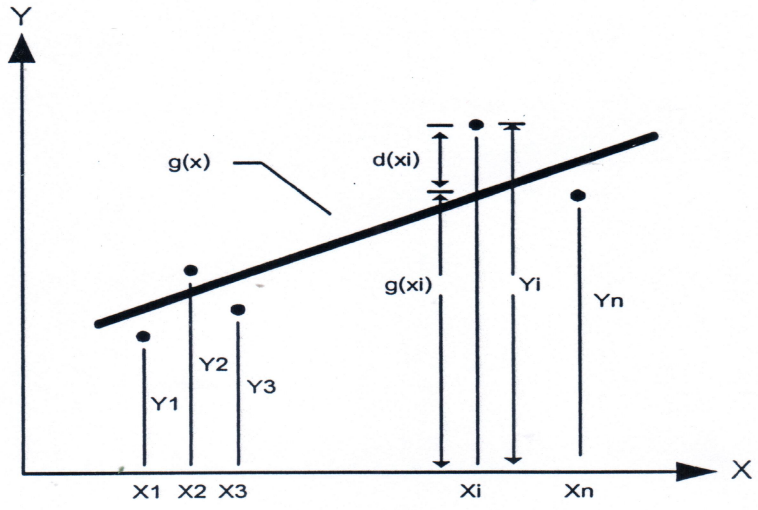
1. การถดถอยเชิงเส้น (linear regression)
2. การประยุกต์การถดถอยแบบเชิงเส้นกับข้อมูลไม่เชิงเส้น
3. การถดถอยแบบพหุนาม (polynomial regression)

2.9.1 การถดถอยแบบเชิงเส้น (linear regression)

การถดถอยแบบเชิงเส้น(linear regression) อาจจัดได้ว่าเป็นระเบียบวิธีที่ง่ายที่สุดที่ใช้ในการประดิษฐ์ฟังก์ชันเส้นตรงสำหรับชุดของข้อมูลที่กำหนดมาให้ รูปที่ 2.14 แสดงชุดของข้อมูลที่ประกอบด้วย $x_i, y_i, i = 1,2,\ldots,n$ นั่นคือมีจำนวนข้อมูลทั้งสิ้น n ข้อมูล ในที่นี้เราจะประดิษฐ์สมการเส้นตรงในรูปแบบของฟังก์ชัน

$$g(x) = a_0 + a_1 x \tag{2.24}$$

โดย a_0 และ a_1 เป็นค่าคงตัวที่ไม่รู้ค่าและจำเป็นต้องคำนวณหาจากเงื่อนไขที่ว่า สมการเส้นตรงที่ประดิษฐ์ขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุดจากข้อมูลทุกข้อมูลที่กำหนดมาให้



รูปที่ 2.15 การถดถอยแบบเชิงเส้นโดยการประดิษฐ์ฟังก์ชันเส้นตรงจากชุดของข้อมูลที่กำหนดมาให้

จากรูป 2.15นี้ เราจะเห็นได้ว่า ณ ตำแหน่ง x_i ของข้อมูล i ใดๆ ค่าของฟังก์ชัน $g(x)$ ที่เราจะประดิษฐ์ขึ้นจะมีค่าที่แตกต่างไปจากค่าของข้อมูล y_i เท่ากับ $d(x_i)$ ที่ตำแหน่งนั้น นั่นหมายถึงว่า ค่าความผิดพลาด E ทั้งหมดที่เกิดขึ้นจากข้อมูลทั้งหมด n ข้อมูล อาจเขียนให้อยู่ในรูปแบบ ดังนี้

$$E = \sum_{i=1}^{n} [d(x_i)]^2 \tag{2.25}$$

ซึ่งในที่นี้เราทำการยกกำลังสองของค่าแตกต่าง $d(x_i)$ ก็เพื่อกำจัดค่าที่อาจมีเครื่องหมายเป็นลบ ดังนั้นสมการ 2.25 จะให้ความหมายของค่าความผิดพลาดทั้งหมด สมการ 2.25 สามารถเขียนได้ว่า

$$E = \sum_{i=1}^{n} [y_i - g(x_i)]^2 \tag{2.26}$$

แทนสมการ 2.24 ที่ $x = x_i$ ลงในสมการ 2.26 นี้จะได้

$$E = \sum_{i=1}^{n} [y_i - (a_0 - a_1 x_i)]^2 \tag{2.27}$$

จากสมการ 2.27 นี้ เราสามารถคำนวณหาตัวไม่รู้ค่า a_0 และ a_1 ที่ต้องการได้โดยวิธีกำลังสองน้อยที่สุด (least – squares) ซึ่งทำจากวิธีการหาค่าต่ำ(minimization) ของค่าความผิดพลาดโดยเกี่ยวข้องกับตัวไม่รู้ค่า นั่นคือ

$$\frac{\partial E}{\partial a_0} = 0 \tag{2.28}$$

และ $$\frac{\partial E}{\partial a_1} = 0 \qquad (2.29)$$

และเงื่อนไขในสมการ 2.28 ให้ผลดังนี้

$$2\sum_{i=1}^{n}\left[y_i - (a_0 + a_1 x_i)\right](-1) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$$

$$na_0 + (\sum_{i=1}^{n} x_i)a_1 = \sum_{i=1}^{n} y_i \qquad (2.30)$$

และเงื่อนไขในสมการ 2.29 ให้ผลดังนี้

$$2\sum_{i=1}^{n}\left[y_i - (a_0 + a_1 x_i)\right](-x_i) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} a_0 x_i - \sum_{i=1}^{n} a_1 x_i^2 = 0$$

$$(\sum_{i=1}^{n} x_i)a_0 + (\sum_{i=1}^{n} x_i^2)a_1 = \sum_{i=1}^{n} x_i y_i \qquad (2.31)$$

ทั้งสองสมการ 2.30 และ 2.31 นี้ สามารถเขียนให้อยู่ในรูปแบบของเมตริกซ์ได้ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{Bmatrix} \qquad (2.32)$$

ซึ่งเราสามารถใช้กฎของคราเมอร์ ในการแก้ระบบสมการนี้ เพื่อหาค่าคงตัว a_0 และ a_1 ได้ดังนี้

$$a_0 = \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i y_i)(\sum_{i=1}^{n} x_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i^2)} \qquad (2.33)$$

$$a_1 = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i^2)} \qquad (2.34)$$

ค่าคงตัว a_0 และ a_1 ที่คำนวณได้นี้ เมื่อแทนกลับลงในสมการ 2.24 ก็จะได้สมการเส้นตรงที่แสดงการถดถอยแบบเชิงเส้นที่ต้องการ

2.9.2 การประยุกต์การถดถอยแบบเชิงเส้นกับข้อมูลไม่เป็นเชิงเส้น

สำหรับปัญหาในทางปฏิบัติส่วนใหญ่ ลักษณะการกระจายของข้อมูลมักไม่อยู่ในรูปแบบของเชิงเส้น การกระจายของข้อมูลมักจะเบี่ยงเบนไปมาในลักษณะไม่เชิงเส้น(nonlinear) ซึ่งควรใช้สมการในรูปแบบของฟังก์ชันพหุนาม (polynomial function) จึงจะให้ผลลัพธ์ที่เที่ยงตรงมากขึ้น เราจะศึกษาการใช้สมการพหุนามในหัวข้อต่อไป ซึ่งจะนำไปอยู่ระบบสมการที่ประกอบด้วยหลายสมการย่อย แต่ก่อนจะถึงหัวข้อการใช้สมการพหุนามนั้นเราจะมาศึกษาการประยุกต์การถดถอยแบบเชิงเส้นเข้ากับข้อมูลที่อยู่ในรูปแบบไม่เชิงเส้นก่อน ประโยชน์หลักของการประยุกต์นี้ก็คือระบบสมการที่เกิดขึ้นจากวิธีการนี้สามารถแก้ไขได้โดยง่ายในหัวข้อที่แล้ว มีปัญหาทางวิศวกรรมศาสตร์จำนวนมากที่ลักษณะการกระจายของข้อมูลนั้นอยู่ในรูปแบบของสมการกำลัง (power equation) ซึ่งมีลักษณะการกระจายในรูปแบบของสมการกำลัง ดังนี้

$$\bar{y} = a\bar{x}^{\,b} \tag{2.35}$$

ซึ่งมีลักษณะการกระจายแบบไม่เชิงเส้นดังแสดงในรูป 2.16 สมการลอการิทึมที่สืบเนื่องมาจากสมการ 2.35 นี้คือ

$$\log \bar{y} = \log a + b \log \bar{x} \tag{2.36}$$

ซึ่งอาจสามารถเขียนให้อยู่ในรูปแบบดังนี้

$$y = a_0 + a_1 x \tag{2.37}$$

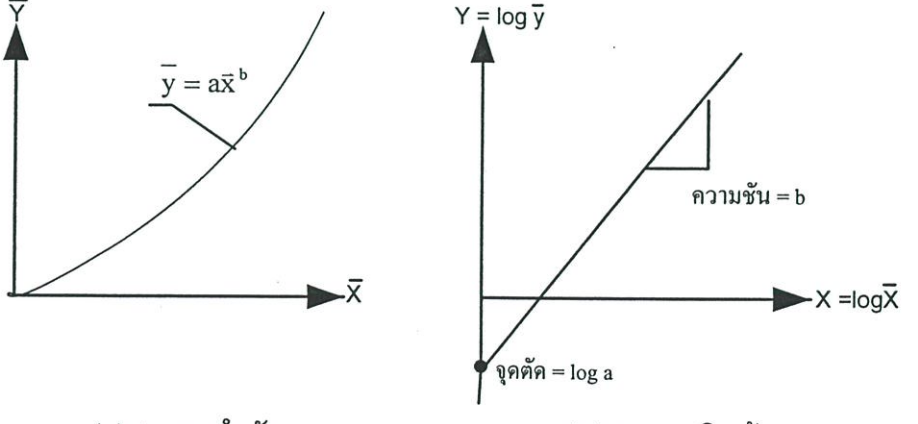
นั่นคือ อยู่ในรูปแบบของสมการเชิงเส้นดังเช่นสมการ 2.24 และแสดงในรูป 2.16 b โดย

$$x = \log \bar{x} \tag{2.38}$$

$$y = \log \bar{y} \tag{2.39}$$

$$a_0 = \log a \tag{2.40}$$

$$a_1 = b \tag{2.41}$$



(a) สมการกำลัง (b) สมการเชิงเส้น

รูปที่ 2.16 การแปลงสมการกำลังไปสู่สมการเชิงเส้นโดยใช้ลอการิทึมเพื่อใช้วิธีถดถอยแบบเชิงเส้น

เมื่อเราเปลี่ยนข้อมูลที่มีการกระจายจากรูปแบบของสมการกำลัง 2.35 ไปสู่รูปแบบเชิงเส้นในสมการ 2.37 แล้ว เราก็สามารถใช้วิธีถดถอยแบบเชิงเส้นคำนวณหาค่าตัวคงที่ a_0 และ a_1 ได้ และจากนั้นเราจึงสามารถแปลงกลับโดยใช้สมการ2.38 - 2.41 เพื่อหาค่าสัมประสิทธิ์ a และ b สำหรับสมการกำลัง 2.35 ที่ต้องการได้

นอกจากนี้ ยังมีรูปแบบของสมการไม่เชิงเส้นในลักษณะอื่นๆ ซึ่งเราสามารถประยุกต์การถดถอยแบบเชิงเส้นมาใช้ได้เช่นกัน รูปแบบของสมการอีกชนิดหนึ่งที่มีลักษณะของการกระจายในทำนองเดียวกันกับสมการกำลังก็คือ รูปแบบของสมการเอกซ์โพเนนเชียล(exponential model) ดังนี้

$$\bar{y} = ae^{b\bar{x}} \tag{2.42}$$

เราสามารถแปลงสมการนี้ให้ไปอยู่ในรูปแบบของสมการเชิงเส้นโดยการประยุกต์ลอการิทึม ดังนี้

$$\ln \bar{y} = \ln a + b\bar{x} \ln e$$

แต่เนื่องจาก $\ln e = 1$ ดังนั้น

$$\ln \bar{y} = \ln a + b\bar{x} \tag{2.43}$$

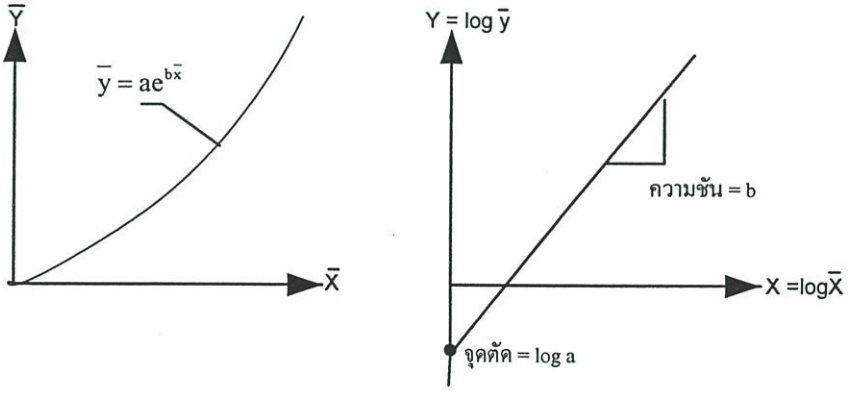
ซึ่งผลที่ได้อยู่ในรูปแบบของสมการเชิงเส้น คือ

$$y = a_0 + a_1 x \tag{2.44}$$

โดย

$$y = \ln \bar{y} \quad ; \quad x = \bar{x}$$
$$a_0 = \ln a \quad ; \quad a_1 = b \tag{2.45}$$

ลักษณะการแปลงสมการเอกซ์โพเนนเชียล (2.42) ให้ไปอยู่ในรูปแบบของสมการเชิงเส้น (2.44) สามารถทำความเข้าใจได้ดียิ่งขึ้นโดยพิจารณารูป 2.17



(a) สมการเอ็กโพแนนเชียล (b) สมการเชิงเส้น

รูปที่ 2.17 การแปลงสมการเอกซ์โพเนนเชียลไปสู่สมการเชิงเส้นโดยใช้ลอการิทึมเพื่อใช้วิธีถดถอยแบบเชิงเส้น

จากนั้นจึงทำการคำนวณหาค่าสัมประสิทธิ์ a_0 และ a_1 สำหรับชุดข้อมูลที่กำหนดมาให้และเมื่อได้ค่าสัมประสิทธิ์เหล่านี้แล้วจึงนำไปหาค่าสัมประสิทธิ์ a และ b จากสมการ 2.45 เพื่อก่อให้เกิดสมการเอกซ์โพเนนเชียล 2.42 ที่ต้องการ

สมการไม่เชิงเส้นอีกชนิดหนึ่งที่มีลักษณะการกระจายซึ่งพบเห็นจากการทดลองอยู่เสมอเป็นสมการอัตราการเพิ่มสู่จุดอิ่มตัว (saturation-growth-rate equation) ซึ่งอยู่ในรูปแบบดังนี้

$$\bar{y} = a\frac{\bar{x}}{b+\bar{x}} \tag{2.46}$$

หรือ

$$\frac{b+\bar{x}}{a\bar{x}} = \frac{1}{\bar{y}}$$

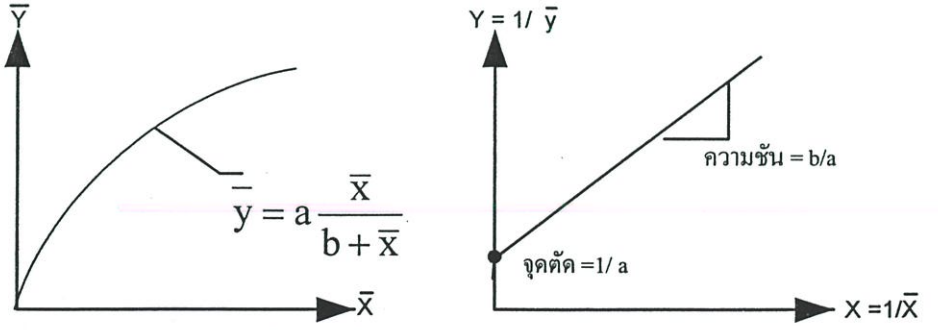
$$\frac{1}{\bar{y}} = \frac{1}{a} + \frac{b}{a}\frac{1}{\bar{x}} \tag{2.47}$$

ซึ่งผลที่ได้อยู่ในรูปแบบของสมการเชิงเส้น คือ

$$y = a_0 + a_1 x \tag{2.48}$$

โดย

$$y = 1/\bar{y} \quad ; \quad x = 1/\bar{x}$$
$$a_0 = 1/a \quad ; \quad a_1 = b/a \tag{2.49}$$



(a) สมการอัตราการเพิ่มสู่จุดอิ่มตัว (b) สมการเชิงเส้น

รูปที่ 2.18 การแปลงสมการอัตราการเพิ่มสู่จุดอิ่มตัวไปสู่สมการเชิงเส้นเพื่อใช้วิธีถดถอยแบบเชิงเส้น

จากนั้นจึงสามารถใช้วิธีถดถอยแบบเชิงเส้นเพื่อหาค่าสัมประสิทธิ์ a_0 และ a_1 จากชุดของข้อมูลที่กำหนดมาให้ แล้วจึงใช้สมการ 2.49 เพื่อคำนวณสัมประสิทธิ์ a และ b สำหรับสมการอัตราการเพิ่มจุดอิ่มตัว 2.46 ซึ่งอยู่ในรูปแบบไม่เชิงเส้นดังกล่าว

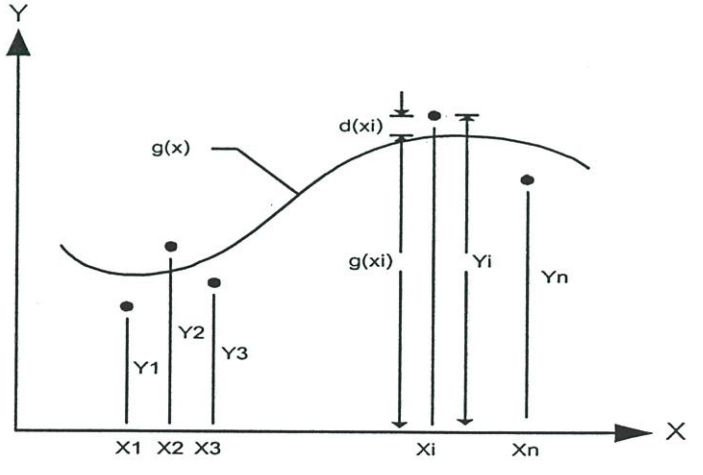
จากสมการไม่เชิงเส้นที่อยู่ในรูปแบบจำเพาะ ไม่ว่าจะเป็นสมการกำลัง 2.35 สมการเอกซ์โพเนนเชียล 2.42 หรือสมการอัตราการเพิ่มสู่จุดอิ่มตัว 2.46 เราจะเห็นได้ว่าหากเรามีความเข้าใจในลักษณะรูปแบบของการกระจายของสมการเหล่านี้โดยแท้จริง และหากเราพบข้อมูลที่มีการกระจายในลักษณะรูปแบบของการกระจายในลักษณะรูปแบบไม่เชิงเส้นเหล่านี้ได้โดยตรง ผลประโยชน์ที่ได้รับก็คือ ขั้นตอนของการคำนวณนั้นค่อนข้างง่าย

2.9.3 การถดถอยแบบพหุนาม

การถดถอยแบบพหุนาม(polynomial regression) เป็นระเบียบวิธี ที่ใช้ประดิษฐ์ฟังก์ชันพหุนามสำหรับข้อมูลที่มีการกระจายโดยทั่วไปที่ไม่อยู่ในรูปแบบของเชิงเส้นหรือสมการกำลัง กระบวนการ ในการประดิษฐ์ฟังก์ชันพหุนามโดยระเบียบวิธีการถดถอยแบบพหุนาม นั้นประกอบด้วยขั้นตอนเช่นเดียวกันกับที่ในระเบียบวิธีการถดถอยเชิงเส้น รูปที่ 2.19 แสดงการใช้ฟังก์ชันพหุนามในรูปแบบทั่วไปกับชุดของข้อมูลที่ไม่ได้อยู่ในรูปแบบเชิงเส้นชุดหนึ่งที่กำหนดมาให้

ชุดของข้อมูลในรูป 2.19 นี้ ประกอบด้วย $x_i, y_i, i = 1,2,\ldots,n$ กล่าวคือมีจำนวนข้อมูลทั้งสิ้น n ข้อมูล ในที่นี้เราจะประดิษฐ์ฟังก์ชันพหุนามอันดับ m สำหรับข้อมูลชุดนี้

$$g(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m \qquad (2.50)$$



รูปที่ 2.19 การถดถอยแบบพหุนามโดยการประดิษฐ์ฟังก์ชันพหุนามจากชุดของข้อมูล

โดย $a_0, a_1, a_2, \ldots, a_m$ เป็นค่าคงตัวที่ไม่รู้ค่าซึ่งจะคำนวณหาจากเงื่อนไขที่ว่า สมการพหุนามที่จะประดิษฐ์ขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุดจากข้อมูลทั้งหมดที่

กำหนดมาให้ ขั้นตอนในการประดิษฐ์สมการพหุนามนี้ เริ่มจากการหาค่าความผิดพลาด E ทั้งหมดที่เกิดขึ้นจาก n ข้อมูลในรูปแบบดังนี้

$$E = \sum_{i=1}^{n} \left[d(x_i) \right]^2 \tag{2.51}$$

ซึ่งสามารถเขียนให้ประกอบด้วยฟังก์ชันพหุนามดังแสดงในสมการ 2.50 ได้คือ

$$E = \sum_{i=1}^{n} \left[y_i - g(x_i) \right]^2 \tag{2.52}$$

$$E = \sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m) \right]^2 \tag{2.53}$$

ในการหาตัวไม่รู้ค่า $a_0, a_1, a_2, \ldots, a_m$ รวมทั้งสิ้น $m+1$ ค่านั้น เราจะใช้วิธีกำลังสองน้อยสุด (least-squares) ซึ่งทำจากการหาค่าต่ำที่สุด(minimization) ของค่าความผิดพลาด E โดยเกี่ยวข้องกับตัวไม่รู้ค่า ก่อให้เกิดระบบสมการที่ประกอบด้วย $m+1$ สมการย่อย นั่นคือ

$$\left.\begin{aligned} \frac{\partial E}{\partial a_0} &= 0 \\ \frac{\partial E}{\partial a_1} &= 0 \\ \frac{\partial E}{\partial a_2} &= 0 \\ \vdots &= \vdots \\ \frac{\partial E}{\partial a_m} &= 0 \end{aligned}\right\} m+1 \text{ สมการ} \tag{2.54}$$

ดังตัวอย่าง เช่น สมการแรกในระบบสมการนี้สามารถประดิษฐ์ขึ้นได้ดังนี้

$$2\sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_m x_i^m \right](-1) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} x_i a_1 - \sum_{i=1}^{n} x_i^2 a_2 - \ldots - \sum_{i=1}^{n} x^m{}_i a_m = 0$$

$$n a_0 + (\sum_{i=1}^{n} x_i) a_1 + (\sum_{i=1}^{n} x_i^2) a_2 + \ldots + (\sum_{i=1}^{n} x_i^m) a_m = \sum_{i=1}^{n} y_i$$

และเช่นเดียวกันกับสมการที่สอง ซึ่งคือ

$$2\sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_m x_i^m \right](-x) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i a_0 - \sum_{i=1}^{n} x_i^2 a_1 - \sum_{i=1}^{n} x_i^3 a_2 - \ldots - \sum_{i=1}^{n} x^{m+1}{}_i a_m = 0$$

$$(\sum_{i=1}^{n} x_i) a_0 + (\sum_{i=1}^{n} x_i^2) a_1 + (\sum_{i=1}^{n} x_i^3) a_2 + \ldots + (\sum_{i=1}^{n} x^{m+1}{}_i) a_m = \sum_{i=1}^{n} x_i y_i$$

และสมการอื่นๆที่เหลือก็สามารถทำได้ในทำนองเดียวกัน สมการทั้งหมดที่ประดิษฐ์ขึ้นมาได้นี้สามารถเขียนให้อยู่ในรูปแบบของระบบสมการที่ประกอบด้วย $m+1$ สมการย่อยได้ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots\cdots & \sum_{i=1}^{n} x_i^m \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \cdots\cdots & \sum_{i=1}^{n} x_i^{m+1} \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \cdots\cdots & \sum_{i=1}^{n} x_i^{m+2} \\ \vdots & \vdots & \vdots & & \\ \vdots & \vdots & \vdots & & \\ \sum_{i=1}^{n} x_i^m & \sum_{i=1}^{n} x_i^{m+1} & \sum_{i=1}^{n} x_i^{m+2} & \cdots\cdots & \sum_{i=1}^{n} x_i^{2m} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_m \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \\ \vdots \\ \vdots \\ \sum_{i=1}^{n} x_i^m y_i \end{Bmatrix} \quad (2.55)$$

โดยเมตริกซ์จัตุรัสขนาด $(m+1)\times(m+1)$ ทางด้านซ้ายของระบบสมการนี้เป็นเมตริกซ์สมมาตรที่รู้ค่า และเวกเตอร์ขนาด $(m+1)\times 1$ ทางด้านขวาของระบบสมการก็รู้ค่าเช่นกัน ดังนั้นตัวไม่รู้ค่า $a_0, a_1, a_2, \ldots, a_m$ ทั้งหมด $m+1$ ค่า สามารถคำนวณหาได้จากระบบสมการนี้โดยใช้ระเบียบวิธีการแก้ระบบสมการวิธีใดวิธีหนึ่ง

2.10 การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชั่นพลังงานเสียง (Energy function normalize)

เพื่อให้ฟังก์ชั่นพลังงานเสียงที่จะนำไปทำการทดสอบบรรทัดฐานค่าพลังงานเดียวกัน ทำได้โดยการกำหนดให้พื้นที่ใต้ฟังก์ชั่นพลังงานเสียงมีค่าเท่ากับหนึ่ง ทุกเสียงทดสอบ ซึ่งแสดงได้ดังสมการต่อไปนี้

กำหนดให้พื้นที่ใต้ฟังก์ชันพลังงานเสียง $E(n)$ มีค่าเท่ากับ 1

$$\sum_{N}^{M} AE(n) = 1$$

โดยที่ A คือ normalize factor

N คือ ค่าพลังงานเสียงเริ่มต้น , M ค่าพลังงานเสียงสุดท้าย

สามารถหา normalize factor ได้คือ

$$A = \frac{1}{\sum_{N}^{M} E(n)}$$

ดังนั้นจะได้ค่าพลังงานเสียงที่ผ่านการปรับบรรทัดฐานพลังงานเสียง ดังนี้

$$Qn = AEn \tag{2.56}$$

เมื่อ Qn คือ ค่าพลังงานเสียงที่ผ่านการปรับบรรทัดฐานพลังงาน

บทที่ 3
วิธีดำเนินการวิจัย

3.1 การสร้างรายการคำทดสอบ

การสร้างรายการคำทดสอบในวิทยานิพนธ์นี้ จะใช้คนทั้งหมด 10 คน ประกอบด้วยผู้ชาย 5 คน ผู้หญิง 5 คน โดยให้ผู้ทดสอบแต่ละคนอ่านออกเสียงสระ 24 รูปเสียง ที่ผันตาม พยัญชนะ 21 รูปเสียง โดยที่ผู้ทดสอบแต่ละ คนจะอ่านออกเสียง 504 คำ 2 รอบ ดังนั้นวิทยานิพนธ์นี้จะมีจำนวนคำที่ใช้ทั้งหมด 10,080 คำ ซึ่งนำไปสร้างเป็นแบบเสียงอ้างอิงจำนวน 5,040 คำ และสร้างแบบเสียงทดสอบ จำนวน5,040 คำ

3.2 การคัดเลือกผู้บอกภาษา

การคัดเลือกผู้บอกภาษาในวิทยานิพนธ์นี้ จะใช้ผู้บอกภาษาที่เป็นคนไทยมีอายุระหว่าง 20 – 30 ปี เป็นเพศชาย 5 คน และ เพศหญิง 5 คน โดยภาษาที่ใช้จะใช้เสียงภาษาไทยของคนภาคกลาง

3.3 การเก็บข้อมูล

การเก็บข้อมูลในวิทยานิพนธ์นี้ จะใช้คำพูดที่เป็นภาษาไทยโดยจะให้ผู้ทดสอบแต่ละคนอ่านออกเสียง 504 คำ 2 รอบ โดยการอ่านจะให้อ่านออกเสียงแบบไม่ต่อเนื่องคือจะให้อ่านแล้วหยุดสลับกันเพื่อให้คำที่พูดมีลักษณะไม่ต่อเนื่องกัน คำที่จะอ่านจะใช้เสียงพยัญชนะ 1 ตัว แล้วเปลี่ยนเสียงสระไปจนครบ 24 เสียง จากนั้นก็เปลี่ยนพยัญชนะใหม่จนครบ 21 เสียง ในการบันทึกเสียงซึ่งข้อมูลที่ถูกบันทึกจะบันทึกอยู่ในรูปแบบของไฟล์ข้อมูล ".wav" ซึ่งข้อมูล 1 ตัวอย่างของเสียงจะถูกแทนด้วยข้อมูลขนาด 8 บิท โดยใช้ความถี่ในการแซมปลิ้งเสียงเท่ากับ 11.025 KHz ไฟล์ข้อมูลเสียง ".wav"

3.4 เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียง

เครื่องมือและอุปกรณ์ในการวิเคราะห์สัญญาณเสียงในวิทยานิพนธ์นี้ จะใช้คอมพิวเตอร์ 1 เครื่อง ไมโครโฟน 1 ตัว ลำโพง 1 ชุด โปรแกรมบันทึกเสียงใช้โปรแกรม Sound Blaster Wave Studio และโปรแกรมที่ใช้ในการทำงานวิจัยใช้โปรแกรม Borland C++ Builder

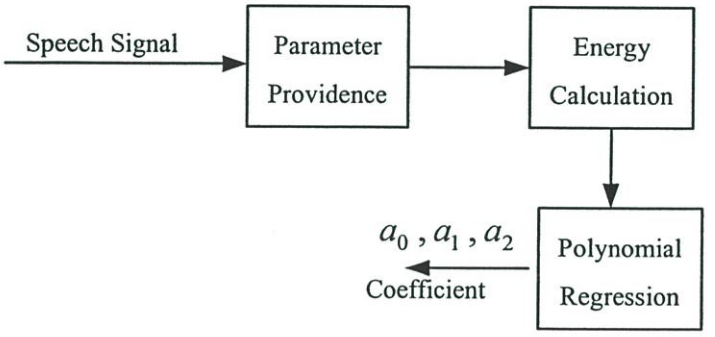
3.5 ข้อจำกัดในการดำเนินการวิจัย

ข้อจำกัดในการดำเนินการวิจัยในวิทยานิพนธ์นี้ ผู้บอกภาษาจะต้องออกเสียงภาษาไทยได้ชัดเจน ต้องพูดภาษาไทย ภาคกลางได้ชัดเจน และ ต้องพูดในจังหวะที่เป็นปกติเหมือนกันกับการพูด โดยทั่ว ๆไป

บทที่ 4
การทดลองและผลการทดลอง

4.1 กล่าวนำ

ขั้นตอนการแยกแยะสระเสียงสั้นและสระเสียงยาว แบ่งออกเป็น 3 ขั้นตอนดังรูป

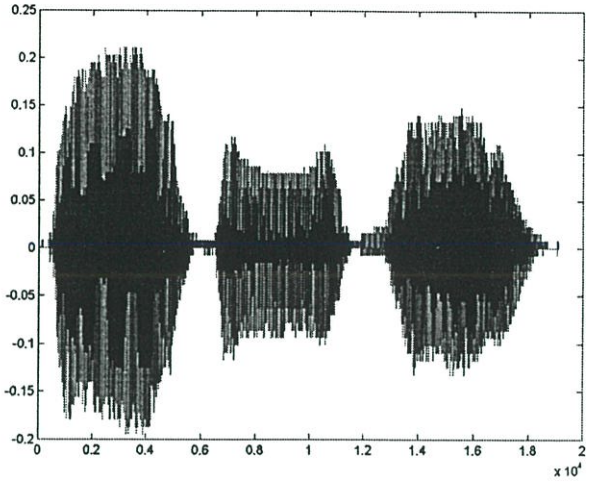


รูปที่ 4.1 ขั้นตอนการแยกสระเสียงสั้นและสระเสียงยาว

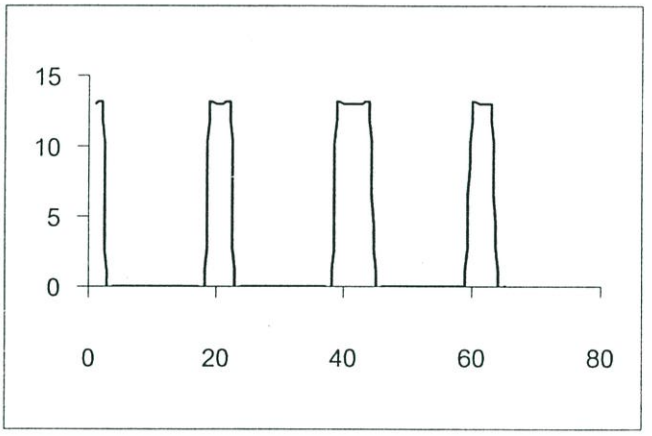
ในบทนี้จะกล่าวถึงวิธีการทดลองและผลการทดลองในขั้นตอนต่างของการแยกแยะสระเสียงสั้นและสระเสียงยาว

4.2 การกำหนดขอบเขตของพยางค์

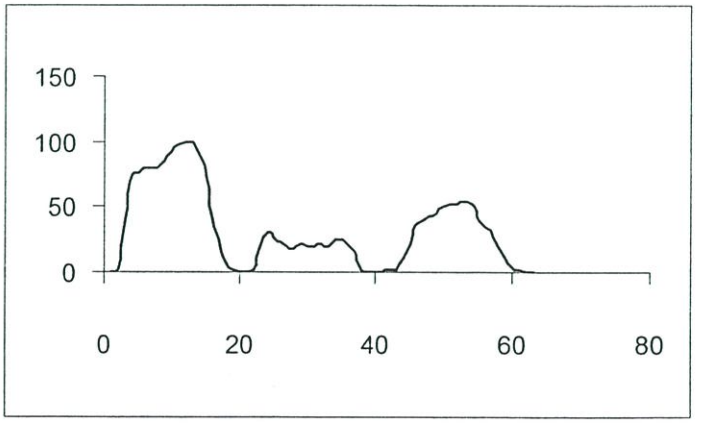
หน่วยเสียงสระที่ลักษณะเป็นเสียงก้อง)voice) และมีลักษณะสัญญาณเป็นคาบ ซึ่งเป็นลักษณะที่เห็นได้ชัดในโดเมนของเวลา ดังนั้นการกำหนดขอบเขตของคำเพื่อที่จะวิเคราะห์หน่วยเสียงสระ จึงกำหนดขอบเขตภายในช่วงที่มีความเป็นคาบทั้งหมดของคำหรือพยางค์นั้นๆ รวมถึงการใช้ค่าระดับพลังงานร่วมด้วย โดยเริ่มจากสัญญาณเสียงที่ถูกเก็บมาในลักษณะที่ให้ผู้พูดพูดคำโดดโดยแต่ละคำมีช่วงห่างกันเล็กน้อย ดังรูปที่ 4.)1a) จากนั้นทำการคำนวณหาค่าพลังงานของเสียงดังรูปที่ 4.1(b) และค่าขอบเขตความถี่มูลฐานดังรูปที่ 4.)1c) นำค่าขอบเขตของพลังงานและขอบเขตของความถี่มาทับซ้อนกันดังรูปที่ 4.1(d) เพื่อใช้ในการตัดเสียงจะได้เสียงที่ตัดออกมาดังรูปที่ 4.)1e)



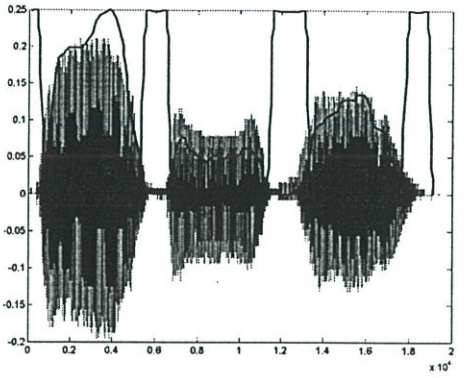
(a) แสดงตัวอย่างเสียงสระเสียง "กอ ขอ งอ"



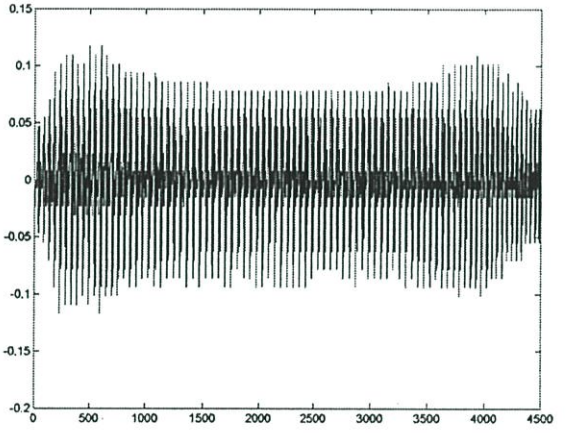
(b) ขอบเขตของค่าพลังงาน



(c) ขอบเขตของค่าความถี่มูลฐาน



(d) ขอบเขตของทั้งสองวิธี โดยบริเวณที่ใช้คือบริเวณที่มีการทับซ้อนกันของทั้งสองวิธี

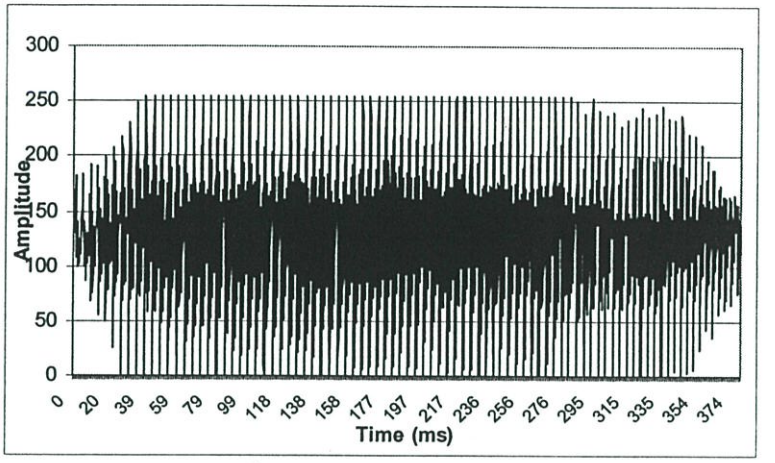


(e) แสดงตัวอย่างเสียงสัญญาณที่ผ่านการตัดเสียง

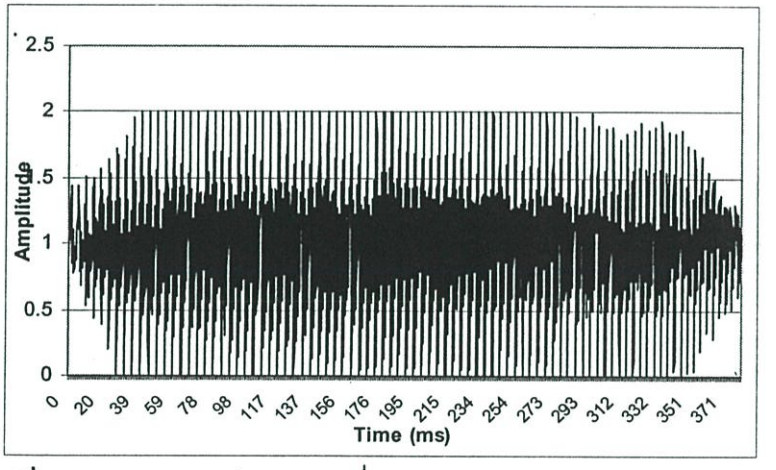
รูปที่ 4.2 การตัดเสียงโดยใช้ค่าพลังงานและความถี่มูลฐาน

4.3 การเตรียมข้อมูลเสียง

การปรับบรรทัดฐานแอมปลิจูดใช้สมการที่ 2.1 โดยรูปที่ 4.3 และ รูปที่ 4.4 เป็นตัวอย่างสัญญาณเสียงที่เข้ามา และสัญญาณเสียงที่ผ่านการปรับบรรทัดฐาน

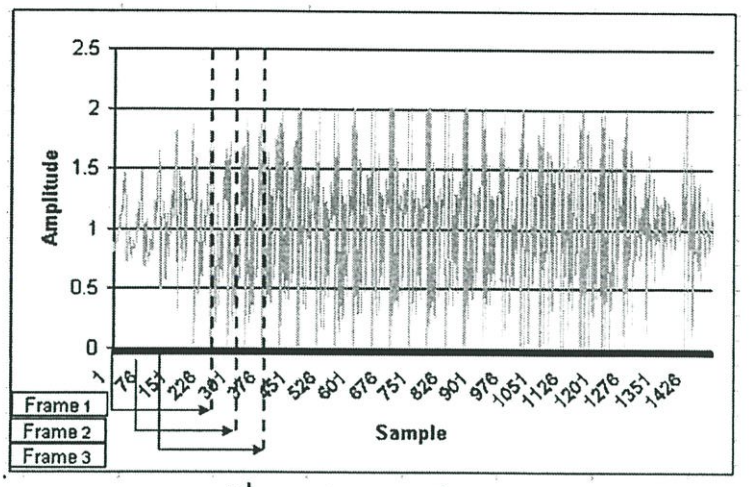


รูปที่ 4.3 สัญญาณเสียง "ขอ"

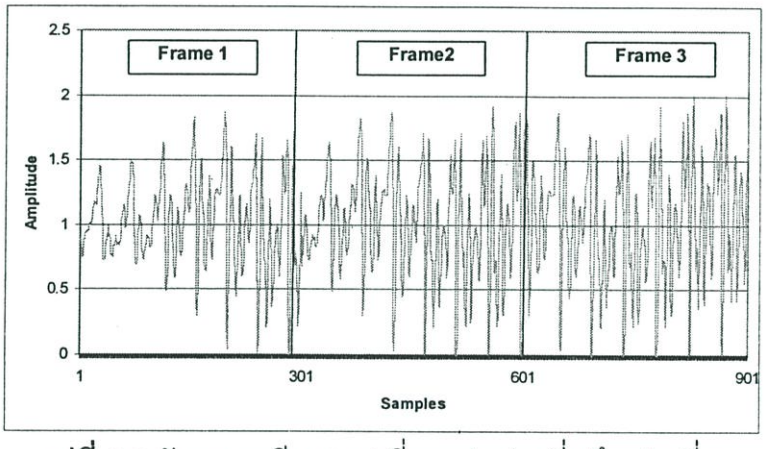


รูปที่ 4.4 สัญญาณเสียง "ขอ" ที่ผ่านการปรับบรรทัดฐาน Amplitude

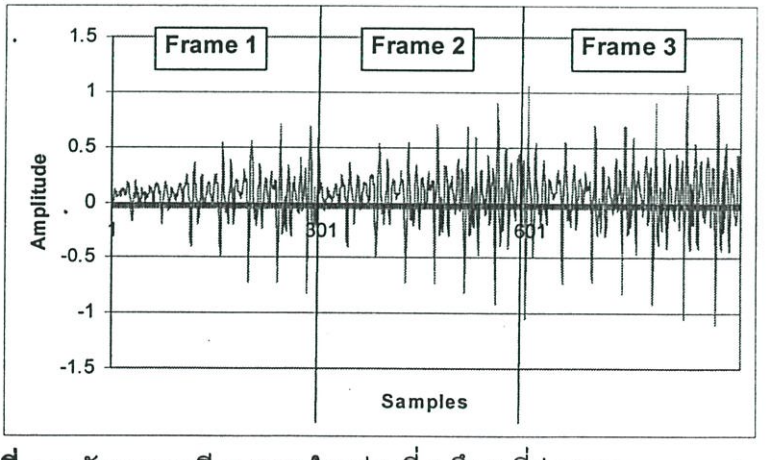
การคำนวณหาค่าพลังงานเสียง เริ่มต้นจากการแบ่งช่วงสัญญาณออกเป็นเฟรม เฟรมละ 300 samples และการวิเคราะห์แต่ละเฟรมจะขยับไปครั้งละ 75 samples(overlap ¾) ตามรูปที่ 4.5 และรูปที่ 4.6 หลังจากนั้นเป็นการปรีเอ็มฟาซิสให้ความถี่สูงมีขนาดสูงขึ้นซึ่งจะได้เป็นดังรูปที่ 4.7 แล้วจึงทำการวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณบริเวณขอบโดยการใช้แฮมมิ่งวินโดว์ซึ่งจะได้ผลดังรูปที่ 4.8 จากนั้นจึงนำสัญญาณที่ผ่านแฮมมิ่งวินโดว์ ไปคำนวณหาค่าพลังงานเสียงในรูปแบบต่างๆ ตามที่ได้กล่าวมาใน หัวข้อที่ 2.7



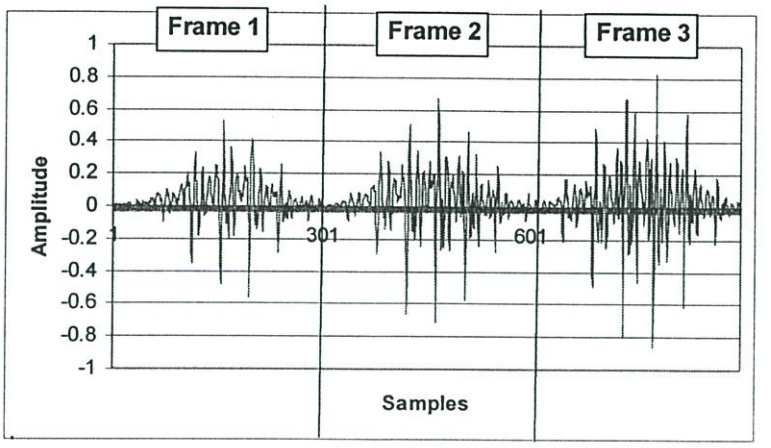
รูปที่ 4.5 สัญญาณเสียง "ขะ"



รูปที่ 4.6 สัญญาณเสียง "ขะ" ที่ถูกแบ่งเฟรมที่ 1ถึงเฟรมที่ 3



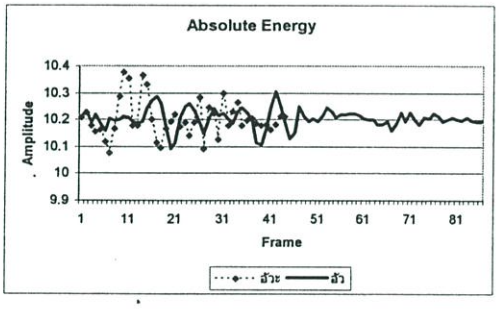
รูปที่ 4.7 สัญญาณเสียง "ขะ" ในเฟรมที่ 1 ถึง 3 ที่ผ่านการ preemphasis



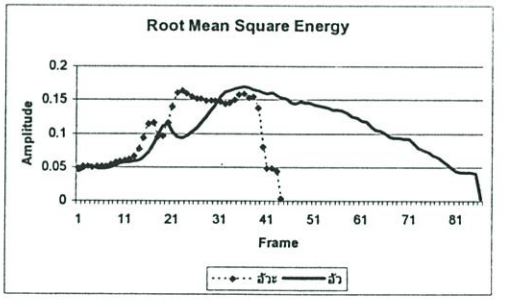
รูปที่ 4.8 สัญญาณเสียง "ขะ" ในเฟรมที่ 1 ถึง 3 ที่ผ่านฟังก์ชั่นวินโดว์แบบแฮมมิ่ง

4.4 การหาค่าพลังงานเสียง

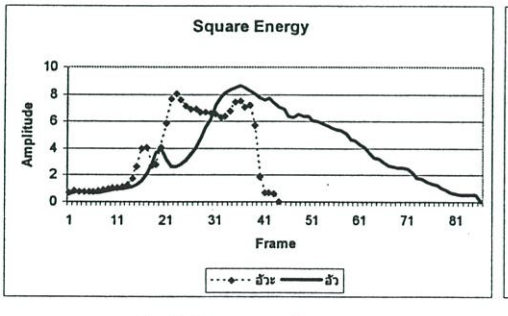
การคำนวณหาค่าพลังงานเสียง จะทำทีละกรอบเสียงพูด โดยมีตัวอย่างกรอบเสียงพูดละ 300 Samples และค่าพลังงานเสียงในกรอบต่อไปจะถูกเลื่อนไปอีก 75 Samples ซึ่งการคำนวณหาค่าพลังงานของสัญญาณเสียง จะได้ลักษณะพลังงานเสียงในรูปแบบต่างๆ ดังนี้



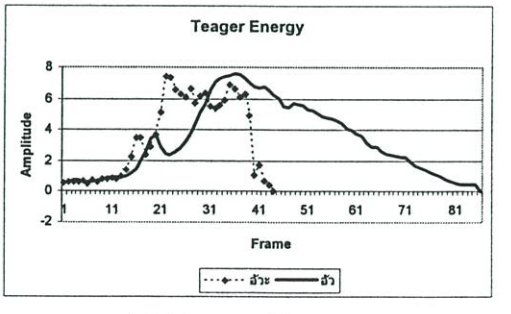
(a) Absolute Energy



(b) Root Mean Square Energy



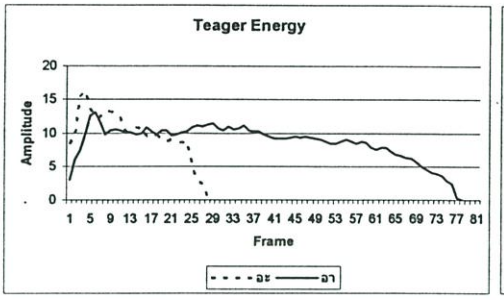
(c.) Square Energy



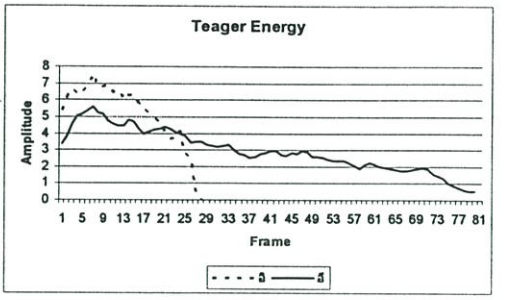
(d) Teager Energy

รูปที่ 4.9 ลักษณะพลังงานเสียง

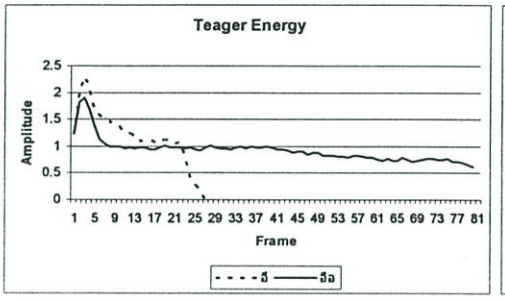
ลักษณะพลังงานเสียง แยกตาม เสียงสระต่าง แสดงดังรูปที่ 4.10



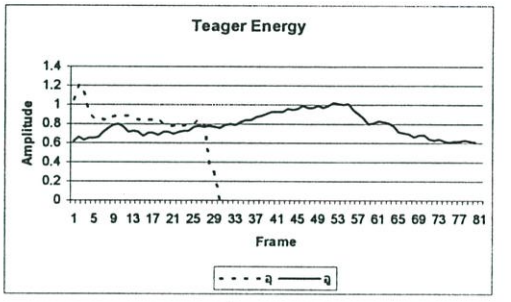
(a) เสียง "อะ-อา"



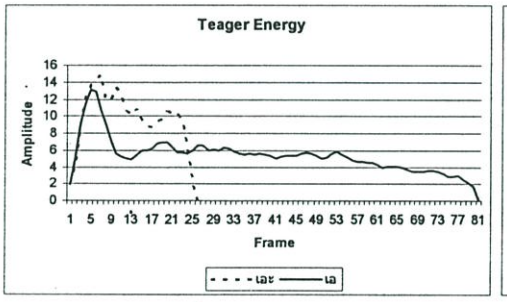
(b) เสียง "อิ-อี"



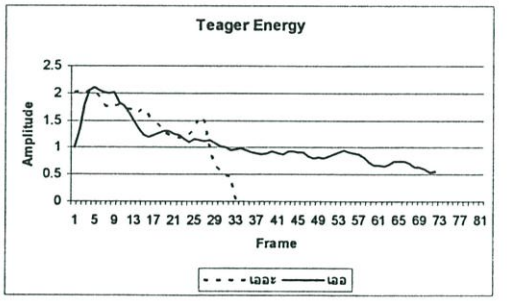
(c) เสียง "อึ-อือ"



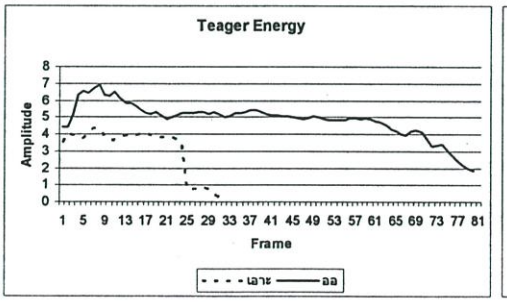
(d) เสียง "อุ-อู"



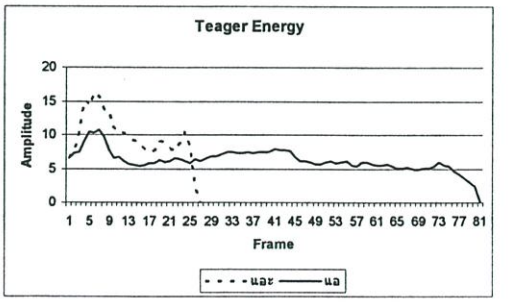
(e) เสียง "เอะ-เอ"



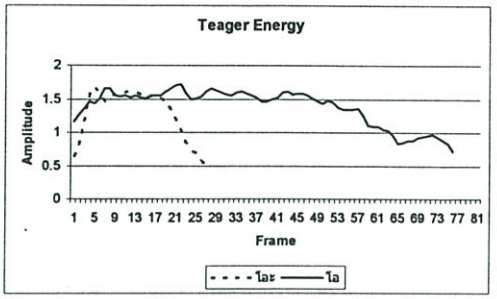
(f) เสียง "เออะ-เออ"



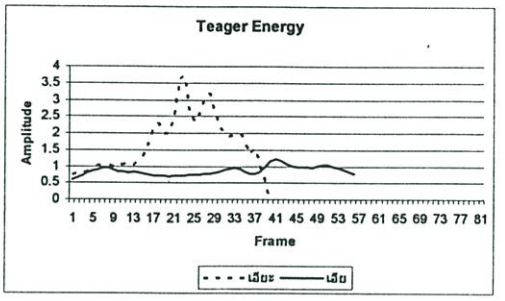
(g) เสียง "เอาะ-ออ"



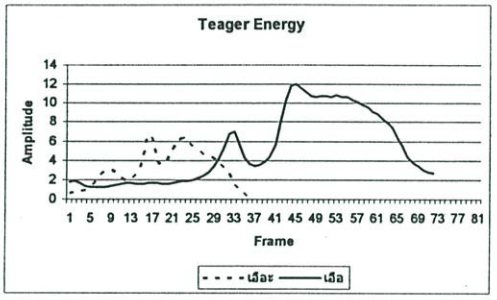
(h) เสียง "แอะ-แอ"



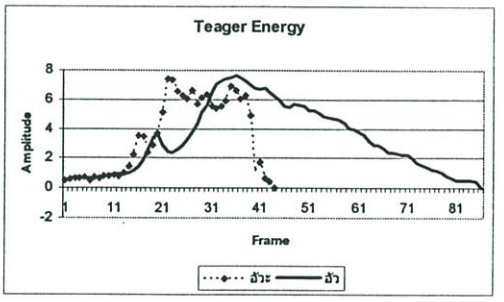
(i) เสียง "โอะ-โอ"



(j) เสียง "เอียะ-เอีย"



(k) เสียง "เออะ-เออ"



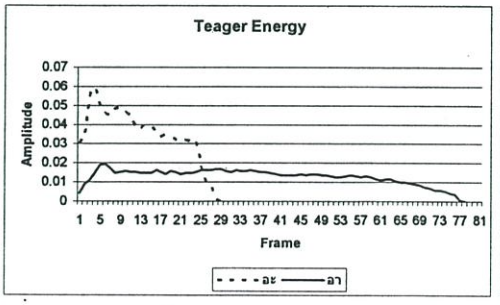
(l) เสียง "อัวะ-อัว"

รูปที่ 4.10 ลักษณะพลังงานเสียงแบบ Teager Energy ของเสียงสระต่างๆ

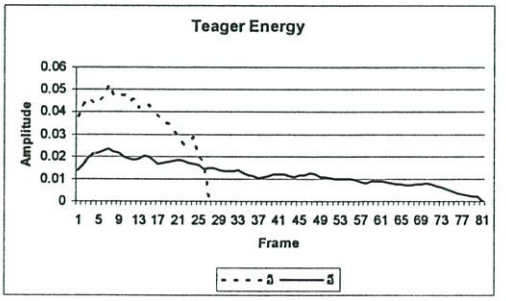
จากรูปลักษณะพลังงานของเสียงสระ จะเห็นได้ว่าลักษณะการหยุดเสียงของสระเสียงสั้นนั้นจะเป็นการหยุดแบบทันทีทันใด ส่วนสระเสียงยาวจะมีการหยุดเสียงแบบค่อยๆ ลดระดับลงช้าๆ

4.5 การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชั่นพลังงานเสียง (Energy function normalize)

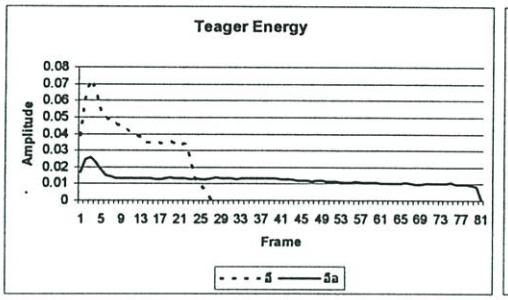
เมื่อได้ลักษณะพลังงานเสียง นำพลังงานเสียงที่ได้มาปรับบรรทัดฐานพลังงาน โดยให้พื้นที่ใต้ฟังก์ชั่นพลังงานเสียง มีค่า เท่ากับ 1 ทุกๆเสียง ตามหัวข้อ 2.10 ซึ่งจะทำให้เห็นความแตกต่าง พลังงานเสียงสระเสียงสั้น และ สระเสียงยาว มากขึ้น ดังรูปที่ 4.11



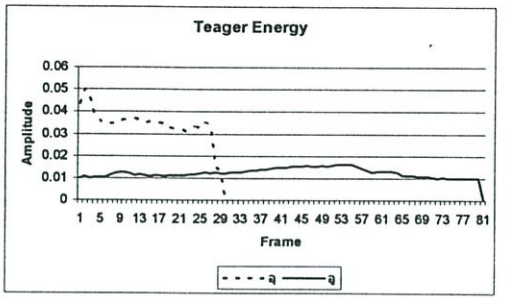
(a) เสียง "อะ-อา"



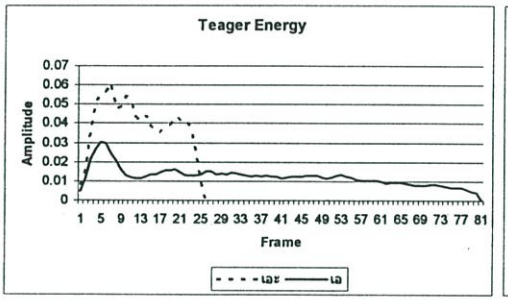
(b) เสียง "อิ-อี"



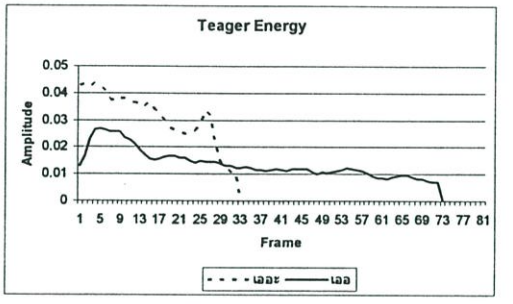
(c) เสียง “อึ-อือ”



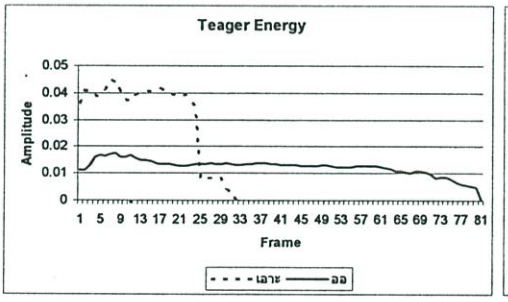
(d) เสียง “อุ-อู”



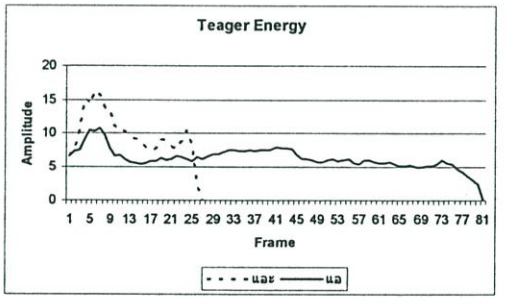
(e) เสียง “เอะ-เอ”



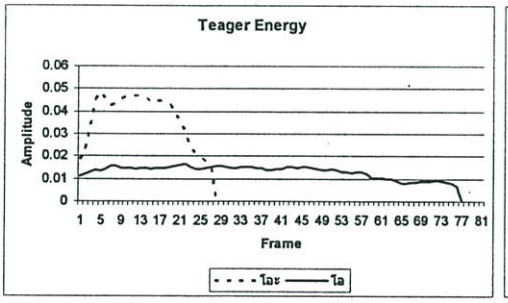
(f) เสียง “เออะ-เออ”



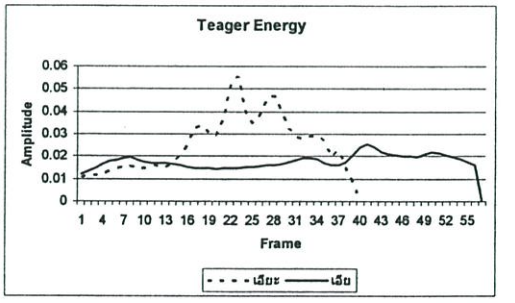
(g) เสียง “เอาะ-ออ”



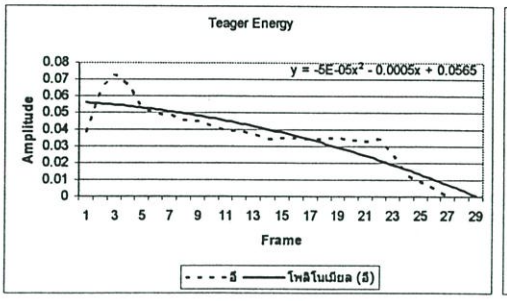
(h) เสียง “แอะ-แอ”



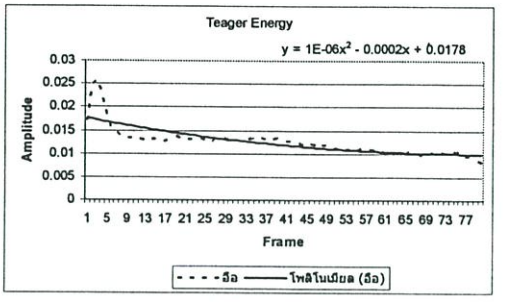
(i) เสียง “โอะ-โอ”



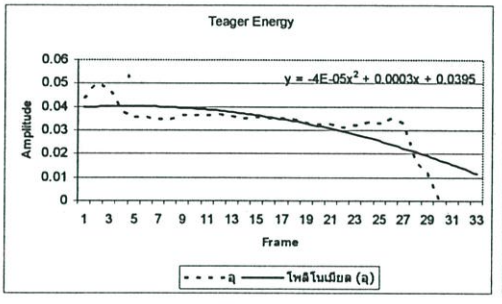
(j) เสียง “เอียะ-เอีย”



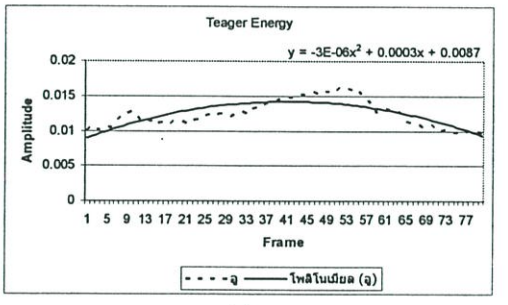
(d) เสียง “อี”



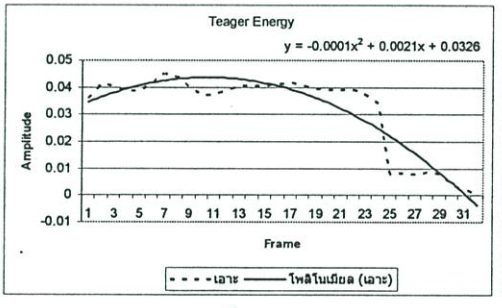
(e) เสียง “อือ”



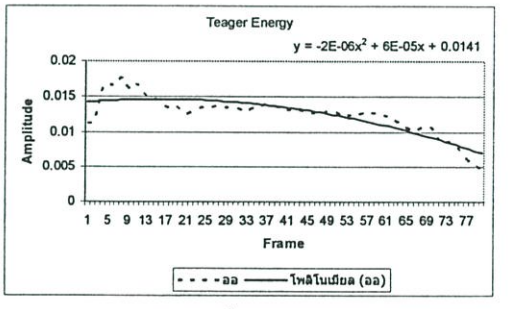
(f) เสียง “อุ”



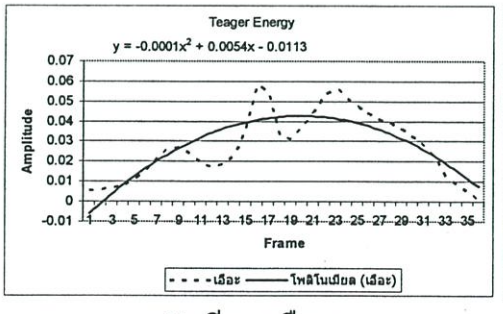
(g) เสียง “อู”



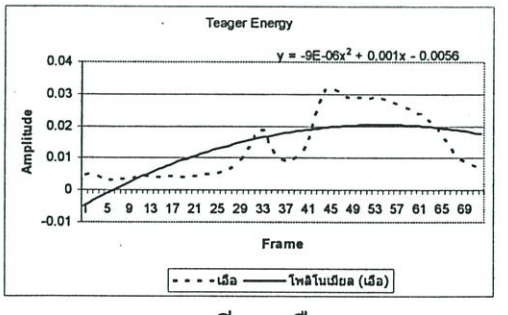
(h) เสียง “เอาะ”



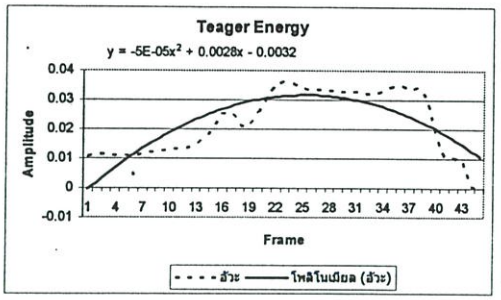
(i) เสียง “ออ”



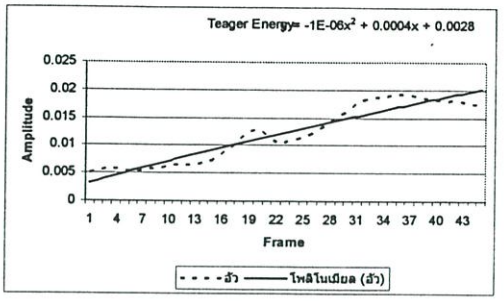
(j) เสียง “เออะ”



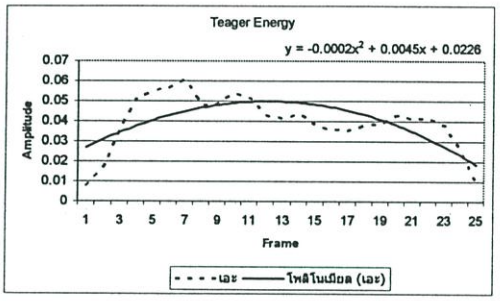
(k) เสียง “เอือ”



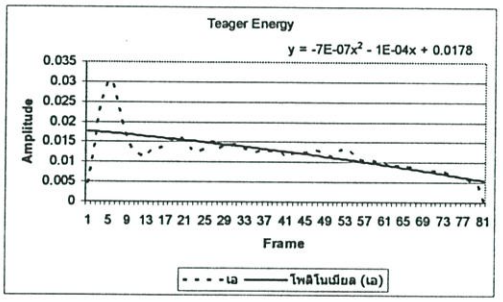
(l) เสียง “อัวะ”



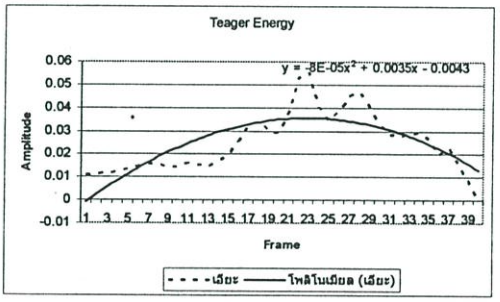
(m) เสียง “อัว”



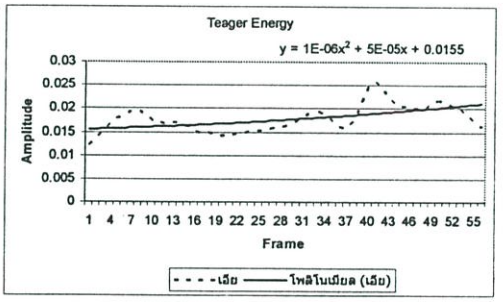
(n) เสียง “เอะ”



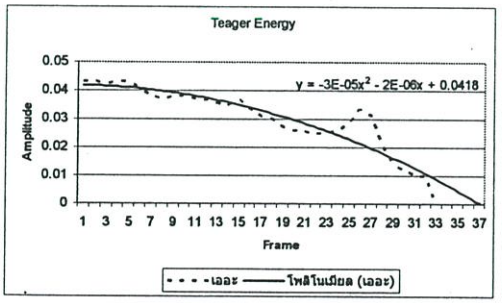
(o) เสียง “เอ”



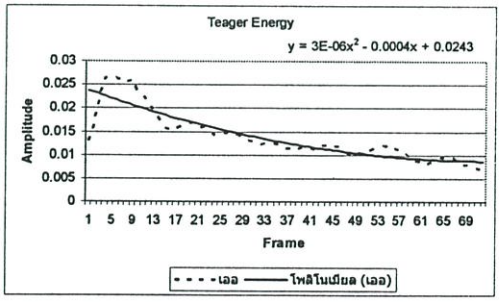
(p) เสียง “เอียะ”



(q) เสียง “เอีย”



(r) เสียง “เออะ”



(s) เสียง “เออ”

ตาราง 4.1 การแจกแจง สัมประสิทธิ์การถดถอยพหุนามอันดับสอง (a_2)

a2 x 5000	Short Vowel	Long Vowel
1.4	0	0
1.2	1	0
1	2	0
0.8	1	0
0.6	6	0
0.4	2	3
0.2	12	92
0	39	2103
-0.2	84	266
-0.4	120	36
-0.6	***135***	***2***
-0.8	214	0
-1	202	0
-1.2	198	0
-1.4	142	0
-1.6	172	0
-1.8	142	0
-2	120	0
-2.2	112	0
-2.4	104	0
-2.6	78	0
-2.8	94	0
-3	76	0
-3.2	40	0
-3.4	52	0
-3.6	52	0
-3.8	24	0
-4	26	0
-4.2	18	0
-4.4	32	0
-4.6	18	0
-4.8	32	0
-5	20	0
-5.2	16	0
-5.4	4	0
-5.6	22	0
-5.8	4	0
-6	10	0
-6.2	10	0
-6.4	2	0
-6.6	8	0
-6.8	2	0
-7	2	0
-7.2	8	0
-7.4	4	0
-7.6	4	0
-7.8	0	0

จากรูป 4.10 จะเห็นได้ว่า เมื่อทำการแจกแจงค่าสัมประสิทธิ์ a_2 ของสระเสียงสั้นและสระเสียงยาว พบว่าสามารถใช้ค่า a_2 ค่าๆหนึ่งที่แบ่งแยก สระเสียงสั้นและสระเสียงยาวออกจากกันได้ โดยในที่นี้ใช้ค่าสัมประสิทธิ์ a_2 ที่ใช้ในการแยกสระเสียงสั้นและสระเสียงยาว คือ -0.6 และเมื่อนำรูปแบบพลังงานแบบต่างๆ มาหาค่าสัมประสิทธิ์ a_2 จะได้อัตราการแยกแยะสระเสียงสั้นและสระเสียงยาวตามตารางที่ 4.2 จากตัวอย่างเสียง 5,040 เสียง (สระเสียงสั้น 2,520 , สระเสียงยาว 2,520 เสียง)

ตารางที่ 4.2 แสดงอัตราการแยกแยะ (แบบอ้างอิง)

Energy Algorithm	**สระเสียงสั้น**	**สระเสียงยาว**	**เปอร์เซ็นต์**
Teager Energy	2273	2437	93.45%
Square Energy	2331	2446	94.78%
Root Mean Square Energy	2133	2471	91.35%

จากผลการทดสอบเสียงอ้างอิงเพื่อหาค่า จะเห็นได้ว่าหากใช้การคิดค่าพลังงานแบบ Square ที่สัมประสิทธิ์ a_2 = -0.0012 จะได้อัตราการแยกสัญญาณเสียงสั้น และ เสียงยาวมากที่สุดที่ 94.78 % จึงได้ทดสอบกับสัญญาณเสียงสระอีกชุดในจำนวนเสียงเท่ากัน คือ 5,040 เสียง และเป็นผู้พูดคนเดียวกันกับข้อมูลเสียงชุดแรก ได้อัตราการแยกแยะดังนี้

ตารางที่ 4.3 แสดงอัตราการแยกแยะ (แบบทดสอบ)

Energy Algorithm	**สระเสียงสั้น**	**สระเสียงยาว**	**เปอร์เซ็นต์**
Square Energy	2356	2420	94.76%

วิธีการแยกแยะ สระเสียงสั้น และสระเสียงยาวที่ได้นำเสนอไปในข้างต้น ให้อัตราการแยกแยะได้ที่ 94.76 % จากการทดสอบสัญญาณเสียงทั้งสิ้น 5,040 เสียงแบ่งเป็นสระเสียงสั้นจำนวน 2,520 เสียง สระเสียงยาว 2,520 จากผู้ออกเสียงชาย 5 คน ผู้ออกเสียงหญิง 5 คน

บทที่ 5
สรุปผลและข้อเสนอแนะ

วิทยานิพนธ์นี้เป็นการเสนอการแยกแยะสระเสียงสั้นและสระเสียงยาว ในในภาษาไทย โดยใช้คุณลักษณะทางพลังงานของเสียงที่เปล่งออกมา โดยพิจารณาที่ลักษณะความแตกต่างการหยุดเสียงของสระเสียงสั้น,สระเสียงยาว เป็นตัวบ่งชี้ โดยใช้การค่าสัมประสิทธิ์การถดถอยพหุนามอันดับสอง (ค่า a_2).ที่ได้จากลักษณะพลังงานเสียง เป็นตัวแยกแยะ

5.1 การทดลอง

สัญญาณเสียงเป็นข้อมูลดิบที่ใช้ในการวิเคราะห์นั้นจะเป็นเสียงสระเสียงสั้นและเสียงสระเสียงยาว ที่ผันตามพยัญชนะ 21 รูป จากผู้ออกเสียงชาย จำนวน 5 คน และ หญิงจำนวน 5 คน รวมเสียงที่ใช้ในการวิเคราะห์ทั้งสิ้น 5,040 เสียงสระ จากนั้นจะนำเสียงสระนั้นมาทำการเตรียมข้อมูลเสียง โดยผ่าน กระบวนการปรับบรรทัดฐานเชิงขนาด (Amplitude Normalization) เพื่อให้ข้อมูลเสียงนั้นง่ายต่อการกำหนดค่าพารามิเตอร์ , กรรมวิธีการเน้นล่วงหน้า (Preeamphasis) เพื่อให้อัตราส่วนของเสียงพูดต่อสัญญาณรบกวนมีค่าสูงขึ้น, การแบ่งช่วงสัญญาณ (Frame Blocking) เพื่อข้อมูลเสียงมีความเสถียรและไม่เปลี่ยนแปลงตาม ,กรรมวิธีการวางกรอบหน้าต่าง (Windowing) เพื่อป้องกันการเปลี่ยนแปลงอย่างกะทันหันที่บริเวณปลายของกรอบเสียงพูด

จากนั้นจึงนำข้อมูลเสียงมาหาลักษณะทางพลังงานของเสียงนั้น โดยใช้รูปแบบการหาพลังงานเสียง 4 รูปแบบคือ พลังงานสัมบูรณ์ (Absolute Energy), พลังงานเฉลี่ย (Root Mean square Energy), พลังงานกำลังสอง (Square Energy) และพลังงานของ Teager (Teager Energy) โดยจะเปรียบเทียบลักษณะพลังงานเสียงทั้ง 4 แบบว่าพลังงานเสียงแบบใดที่ จะบ่งชี้ลักษณะเสียงสั้น และ เสียงยาวได้เด่นชัดที่สุด ซึ่งจากการทดลองพบว่าพลังงานเสียงแบบพลังงานกำลังสอง (Square) จะให้อัตราการแยกยะ สระเสียงสั้น และสระเสียงยาวมากที่สุด

จากนั้นจึงนำลักษณะพลังงานเสียงมาหาค่าสัมประสิทธิ์การถดถอยพหุนามอันดับสอง (a_2) โดยค่าสัมประสิทธิ์ a_2 นี้จะใช้เป็นตัวแยกแยะลักษณะสระเสียงสั้น และ สระเสียงยาว โดยถ้าเสียงพูดเสียงนั้นเมื่อหาค่าสัมประสิทธิ์ a_2 ได้ค่า <u>มากกว่า -0.0012</u> จะได้ผลเป็นสระเสียงยาว และถ้าค่าสัมประสิทธิ์ a_2 ได้ค่า<u>น้อยกว่าหรือเท่ากับ -0.0012</u> จะได้ผลเป็นสระเสียงสั้น

จากวิธีการแยกแยะสระเสียงสั้นและสระเสียงยาวดังกล่าว สามารถแยกแยะสระเสียงสั้นและเสียงยาวสูงสุดที่ <u>94.76 %</u> โดยที่ความผิดพลาดส่วนใหญ่เกิดจากเสียงรบกวนที่มีมากในการ

บันทึกเสียง และ ในสระเสียงผสม ซึ่งลักษณะของเนื้อเสียงที่เปล่งออกมาจะมีการเปลี่ยนแปลงของพลังงานตลอดช่วงการเปล่งเสียง มากกว่า สระเสียงเดี่ยว

5.2 ข้อสังเกต และปัญหาที่พบในการทดลอง

- การตัดคำมีความสำคัญต่อผลการแยกแยะเป็นอย่างมากเนื่องจากข้อมูลเสียงที่ใช้ในการวิเคราะห์มีจำนวนมาก การคัดเลือกเสียง จึงจำเป็นต้องอาศัยการเขียนโปรแกรม ซึ่งบางครั้งโปรแกรมตัดคำที่จัดทำขึ้นอาจจะตัดคำได้ไม่สมบูรณ์ โดยเฉพาะ สระเสียงสั้น หากการตัดเสียง ไม่ครบข้อมูลเสียงที่เปล่งออกมา ทำให้ค่าสัมประสิทธิ์ a_2 ที่ได้ผิดพลาดไป นอกจากนี้การบันทึกเสียง ก็ยังมีผลต่อการตัดคำด้วย คือถ้าอยู่ในสภาพแวดล้อมที่มีเสียงรบกวนมาก ทำให้การตัดคำมีโอกาสเกิดการผิดพลาดได้

- การใช้คุณลักษณะทางพลังงานเสียงใน วิทยานิพนธ์นี้ เน้นที่ ลักษณะทางพลังงานเสียงในแกนเวลา ทำให้ความเสถียรของพลังงานเสียงมีน้อย โดยยังมีรูปแบบการหาพลังงานเสียง อีกหลายรูปแบบที่อยู่ในแกนความถี่ที่สามารถนำมาเป็นตัวแยกแยะได้ด้วยเช่นกัน

- การเลือกอันดับการถดถอยพหุนาม สามารถทำได้ในอันดับที่สูงกว่า อันดับสอง ซึ่ง ยิ่งถ้าลักษณะของเสียงนั้นเกิดความไม่เสถียรมากเท่าไหร่ (จากเสียงรบกวนในขั้นตอนการบันทึกเสียง) พลังงานที่ได้ก็จะมีการเปลี่ยนแปลงอย่างมาก ตลอดย่านเสียงที่เปล่งออกมา หากให้สัมประสิทธิ์การถดถอยพหุนามอันดับสองจะทำให้เกิดข้อผิดพลาดได้มาก ซึ่งต้องใช้ อันดับของสัมประสิทธิ์การถดถอยพหุนามเพิ่มขึ้น

5.3 ข้อเสนอแนะ

เสียงสระที่ใช้ในการทดสอบ เป็นเสียงสระที่ผันตามพยัญชนะ 21 รูป ยังไม่รวมถึง เสียงที่มีการผันตามวรรณยุกต์ และ เสียงที่มีพยัญชนะท้าย ซึ่งมีแนวทางในการพัฒนาดังนี้

- เสียงที่นำมาพิจารณาในวิทยานิพนธ์นี้ จะมีเสียงของพยัญชนะต้นรวมกับเสียงสระ หากใช้วิธีการแยก พยัญชนะต้น ออกจากเสียงสระ แล้วจะทำให้ความแตกต่างของพลังงานเสียงสระเด่นชัดขึ้น ทั้งนี้ยังรวมถึงในกรณีที่ เสียงสระที่ มีการผันตามวรรณยุกต์ และมี พยัญชนะท้าย

- วิธีการแยกแยะในวิทยานิพนธ์นี้ ใช้ลักษณะข้อมูลทางสถิติของสัมประสิทธิ์การถดถอยพหุนามในการแยกแยะ ซึ่งหากใช้ข้อมูลเสียงมีมากขึ้นโอกาส ที่ค่าสัมประสิทธิ์การถดถอยพหุนาม ที่ตั้งไว้อาจเปลี่ยนแปลงได้ หรือ อัตราการแยกแยะอาจลดต่ำลงได้ ดังนั้นหาก นำค่าสัมประสิทธิ์การถดถอยพหุนาม มาสร้างเป็นโมเดลการรู้จำ ก็จะเกิดความแน่นอนในการแยกแยะมากขึ้น ซึ่งเหมาะกับในกรณีที่เลือกใช้ สัมประสิทธิ์การถดถอยพหุนามในอันดับสูงกว่า อันดับสอง

บรรณานุกรม

[1] ธันวา ศรีประมง. "การวิเคราะห์เสียงพูดภาษาไทยในแกนความถี่ฮาร์โมนิค", วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.2537.

[2] ณัฐกร ทับทอง. "การรู้จำคำพูดภาษาไทย โดยใช้ลักษณะบ่งความต่างของหน่วยเสียง", วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2538.

[3] ณัฏฐชา จิตติวรางกูล "ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย", วิทยานิพนธ์วิศวกรรมศาสตร์มหบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัยง 2541.

[4] จิตรลดา จารุมิศร์. "การออกแบบจำลองในการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทยโดยใช้เทคนิคการควอนไตซ์พิทช์ และ Hidden Makov Modelling", วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542.

[5] วรา คงคาวิฑูรย์. "การรู้จำเสียงสระเสียงเดี่ยวในภาษาไทยโดยใช้ทรานส์เฟอร์ฟังก์ชันบนสเกลบาร์ก", วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีเจ้าคุณทหารลาดกระบัง. 2542.

[6] ทัศเวท วีระวัฒน์. "การรู้จำเสียงคำไทยเฉพาะบุคคล" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2541.

[7] เบญจวรรณ อินทรารักษ์สกุล , ปาริฉัตต์ ลิ่วิระยะ , วิภาวดี อ่วมเจริญ , "การประมาณค่าด้วยสมการพหุนาม " วิทยานิพนธ์วิทยาศาสตรบัณฑิต คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

[8] วิไลวรรณ ขนิษฐานันท์. "ภาษาและภาษาศาสตร์", พิมพ์ครั้งที่ 5. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์. 2533.

[9] สุนทร อรอินทร์, อัฐ เครือฟัก, "การประมวลเสียงพูดโดยการประมาณเชิงเส้น", วิทยานิพนธ์ภาควิศวกรรมไฟฟ้า สาขาวิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2538.

[10] ปราโมทย์ เดชะอำไพ "ระเบียบวิธีเชิงตัวเลขในงานวิศวกรรม", กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย. 2538.

[11] เปรมใจ ตรีสรนุวัฒนา "การวิเคราะห์การถดถอย" เอกสารประกอบการเรียน ภาควิชาสถิติ มหาวิทยาลัยเกษตรศาสตร์.2532.

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

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

[14] E. Zwicker, H. Fastl, "Psychoacoustics : Facts and Models Second Edition.", Springer, 1999.

ภาคผนวก ก.
ผลงานวิจัยที่ได้รับการตีพิมพ์

1. กัญน์ สิทธิประเสริฐ , ไกรสิน ส่งวัฒนา "การแยกสระเสียงสั้น และ สระเสียงยาวในภาษาไทย โดยใช้การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง", วิศวกรรมสารลาดกระบัง, ปีที่ 21 ฉบับวันที่ 2 มิถุนายน 2547.

การแยกสระเสียงสั้น และ สระเสียงยาวในภาษาไทยโดยใช้การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง

Short and Long Vowels Classification in Thai Spoken Language Using 2nd Order Polynomial Curve Fitting on the Voice Energy Function

กัญน์ สิทธิประเสริฐ ไกรสิน ส่งวัฒนา

ภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

บทความนี้เสนอ วิธีการแยกสระเสียงสั้น และ สระเสียงยาวในภาษาไทย โดยใช้ ลักษณะของพลังงานเสียง เป็นตัวบ่งชี้ โดยนำสัญญาณเสียงแต่ละเสียงมาคำนวณหาค่าพลังงานของเสียงในรูปแบบต่างๆ และนำลักษณะของพลังงานเสียง มาประดิษฐ์ฟังก์ชันพหุนามอันดับสอง เพื่อนำค่าสัมประสิทธิ์ ของฟังก์ชั่นพหุนามไปใช้ในการแยกสระเสียงสั้นและเสียงยาว

สระในภาษาไทยประกอบด้วยสระเดี่ยว เสียงสั้นจำนวน 9 เสียง(อะ อิ อึ อุ โอะ เอะ แอะ เออะ เอาะ) สระเดี่ยวเสียงยาว จำนวน 9 เสียง (อา อี อือ อู โอ เอ แอ เออ ออ) สระผสมเสียงสั้นจำนวน 3 เสียง (เอียะ เอือะ อัวะ) และสระผสมเสียงยาวจำนวน 3 เสียง (เอีย เอือ อัว) โดยใช้ข้อมูลเสียงจากเพศชาย 5 คน และ เพศหญิง 5 คน รวมทั้งสิ้น 5,040 เสียง

Abstract

This paper presents method short and long vowels classifying in Thai spoken language by using voice energy feature. The variation in energy of vowels is calculated and the feature of energy is fitted with 2nd order polynomial regression function. The coefficient of polynomial regression is used to classify the short and long vowels.

The vowels in Thai spoken language consist of 9 unmixed short vowels (a, i,ω, u, o, e, ɛ, γ, ɔ̆) , 9 unmixed long vowels (aa, ii, ωω, uu, oo, ee, ɛɛ, γγ, ɔ̆ɔ̆) , 3 mixed short vowels (ia, ωa, ua) and 3 mixed long vowels (iia, ωωa, uua). The voice data samples consist of 5,040 words from 5 male and 5 female.

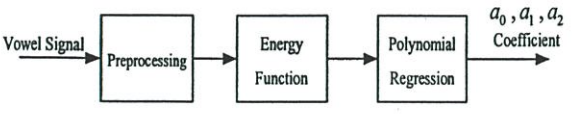
1.บทนำ

ในระบบการรู้จำเสียงพูดภาษาไทยนั้น มีความซับซ้อนมาก เนื่องจากในหนึ่งคำ หรือ พยางค์ ประกอบด้วยเสียงวรรณยุกต์ เสียงพยัญชนะ และเสียงสระ ซึ่งในส่วนของ การรู้จำเสียงสระนั้น ยังไม่สามารถระบุเสียงสั้น และเสียงยาวได้ เนื่องจากความแตกต่างของสระเสียงสั้น และสระเสียงยาวที่เห็นได้ชัดที่สุดคือ ระยะเวลาในการออกเสียง แต่ความแตกต่างนี้ไม่ทำให้สามารถแยกแยะ เสียงที่สั้นและยาวระหว่างบุคคลได้ โดยในบทความนี้จะแยกสระเสียงสั้น และสระเสียงยาว โดยใช้ ลักษณะการหยุดของเสียงแบบทันทีทันใดในสระเสียงสั้น และ การหยุดของเสียงแบบค่อยๆ ลดระดับเสียงลงอย่างช้าๆ ในสระเสียงยาว

โดยขั้นตอนจะเริ่มจาก นำเสียงสระจากตัวอย่างเสียง ซึ่งประกอบด้วยสระเสียงเดี่ยว 9 คู่ และ สระผสม 3คู่ ที่ผันตามเสียงพยัญชนะ 21 หน่วยเสียง มาแยกเสียงสระ ออกจากเสียงพยัญชนะ โดยวิธี [4] แล้วหาระดับพลังงานของเสียงสระ ใน 4 รูปแบบพลังงานเสียง จากนั้นนำลักษณะพลังงานเสียง มาประดิษฐ์ฟังก์ชันการถดถอยพหุนามกำลังสอง และใช้ค่าสัมประสิทธิ์ฟังก์ชันการถดถอยพหุนามมาเป็นจุดแยกของสระเสียงสั้นและ สระเสียงยาว

2. ขั้นตอนการแยกสระเสียงสั้น และ สระเสียงยาว

ในการแยกแยะเสียงสระเสียงสั้น และ เสียงสระเสียงยาวนั้นแบ่งออกเป็นแบ่งออกเป็น 3 ขั้นตอนดังแสดงในรูปที่ 1



รูปที่ 1 แสดงขั้นตอนการแยกแยะเสียงสระเสียงสั้นและสระเสียงยาว

2.1 กระบวนการเตรียมสัญญาณ (Preprocessing)

โดยธรรมชาติสัญญาณเสียงพูดจะไม่เสถียร และเปลี่ยนแปลงตามเวลา (Non Stationary) ดังนั้น เมื่อต้องการนำสัญญาณเสียงพูดมาประมวลผลสัญญาณดิจิตอล (Digital Signal Processing) จึงจำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นช่วงเวลาสั้นๆ โดยนำสัญญาณเสียงที่ได้จากการ sampling ที่ความถี่ 11.025 kHz แบ่งสัญญาณออกเป็นช่วงๆหรือเฟรม (frame) โดยในแต่ละเฟรมกำหนดให้มีตัวอย่างสัญญาณ 300 samples การวิเคราะห์จะทำการวิเคราะห์ทีละเฟรม โดยกำหนดให้มีช่วงของการเลื่อนเฟรมครั้งละ 75 samples นั่นคือแต่ละเฟรมจะมีช่วงของการซ้อนทับกัน 3 ใน 4 เฟรม ซึ่งกรรมวิธีการวิเคราะห์สัญญาณเสียงพูดประกอบด้วย [1]

2.1.1 การปรับบรรทัดฐานแอมปลิจูด (Amplitude Normalization)

การปรับบรรทัดฐานของสัญญาณเสียงเพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน ดังสมการที่ (1)

$$\tilde{S}[i] = \frac{S[i]}{2^{N-1}} \qquad (1)$$

เมื่อ N คือจำนวนบิต(Bit) ที่ใช้แทนค่า ของสัญญาณเสียงพูด

2.1.2 กรรมวิธีการเน้นล่วงหน้า (Pre-emphasis)

กรรมวิธีการเน้นล่วงหน้าจะทำให้ความลาดเอียงในเชิงความถี่แบนราบลงและทำให้อัตราส่วนของสัญญาณเสียงพูดต่อสัญญาณรบกวนมีค่าสูงขึ้น ดังสมการที่ (2)

$$S'[i] = S[i] - aS(i-1) \qquad (2)$$

เมื่อ a คือสัมประสิทธิ์ของตัวกรอง (0.95)

2.1.3 กรรมวิธีการวางกรอบหน้าต่าง (Windowing)

เพื่อป้องกันการเปลี่ยนแปลง อย่างทันทีทันใดที่บริเวณปลายของกรอบเสียงพูด ดังสมการที่ (3) และสมการที่ (4)

$$W_n = 0.54 - 0.46\cos[\frac{2\pi n}{N-1}] \quad (3)$$

$$S_n'[i] = S_n[i] * W_n \quad (4)$$

เมื่อ n= 0,1,…..,N-1

2.2 การวิเคราะห์หาพลังงานเสียงพูด (Energy Function)

การคำนวณหาค่าพลังงานเสียง จะทำทีละกรอบเสียงพูด โดย En คือ ค่าพลังงานของกรอบเสียงพูดที่ n , $s_n[i]$ คือสัญญาณเสียงพูดที่ i ในกรอบเสียงพูด n และในแต่ละกรอบเสียงพูดจะมีสัญญาณเสียงจำนวน K ซึ่งการคำนวณหาค่าพลังงานของสัญญาณเสียงพูดมีดังนี้คือ

2.2.1 พลังงานสัมบูรณ์ (Absolute Energy)

เป็นการหาผลรวมของสัญญาณเสียงสัมบูรณ์ในแต่ละกรอบเสียงพูด ดังสมการที่ (5)

$$En = \sum_{i=1}^{K} |S_n[i]| \quad (5)$$

2.2.2 พลังงานรากของกำลังสองเฉลี่ย (Root Mean Square Energy)

เป็นการหาพลังงานของสัญญาณเสียงพูดจากรากที่สองของผลรวมกำลังสองเฉลี่ย ดังสมการที่ (6)

$$En = \left[\frac{1}{K}\sum_{i=1}^{K} S_n^2[i]\right]^{1/2} \quad (6)$$

2.2.3 พลังงานกำลังสอง (Square Energy)

เป็นการวัดค่าพลังงานจากสัญญาณเสียงพูดยกกำลังสอง ทำให้ค่าพลังงานมีความไวต่อสัญญาณที่มีขนาดใหญ่ ดังสมการที่ (7)

$$En = \left[\sum_{i=1}^{K} S_n^2[i]\right] \quad (7)$$

2.2.4 พลังงานของ Teager (Teager Energy)

เป็นการหาพลังงานจากผลรวมของพลังงานรายคาบ (Sinusoidal) ซึ่งจะเท่ากับขนาดของสัญญาณกำลังสองคูณกับความถี่กำลังสอง หรือหาได้จากสัญญาณของตัวเอง ลบด้วยผลคูณของสัญญาณข้างหน้า กับสัญญาณข้างหลัง ดังสมการที่ (8) และ พลังงานของแต่ละกรอบเสียงพูดสามารถหาได้จากสมการที่ (9)

$$S^2[i] - S[i+1]S[i-1] = A^2\Omega^2 \quad (8)$$

$$En = \sum_{i=1}^{K} E[i] \quad (9)$$

เมื่อ $\Omega = 2\pi f / f_S$ ซึ่ง f คือความถี่แอนาลอก และ f_S คืออัตราการสุ่มตัวอย่าง

A คือขนาดของสัญญาณเสียงพูด

ข้อมูลพลังงานของแต่ละเสียงสระ จะถูกนำมา Normalize ให้ค่าพื้นที่ใต้ function เป็น 1 ด้วยสมการที่ (10) ก่อน ที่จะนำไปทำการทดสอบ

$$Q_n = AX_n \quad (10)$$

$$\sum_{N}^{M} AX_n = 1 \quad (11)$$

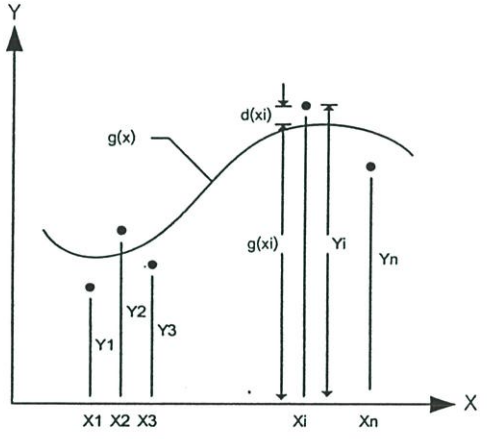
$$A = \frac{1}{\sum_{N}^{M} X_n} \quad (12)$$

เมื่อ M คือจำนวน , A คือ Normalize factor

Q_n คือค่า X_n ที่ ผ่านการ Normalize

2.3 การถดถอยแบบพหุนาม .

การถดถอยแบบพหุนาม (polynomial regression) เป็นระเบียบวิธีที่ใช้ประดิษฐ์ ฟังก์ชั่นพหุนามสำหรับข้อมูลที่มีการกระจายโดยทั่วไปที่ไม่อยู่ในรูปของเชิงเส้น ซึ่งชุดข้อมูลดังรูปที่2 สามารถประดิษฐ์ฟังก์ชั่นพหุนามได้ดังสมการที่ (13)



รูปที่ 2 ชุดข้อมูลการประดิษฐ์ฟังก์ชันพหุนาม

$$g(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m \qquad (13)$$

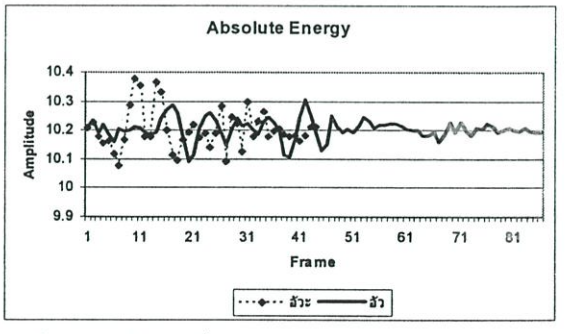
โดย $a_0, a_1, \ldots, a_m$ เป็นค่าคงตัวที่ไม่รู้ค่าซึ่งจะคำนวณหาจากเงื่อนไขที่ว่า สมการพหุนามที่จะประดิษฐ์ขึ้นมานี้ จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุด ซึ่งฟังก์ชั่นการถดถอยพหุนาม เขียนให้อยู่ในรูปแบบของระบบสมการที่ประกอบด้วย m+1 สมการย่อยได้ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^m \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \cdots & \sum_{i=1}^{n} x_i^{m+1} \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \cdots & \sum_{i=1}^{n} x_i^{m+2} \\ \vdots & \vdots & \vdots & & \\ \sum_{i=1}^{n} x_i^m & \sum_{i=1}^{n} x_i^{m+1} & \sum_{i=1}^{n} x_i^{m+2} & \cdots & \sum_{i=1}^{n} x_i^{2m} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^m y_i \end{Bmatrix}$$

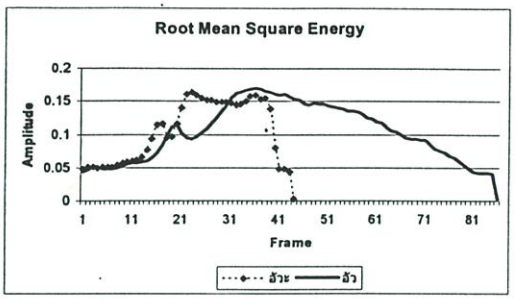
โดยในบทความนี้ใช้การประมาณฟังก์ชั่น ของข้อมูลพลังงานเสียงสระ อันดับสอง ซึ่งจะทำให้ได้ค่าสัมประสิทธิ์เป็น $a_0, a_1,$ และ a_2 มาเป็นตัวแยกแยะเสียงสระเสียงสั้น และ เสียงสระเสียงยาว

3. ผลการทดลอง

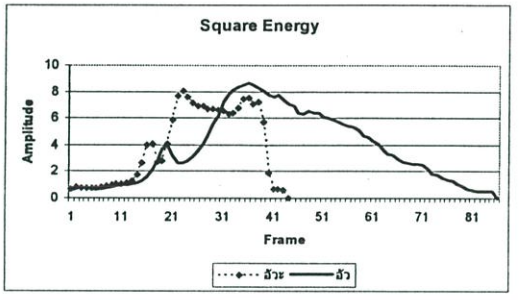
เมื่อนำสัญญาณเสียงสระ มาผ่านการเตรียมข้อมูลสัญญาณเสียงเพื่อนำไปคำนวณหาค่าพลังงานเสียงในรูปแบบ 4 รูปแบบ จะได้ลักษณะของพลังงานเสียง ดังนี้



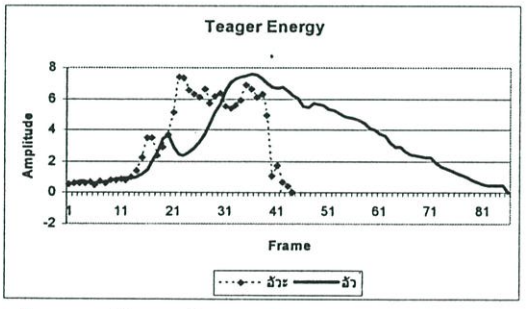
(a) ลักษณะพลังงานเสียงแบบ Absolute Energy



(b) ลักษณะพลังงานเสียงแบบ Root Mean Square Energy



(c) ลักษณะพลังงานเสียงแบบ Square Energy

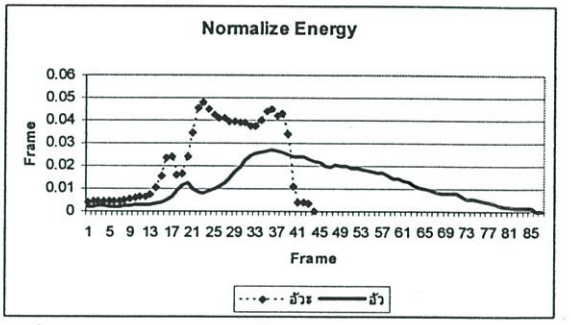


(d) ลักษณะพลังงานเสียงแบบ Teager Energy

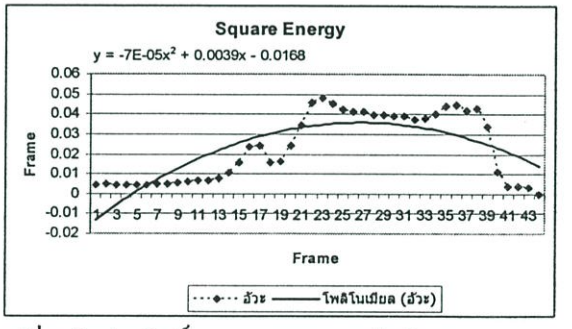
รูปที่ 3 แสดงการหาค่าพลังงานเสียงแบบต่างๆ

จากรูปที่ 3 (a)-(d) แสดงการหาค่าพลังงานเสียงทั้ง 4 รูปแบบจะเห็นได้ว่าลักษณะการหาพลังงานเสียงแบบ Absolute Energy ไม่เหมาะที่จะนำมาใช้ในการแยกสระเสียงสั้นและสระเสียงยาว เนื่องจากไม่สามารถบ่งชี้ ลักษณะการหยุดเสียงของเสียงสระได้ และจากรูปจะเห็นได้ว่า ลักษณะการหยุดเสียงของสระเสียงสั้นนั้น จะเป็นการหยุดแบบทันทีทันใด ส่วนสระเสียงยาวจะมีลักษณะการหยุดเสียงแบบค่อยๆลดระดับลงช้าๆ

เมื่อได้ลักษณะของพลังงานเสียงแล้ว ก็นำข้อมูลพลังงานเสียง มา Normalize ดังรูปที่ 4 และหาสัมประสิทธิ์การถดถอยพหุนามอันดับสอง ซึ่งจะได้ลักษณะดังรูปที่ 5

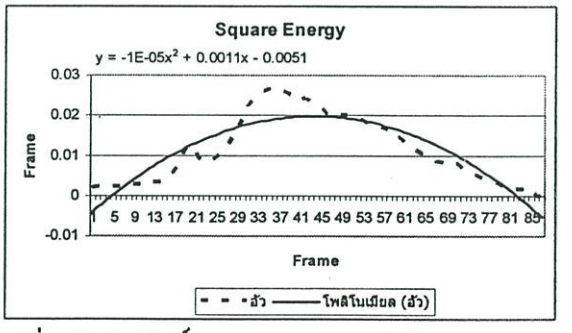


รูปที่ 4 แสดงพลังงานเสียงที่ผ่านการ Normalize



รูปที่ 5 สัมประสิทธิ์การถดถอยพหุนามอันดับสองของสระเสียงสั้น

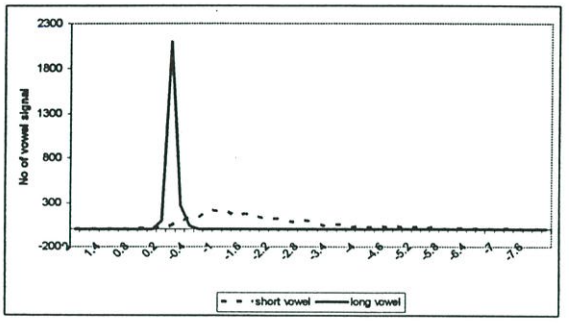
จากรูปที่ 5 จะได้ a_0 =-0.0168, a_1 = 0.0039 , a_2 = -0.000074



รูปที่ 6 สัมประสิทธิ์การถดถอยพหุนามอันดับสองของสระเสียงยาว

จากรูปที่6 จะได้ a_0 = -0.0051, a_1 = 0.0011 , a_2 = -0.000013

เมื่อนำค่าสัมประสิทธิ์การถดถอยพหุนามมาทำการแจกแจง จะได้ดังรูปที่ 7 .



รูปที่ 7 การแจกแจงค่าสัมประสิทธิ์การถดถอยพหุนาม

เพื่อให้ง่ายต่อการพิจารณา ค่า a_2 ที่นำมาพิจารณาในขั้นตอนนี้ จะทำการปรับ สเกล ให้เพิ่มขึ้น เป็น a_2 x5,000

จากการทดลอง กับตัวอย่างเสียง 5,040 เสียง (สระเสียงสั้น 2,520 เสียง, สระเสียงยาว 2,520 เสียง) ค่าสัมประสิทธิ์ a_2 สามารถที่จะใช้เป็นตัวแยกแยะ เสียงสระเสียงสั้น และ เสียงสระเสียงยาวได้ดีที่สุด คือ a_2 = -0.6 โดยเมื่อคำนวณ ตามรูปแบบของพลังงานเสียง 3 รูปแบบ จะได้อัตราการแยกแยะเสียงดังตารางที่1

ตารางที่ 1 อัตราการแยกแยะสระเสียงสั้นและสระเสียงยาว (แบบเสียงอ้างอิง)

Energy Algorithm	สระเสียงสั้น	สระเสียงยาว	เปอร์เซ็นต์
Teager	2273	2437	93.45%
Square	2331	2446	94.78%
Root Mean Square	2133	2471	91.35%

จากผลการทดลองจะเห็นได้ว่าถ้าใช้พลังงานเสียงแบบ Square Energy จะได้เปอร์เซ็นต์การแยกแยะมากที่สุด ดังนั้น จึงได้ทำการทดสอบกับ สัญญาณเสียงสระอีกชุดในจำนวนเสียงเท่ากัน คือ 5,040 และ เป็นผู้พูดคนเดียวกัน ได้เปอร์เซ็นต์การแยกแยะ ดังตารางที่ 2

ตารางที่ 2 อัตราการแยกแยะสระเสียงสั้นและสระเสียงยาว (แบบเสียงทดสอบ)

Energy Algorithm	สระเสียงสั้น	สระเสียงยาว	เปอร์เซ็นต์
Square	2356	2420	94.76%

4. สรุป

บทความนี้เป็นการนำเสนอวิธีการแยกแยะ เสียงสระเสียงสั้น และ เสียงสระเสียงยาว โดยค่าสัมประสิทธิ์ a_2 ของการถดถอยพหุนามอันดับสอง ในรูปแบบพลังงานเสียงแบบ Square Energy ให้เปอร์เซ็นต์การแยกแยะได้สูงที่สุด คือให้ค่าความถูกต้องอยู่ที่ 94.76 % จากการทดสอบสัญญาณเสียงทั้งสิ้น 5,040 เสียงอ้างอิง และ 5,040 เสียงทดสอบ

5. เอกสารอ้างอิง

[1] ณัฏฐชา จิตติวรางกูล , "ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย", มหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, 2541

[2] เบฐจวรรณ อินทรารักษ์สกุล , " การประมาณค่าด้วยสมการพหุนาม" , บัณฑิต , สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,2541

[3] มนัส สังวรศิลป์ , "คู่มือการใช้งาน Matlab"

[4] S. H. Chen and J. F. Wang, "Application of wavelet transforms for C/V segmentation on speech signals", *IEEE Trans. on Signal Processing*., vol. 148, No. 2, pp. 133-139 April 2001.

[5] J. F. Wang, C. H. Wu, "A Hierarchical Neural Network Model Based on a C/V Segmentation Algorithm for Isolated Mandarin Speech Recognition, " *IEEE Trans. On Signal Processing,* vol.39, No.9, pp.2141-2146, September 1991.

ประวัติผู้เขียน

นายกัญน์ สิทธิประเสริฐ เกิดเมื่อวันที่ 2 กรกฎาคม พ.ศ. 2519 ที่จังหวัดเชียงราย สำเร็จการศึกษาปริญญาตรี คณะครุศาสตร์อุตสาหกรรม สาขาวิชาวิศวกรรมโทรคมนาคม จากสถาบันเทคโนโลยีเจ้าคุณทหารลาดกระบัง ปีการศึกษา 2541